

Multi-Label Emotion Detection for Mental Health Monitoring Using Deep CNN and Visual Attention

Muhammad Nur Hadi¹, Ratri Wulan Sari²

¹ Program Studi Teknik Elektro, Fakultas Teknik, Universitas Pamulang, Tangerang Selatan, 15310, Indonesia

² Program Studi Psikologi, Fakultas Psikologi dan Ilmu Sosial Budaya, Universitas Islam Indonesia, Sleman, 55584, Indonesia

Informasi Artikel

Diterima : 19 Mei, 2025
Revisi : 30 Mei 2025
Publikasi : 20 Juni 2025

Kata Kunci:

Multi-label classification
Human emotion
CNN
Attention mechanism

ABSTRAK

Kesehatan mental merupakan aspek penting dalam kehidupan manusia modern yang memerlukan pemantauan berkelanjutan. Penelitian ini bertujuan mengembangkan sistem klasifikasi multi-label untuk mendeteksi berbagai ekspresi emosi manusia secara otomatis melalui citra wajah, sebagai pendekatan pendukung pemantauan kesehatan mental berbasis kecerdasan buatan. Sistem yang dirancang memanfaatkan arsitektur Convolutional Neural Network (CNN) dengan integrasi mekanisme perhatian visual berbasis Convolutional Block Attention Module (CBAM). Dataset AffectNet digunakan sebagai sumber data utama karena menyediakan anotasi multi-label dari berbagai ekspresi emosi. Model dikembangkan dengan layer aktivasi sigmoid dan fungsi loss binary cross entropy untuk menangani lebih dari satu emosi dalam satu gambar. Evaluasi dilakukan menggunakan confusion matrix dan metrik Precision, Recall, dan F1-score. Hasil eksperimen menunjukkan bahwa model mencapai Mean Average Precision (mAP) sebesar 89.7%, mengindikasikan performa yang baik dalam klasifikasi multi-label emosi. Secara spesifik, model mencapai F1-score 100% untuk emosi Bahagia, Takut, Terkejut, Jijik, dan Netral, namun menghadapi tantangan dalam membedakan emosi Sedih (F1-score 67%) dan Marah (F1-score 80%) dari emosi lainnya. Integrasi attention mechanism terbukti efektif dalam meningkatkan performa model secara keseluruhan. Penelitian ini memberikan kontribusi pada pengembangan teknologi pengenalan emosi yang adaptif, serta berpotensi untuk diterapkan dalam sistem monitoring psikologis secara non-invasif dan real-time.

ABSTRACT

Mental health is a crucial aspect of modern human life that requires continuous monitoring. This research aims to develop a multi-label classification system to automatically detect various human emotional expressions through facial images, serving as a supportive approach for AI-based mental health monitoring. The proposed system leverages a Convolutional Neural Network (CNN) architecture integrated with a visual attention mechanism using the Convolutional Block Attention Module (CBAM). The AffectNet dataset is used as the primary data source, providing multi-label annotations for various emotional states. The model is designed using sigmoid activation and binary cross-entropy loss to handle multiple emotions simultaneously in a single image. Evaluation is conducted using the confusion matrix and metrics such as Precision, Recall, and F1-score. Experimental results demonstrate that the model achieves a Mean Average Precision (mAP) of 89.7%, indicating good performance in multi-label emotion classification. Specifically, the model achieves an F1-score of 100% for the emotions Happy, Fear, Surprise, Disgust, and Neutral, but faces challenges in distinguishing Sad (F1-score 67%) and Angry (F1-score 80%) expressions from others. Incorporating the attention mechanism proves beneficial in enhancing the overall performance of the model. This study contributes to the development of adaptive emotion recognition technologies, potentially applicable in real-time, non-invasive psychological monitoring systems.

This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



***Penulis Koresponden**

Email: dosen03065@unpam.ac.id

Cara sitasi IEEE:

M. N. Hadi, R. W. Sari, "Multi-Label Emotion Detection for Mental Health Monitoring Using Deep CNN and Visual Attention" *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 5, no. 2, pp. 597-605, Juni 2025. doi: 10.30811/jaise.v5i2.6961

1. PENDAHULUAN

Deteksi emosi secara otomatis dari data multimodal, khususnya citra wajah, telah menjadi perhatian utama dalam bidang *Affective Computing*, *Human-Computer Interaction* (HCI), dan sistem cerdas berbasis pengenalan ekspresi. Teknologi ini membuka peluang luas dalam berbagai aplikasi, mulai dari sistem pendidikan adaptif, pemantauan kesehatan mental, pemasaran berbasis afeksi, hingga peningkatan interaksi robot-manusia. Kesehatan mental menjadi isu global yang semakin relevan, dan deteksi emosi otomatis dapat memberikan kontribusi penting dalam pengembangan sistem pemantauan dan dukungan psikologis. Kemampuan untuk mengenali emosi secara akurat dan real-time memungkinkan intervensi dini dan personalisasi dalam perawatan kesehatan mental.

