

# **Pengenalan Ekspresi Mikro Spontan Menggunakan Fitur Aliran Spasial- Temporal Deep Learning Pada Urutan Video Panjang Untuk Analisis Emosi**

## **RINGKASAN DISERTASI**

**Budhi Irawan**

**NIM: 33220019**

**(Program Studi Doktor Teknik Elektro dan Informatika)**



**Institut Teknologi Bandung  
Juli 2025**

# **Pengenalan Ekspresi Mikro Spontan Menggunakan Fitur Aliran Spasial- Temporal Deep Learning Pada Urutan Video Panjang Untuk Analisis Emosi**

Disertasi ini dipertahankan pada Sidang Terbuka Sekolah  
Pascasarjana sebagai salah satu syarat untuk memperoleh gelar  
Doktor Institut Teknologi Bandung

Juli 2025

**Budhi Irawan**

NIM: 33220019

(Program Studi Doktor Teknik Elektro dan Informatika)



Promotor : Dr. Ir. Rinaldi Munir, M.T.

Ko-promotor : Nugraha Priya Utama, S.T, M.A, Ph.D.  
Dr. Eng. Ayu Purwarianti, S.T, M.T.

Institut Teknologi Bandung

Juli 2025

# Pengenalan Ekspresi Mikro Spontan Menggunakan Fitur Aliran Spasial-Temporal Deep Learning Pada Urutan Video Panjang Untuk Analisis Emosi

Budhi Irawan  
NIM: 33220019

## 1. Latar Belakang

Ekspresi mikro merupakan ekspresi wajah yang berlangsung sangat singkat dan spontan, dengan durasi kurang dari satu detik. Perubahan yang terjadi pada ekspresi mikro sangat halus dan hanya terlihat pada area tertentu di wajah, menjadikannya sulit dikenali, bahkan oleh ahli sekalipun. Ekspresi mikro ini sering kali mencerminkan emosi sejati seseorang dan memiliki aplikasi penting dalam berbagai bidang, seperti keamanan, interogasi kriminal, dan psikoanalisis. Namun, tantangan utama dalam pengenalan ekspresi mikro adalah durasi yang sangat singkat dan perubahan otot wajah yang halus, yang membuatnya hampir tidak terlihat tanpa alat bantu profesional.

Seiring berkembangnya teknologi, terutama di bidang pengenalan wajah menggunakan *deep learning* dan *computer vision*, pendekatan baru mulai diterapkan untuk mengatasi kesulitan dalam mendeteksi ekspresi mikro. Salah satunya adalah penggunaan video panjang dengan *frame rate* tinggi, yang memungkinkan analisis spasial-temporal yang lebih akurat. Dengan pendekatan ini, berbagai tahapan ekspresi mikro dapat dianalisis dengan lebih efektif, mulai dari kondisi netral hingga puncak ekspresi dan kembali ke kondisi netral. Meskipun teknologi ini telah menunjukkan hasil yang menjanjikan, pengenalan ekspresi mikro spontan masih menghadapi kendala seperti ketidakseimbangan distribusi kelas emosi dalam *dataset* dan keterbatasan jumlah sampel data.

Sebagai solusi atas permasalahan tersebut, penelitian ini mengusulkan *pipeline* inovatif dalam pengenalan ekspresi mikro spontan. Pendekatan ini mencakup tahapan *image preprocessing* citra dan *data preprocessing* untuk meningkatkan variasi *dataset* serta mengatasi ketidakseimbangan distribusi kelas emosi. Selain itu, model *deep learning* dengan konvolusi 3D yang dilengkapi dengan *attention mechanism* digunakan untuk mengolah potongan area citra wajah yang relevan dengan ekspresi mikro. Pendekatan ini memungkinkan model untuk menangkap fitur spasial dan temporal secara lebih efektif, sehingga dapat mengenali ekspresi mikro yang sangat cepat dan perubahan kecil yang sering terlewatkan. *Pipeline* ini diharapkan dapat mengatasi tantangan utama dalam pengenalan ekspresi mikro spontan, serta meningkatkan akurasi dan efisiensi dalam proses pengenalan.

## 2. Tujuan dan Sasaran Penelitian

Penelitian disertasi ini bertujuan mengembangkan *pipeline* untuk mengenali ekspresi mikro spontan dengan tingkat akurasi yang lebih tinggi dibandingkan penelitian *state of the art*. *Pipeline* ini dirancang untuk mengatasi beberapa permasalahan, seperti keterbatasan jumlah dan variasi data, serta ketidakseimbangan distribusi kelas emosi pada *dataset*. Peningkatan akurasi dicapai dengan mengintegrasikan model jaringan konvolusi 3D dan Attention Mechanism, yang berfungsi untuk meningkatkan fokus pada fitur spasial-temporal yang relevan.

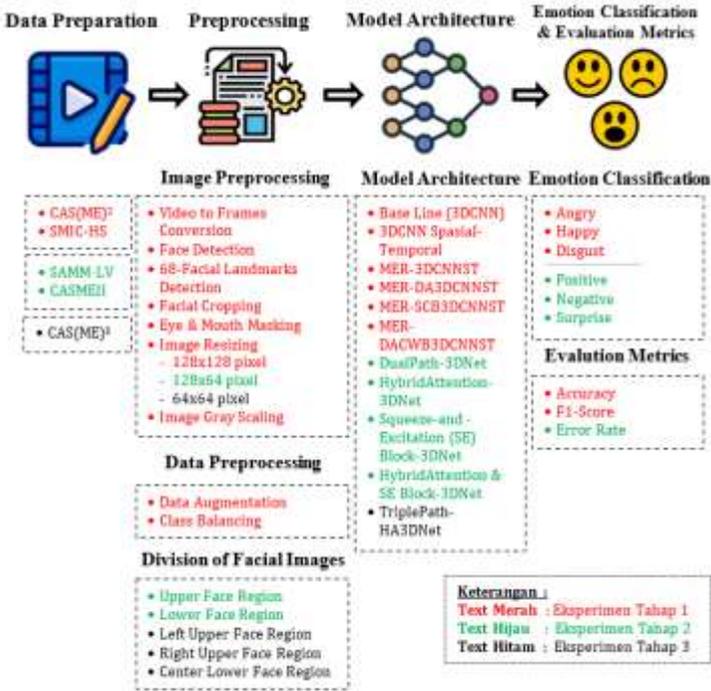
## 3. Metode Penelitian

Pada penelitian disertasi ini, dikembangkan *pipeline* untuk pengenalan ekspresi mikro spontan dari urutan video panjang, dengan fokus pada tantangan utama yang dihadapi dalam pengenalan ekspresi mikro spontan, seperti durasi ekspresi yang sangat singkat, kebutuhan *dataset* yang besar dan bervariasi, dan distribusi sampel kelas yang seimbang. Untuk mengatasi tantangan ini, diusulkan arsitektur model *Triple Path-HA3DNet* yang dilengkapi dengan mekanisme *hybrid attention*. Pendekatan ini dirancang untuk meningkatkan akurasi pengenalan ekspresi mikro spontan dengan memanfaatkan gabungan jenis *attention mechanism*, yaitu *spatial* dan *temporal*, yang memungkinkan model untuk fokus pada bagian-bagian video yang paling relevan dan dibutuhkan.

Secara lebih rinci, arsitektur model *Triple Path-HA3DNet* mengadopsi tiga jalur pemrosesan paralel yang masing-masing berfokus pada wilayah-wilayah penting di bagian area wajah, yaitu wilayah sekitar mata dan alis kiri, wilayah sekitar mata dan alis kanan, serta wilayah sekitar mulut. Dengan pendekatan ini, model dapat menangkap ekspresi mikro spontan dari berbagai bagian wajah secara lebih mendalam. Lebih lanjut, *Hybrid Attention* yang terdiri dari *Spatial Attention* dan *Temporal Attention* memungkinkan model untuk lebih fokus pada fitur wajah yang relevan baik secara spasial atau rearea tertentu di wajah maupun temporal atau perubahan ekspresi wajah dari waktu ke waktu. Pendekatan ini sangat penting untuk menangani ekspresi mikro yang cepat dan seringkali halus, yang merupakan tantangan besar dalam pengenalan ekspresi wajah.

*Pipeline* yang dikembangkan dalam penelitian ini juga mencakup serangkaian proses yang terstruktur, mulai dari *data preparation*, *image preprocessing*, *data preprocessing* hingga klasifikasi menggunakan model *deep learning*. Setiap tahapan dalam *pipeline* ini dirancang untuk memastikan bahwa data diproses secara sistematis dan konsisten, sehingga menghasilkan pengenalan ekspresi mikro spontan yang akurat dan efisien. Dengan demikian, *pipeline* ini mampu memberikan kontribusi signifikan terhadap efisiensi pengenalan ekspresi mikro spontan pada urutan video panjang.

Secara keseluruhan, penelitian ini dibagi menjadi tiga tahapan eksperimen yang bertujuan untuk terus menyempurnakan *pipeline* melalui uji coba dengan berbagai desain dan skenario eksperimen. Tahapan eksperimen tersebut terdiri dari eksperimen awal, eksperimen pengembangan, dan eksperimen final, yang kesemuanya bertujuan untuk mendapatkan *pipeline* dengan arsitektur model terbaik sehingga dapat mencapai tingkat akurasi terbaik dalam pengenalan ekspresi mikro spontan. Hasil akhir penelitian diharapkan dapat memberikan kontribusi besar dalam bidang ini. Rancangan *pipeline* keseluruhan dapat dilihat pada gambar 1.



Gambar 1. Rancangan Umum Usulan *Pipeline* Keseluruhan

**3.1. Data Preparation**

*Dataset* yang digunakan dalam penelitian disertasi ini terdiri dari lima *dataset*, yaitu CAS(ME)<sup>2</sup>, SMIC-HS, SAMM-LV, CASME II, dan CAS(ME)<sup>3</sup>. Masing-masing *dataset* memiliki kelas emosi yang berbeda-beda. *Dataset* CAS(ME)<sup>2</sup> memiliki tiga kelas emosi, yaitu *angry*, *happy*, dan *disgust*, sementara *dataset* SMIC-HS memiliki tiga kelas emosi, yakni *negative*, *positive*, dan *surprise*. *Dataset* SAMM-LV, CASME II, dan CAS(ME)<sup>3</sup>, memiliki jumlah kelas yang lebih banyak dan berbeda dibandingkan dengan *dataset* sebelumnya. *Dataset* SAMM-LV mengandung

delapan kelas emosi, yaitu *anger*, *contempt*, *disgust*, *fear*, *happiness*, *sadness*, *surprise*, dan *other*. *Dataset* CASME II memiliki tujuh kelas emosi, yakni *disgust*, *fear*, *happiness*, *repression*, *sadness*, *surprise*, dan *others*. Sementara itu, *dataset* CAS(ME)<sup>3</sup> mencakup tujuh kelas emosi, yaitu *anger*, *disgust*, *fear*, *happiness*, *sadness*, *surprise*, dan *others*.

*Dataset* yang digunakan dalam penelitian ini memiliki beberapa keunggulan yang sangat mendukung tujuan eksperimen pengenalan ekspresi mikro spontan. Salah satu alasan utama pemilihan *dataset* ini adalah karena semuanya merupakan *dataset* ekspresi mikro spontan, yang mana penting untuk penelitian ini karena mencerminkan ekspresi wajah alami yang tidak dipengaruhi oleh pelatihan atau pengarahan sebelumnya. Selain itu, semua *dataset* tersebut berupa video berdurasi panjang, memungkinkan untuk menangkap dinamika ekspresi yang lebih kaya dan lebih kompleks, serta memberikan gambaran yang lebih realistis tentang ekspresi mikro dalam konteks kehidupan nyata. Keberadaan anotasi yang jelas pada setiap *dataset* juga sangat membantu dalam proses pemberian label dan pengembangan model yang akurat.

Pada eksperimen ini, untuk *dataset* SAMM-LV, CASME II, dan CAS(ME)<sup>3</sup>, kelas emosi disederhanakan agar sejalan dengan kelas-kelas yang ada pada *dataset* sebelumnya. Konversi dilakukan dengan mengelompokkan kelas-kelas emosi tersebut menjadi tiga kategori utama, yaitu: *negative* termasuk (*'Repression'*, *'Anger'*, *'Contempt'*, *'Disgust'*, *'Fear'*, dan *'Sadness'*), *positive* (*'Happiness'*), dan *surprise* (*'Surprise'*) (See dkk., 2019), (Goh dkk., 2020), (L. Zhou dkk., 2021), (Xie dkk., 2023), (Zebiao Wang dkk., 2024), (H. Zhang dkk., 2024).

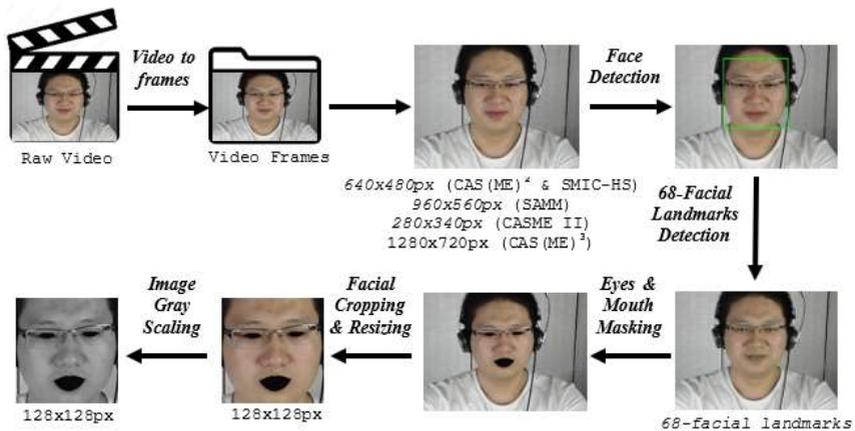
Pembagian kelas emosi menjadi tiga kategori utama yaitu *Positive*, *Negative*, dan *Surprise*, dalam analisis ekspresi mikro didasarkan pada pertimbangan ilmiah dan teknis. Mengingat bahwa ekspresi mikro adalah gerakan wajah yang sangat cepat dan halus, karakteristik tersebut menyulitkan identifikasi emosi yang lebih kompleks. Oleh karena itu, menyederhanakan klasifikasi menjadi tiga kategori utama membantu mempermudah proses pengolahan dan pengenalan emosi secara lebih efektif (Zhao dan Li, 2019).

### **3.2. Image Preprocessing**

*Image preprocessing* merupakan tahap awal dalam pengolahan citra untuk model pengenalan ekspresi mikro spontan. Pada dasarnya, tujuan utama dari *image preprocessing* adalah mengubah data mentah, seperti video atau citra, menjadi format yang lebih terstruktur dan siap untuk dianalisis oleh model *deep learning*. Langkah ini sangat penting untuk memastikan bahwa data yang diolah bebas dari gangguan atau *noise* dan elemen-elemen yang tidak relevan, serta menonjolkan fitur yang esensial guna meningkatkan performa dan akurasi model dalam memprediksi

ekspresi mikro spontan yang tersembunyi. Selain itu, *image preprocessing* berperan dalam meningkatkan kualitas citra dengan mengurangi distorsi visual dan memperbaiki ketajaman informasi yang dapat digunakan oleh model. Tahapan *image preprocessing* ditunjukkan pada gambar 2.

Pada tahap ini, konversi video mentah menjadi rangkaian *frame* citra berurutan dilakukan, di mana resolusi awal setiap *frame* disesuaikan dengan *dataset* yang digunakan, seperti 640x480 piksel untuk CAS(ME)<sup>2</sup> dan SMIC-HS, 960x560 piksel untuk SAMM-LV, 280x340 piksel untuk CASME II, serta 1280x720 piksel untuk CAS(ME)<sup>3</sup>. Proses ini bertujuan untuk menstandarisasi ukuran citra agar sesuai dengan kebutuhan model *learning*. Selanjutnya, serangkaian langkah pengolahan lebih lanjut seperti *face detection*, deteksi *68-facial landmark*, hingga pemrosesan bagian area sekitar mata dan mulut, diterapkan untuk mengekstraksi fitur-fitur penting yang akan digunakan dalam analisis ekspresi mikro spontan. Proses-proses ini dilakukan dengan tujuan untuk memperkecil dimensi citra tanpa mengurangi informasi yang signifikan, serta untuk memperjelas bagian-bagian wajah yang akan menjadi fokus dalam analisis ekspresi mikro spontan.



Gambar 2. *Image Preprocessing*

### 3.3. Data Preprocessing

Tahapan selanjutnya dalam penelitian disertasi ini adalah *data preprocessing*, yang mencakup augmentasi data dan penyeimbangan distribusi kelas emosi. Augmentasi data dilakukan untuk meningkatkan jumlah serta variasi *dataset* dengan menerapkan transformasi seperti rotasi, penyesuaian kontras, dan pengaturan kecerahan. Langkah ini bertujuan untuk memperkaya keragaman data sehingga model yang dilatih menjadi lebih *robust* dan mampu mengenali pola yang lebih beragam. Sebagaimana dijelaskan oleh Shorten dan Khoshgoftaar, (2019), augmentasi data

dapat meningkatkan kinerja model dengan memperluas cakupan pola dalam data pelatihan.

Selain itu, dilakukan penyeimbangan distribusi kelas emosi dalam *dataset* untuk mengatasi masalah ketidakseimbangan jumlah sampel antar kelas. Salah satu teknik yang digunakan adalah *class weighting*, yaitu pemberian bobot lebih besar pada kelas dengan jumlah sampel yang lebih kecil, sehingga model dapat mempelajari setiap kelas dengan lebih proporsional. Penelitian oleh Buda dkk., (2018b) menunjukkan bahwa penyeimbangan kelas menggunakan metode ini dapat meningkatkan akurasi model, khususnya pada *dataset* yang tidak seimbang. Dengan langkah-langkah tersebut, kualitas *dataset* diharapkan meningkat, sehingga model memiliki performa yang lebih optimal dan generalisasi yang lebih baik.

### **a. Data Augmentation**

*Data augmentation* data atau augmentasi data adalah teknik yang digunakan dalam *machine learning* dan *deep learning* untuk meningkatkan jumlah data pelatihan dengan memanipulasi atau mengubah data yang ada. Dalam pengenalan ekspresi mikro, augmentasi data dapat meningkatkan kinerja model dengan memperbesar *dataset*, seperti yang dijelaskan oleh Shorten dan Khoshgoftaar, (2019), dengan membuat variasi baru dari *frame* citra yang ada melalui transformasi seperti rotasi, *cropping*, perubahan kontras, dan kecerahan. Hal ini meningkatkan keragaman data pelatihan dan membantu model mengenali variasi data yang lebih besar. Melatih model pada variasi data yang lebih banyak juga memungkinkan model untuk menggeneralisasi lebih baik pada data yang belum pernah dilihat sebelumnya, yang penting dalam pengenalan ekspresi mikro spontan. Zhang dan Chai, (2024) menjelaskan bahwa augmentasi data meningkatkan kualitas model dengan memberikan kesempatan kepada model untuk mempelajari lebih banyak fitur dan pola dalam data, yang pada gilirannya membantu model memahami perubahan halus dalam ekspresi mikro wajah dan meningkatkan akurasi pengenalan ekspresi mikro spontan. Metode augmentasi yang diterapkan pada penelitian ini ada tiga metode yaitu: *Rotation*, *Cropping* dan *Brightness and Contrast Changes*.

### **b. Class Balancing**

*Class balancing* adalah strategi yang digunakan untuk menyelaraskan distribusi jumlah sampel atau observasi antara kelas-kelas dalam *dataset*. Ketidakseimbangan kelas muncul ketika terdapat disparitas signifikan antara jumlah sampel pada satu kelas dibandingkan kelas lainnya. Hal ini dapat menjadi masalah serius dalam *machine learning* atau *deep learning* karena model yang dilatih pada *dataset* tidak seimbang cenderung menghasilkan prediksi yang memihak pada kelas mayoritas dan mengabaikan kelas minoritas. Salah satu metode untuk mengatasi masalah ini adalah dengan menggunakan pendekatan seperti *oversampling* untuk meningkatkan jumlah sampel dari kelas minoritas atau dengan menggunakan *weighted loss* untuk

memberikan bobot lebih besar pada kelas-kelas yang kurang terwakili. Fungsi utama dari ketidakseimbangan kelas dalam konteks ini adalah untuk memastikan bahwa model memiliki kapasitas yang baik untuk mengenali dan memprediksi berbagai ekspresi mikro spontan tanpa bias yang signifikan terhadap kelas-kelas tertentu.