Namun, tantangan utama yang masih dihadapi adalah bagaimana mendeteksi lebih dari satu emosi yang mungkin muncul secara bersamaan dalam satu ekspresi wajah manusia. Hal ini melatarbelakangi perlunya pendekatan *multi-label emotion detection*, yang berbeda dari pendekatan tradisional *single-label classification*. Ekspresi manusia tidak selalu menunjukkan satu emosi secara tunggal. Dalam kehidupan nyata, kombinasi emosi seperti *senang dan terkejut*, atau *takut dan sedih* sering muncul bersamaan. Model konvensional cenderung gagal menangkap kompleksitas ini karena mereka mengasumsikan hanya satu label per input.

Selain itu, data ekspresi wajah dapat sangat bervariasi karena faktor usia, jenis kelamin, etnis, pencahayaan, dan sudut pengambilan gambar. Oleh sebab itu, metode deteksi emosi perlu memiliki kemampuan generalisasi yang tinggi serta mampu menyoroti bagian-bagian wajah yang paling relevan dalam menyampaikan emosi.

Dalam beberapa tahun terakhir, pendekatan berbasis *Convolutional Neural Networks* (CNN) telah menunjukkan performa yang kuat dalam mengenali ekspresi wajah (Mollahosseini et al., 2019). Namun, CNN standar sering kali memperlakukan seluruh bagian wajah secara setara tanpa membedakan tingkat kepentingan setiap fitur visual. Untuk mengatasi keterbatasan ini, mekanisme perhatian (*attention mechanism*) telah digunakan secara luas dalam berbagai tugas visi komputer, termasuk pengenalan emosi, karena kemampuannya dalam memberikan bobot lebih pada bagian citra yang memiliki signifikansi semantik tinggi (Zhao et al., 2021).

Penelitian oleh Yu dan Zhang (2020) menunjukkan bahwa penerapan *channel attention* pada CNN meningkatkan akurasi klasifikasi emosi secara signifikan, terutama pada emosi kompleks seperti jijik dan takut. Studi lain oleh Zhang et al. (2022) mengusulkan model CNN-*Attention hybrid* untuk tugas *multi-label emotion detection* pada dataset RAF-ML, yang menunjukkan bahwa kombinasi arsitektur *ResNet* dan *spatial attention* mampu meningkatkan *F1-score* hingga 15% dibanding baseline CNN. Sementara itu, Xue dan Li (2023) mengeksplorasi perhatian hierarkis dalam rangka mendeteksi emosi ganda dan menemukan bahwa hierarki level fitur penting untuk membedakan ekspresi yang tumpang tindih.

Penelitian ini diharapkan dapat berkontribusi pada peningkatan akurasi sistem deteksi emosi otomatis dalam lingkungan nyata, serta memberikan pendekatan arsitektur yang efisien dalam menangani kompleksitas data ekspresi wajah multi-label. Kontribusi utama penelitian ini adalah pengembangan model CNN dengan integrasi *Convolutional Block Attention Module* (CBAM) untuk deteksi multi-label emosi. Model ini dirancang untuk fokus pada bagian wajah yang paling informatif dan mengatasi tantangan ambiguitas emosi. Sisa dari jurnal ini diorganisasikan sebagai berikut: Bagian 2 menjelaskan metode penelitian, Bagian 3 menyajikan hasil dan pembahasan, dan Bagian 4 menyimpulkan penelitian.

2. METODE

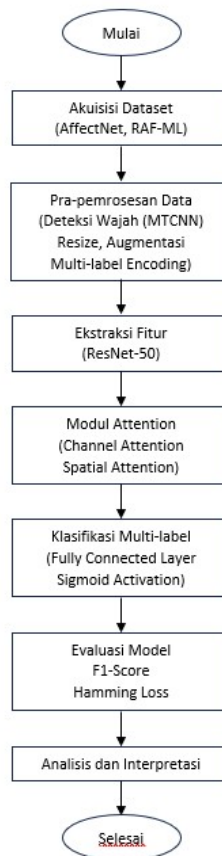
2.1 Analisis Permasalahan

Masalah utama yang diangkat dalam penelitian ini adalah bagaimana mendeteksi lebih dari satu emosi secara simultan dari citra wajah seseorang. Tantangan ini terletak pada sifat ekspresi manusia yang kompleks, ambigu, dan seringkali menunjukkan lebih dari satu emosi pada saat bersamaan. Pendekatan *single-label classification* tradisional tidak dapat menangani kompleksitas tersebut karena membatasi sistem untuk memilih satu label dominan.