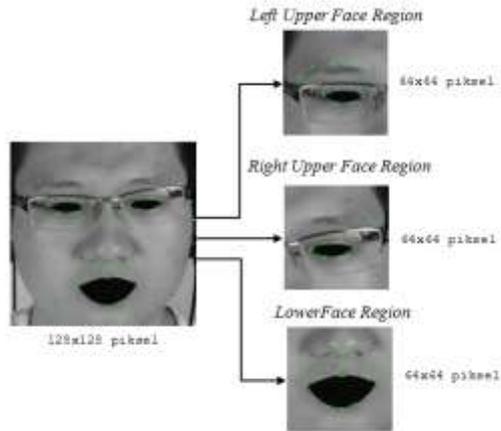
Metode *class weight* dapat membantu pengenalan ekspresi mikro di antaranya adalah untuk mengatasi ketidakseimbangan distribusi kelas emosi dalam *dataset*. Ketidakseimbangan kelas ini menjadi salah satu masalah umum dalam pengenalan ekspresi mikro, di mana beberapa kelas memiliki lebih banyak sampel daripada yang lain. Dengan memberikan bobot kelas, maka akan diberikan nilai lebih besar pada kelas yang kurang umum, sehingga model lebih cenderung untuk belajar dengan baik dari kelas-kelas minoritas. Ini sangat membantu di dalam mengatasi masalah ketidakseimbangan kelas dan meningkatkan kemampuan model dalam mengenali kelas minoritas. Selain itu kelas yang kurang umum seperti ekspresi mikro yang jarang muncul cenderung memiliki akurasi yang lebih rendah karena jumlah sampel yang terbatas (Putra dkk., 2022). Dengan memberikan bobot kelas, maka akan diberikan insentif yang lebih besar bagi model untuk mengenali dengan benar kelas-kelas ini, yang dapat meningkatkan akurasi pada kelas-kelas minoritas. Pemberian bobot kelas juga diharapkan dapat membantu mengurangi *overfitting* pada kelas mayoritas. Dengan memberikan bobot kelas, maka akan mencegah model terlalu fokus pada kelas mayoritas dan membantu model untuk lebih seimbang dalam mempelajari semua kelas. Persamaan untuk menghitung bobot kelas dalam konteks ketidakseimbangan kelas seringkali mengikuti pendekatan sederhana yang melibatkan proporsi atau perbandingan jumlah sampel antar kelas (Putra dkk., 2022).

### **c. Division of Facial Images**

Dalam penerapan model klasifikasi untuk pengenalan ekspresi mikro spontan, citra wajah hasil dari *data preprocessing* yang berukuran 128x128 piksel dibagi menjadi beberapa bagian untuk mengakomodasi distribusi AU ekspresi mikro spontan. Pembagian ini dilakukan untuk memungkinkan model menangkap variasi ekspresi wajah yang relevan di setiap bagian wajah, karena ekspresi mikro umumnya terlokalisasi pada area tertentu dari wajah. Dengan cara ini, model dapat fokus pada pergerakan otot wajah yang terjadi di bagian tertentu, seperti area sekitar mata atau mulut, yang masing-masing memiliki kontribusi penting dalam pengenalan ekspresi mikro spontan.

Citra wajah dibagi menjadi tiga bagian: *left upper face region*, *right upper face region*, dan *lower face region*, dengan masing-masing ukuran 64x64 piksel. Setiap bagian difokuskan pada area dengan konsentrasi AU yang lebih tinggi. Pembagian ini memungkinkan model memproses setiap wilayah secara terpisah, sehingga fitur spasial-temporal terkait ekspresi mikro dapat lebih mendalam. Bagian atas wajah,

baik kiri maupun kanan, cenderung merepresentasikan emosi positif seperti *happy* dan *surprise* (Ekman & Friesen, 1969). Sementara itu, bagian bawah wajah, terutama sekitar mulut, sering merepresentasikan emosi *Negative* seperti *sadness* seperti yang ditunjukkan pada gambar 3.



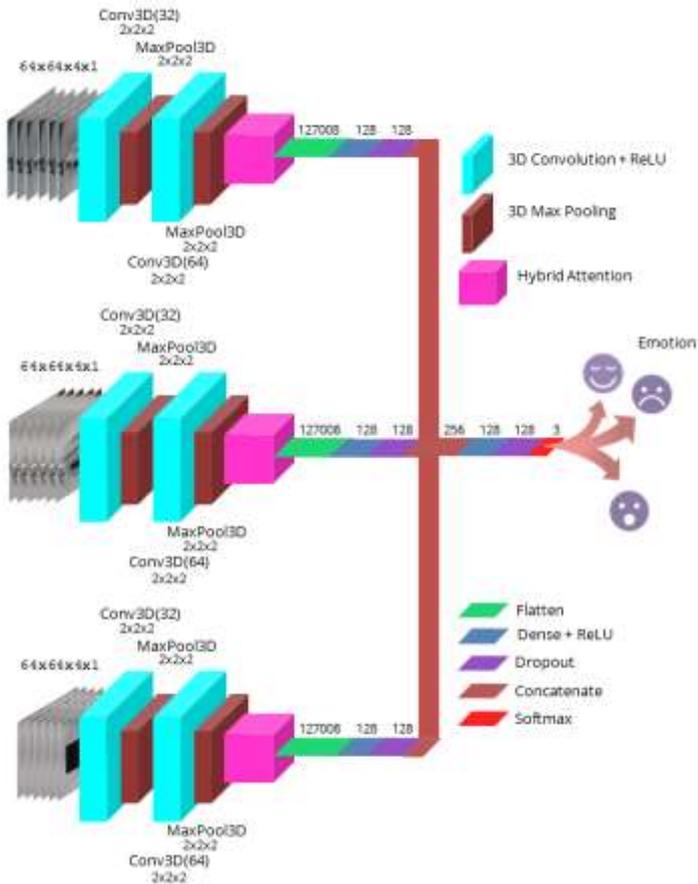
Gambar 3. Pembagian citra wajah menjadi *Left Upper Face Region*, *Right Upper Face Region* dan *Lower Face Region*

### 3.4. Arsitektur Model TriplePath-HA3DNet

Arsitektur model *Triple Path – Hybrid Attention (HA) 3DNet* dirancang untuk meningkatkan pengenalan ekspresi mikro dengan memanfaatkan tiga jalur pemrosesan paralel, yang masing-masing berfokus pada tiga area wajah: *wilayah sekitar mata dan alis kiri*, *wilayah sekitar mata dan alis kanan*, serta *wilayah sekitar mulut*. Setiap jalur menerima data input dengan dimensi 64×64 piksel, yang memungkinkan model ini untuk menangkap ekspresi mikro dengan lebih spesifik dan rinci pada setiap bagian wajah. Keunggulan utama dari model ini adalah kemampuannya untuk memproses informasi dari tiga area wajah secara paralel, memberikan pemahaman yang lebih mendalam dan lebih akurat tentang ekspresi wajah. Arsitektur Model *Triple Path-HA3Dnet* ditunjukkan seperti pada gambar 4.

Pada setiap jalur, pertama-tama data diproses melalui lapisan *3D Conv* dengan kernel 3×3×3 dan 32 filter, yang berfungsi untuk mengekstraksi fitur spasial dan temporal. Kemudian, lapisan *MaxPooling 3D* digunakan untuk mereduksi dimensi spasial data, mempercepat komputasi tanpa mengorbankan informasi penting. Setelah proses ekstraksi fitur spasial-temporal, model ini mengaplikasikan *Hybrid Attention*, yang terdiri dari *Spatial Attention* dan *Temporal Attention*. *Spatial Attention* membantu model dalam memfokuskan perhatian pada bagian wajah yang paling relevan, seperti mata atau bibir, sementara *Temporal Attention*

memperhatikan perubahan ekspresi wajah seiring waktu, memungkinkan model untuk mengenali ekspresi mikro spontan yang sangat singkat.



Gambar 4. Arsitektur Model *Triple Path (TP) – Hybrid Attention (HA) 3DNet*

Model ini menggabungkan tiga jalur, yang meningkatkan kemampuan model untuk menangkap variasi ekspresi wajah dari berbagai sudut pandang. Dengan tiga jalur ini, model lebih sensitif terhadap ekspresi mikro spontan yang mungkin tidak terlihat jika hanya mengandalkan satu atau dua wilayah wajah. Setelah proses *Hybrid Attention*, data diproses melalui lapisan *Flatten*, yang mengubah data menjadi vektor satu dimensi, dan dilanjutkan dengan lapisan *Dense* yang menggunakan fungsi aktivasi *ReLU* serta lapisan *Dropout* untuk mencegah *overfitting*.

Akhirnya, informasi dari ketiga jalur digabungkan menggunakan lapisan *Concatenate*. Output akhir model ini diproses melalui lapisan *Softmax*, yang mengonversi hasil menjadi probabilitas untuk masing-masing kategori emosi. Dibandingkan dengan model-model sebelumnya yang hanya mengandalkan dua jalur atau satu jalur pemrosesan, Arsitektur model *Triple Path-HA3DNet* menunjukkan kemampuan yang lebih baik dalam menangkap ekspresi mikro yang cepat dan halus dengan memberikan perhatian lebih pada berbagai bagian wajah, baik secara spasial maupun temporal.

Secara keseluruhan, keunggulan utama dari arsitektur model ini terletak pada kemampuan untuk memproses lebih banyak informasi dengan menangkap fitur spasial dan temporal secara lebih mendalam dan terperinci dari tiga bagian wajah, serta meningkatkan akurasi pengenalan dengan *Hybrid Attention* yang memungkinkan model untuk lebih fokus pada bagian wajah yang paling relevan.

### **3.5. Ruang Lingkup Eksperimen**

Pada bagian ini, dijelaskan secara rinci mengenai lingkungan dan tujuan eksperimen, yang mencakup berbagai kebutuhan perangkat keras dan lunak serta skenario eksperimen yang telah dirancang untuk menguji efektivitas dan performa *pipeline* dan arsitektur model yang digunakan. Proses eksperimen dimulai dari penyusunan berbagai *pipeline* untuk arsitektur model jaringan konvolusi 3D, yang melibatkan berbagai teknik *image preprocessing* dan *data preprocessing*, termasuk pengujian beberapa usulan arsitektur model seperti *Dual Path* dan *Triple Path* dengan menerapkan *Hybrid Attention*, yang menggunakan beberapa *dataset* ekspresi wajah seperti CAS(ME)<sup>2</sup>, SMIC-HS, SAMM-LV, CASME II dan CAS(ME)<sup>3</sup>. Selanjutnya, dilakukan analisis terhadap hasil eksperimen, yang mencakup evaluasi kinerja setiap model berdasarkan parameter yang ditentukan sebelumnya, serta pembahasan terkait keunggulan dan keterbatasan dari setiap usulan yang diimplementasikan dalam eksperimen ini.