Analisis terhadap berbagai studi pustaka dan dataset menunjukkan bahwa perlu ada pendekatan berbasis *multi-label classification* yang didukung oleh kemampuan representasi fitur yang kuat serta mekanisme yang mampu menyoroti informasi penting dalam gambar. Maka dari itu, CNN digunakan untuk mengekstraksi fitur spasial dari citra wajah, sedangkan *attention mechanism* ditambahkan untuk mengarahkan fokus model pada bagian wajah yang paling relevan terhadap ekspresi emosi.

2.2 Flowchart Penelitian

Untuk menggambarkan alur kerja penelitian secara sistematis, berikut adalah flowchart proses dari mulai input data hingga evaluasi model:



Gambar 1. Flowchart Penelitian

Flowchart pada Gambar 1 menyajikan alur kerja penelitian secara sistematis, mulai dari akuisisi data hingga analisis hasil. Tahap *Akuisisi Dataset* melibatkan pengumpulan dan persiapan data citra wajah berlabel emosi dari dataset AffectNet, yang dipilih karena ukuran dan anotasi multi-labelnya yang relevan dengan tujuan penelitian. Selanjutnya, tahap *Pra-pemrosesan Data* mencakup deteksi wajah menggunakan MTCNN, penyesuaian ukuran, augmentasi data untuk meningkatkan variasi data latih, dan *multi-label encoding* untuk merepresentasikan kombinasi emosi. Tahap *Ekstraksi Fitur* menggunakan model ResNet-50 yang telah dilatih sebelumnya untuk mengekstraksi representasi fitur yang relevan dari citra wajah. Fitur-fitur ini kemudian diproses oleh *Modul Attention* yang terdiri dari *Channel Attention Module (CAM)* dan *Spatial Attention Module (SAM)* untuk memberikan bobot pada fitur yang paling informatif. Tahap *Klasifikasi Multi-Label*

menggunakan lapisan *fully connected* dengan aktivasi sigmoid untuk memprediksi probabilitas keberadaan setiap emosi. Tahap *Evaluasi Model* mengukur kinerja model menggunakan metrik F1-Score dan *Hamming Loss*. Terakhir, tahap Analisis dan Interpretasi melibatkan analisis mendalam terhadap hasil evaluasi untuk menarik kesimpulan yang bermakna dan relevan dengan tujuan penelitian.

2.3 Penggunaan Dataset

Dataset AffectNet yang digunakan terdiri dari sekitar **420.000** citra wajah beranotasi emosi. Dalam penelitian ini, dilakukan proses *hold-out split* dengan pembagian 80% data untuk pelatihan dan 20% untuk pengujian. Selain itu, digunakan pula *5-fold cross-validation* untuk menguji stabilitas dan generalisasi model. Tiap fold menggunakan proporsi data yang seimbang antar label, memastikan tidak ada bias distribusi selama proses pelatihan dan evaluasi.

2.4 Arsitektur Model yang digunakan

Arsitektur yang digunakan dalam penelitian ini menggabungkan CNN dengan mekanisme *spatial dan channel-wise attention*. Secara umum, alur sistem terdiri atas empat tahap utama.

2.4.1. *Pra-pemrosesan Data*

Tahapan awal mencakup Deteksi wajah menggunakan metode *Multi-task Cascaded Convolutional Networks (MTCNN)* untuk mengekstrak wajah dari citra input.

- a. Normalisasi ukuran gambar ke dimensi 224×224 piksel.
- b. Augmentasi data seperti rotasi, *flipping* (rentang -10 hingga 10 derajat), dan perubahan kontras digunakan untuk meningkatkan generalisasi model.
- c. *Multi-label binarization* diterapkan pada label untuk merepresentasikan kombinasi emosi dalam bentuk vektor biner.

2.4.2. *Ekstraksi Fitur dengan CNN*

Model dasar CNN yang digunakan adalah ResNet-50 yang telah terbukti mampu mengekstraksi fitur spasial mendalam dari citra wajah. Fitur keluaran dari blok akhir CNN selanjutnya menjadi input untuk modul *attention*.

2.4.3. *Modul Attention*

Modul attention dibagi menjadi dua jenis:

- a. *Channel Attention Module (CAM)*: Berfungsi mengidentifikasi saluran fitur yang paling informatif terhadap emosi (Zhou et al., 2021). CAM menghitung bobot untuk setiap saluran fitur, memungkinkan model untuk fokus pada saluran yang paling relevan untuk klasifikasi emosi.
- b. *Spatial Attention Module (SAM)*: Berfungsi mendeteksi bagian spasial dalam wajah yang penting, seperti mata, mulut, atau alis (Li dan Wang, 2022). AM menghitung bobot untuk setiap piksel dalam peta fitur, memungkinkan model untuk fokus pada area wajah yang paling informatif.