#### **3.5.1. Lingkungan dan Perangkat Eksperimen**

Lingkungan eksperimen pada penelitian disertasi ini sangat berperan dalam mendukung keberhasilan pengenalan ekspresi mikro spontan. Pemilihan komponen perangkat keras dan perangkat lunak yang sesuai sangat penting untuk memastikan kinerja model yang optimal. Komponen ini dirancang untuk menangani kebutuhan pemrosesan data yang besar dengan efisiensi tinggi serta mendukung pelatihan arsitektur model berbasis *deep learning* yang membutuhkan sumber daya komputasi tinggi. Eksperimen ini melibatkan proses *image preprocessing*, *data preprocessing*, pelatihan model klasifikasi, serta evaluasi hasil prediksi yang semuanya memerlukan waktu pemrosesan yang cepat dan akurat. Dengan demikian, konfigurasi perangkat keras dan perangkat lunak yang digunakan harus memiliki kemampuan yang andal untuk menangani kompleksitas data dan model.

### a. Perangkat Keras dan Perangkat Lunak

Perangkat keras yang digunakan dalam eksperimen ini dirancang untuk mendukung eksperimen deep learning dengan arsitektur model *Triple Path-HA3DNet* yang cukup kompleks. Konfigurasi yang dipilih meliputi *CPU Intel Core i7 10th Gen* dengan *8 core* dan *16 thread*, *GPU Nvidia GeForce RTX 4070 Ti 16GB*, *RAM 32GB DDR4*, dan penyimpanan *SSD 1TB*. Sedangkan perangkat lunak yang digunakan dalam eksperimen ini dipilih berdasarkan kestabilan, kompatibilitas, dan kemampuannya dalam mendukung berbagai tahapan penelitian. Sistem operasi yang dipilih adalah *Windows 10 Pro*, yang menawarkan kestabilan dan kompatibilitas baik dengan perangkat lunak yang diperlukan. Bahasa pemrograman *Python 3.8* dipilih karena kemudahan penggunaannya dalam analisis data dan pengembangan model *deep learning*. *Library* seperti *TensorFlow 2.18.0*, *Keras 3.8.0*, *NumPy*, *Pandas*, *OpenCV*, dan *Matplotlib* digunakan untuk berbagai keperluan, mulai dari pembangunan model *deep learning* hingga pengolahan citra dan visualisasi data.

### b. Runtime, Hardware Accelerator dan Storage

*Google Colab* digunakan sebagai *platform* utama dalam eksperimen ini karena kemampuan akselerasi komputasi berbasis *cloud* yang disediakannya, termasuk akses ke perangkat keras yang mumpuni seperti *GPU* dan *TPU* yang sangat dibutuhkan untuk pelatihan model *deep learning*. Akselerator dapat dipilih sesuai kebutuhan, seperti *GPU Nvidia Tesla T4* atau *TPU*, yang mempercepat pemrosesan data besar dan pelatihan model. Untuk mendukung eksperimen ini, digunakan *Google Colab Pro+* yang memberikan akses prioritas ke perangkat keras lebih kuat dan kestabilan *runtime* yang lebih lama, penting untuk menghindari gangguan selama pelatihan. Penyimpanan data dilakukan melalui *Google Drive* dengan kapasitas *200 GB*, memungkinkan integrasi yang efisien dengan *Google Colab* untuk proses *preprocessing*, pelatihan, dan pengambilan hasil.

### 3.5.2. Tujuan Eksperimen

Tujuan eksperimen dalam penelitian ini adalah untuk menganalisis pengaruh penerapan arsitektur model yang diusulkan yaitu *TriplePath-HA3DNet*, terhadap peningkatan akurasi pengenalan ekspresi mikro spontan pada lima *dataset* yang berbeda. Eksperimen ini bertujuan untuk mengevaluasi bagaimana setiap model dapat menangani perbedaan karakteristik ekspresi mikro dalam berbagai *dataset* yang memiliki tingkat kompleksitas dan variasi yang berbeda. Selain itu, eksperimen ini juga akan mengkaji dampak dari penerapan tahapan *pipeline* yang meliputi *image preprocessing*, augmentasi data, dan *class balancing*, yang berperan dalam meningkatkan kualitas dan keberagaman data yang digunakan dalam pelatihan model.

Skenario eksperimen adalah serangkaian rencana dan langkah-langkah penelitian yang dilakukan secara iteratif dan bergantian untuk mencapai hasil yang sesuai

dengan tujuan. Dalam penelitian disertasi ini, skenario eksperimen dirancang berdasarkan *pipeline* dan model klasifikasi yang diusulkan yaitu *Triple Path-HA3Dnet*. Setiap skenario eksperimen menggunakan berbagai *dataset* dengan menerapkan beberapa tahapan, yaitu *image preprocessing*, *data preprocessing*, pemodelan klasifikasi, serta pengaturan *hyperparameter*. Pada setiap *dataset*, eksperimen dilakukan beberapa kali untuk menghasilkan total eksperimen yang mencakup seluruh variasi skenario. Proses pelaksanaan eksperimen dilakukan secara bertahap sesuai dengan variasi skenario yang telah ditentukan.

Tabel 1. Skenario Eksperimen ke-14 Menggunakan *Pipeline* dengan Arsitektur Model *Triple Path-HA3Dnet* pada *Dataset* CAS(ME)<sup>2</sup>, SMIC-HS, SAMM-LV, CASME II dan CAS(ME)<sup>3</sup>

No.	Image Pre	Data Pre	Hybrid Att	Data Split	Batch Size	Epoch
1	✓	✓	✓	80:10:10	32	128
2	✓	✓	✓	80:10:10	32	256
4	✓	✓	✓	80:10:10	64	128
4	✓	✓	✓	80:10:10	64	256
5	✓	✓	✓	80:10:10	128	128
6	✓	✓	✓	80:10:10	128	256
7	✓	✓	✓	70:15:15	32	128
8	✓	✓	✓	70:15:15	32	256
9	✓	✓	✓	70:15:15	64	128
10	✓	✓	✓	70:15:15	64	256
11	✓	✓	✓	70:15:15	128	128
12	✓	✓	✓	70:15:15	128	256

Salah satu skenario yang diterapkan ditunjukkan pada tabel 1 yaitu menerapkan arsitektur model *Triple Path-HA3Dnet* dengan *dataset* CAS(ME)<sup>2</sup>, SMIC-HS, SAMM-LV, CASME II dan CAS(ME)<sup>3</sup>. Proses diawali dengan inputan tiga aliran data hasil dari *image preprocessing* dan *data preprocessing*. Kemudian dilanjutkan dengan klasifikasi menggunakan arsitektur model *Triple Path-HA3Dnet*. Eksperimen dilakukan sebanyak 60 kali untuk mengevaluasi kemampuan model dalam mengolah *dataset* yang lebih kompleks dan beragam serta mengetahui efek dari tiga aliran inputan serta penerapan *hybrid attention* terhadap peningkatan akurasi dan *F1-Score* yang dihasilkan.

### 3.5.3. Hyperparameter Tuning

*Hyperparameter tuning* dilakukan untuk menentukan kombinasi parameter yang optimal agar model dapat mengenali ekspresi mikro spontan dengan akurasi yang lebih baik. Beberapa parameter yang diuji dalam eksperimen ini meliputi *batch size*, *epoch*, *data splitting*, *optimizer*, dan *loss function*. *Batch size* yang diuji adalah 32, 64, dan 128 untuk mencari keseimbangan antara kecepatan pelatihan dan akurasi model. Jumlah *epoch* yang digunakan adalah 128 dan 256 untuk memastikan model

mempelajari pola data secara mendalam tanpa *overfitting*. Data splitting dilakukan dengan membagi dataset menjadi training, testing, dan validation untuk memastikan evaluasi model yang efektif. *Optimizer* ADAM dipilih karena kemampuannya dalam menangani fluktuasi gradien dan mempercepat konvergensi, dengan learning rate yang diuji pada nilai 0.001. Loss function yang digunakan adalah *Categorical Cross-Entropy*, yang dipilih untuk tugas klasifikasi multi-kategori guna mengoptimalkan akurasi model dalam mengenali ekspresi mikro.

### 3.5.4. Evaluation Metrics

Metrik evaluasi atau *Evaluation metrics* adalah ukuran atau metrik yang digunakan untuk menilai kinerja model *deep learning*. Pemilihan metrik evaluasi yang tepat sangat penting karena menentukan seberapa baik model melakukan tugas yang diberikan dan membantu dalam membandingkan kinerja berbagai model. Akurasi mengukur sejauh mana model klasifikasi benar dalam mengklasifikasikan seluruh *dataset* (Gan dkk., 2024). Akurasi dihitung dengan membagi jumlah prediksi yang benar ( $TP + TN$ ) oleh jumlah total sampel.

$$Akurasi = \frac{TP+TN}{Total\ Sampel} \dots\dots\dots (1)$$

Sedangkan *F1-Score* adalah metrik evaluasi yang menggabungkan presisi atau *precision* dan *recall*. *Precision* mengukur seberapa baik model memprediksi positif ketika sebenarnya positif, sedangkan *recall* mengukur seberapa baik model mendeteksi semua *instance* positif yang sebenarnya.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \dots\dots\dots (2)$$

*Recall* mengukur seberapa baik model mendeteksi semua *instance* positif yang sebenarnya. *Recall* dihitung dengan membagi jumlah prediksi benar positif (*true positives*) oleh jumlah total *instance positif* yang sebenarnya ( $TP + FN$ ). Nilai *recall* yang tinggi menunjukkan bahwa model berhasil mengidentifikasi hampir semua *instance* positif yang ada di *dataset*.

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

*Precision* mengukur seberapa baik model memprediksi positif ketika sebenarnya positif. *Precision* dihitung dengan membagi jumlah prediksi benar positif atau *TP* oleh jumlah total *instance* yang diprediksi positif ( $TP + FP$ ). Nilai *precision* yang tinggi menunjukkan bahwa model jarang memberikan prediksi positif yang salah atau *FP*.