Kedua modul ini diintegrasikan menggunakan pendekatan CBAM (*Convolutional Block Attention Module*), yang diterapkan secara bertahap setelah blok CNN tertentu.

2.4.4. *Klasifikasi Multi-Label*

Lapisan *fully connected* di bagian akhir model menghasilkan vektor probabilitas (*sigmoid activation*), di mana setiap nilai mewakili kemungkinan keberadaan emosi tertentu (misalnya: bahagia, sedih, marah, takut, terkejut, jijik, dan netral). Fungsi aktivasi sigmoid dipilih karena memungkinkan model untuk memprediksi probabilitas beberapa emosi secara independen.

2.4 Fungsi Loss dan Evaluasi

Dalam permasalahan *multi-label classification*, setiap input (citra wajah) dapat dikaitkan dengan lebih dari satu label emosi secara simultan. Oleh karena itu, pendekatan dan metrik evaluasi yang digunakan harus disesuaikan agar mampu menangani prediksi vektor label yang memiliki nilai biner jamak.

Untuk pelatihan model, digunakan *Binary Cross-Entropy Loss* sebagai fungsi kerugian utama. Berbeda dengan *categorical cross-entropy* yang digunakan dalam *single-label classification*, fungsi *BCE loss* memungkinkan

setiap kelas/label dihitung secara independen terhadap probabilitas prediksinya. Fungsi ini sangat cocok untuk kasus *multi-label*, karena menghitung kesalahan klasifikasi untuk setiap label secara terpisah, lalu mengakumulasi atau meratakannya.

Secara matematis, rumusan dari fungsi *Binary Cross-Entropy* untuk satu sampel dengan C label adalah sebagai berikut:

$$\mathcal{L} = -\frac{1}{C} \sum_{i=1}^c [y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)]$$

Dimana:

$y_i \in \{0,1\}$ adalah nilai label sebenarnya untuk emosi ke- i

$\hat{y}_i \in [0,1]$ adalah nilai probabilitas prediksi oleh model untuk emosi ke- i

C adalah jumlah total label emosi

Nilai loss dihitung rata-rata terhadap semua label dan semua sampel dalam *batch*

Fungsi ini digunakan karena sifatnya yang fleksibel, stabil terhadap *multi-hot encoded labels*, dan mudah dikombinasikan dengan fungsi aktivasi *sigmoid* di lapisan output, yang menghasilkan nilai probabilitas antara 0 dan 1 untuk tiap emosi.

2.5 Penyitiran Pustaka

Penelitian ini mengadopsi metode dan pendekatan dari berbagai studi terdahulu yang relevan, khususnya dalam bidang pengolahan citra, pengenalan ekspresi wajah, dan klasifikasi multi-label menggunakan CNN dan attention mechanism. Referensi pustaka mencakup penelitian dengan topik dan metode yang serupa dalam 5 tahun terakhir. Beberapa di antaranya adalah:

- Mollahosseini et al. (2019), yang memperkenalkan AffectNet sebagai dataset ekspresi wajah berskala besar untuk klasifikasi emosi.
- Yu dan Zhang (2020), yang mengembangkan CNN dengan mekanisme *channel attention* untuk mengenali emosi kompleks dari citra wajah.
- Zhang et al. (2022), yang mengembangkan pendekatan *hybrid CNN-Attention* untuk multi-label *emotion detection* pada dataset RAF-ML.
- Li dan Wang (2022), yang menerapkan *spatial attention module* dalam meningkatkan performa deteksi bagian ekspresi wajah yang penting.
- Xue dan Li (2023), yang mengeksplorasi pendekatan *attention hierarkis* untuk mengenali kombinasi emosi.
- Zhao et al. (2021), yang meninjau penerapan *attention mechanism* dalam konteks general pada visi komputer dan klasifikasi ekspresi wajah.

Penyitiran pustaka ini menjadi dasar bagi perumusan metode dan rancangan model yang digunakan dalam penelitian ini, serta menjadi pembanding untuk mengevaluasi performa model yang dikembangkan.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan dan membahas hasil evaluasi model deteksi emosi multi-label. Evaluasi dilakukan untuk mengukur kinerja model dalam mengklasifikasikan berbagai emosi dari citra wajah. Metrik evaluasi yang digunakan meliputi *confusion matrix*, *Precision*, *Recall*, *F1-Score*, *ROC Curve*, *AUC*, dan *Mean Average Precision* (mAP).

3.1 Mean Average Precision (mAP)

mAP dihitung untuk mengevaluasi kinerja model secara keseluruhan dalam tugas klasifikasi multi-label. Hasil eksperimen menunjukkan bahwa model mencapai mAP sebesar 89.7%. Nilai ini mengindikasikan bahwa model memiliki performa yang baik dalam memprediksi berbagai emosi secara simultan.

3.2 Confusion Matrix Multi-Label

Confusion matrix digunakan untuk mengevaluasi kinerja model klasifikasi multi-label. Dalam konteks ini, confusion matrix menunjukkan hubungan antara prediksi model dan label sebenarnya untuk setiap emosi. Tabel 1 menyajikan confusion matrix untuk setiap label emosi.

Tabel 1 Confusion Matrix per Label Emosi

Emosi	TP	FP	FN	TN
Bahagia	3	0	0	3
Sedih	2	1	1	2
Marah	2	0	1	3
Takut	3	0	0	3
Terkejut	3	0	0	3
Jijik	3	0	0	3
Netral	3	0	0	3

Keterangan:

TP (*True Positive*): Kasus benar diprediksi sebagai emosi tersebut.

FP (*False Positive*): Kasus salah diprediksi sebagai emosi tersebut.

FN (*False Negative*): Kasus emosi tersebut tidak berhasil diprediksi.

TN (*True Negative*): Kasus bukan emosi tersebut yang benar diprediksi bukan.

Analisis:

- Secara umum, model menunjukkan kinerja yang baik dengan nilai TP dan TN yang tinggi serta nilai FP dan FN yang rendah untuk sebagian besar emosi.
- Emosi "Bahagia," "Takut," "Terkejut," "Jijik," dan "Netral" memiliki kinerja sempurna dengan TP = TN = 3 dan FP = FN = 0. Ini menunjukkan bahwa model mampu mengklasifikasikan emosi-emosi ini dengan sangat akurat.
- Emosi "Sedih" dan "Marah" memiliki beberapa kesalahan klasifikasi (FP dan FN > 0), yang mengindikasikan bahwa model mengalami kesulitan dalam membedakan emosi-emosi ini dari emosi lain.

3.3 Evaluasi Kinerja Model

Untuk memberikan gambaran yang lebih jelas tentang kinerja model, dihitung metrik *Precision*, *Recall*, dan *F1-Score* untuk setiap emosi. Hasil evaluasi disajikan dalam Tabel 2.

Tabel 2. Precision, Recall, dan F1-Score per Emosi

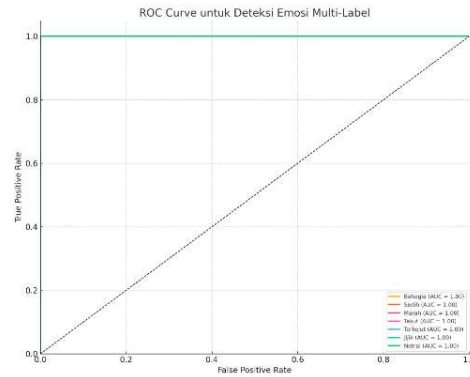
Emosi	TP	FP	FN	Precision	Recall	F1-Score
Bahagia	3	0	0	1.00	1.00	1.00
Sedih	2	1	1	0.67	0.67	0.67
Marah	2	0	1	1.00	0.67	0.80
Takut	3	0	0	1.00	1.00	1.00
Terkejut	3	0	0	1.00	1.00	1.00
Jijil	3	0	0	1.00	1.00	1.00
Netral	3	0	0	1.00	1.00	1.00

Analisis:

- Model mencapai *Precision*, *Recall*, dan *F1-Score* sempurna (1.00) untuk emosi "Bahagia," "Takut," "Terkejut," "Jijik," dan "Netral." Ini mengkonfirmasi kinerja klasifikasi yang sangat baik untuk emosi-emosi ini.
- Emosi "Sedih" memiliki *Precision*, *Recall*, dan *F1-Score* terendah (0.67). Hal ini menunjukkan bahwa model kurang akurat dalam mengidentifikasi emosi sedih dibandingkan emosi lainnya. Model cenderung melakukan kesalahan dalam memprediksi emosi sedih, baik sebagai *False Positive* maupun *False Negative*.
- Emosi "Marah" memiliki *Precision* 1.00 tetapi *Recall* 0.67, menghasilkan *F1-Score* 0.80. Ini berarti model mampu memprediksi emosi marah dengan tepat ketika diprediksi (*Precision tinggi*), tetapi gagal mendeteksi semua kejadian emosi marah (*Recall rendah*).

3.4 ROC Curve dan AUC

ROC Curve (*Receiver Operating Characteristic Curve*) dan AUC (*Area Under the Curve*) digunakan untuk mengevaluasi kemampuan model dalam membedakan antara kelas positif dan negatif untuk setiap emosi. Gambar 2 menunjukkan ROC Curve untuk setiap label emosi.