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots (4)$$

*Error rate* mengukur seberapa sering model melakukan kesalahan dalam mengklasifikasikan sampel. *Error rate* dihitung dengan membagi jumlah prediksi

salah ( $FP + FN$ ) oleh jumlah total sampel. Nilai *error rate* yang rendah menunjukkan bahwa model jarang membuat kesalahan klasifikasi.

$$Error\ Rate = \frac{FP+FN}{Total\ Samples} \dots\dots\dots (5)$$

**4. Hasil dan Pembahasan**

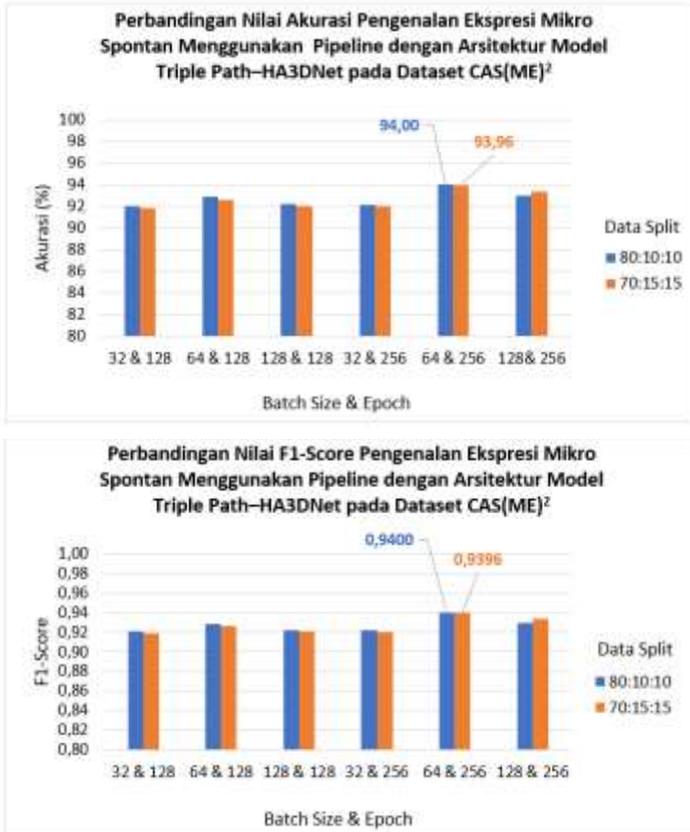
Setiap tahapan eksperimen dilakukan dengan perhitungan metrik evaluasi berupa nilai akurasi, *F1-Score*, dan *error rate*. Grafik yang disajikan mencakup informasi dari semua skenario eksperimen yang dilakukan, termasuk *dataset* yang digunakan, penerapan *image preprocessing* dan *data preprocessing*, serta pemilihan *hyperparameter tuning* yang tepat. Jenis grafik yang ditampilkan adalah grafik garis yang menunjukkan nilai akurasi dan *F1-Score* tertinggi pada setiap eksperimen berdasarkan jenis *dataset* yang digunakan. Dengan demikian, hasil perhitungan akurasi dan *F1-Score* dari setiap skenario eksperimen dapat diketahui secara jelas melalui grafik yang dilengkapi dalam hasil eksperimen.

Metrik evaluasi yang digunakan dalam penelitian ini adalah akurasi dan *F1-Score*, yang keduanya memberikan gambaran yang berbeda mengenai performa model dalam klasifikasi ekspresi mikro. Akurasi mengukur proporsi prediksi yang benar dari seluruh prediksi yang dilakukan oleh model, tetapi metrik ini lebih kurang relevan pada kasus dengan kelas yang tidak seimbang. Sebaliknya, *F1-Score* mengandung kedua elemen penting dalam evaluasi, yaitu *precision* dan *recall*, sehingga memberikan gambaran yang lebih seimbang mengenai performa model. *F1-Score* sangat relevan dalam konteks pengenalan ekspresi mikro spontan, karena mempertimbangkan keseimbangan antara kemampuan model dalam mendeteksi ekspresi mikro benar dan tidak keliru dalam mengidentifikasi data yang relevan.

Pada eksperimnen ini digunakan *pipeline* dengan arsitektur model *Triple Path-Hybrid Attention 3Dnet (TP-HA3Dnet)* untuk meningkatkan pengenalan ekspresi mikro spontan dengan menyoroti fitur spasial-temporal secara lebih spesifik. Model ini memproses input melalui tiga jalur, yaitu area sekitar mata dan alis kiri, area mata dan alis kanan, serta area sekitar mulut, masing-masing berukuran 64 x 64 piksel. Data input berasal dari lima *dataset*, diantaranya CAS(ME)<sup>2</sup>, SMIC-HS, SAMM-LV, CASME II dan CAS(ME)<sup>3</sup> yang berupa rangkaian *frame* citra hasil *preprocessing*. Setiap eksperimen melibatkan evaluasi menggunakan metrik akurasi dan *F1-Score*, yang hasilnya disajikan dalam grafik garis untuk menunjukkan nilai tertinggi berdasarkan jenis *dataset*, serta tabel hasil lengkap dari berbagai skenario eksperimen. Melalui analisis grafik dan tabel, kesimpulan diperoleh dari setiap tahapan eksperimen yang dilakukan.

#### 4.1 Pipeline dengan Arsitektur Model Triple Path-HA3DNet pada *Dataset CAS(ME)<sup>2</sup>*

Berdasarkan kepada grafik perbandingan akurasi dan *F1-Score* dari arsitektur model *TP-HA3Dnet* pada gambar 5 menunjukkan pengaruh signifikan dari penerapan *layer hybrid attention* yang didukung tahapan *image preprocessing* berjenjang, augmentasi data, dan *class balancing*.



Gambar 5. Grafik Perbandingan Nilai Akurasi dan *F1-Score* Tertinggi Pengenalan Ekspresi Mikro Spontan Menggunakan *Pipeline* dengan Arsitektur Model *Triple Path-HA3DNet* pada *Dataset CAS(ME)<sup>2</sup>*.

Pemrosesan input melalui tiga jalur yaitu *left upper face region*, *right upper face region*, dan *lower face region* yang memungkinkan model untuk mengekstraksi fitur spesifik dari area sekitar mata, alis, dan mulut. *Layer hybrid attention* berperan

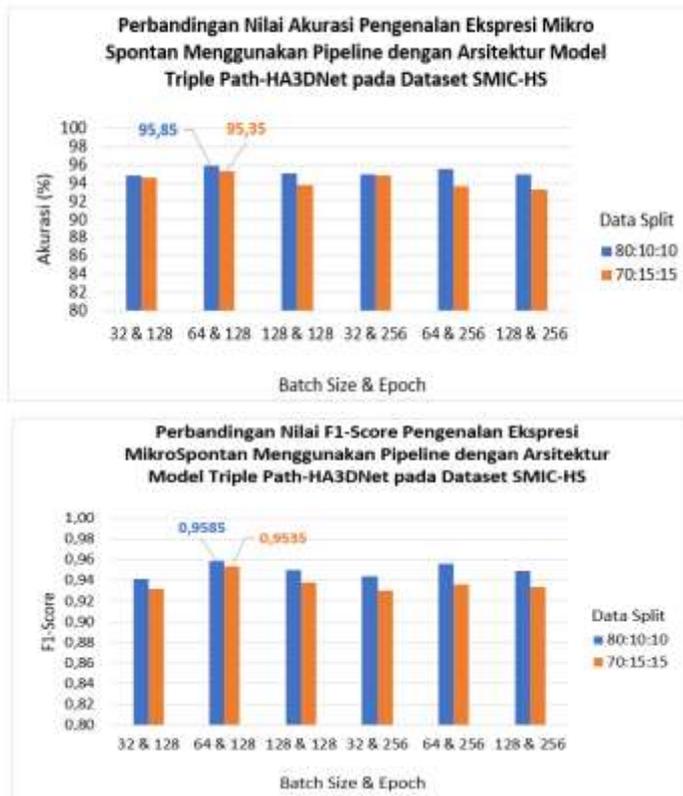
dalam meningkatkan fokus pada fitur spasial-temporal yang relevan, membantu model mendeteksi perubahan mikro yang sulit teridentifikasi. Proses *image preprocessing* memastikan kualitas input citra tetap optimal, sementara augmentasi data dan *class balancing* meningkatkan kemampuan model untuk belajar dari data yang lebih bervariasi dan seimbang.

Dari grafik tersebut, performa terbaik dicapai pada *batch size* 64 dan *epoch* 256 dengan pembagian data 80:10:10 yaitu akurasi sebesar 94,00% dan *F1-Score* sebesar 0,9400 serta 70:15:15 yaitu akurasi sebesar 93,96% dan *F1-Score* sebesar 0,9396. Hasil ini menegaskan bahwa *layer hybrid attention* dalam arsitektur model *TP-HA3DNet* mampu menonjolkan fitur penting dari ketiga jalur input sekaligus mengurangi *attention* pada informasi yang kurang relevan. Peningkatan akurasi dan *F1-Score* membuktikan bahwa kombinasi *spatial-temporal attention* dengan arsitektur model *3DNet* efektif untuk mengenali ekspresi mikro secara akurat, terutama pada *dataset* kompleks seperti CAS(ME)<sup>2</sup>.

## 4.2 Pipeline dengan Arsitektur Model Triple Path–HA3DNet pada *Dataset* SMIC-HS

Berdasarkan grafik akurasi yang ditampilkan pada Gambar 6, performa terbaik tercapai pada kombinasi *batch size* 64 dan *epoch* 128 dengan pembagian data 80:10:10, yang menghasilkan akurasi 95,85% dan *F1-Score* 0,9585. Sementara itu, pembagian data 70:15:15 menghasilkan akurasi 95,35% dan *F1-Score* 0,9535. Hasil ini menandakan kemampuan model dalam menangkap pola ekspresi mikro secara optimal melalui pengolahan tiga jalur input menggunakan *layer hybrid attention*. Fitur-fitur yang diambil dari area sekitar mata, alis, dan mulut menjadi lebih informatif, sehingga meningkatkan sensitivitas model dalam mendeteksi perubahan halus pada wajah. Tahapan *preprocessing*, augmentasi data, dan penyeimbangan kelas terbukti efektif dalam meningkatkan kualitas data input serta meminimalkan pengaruh kelas dominan.

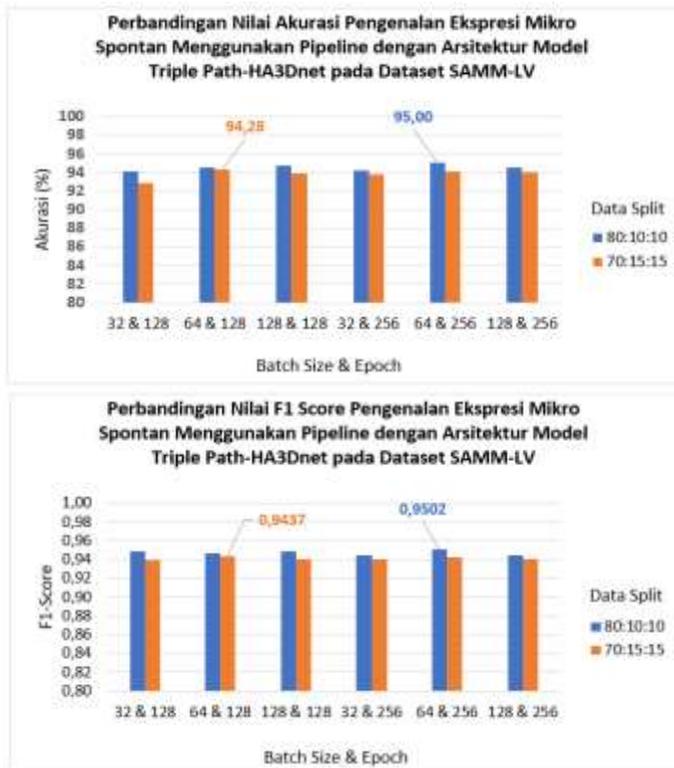
Konsistensi performa pada berbagai kombinasi *batch size* dan *epoch* semakin memperlihatkan keunggulan model *TP-HA3Dnet* dalam mengenali pola mikro halus pada *dataset* SMIC-HS. Meskipun ada variasi parameter, nilai akurasi dan *F1-Score* tetap berada pada kisaran 95%, yang mencerminkan efektivitas tahapan *image preprocessing*, augmentasi data, dan penerapan *layer hybrid attention*. Pendekatan tiga jalur input yang fokus pada area wajah spesifik, dikombinasikan dengan analisis spasial-temporal dari *layer hybrid attention*, berhasil meningkatkan kemampuan model dalam mengenali ekspresi mikro secara signifikan. Dengan demikian, model ini terbukti andal untuk menyelesaikan tugas pengenalan ekspresi mikro spontan, terutama pada *dataset* kompleks seperti SMIC-HS.



Gambar 6. Grafik Perbandingan Nilai Akurasi dan  $F1$ -Score Tertinggi Pengenalan Ekspresi Mikro Spontan Menggunakan *Pipeline* dengan Arsitektur Model *Triple Path-HA3DNet* pada *Dataset SMIC-HS*.

#### 4.3 Pipeline dengan Arsitektur Model Triple Path–HA3DNet pada *Dataset SAMM-LV*

Arsitektur model *Triple Path-HA3DNet* menunjukkan kinerja yang stabil meskipun terdapat variasi pada *batch size* dan *epoch*, dengan fluktuasi yang minim antara pembagian data 80:10:10 dan 70:15:15. Efektivitas layer *hybrid attention* semakin optimal dengan penerapan *class balancing*, yang menjaga distribusi data antar kelas emosi tetap seimbang dan menghindari bias terhadap kelas tertentu. Secara keseluruhan, pendekatan *TP-HA3DNet* yang melibatkan tiga jalur input, preprocessing bertahap, augmentasi data, dan *class balancing* terbukti efektif dalam mengenali ekspresi mikro pada dataset SAMM-LV, dengan akurasi dan  $F1$ -Score yang hampir mencapai nilai maksimal untuk tugas pengenalan ekspresi wajah mikro spontan.

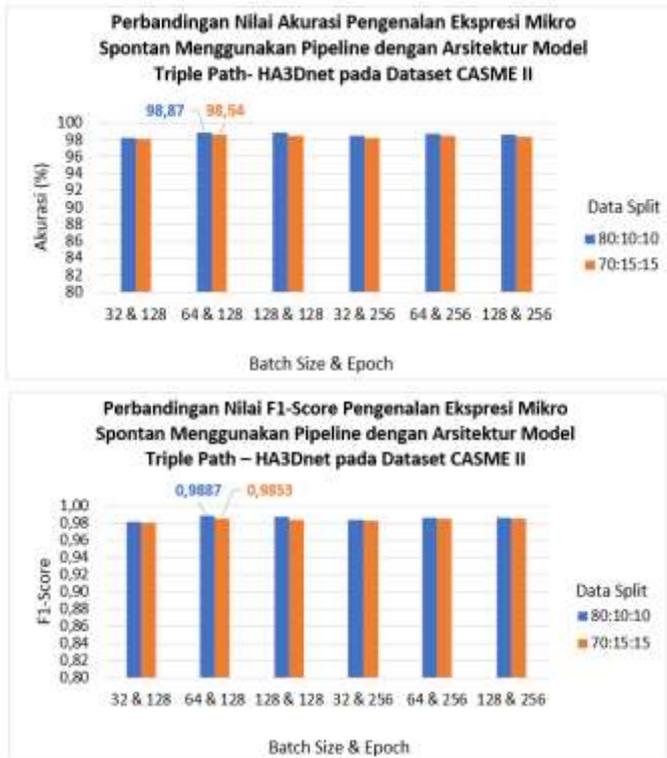


Gambar 7. Grafik Perbandingan Nilai Akurasi dan *F1-Score* Tertinggi Pengenalan Ekspresi Mikro Spontan Menggunakan *Pipeline* dengan Arsitektur Model *Triple Path-HA3DNet* pada *Dataset SAMM-LV*.

Mengacu pada grafik gambar 7, performa terbaik model tercapai pada *batch size* 64 dan *epoch* 256 dengan pembagian data 80:10:10, menghasilkan akurasi 95,00% dan *F1-Score* 0,9502. Sementara itu, untuk pembagian data 70:15:15, akurasi tertinggi tercatat pada *batch size* 64 dan *epoch* 128, dengan nilai akurasi 94,28% dan *F1-Score* 0,9437. Tingginya akurasi dan *F1-Score* ini mencerminkan efektivitas layer *hybrid attention* dalam menyoroti fitur mikro yang penting, seperti perubahan halus pada area sekitar mata, alis, dan mulut. Tahapan preprocessing memastikan kualitas input citra yang lebih baik, sedangkan augmentasi data memperkaya pola yang dapat dipelajari oleh model, meningkatkan kemampuan generalisasi terhadap berbagai variasi ekspresi.

#### 4.4 Pipeline dengan Arsitektur Model Triple Path–HA3DNet pada *Dataset CASME II*

Merujuk kepada gambar 8, performa terbaik model tercapai pada *batch size* 64 dan *epoch* 128 dengan pembagian data 80:10:10, menghasilkan akurasi 98,87% dan *F1-Score* 0,9887. Untuk pembagian data 70:15:15, parameter yang sama memberikan akurasi 98,54% dan *F1-Score* 0,9853. Hasil ini menandakan kemampuan layer *hybrid attention* dalam menyoroti fitur penting pada area wajah, seperti sekitar mata, alis, dan mulut, yang menjadi fokus utama dalam perubahan ekspresi mikro. Dengan penerapan *class balancing*, distribusi data antar kelas menjadi lebih seimbang, memungkinkan model untuk belajar lebih optimal dan mengurangi bias terhadap kelas tertentu, sehingga menghasilkan performa yang stabil dan generalisasi yang lebih baik.



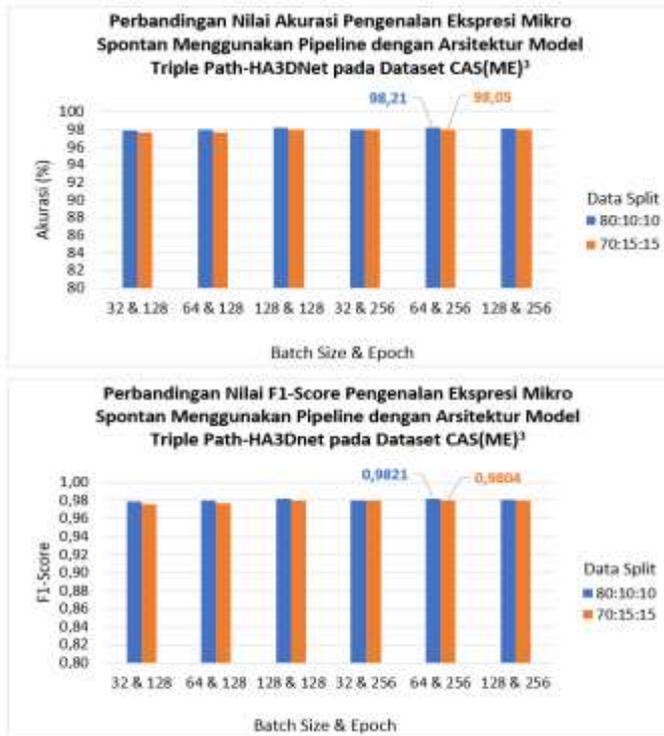
Gambar 8 Grafik Perbandingan Nilai Akurasi dan *F1-Score* Tertinggi Pengenalan Ekspresi Mikro Spontan Menggunakan *Pipeline* dengan Arsitektur Model *Triple Path-HA3DNet* Menggunakan *Dataset CASME II*.

Keunggulan arsitektur model *TP-HA3DNet* dalam mengenali ekspresi mikro pada dataset CASME II semakin terbukti melalui konsistensi performa yang terjaga pada berbagai kombinasi *batch size* dan *epoch*. Meskipun ada variasi pada parameter, akurasi dan *F1-Score* model ini tetap stabil di kisaran 98% hingga 99%. Hal ini mencerminkan efektivitas tahapan *image preprocessing*, augmentasi data, dan penerapan layer *hybrid attention*. Pendekatan tiga jalur input yang fokus pada area wajah spesifik, bersama dengan analisis spasial-temporal dari layer *hybrid attention*, secara signifikan meningkatkan kemampuan model dalam mengenali ekspresi mikro. Dengan demikian, model ini terbukti handal dalam menyelesaikan tugas pengenalan ekspresi mikro spontan, terutama pada *dataset* kompleks seperti CASME II.

#### **4.5 Pipeline dengan Arsitektur Model Triple Path–HA3DNet pada *Dataset CAS(ME)*<sup>3</sup>**

Berdasarkan gambar 9, performa tertinggi dicapai pada *batch size* 64 dan *epoch* 256 dalam pembagian data 80:10:10, dengan akurasi sebesar 98,21% dan *F1-Score* sebesar 0,9821. Sebaliknya, pada pembagian data 70:15:15, hasil optimal tercapai pada *batch size* 64 dan *epoch* 256, dengan akurasi sebesar 98,05% dan *F1-Score* sebesar 0,9804. Hasil ini memperkuat peran penting layer *hybrid attention* dalam memusatkan *attention* model pada area wajah yang relevan, seperti mata, alis, dan mulut, yang merupakan pusat perubahan ekspresi mikro. Dengan input tiga jalur yang terstruktur dan fokus pada fitur spasial-temporal, model mampu mengenali pola mikro dengan presisi yang sangat tinggi. Selain itu, *class balancing* memastikan distribusi data antar kelas tetap seimbang, sehingga model dapat belajar secara optimal tanpa bias terhadap kelas tertentu.

Hasil yang konsisten dengan nilai akurasi dan *F1-Score* tinggi pada berbagai kombinasi *batch size* dan *epoch* menegaskan keunggulan arsitektur model *TP-HA3DNet* dalam mengenali ekspresi mikro pada *dataset CAS(ME)*<sup>3</sup>. Meskipun terdapat variasi pada parameter, akurasi dan *F1-Score* tetap stabil di angka 98%, yang menunjukkan keberhasilan layer *hybrid attention* dalam menangkap fitur spasial-temporal dari ketiga jalur input. Penerapan *image preprocessing*, augmentasi data, dan penyeimbangan kelas secara signifikan dapat meningkatkan performa model, menjadikannya lebih handal dalam mengenali ekspresi mikro spontan, khususnya pada *dataset* yang kompleks dan dinamis. Oleh karena itu, pendekatan ini terbukti efektif dalam meningkatkan akurasi sistem pengenalan ekspresi mikro wajah.