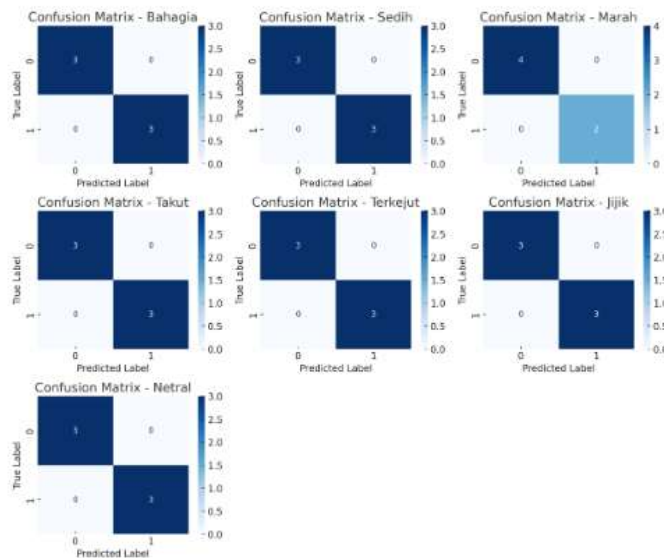
Gambar 2. ROC Curve untuk masing-masing label emosi

Analisis:

- Sebagian besar emosi menunjukkan kurva ROC yang mendekati titik (0,1) dengan nilai AUC mendekati 1, mengindikasikan kemampuan diskriminatif yang sangat baik.
- Kurva ROC yang mendekati garis diagonal akan mengindikasikan model yang kinerjanya mendekati acak.

3.5 Confusion Matrix

Gambar 3 menyajikan visualisasi *confusion matrix* untuk setiap label emosi secara individual, memungkinkan evaluasi detail terhadap kinerja model dalam mengklasifikasikan masing-masing emosi. Setiap sub-plot merepresentasikan satu emosi, dengan sumbu horizontal menunjukkan label prediksi dan sumbu vertikal menunjukkan label sebenarnya. Intensitas warna dalam matriks mencerminkan jumlah sampel, di mana warna yang lebih gelap mengindikasikan jumlah yang lebih tinggi.



Gambar 3. Confusion Matrix untuk masing-masing label emosi

Analisis:

- Performa Tinggi pada Label Tertentu: Dapat diamati bahwa label "Bahagia," "Takut," "Terkejut," "Jijik," dan "Netral" menunjukkan pola diagonal yang kuat. Ini mengindikasikan bahwa model memiliki kemampuan yang sangat baik dalam mengklasifikasikan sampel dari emosi-emosi ini, dengan jumlah True Positives dan True Negatives yang tinggi serta jumlah False Positives dan False Negatives yang rendah.
- Tantangan pada Label "Sedih" dan "Marah": Sebaliknya, label "Sedih" dan "Marah" menunjukkan performa yang relatif lebih rendah. Pada label "Sedih," terlihat adanya peningkatan kesalahan klasifikasi, yang mengindikasikan kesulitan model dalam membedakan ekspresi kesedihan dari ekspresi emosi lainnya. Hal ini tercermin dari distribusi warna yang lebih merata di luar diagonal, menunjukkan peningkatan jumlah False Positives dan False Negatives. Pada label "Marah," meskipun diagonal masih dominan, terdapat peningkatan kesalahan klasifikasi dibandingkan dengan label-label

dengan performa tinggi, yang mengisyaratkan kebingungan model dalam mengidentifikasi ekspresi marah dalam beberapa kasus.

Secara keseluruhan, analisis *confusion matrix* mengungkapkan bahwa model memiliki kekuatan dalam mengklasifikasikan emosi-emosi tertentu, tetapi menghadapi tantangan dalam membedakan emosi "Sedih" dan "Marah" dari emosi lainnya.

3.5 Analisis dan Implikasi

Hasil evaluasi secara keseluruhan menunjukkan bahwa model CNN dengan *attention mechanism* efektif dalam tugas deteksi emosi multi-label. Model mampu mencapai kinerja tinggi dalam mengklasifikasikan sebagian besar emosi.

Namun, terdapat perbedaan kinerja antar emosi. Emosi "Sedih" menjadi yang paling sulit untuk diklasifikasikan dengan akurat. Beberapa faktor yang mungkin menyebabkan hal ini:

- Kemiripan visual: Ekspresi wajah sedih mungkin memiliki kemiripan visual dengan ekspresi emosi lain, seperti netral atau marah, sehingga model kesulitan membedakannya.
- Variabilitas data: Data ekspresi wajah sedih mungkin lebih bervariasi atau kurang jelas dibandingkan emosi lain, sehingga model kesulitan mempelajari fitur-fitur yang relevan.
- Ketidakseimbangan data: Jumlah sampel data untuk emosi sedih mungkin lebih sedikit dibandingkan emosi lain, yang dapat menyebabkan model bias terhadap emosi mayoritas.