Gambar 9. Grafik Perbandingan Nilai Akurasi dan *F1-Score* Tertinggi Pengenalan Ekspresi Mikro Spontan Menggunakan *Pipeline* dengan Arsitektur Model *Triple Path-HA3DNet* pada *Dataset CAS(ME)*<sup>3</sup>.

Berdasarkan tabel 2 yang berisikan hasil pengujian pengenalan ekspresi mikro dengan menggunakan model *Triple Path-HA3DNet* pada lima skenario *pipeline*, dapat diperoleh gambaran yang jelas mengenai pengaruh dari teknik augmentasi, *class balancing*, dan *hybrid attention* dalam meningkatkan akurasi dan *F1-score*. Dari tabel tersebut, terlihat bahwa kombinasi teknik *image preprocessing*, *data preprocessing* berupa teknik augmentasi dan *class balancing*, serta penerapan *hybrid attention* memberikan hasil yang signifikan dalam meningkatkan performa model untuk pengenalan ekspresi mikro.

Pertama, pada eksperimen ke-10 yang menggunakan *image preprocessing* dan arsitektur model *Triple Path-HA3DNet* tanpa menerapkan augmentasi dan *class balancing*, diperoleh akurasi 90,95% dengan *F1-score* sebesar 0,9037 pada *dataset CAS(ME)*<sup>2</sup> dengan pembagian data 80:10:10. Performa ini cenderung lebih rendah

dibandingkan dengan eksperimen ke-11 yang menerapkan augmentasi tanpa *class balancing*, dimana akurasi meningkat menjadi 91,39% dan *F1-Score* menjadi 0,9153 pada *dataset* yang sama. Hal ini menunjukkan bahwa penambahan teknik augmentasi data tanpa *class balancing* meningkatkan kemampuan model untuk mengenali ekspresi mikro dengan lebih akurat.

Tabel 2 Perbandingan Nilai Akurasi dan *F1-Score* Pengenalan Ekspresi Mikro Spontan Menggunakan 5 Skenario *Pipeline* dengan Arsitektur *Model Triple Path-HA3DNet* pada *Dataset* CAS(ME)<sup>2</sup>, SMIC-HS, SMM-LV, CASME II dan CAS(ME)<sup>3</sup>

Dataset	Skenario Eksperimen	Akurasi (%)		F1-Score	
		80:10:10	70:15:15	80:10:10	70:15:15
CAS(ME) <sup>2</sup>	Eksperimen ke-10	90,95	89,98	0,9037	0,9004
	Eksperimen ke-11	91,39	90,44	0,9153	0,9048
	Eksperimen ke-12	92,73	91,80	0,9217	0,9172
	Eksperimen ke-13	88,27	87,68	0,8815	0,8755
	Eksperimen ke-14	94,00	93,96	0,9400	0,9396
SMIC-HS	Eksperimen ke-10	93,48	92,89	0,9342	0,9273
	Eksperimen ke-11	94,32	93,69	0,9426	0,9369
	Eksperimen ke-12	93,96	93,35	0,9384	0,9317
	Eksperimen ke-13	90,94	90,19	0,9078	0,9004
	Eksperimen ke-14	95,85	95,35	0,9585	0,9535
SMM-LV	Eksperimen ke-10	92,55	91,72	0,9273	0,9174
	Eksperimen ke-11	93,42	92,56	0,9337	0,9249
	Eksperimen ke-12	93,07	92,18	0,9286	0,9215
	Eksperimen ke-13	89,07	87,94	0,8881	0,8796
	Eksperimen ke-14	95,00	94,28	0,9502	0,9437
CASME II	Eksperimen ke-10	96,61	96,14	0,9621	0,9570
	Eksperimen ke-11	97,28	96,16	0,9731	0,9607
	Eksperimen ke-12	96,94	96,44	0,9702	0,9649
	Eksperimen ke-13	94,16	93,56	0,9370	0,9326
	Eksperimen ke-14	98,87	98,54	0,9887	0,9853
CAS(ME) <sup>3</sup>	Eksperimen ke-10	96,06	95,65	0,9607	0,9566
	Eksperimen ke-11	96,72	96,36	0,9672	0,9630
	Eksperimen ke-12	96,19	95,87	0,9616	0,9588
	Eksperimen ke-13	93,50	93,07	0,9336	0,9301
	Eksperimen ke-14	98,21	98,05	0,9821	0,9804

Selanjutnya, pada eksperimen dengan menggunakan teknik augmentasi data, hasil yang konsisten menunjukkan peningkatan performa pada setiap *dataset* yang diuji. Misalnya, pada *dataset* SMIC-HS, eksperimen ke-11 menghasilkan akurasi 94,32% dan *F1-score* 0,9426 pada pembagian data 80:10:10, lebih tinggi dibandingkan dengan eksperimen 10 tanpa *data preprocessing* yang memperoleh akurasi 93,48% dan *F1-score* 0,9342. Augmentasi data membantu meningkatkan keragaman data

latih, sehingga model dapat belajar mengenali berbagai variasi ekspresi yang lebih baik, yang pada akhirnya meningkatkan akurasi dan *F1-score*.

Selain itu, pada eksperimen dengan menggunakan teknik class balancing tanpa menerapkan augmentasi data, hasil yang konsisten menunjukkan peningkatan performa pada setiap *dataset* yang diuji. Misalnya, pada *dataset* SAMM-LV, eksperimen ke-12 menghasilkan akurasi 93,07% dan *F1-score* 0,9286 pada pembagian data 80:10:10, lebih tinggi dibandingkan dengan eksperimen 10 tanpa *data preprocessing* yang memperoleh akurasi 92,55% dan *F1-score* 0,9273. Dengan menerapkan *class balancing*, model dapat mengatasi masalah ketidakseimbangan jumlah sampel antara kelas-kelas yang berbeda dalam *dataset*. Hal ini sangat penting terutama ketika terdapat kelas minoritas yang memiliki jumlah sampel jauh lebih sedikit dibandingkan dengan kelas mayoritas.

#### 4.6 Perbandingan Hasil

Berdasarkan dari Tabel 3 hingga 7 menunjukkan perbandingan performa metode yang diusulkan dengan pendekatan *state of the art* pada lima *dataset*, yaitu CAS(ME)<sup>2</sup>, SMIC-HS, SAMM-LV, CASME II, dan CAS(ME)<sup>3</sup>. Secara umum, *pipeline* dengan arsitektur model yang diusulkan yaitu *TP-HA3DNet* memberikan hasil yang lebih unggul dalam hal akurasi dan *F1-Score* dibandingkan metode lain. Misalnya, pada *dataset* CAS(ME)<sup>2</sup> pada tabel 3, metode *TP-HA3DNet* mencapai akurasi 94% dan *F1-Score* 0.9400, yang lebih tinggi dibandingkan metode sebelumnya, seperti *Deep3DCANN* dengan akurasi 90.00% dan *F1-Score* 88.00%. Hasil serupa juga terlihat pada *dataset* SMIC-HS pada tabel 4, di mana *TP-HA3DNet* mencatatkan akurasi tertinggi sebesar 95.85% dengan *F1-Score* 0.9585, menunjukkan keunggulan signifikan atas metode lainnya.

Selain itu, analisis pada *dataset* SAMM-LV pada tabel 5 memperlihatkan bahwa metode *TP-HA3DNet* mencapai akurasi 95.00% dan *F1-Score* 0.9502, melampaui metode *LAENet* yang sebelumnya memiliki performa tertinggi. Hal serupa juga ditemukan pada *dataset* CASME II pada tabel 6, di mana *TP-HA3DNet* mencatat akurasi 98.87% dan *F1-Score* 0.9887, jauh lebih tinggi dibandingkan metode *BALANCE* dan *MMAM*. *Dataset* CAS(ME)<sup>3</sup> pada tabel 7 juga mengonfirmasi keunggulan metode yang diusulkan, dengan *TP-HA3DNet* mencapai akurasi 98.21% dan *F1-Score* 0.9821. Secara keseluruhan, implementasi *image preprocessing*, *data preprocessing*, dan model arsitektur *TP-HA3DNet* dengan 3 aliran spasial-temporal yang merupakan bagian dari *pipeline* yang diusulkan ini terbukti mampu meningkatkan kemampuan mengenali ekspresi mikro spontan dan menjadikannya solusi yang efektif untuk analisis data kompleks dalam domain ini.

Tabel 3 Perbandingan Akurasi dan *F1-Score* dari Metode yang diusulkan dengan Penelitian Disertasi dan *State of the art* pada *Dataset CAS(ME)*<sup>2</sup>

Tahun	Metode	Akurasi (%)	F1-Score
2022	3D-DSENet	90.32	0.8987
2023	Deep3DCANN	90.00	0.8800
2023	STCN	82.04	0.8103
2024	MAL	85.23	0.8222
2024	MER-3DCNNST	88.25	0.8870
2024	MER-DA3DCNNST	89.35	0.8920
2024	MER-SCB3DCNNST	90.70	0.9073
2024	MER-DACWB3DCNNST	92.75	0.9128
2024	DP-3DNet	93.68	0.9358
2024	HA-3DNet	93.95	0.9395
2024	SE-3DNet	93.76	0.9368
2024	HA-3DSENet	93.88	0.9379
<b>2024</b>	<b>TP-HA3DNet</b>	<b>94.00</b>	<b>0.9400</b>

Tabel 4 Perbandingan Akurasi dan *F1-Score* dari Metode yang diusulkan dengan Penelitian Disertasi dan *State of the art* pada *Dataset SMIC-HS*

Tahun	Metode	Akurasi (%)	F1-Score
2022	AMAN	79.87	0.7708
2022	BDCN	78.69	0.7859
2022	DSCNN	69.56	0.7077
2022	3D-DSENet	92.68	0.9185
2022	ODCNN	74.80	0.7236
2022	SSFR	91.55	0.9150
2022	BALANCE	85.24	0.8429
2022	ML-STFN	84.06	0.8352
2023	STCN	84.51	0.8429
2023	MBERT	83.84	0.8550
2023	MADFN	81.02	0.8179
2024	SRMCL	80.53	0.7946
2024	MMAM	82.04	0.8157
2024	HTNet	79.05	0.8049
2024	MAL	85.19	0.8383
2024	MCFMCN	79.37	0.7848
2024	MER-DA3DCNNST	88.76	0.8820
2024	MER-SCB3DCNNST	90.07	0.9011
2024	MER-DACWB3DCNNST	91.49	0.9032
2024	DP-3DNet	91.55	0.9147
2024	HA-3DNet	93.42	0.9330
2024	SE-3DNet	92.10	0.9198
2024	HA-3DSENet	92.72	0.9263
<b>2024</b>	<b>TP-HA3DNet</b>	<b>95.85</b>	<b>0.9585</b>