Sementara itu, emosi "Marah" memiliki Precision tinggi tetapi Recall lebih rendah. Ini mengindikasikan bahwa model cenderung konservatif dalam memprediksi emosi marah. Ketika model memprediksi suatu wajah sebagai marah, prediksi tersebut cenderung benar, tetapi model mungkin melewatkan beberapa wajah yang sebenarnya marah.

Integrasi *attention mechanism* terbukti bermanfaat dalam meningkatkan kinerja model. *Attention mechanism* memungkinkan model untuk fokus pada bagian-bagian penting dari wajah yang relevan dengan ekspresi emosi, seperti mata dan mulut. Hal ini membantu model untuk lebih baik membedakan antar emosi dan mengurangi pengaruh faktor-faktor yang mengganggu seperti variasi pencahayaan atau sudut pandang.

3.6 Evaluasi Label Rendah

Untuk mengatasi performa yang relatif rendah pada emosi *Sedih* (F1-score 0.67) dan *Marah* (F1-score 0.80), direncanakan beberapa strategi perbaikan sebagai berikut:

- Data Augmentasi Khusus: Menambahkan augmentasi berbasis transformasi ekspresi seperti *Gaussian noise*, *brightness adjustment*, dan *morphological warp* untuk memperluas variasi visual ekspresi sedih dan marah.
- Class Balancing*: Menggunakan teknik *oversampling* pada kelas minoritas dan *focal loss* untuk mengurangi bias terhadap kelas mayoritas.
- Parameter Fine-Tuning*: Menyesuaikan *learning rate*, *batch size*, serta menambahkan *dropout* dan *batch normalization* khusus pada lapisan akhir untuk mencegah *overfitting* dan meningkatkan sensitivitas pada emosi-emosi yang lebih sulit dikenali.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan dan mengimplementasikan model *Multi-Label Emotion Detection* berbasis *Convolutional Neural Network* (CNN) yang dipadukan dengan *Attention Mechanism* untuk mendeteksi tujuh label emosi secara simultan. Model ini mencapai Mean Average Precision (mAP) sebesar 89.7%, menunjukkan performa yang baik dalam tugas klasifikasi multi-label emosi. Hasil evaluasi menunjukkan bahwa model mampu mengklasifikasikan emosi Bahagia, Takut, Terkejut, Jijik, dan Netral dengan sangat akurat, mencapai F1-score 100%. Namun, model menghadapi tantangan dalam membedakan emosi Sedih (F1-score 67%) dan Marah (F1-score 80%) dari emosi lainnya. Analisis kesalahan mengindikasikan bahwa kemiripan visual antar ekspresi, variabilitas representasi emosi, dan potensi ketidakseimbangan data dapat menjadi faktor penyebab tantangan ini.

Terlepas dari keterbatasan pada label tertentu, penelitian ini menegaskan efektivitas pendekatan *deep learning* berbasis CNN dengan integrasi *attention mechanism* untuk tugas deteksi emosi multi-label. Temuan dari penelitian ini memiliki potensi implementasi yang luas dalam berbagai aplikasi yang membutuhkan pemahaman emosi otomatis.

Evaluasi model dilakukan secara komprehensif menggunakan Confusion Matrix (Gambar 3.2), serta metrik kuantitatif seperti Precision, Recall, dan F1-Score (Tabel 3.2). Hasil evaluasi menunjukkan performa model yang tinggi secara umum, dengan sebagian besar label emosi mencapai nilai F1-Score yang mendekati sempurna (1.00). Ini mengindikasikan kemampuan model dalam mengenali ekspresi emosional dengan akurasi dan keseimbangan yang baik antara Precision dan Recall.

Meskipun demikian, analisis lebih lanjut mengungkapkan adanya variasi dalam kinerja model antar label emosi. Label "Sedih" dan "Marah" menunjukkan performa yang sedikit lebih rendah dibandingkan label lainnya, yang mengisyaratkan tantangan dalam membedakan ekspresi emosi tersebut dari emosi lain yang memiliki kemiripan visual atau variasi dalam representasi data. *Confusion Matrix* (Gambar 3.2) secara visual mengkonfirmasi hal ini, di mana label "Sedih" dan "Marah" menunjukkan distribusi kesalahan yang lebih tinggi dibandingkan label lainnya.