Tabel 5 Perbandingan Akurasi dan *F1-Score* dari Metode yang diusulkan dengan Penelitian Disertasi dan *State of the art* pada *Dataset* SAMM-LV

Tahun	Metode	Akurasi (%)	F1-Score
2022	BDCN	79.94	0.8186
2022	SConvLSTM	78.33	0.7892
2022	MMNet	90.22	0.8391
2022	ODCNN	68.69	0.6919
2022	BALANCE	80.73	0.8497
2022	ML-STFN	84.01	0.8531
2023	ADMME	90.98	0.9092
2023	AGAN-TF	76.43	0.7928
2023	DS3DCNN	78.30	0.7554
2023	MBERT	84.75	0.8386
2023	MADFN	82.89	0.8322
2024	SRMCL	88.66	0.8470
2024	HTNet	81.24	0.8131
2024	MAL	81.83	0.8293
2024	MCFMCN	81.38	0.8301
2024	SPCL-MER	85.71	0.7794
2024	DSN-CASC	68.97	0.6782
2024	MER-3DCNNST	91.48	0.9140
2024	MER-DA3DCNNST	91.88	0.9152
2024	MER-SCB3DCNNST	92.20	0.9167
2024	MER-DACWB3DCNNST	92.40	0.9180
2024	DP-3DNet	92.76	0.9008
2024	HA-3DNet	93.61	0.9113
2024	SE-3DNet	93.35	0.9120
2024	HA-3DSENet	93.44	0.9138
<b>2024</b>	<b>TP-HA3DNet</b>	<b>95.00</b>	<b>0.9502</b>

Tabel 6 Perbandingan Akurasi dan *F1-Score* dari Metode yang diusulkan dengan Penelitian Disertasi dan *State of the art* pada *Dataset* CASME II

Tahun	Metode	Akurasi (%)	F1-Score
2022	BDCN	95.16	0.9501
2022	DSCNN	77.95	0.7809
2022	MMNet	95.51	0.9494
2022	ODCNN	86.67	0.8989
2022	BALANCE	97.82	0.9721
2022	ML-STFN	96.25	0.9601
2023	ADMME	92.37	0.9210
2023	AGAN-TF	92.21	0.9024
2023	DS3DCNN	80.61	0.7887
2023	STCN	74.14	0.7489
2023	MBERT	89.14	0.9034

Tahun	Metode	Akurasi (%)	F1-Score
2023	MADFN	89.86	0.9061
2024	SRMCL	96.49	0.9635
2024	MMAM	96.79	0.9578
2024	HTNet	95.16	0.9532
2024	MAL	93.48	0.9049
2024	MCFMCN	94.97	0.9328
2024	SPCL-MER	94.52	0.9312
2024	DSN-CASC	89.95	0.9074
2024	FAUD-STF	89.97	0.9080
2024	MER-3DCNNST	91.30	0.9118
2024	MER-DA3DCNNST	92.67	0.9259
2024	MER-SCB3DCNNST	93.25	0.9354
2024	MER-DACWB3DCNNST	93.66	0.9369
2024	DP-3DNet	93.35	0.9115
2024	HA-3DNet	93.79	0.9203
2024	SE-3DNet	93.38	0.9112
2024	HA-3DSENet	93.52	0.9125
<b>2024</b>	<b>TP-HA3DNet</b>	<b>98.87</b>	<b>0.9887</b>

Tabel 7. Perbandingan Akurasi dan *F1-Score* dari Metode yang diusulkan dengan Penelitian Disertasi dan *State of the art* pada *Dataset CAS(ME)*<sup>3</sup>

Tahun	Metode	Akurasi (%)	F1-Score
2023	MBERT	61.25	0.5604
2024	HTNet	54.15	0.5767
<b>2024</b>	<b>TP-HA3DNet</b>	<b>98.21</b>	<b>0.9821</b>

## 5. Kesimpulan atau Kontribusi Ilmiah

Kesimpulan yang dapat diambil dari hasil eksperimen dan pengujian dalam penelitian Disertasi ini adalah:

- a. Penelitian disertasi ini berhasil mengusulkan *pipeline* berbasis *computer vision* dan *deep learning* yang efektif dalam meningkatkan akurasi pengenalan ekspresi mikro spontan pada urutan video panjang. Meskipun *pipeline* ini menggambarkan proses *end-to-end*, kontribusi utama penelitian ini terletak pada teknik-teknik inovatif yang diterapkan dalam setiap tahapan *pipeline*, bukan hanya pada struktur *pipeline* itu sendiri. Setiap tahapan, mulai dari *data preparation*, *image preprocessing*, *data preprocessing*, hingga klasifikasi, mengintegrasikan teknik-teknik yang bermanfaat seperti penggunaan model jaringan konvolusi *3D Triple Path* dan *Hybrid Attention* untuk mengatasi tantangan utama dalam pengenalan ekspresi mikro spontan, seperti keterbatasan *dataset*, ketidakseimbangan distribusi kelas, dan penanganan informasi penting dalam dimensi spasial dan temporal.
- b. *Pipeline* ini berhasil mengatasi masalah dalam *dataset* melalui tahapan *image preprocessing* berjenjang, *data preprocessing* yang meliputi augmentasi data,

dan teknik *class balancing*. Teknik-teknik ini secara signifikan meningkatkan hasil pengenalan ekspresi mikro spontan, terutama pada kelas minoritas. Kontribusi penelitian ini juga terletak pada penerapan teknik-teknik ini dalam *pipeline* yang terintegrasi, yang memungkinkan model untuk mengenali ekspresi mikro spontan dengan lebih baik.

- c. Keunggulan utama dari *pipeline* yang diusulkan terletak pada kemampuannya untuk menangkap pola spasial-temporal secara efektif dalam berbagai area wajah, seperti area sekitar mata, alis, dan mulut, yang sangat penting dalam mengenali ekspresi mikro spontan dengan lebih tepat. Kontribusi tersebut terletak pada penerapan *Hybrid Attention* dalam tahapan klasifikasi, yang memungkinkan model untuk fokus pada area wajah yang mengandung informasi emosi paling relevan. Teknik ini meningkatkan akurasi secara signifikan dibandingkan dengan penelitian *state of the art*.

## 6. Tindak Lanjut

Kontribusi ilmiah dari penelitian ini membuka cakrawala keilmuan baru yang berharga untuk penelitian di masa depan, khususnya dalam pengenalan ekspresi mikro dengan teknologi *computer vision*. Keberhasilan integrasi *pipeline* berbasis *deep learning*, seperti *3D Triple Path* dan *Hybrid Attention*, memberikan dasar kuat untuk pengembangan aplikasi lebih lanjut. Penelitian ini berpotensi merambah ke bidang lain, seperti psikologi, keamanan, kedokteran atau interaksi manusia-komputer, dengan memanfaatkan teknologi serupa. Pengembangan lebih lanjut dapat meningkatkan akurasi dan membuka peluang aplikasi di berbagai disiplin ilmu.

## Riwayat Hidup

Budhi Irawan meraih gelar Sarjana Strata Satu di bidang Matematika sub minat Ilmu Komputer dari Universitas Padjadjaran pada tahun 1999, dan melanjutkan studinya ke jenjang pascasarjana dengan meraih gelar Magister Teknik dari Prodi Magister Teknik Komputer STEI ITB pada tahun 2006. Saat ini, Budhi aktif sebagai Dosen Tetap di Program Studi Teknik Komputer Fakultas Teknik Elektro di Universitas Telkom.

## Daftar Publikasi Terkait Penelitian

- Irawan, Budhi, & Munir, Rinaldi & Utama, Nugraha Priya & Purwarianti Ayu. (2024). *Improving the Accuracy of Facial Micro-Expression Recognition: Spatio-Temporal Deep Learning with Enhanced Data Augmentation and Class Balancing*, *Interdisciplinary Journal of Information, Knowledge and Management*, DOI: 10.28945/5386.
- Irawan, Budhi, & Munir, Rinaldi & Utama, Nugraha Priya & Purwarianti Ayu. (2025). *Enhancing Micro-Expression Recognition: A Novel Approach with Hybrid Attention 3Dnet*, *Jordanian Journal of Computers and Information Technology* (JJCIT), DOI: 10.5455/jcit.71-1725518992

- Irawan, Budhi, & Munir, Rinaldi & Utama, Nugraha Priya & Purwarianti Ayu. (2023). *Spontaneous Micro-expression Recognition using 3DCNN on Long Videos for Emotion Analysis*. 2023 10th International Conference on Advanced Informatics: COncep, Theory and Application (ICAICTA) 979-8-3503-2991-9/23/S31.00 (c) 2023 IEEE, DOI: 10.1109/ICAICTA59291. 2023.103390247.

### **Ucapan Terima Kasih**

Penulis mengucapkan puji syukur ke hadirat Allah SWT atas segala kekuatan dan kemudahan dalam menyelesaikan disertasi ini. Rasa terima kasih yang mendalam disampaikan kepada almarhum Prof. Ir. Dwi Hendratmo Widyantoro, M.Sc., Ph.D., atas inspirasi dan motivasinya, serta kepada Dr. Ir. Rinaldi Munir, M.T., selaku Promotor, atas bimbingan dan masukan yang sangat berharga. Ucapan terima kasih juga ditujukan kepada co-promotor, Nugraha Priya Utama, S.T., M.A., Ph.D., dan Dr. Eng. Ayu Purwarianti, S.T., M.T., atas dukungan dan arahan selama penelitian. Penulis menghargai bantuan dari Dr. Umar Khayam, S.T., M.T., selaku Ketua Program Studi S3 Teknik Elektro dan Informatika STEI ITB, serta Ibu Nurhayati dan seluruh sivitas akademik ITB, khususnya STEI. Penghargaan khusus diberikan kepada Universitas Telkom atas dukungan pendanaan studi dan penelitian. Terakhir, penulis menyampaikan terima kasih yang sebesar-besarnya kepada keluarga tercinta, khususnya almarhum Bapak dan Ibu, istri, anak-anak, serta rekan-rekan seperjuangan, atas segala doa, cinta, motivasi, dan dukungan yang tulus dalam suka dan duka selama proses penyelesaian disertasi ini. Semoga Allah SWT senantiasa melimpahkan keberkahan kepada kita semua.