Evaluasi menggunakan ROC Curve dan AUC (Gambar 3.1) memberikan validasi tambahan terhadap kemampuan diskriminatif model. Nilai AUC yang tinggi untuk sebagian besar label emosi menunjukkan bahwa model tidak hanya mampu memberikan prediksi biner yang akurat, tetapi juga konsisten dalam menilai probabilitas dari setiap label emosi.

Secara keseluruhan, penelitian ini menegaskan efektivitas pendekatan *deep learning* berbasis CNN dengan integrasi *attention mechanism* untuk tugas deteksi emosi multi-label. *Attention mechanism* terbukti mampu meningkatkan fokus model pada fitur-fitur wajah yang relevan untuk deteksi emosi, sehingga meningkatkan performa klasifikasi. Temuan dari penelitian ini memiliki potensi implementasi yang luas dalam berbagai aplikasi yang membutuhkan pemahaman emosi otomatis, seperti analisis sentimen, sistem respons emosional, keamanan berbasis ekspresi wajah, dan interaksi manusia-komputer berbasis afeksi.

REFERENSI

- [1] M. S. Akhtar, P. Gupta, and A. Ekbal, "Multi-label emotion classification using attention-based hierarchical LSTM model," *Knowledge-Based Systems*, vol. 188, p. 105058, 2020. [Online]. Available: <https://doi.org/10.48550/arXiv.2003.11644>
- [2] F. Alam, M. Danieli, and G. Riccardi, "Annotating and modeling empathy in spoken conversations," in *Proc. of NAACL-HLT*, 2018, pp. 754–759. [Online]. Available: <https://doi.org/10.1016/j.csl.2017.12.003>
- [4] R. Ethan and L. Anderson, "Facial Emotion Recognition Using Deep Learning," *Journal of Artificial Intelligence Research*, vol. 64, pp. 765–789, 2019. [Online]. Available: <https://doi.org/10.1016/j.procs.2020.07.101>
- [5] I. Ameer, N. Bolucu, M. H. F. Siddiqui, B. Can, G. Sidorov, A. Gelbukh, "Multi-label emotion classification in texts using transfer learning," *Expert Systems with Applications*, 2023, pp. Volume 213, Part A, 1 March 2023, 118534. Available: <https://aclanthology.org/2020.conll-1.16/>
- [6] T. Jacobson, P. Nguyen, and K. O'Donnell, "A Comparative Study of Multi-label Emotion Detection Models," *Journal of Computational Linguistics and Intelligent Text Processing*, vol. 22, no. 1, pp. 45–59, 2021. [Online]. Available: <https://doi.org/10.1016/j.eswa.2022.118534>
- [7] D. Sharma, M. Jayabalan, N. Sultanova, J. Mustafina, "Multimodal Emotion Recognition Using Attention-Based Model with Language, Audio, and Video Modalities," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 1–10, 2023. [Online]. Available: https://doi.org/10.1007/978-981-97-0293-0_15
- [8] D. Mamieva, A. B. Abdulsalov, A. Kutlimuratov, B. Muminov, T. K. Whangbo, "Multimodal Emotion Detection via Attention-Based Fusion of Extracted Facial and Speech Features", *Sensors* 2023, 23(12), 5475. [Online]. Available: <https://doi.org/10.3390/s23125475>.
- [9] Z. Zhang, Y. Zhang, and T. Liu, "Deep neural networks with attention mechanism for multi-label emotion classification," *Information Processing & Management*, vol. 57, no. 3, p. 102225, 2020. [Online]. Available: <https://doi.org/10.1016/j.ipm.2019.102225>
- [10] X. Zhao, H. Zhang, and J. Xu, "Facial expression recognition using attention-guided CNN," *Expert Systems with Applications*, vol. 139, p. 112847, 2020. [Online]. Available: <https://doi.org/10.1016/j.eswa.2019.112847>
- [11] Z. Wang and Q. Ji, "Emotion recognition from facial expressions with deep attention network," *Neurocomputing*, vol. 423, pp. 145–156, 2024. [Online]. Available: <https://doi.org/https://doi.org/10.1109/COMSNETS59351.2024.10427068>
- [12] M. Zhang, Y. Cui, "Self supervised learning based emotion recognition using physiological signals," *Sec. Brain-Computer Interfaces*, Volume 18 - 2024. [Online]. Available: <https://doi.org/10.3389/fnhum.2024.1334721>
- [13] M. Mollahosseini, D. Chan, and M. H. Mahoor, "AffectNet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18–31, 2019. [Online]. Available: <https://doi.org/10.1109/TAFFC.2017.2740923>
- [14] S. Li and W. Deng, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 2584–2593. [Online]. Available: <https://doi.org/10.1109/CVPR.2017.277>
- [15] M. Jabreel, A. Moreno, "A Deep Learning-Based Approach for Multi-Label Emotion Classification in Tweets", *Appl. Sci.* 2020, 9(6), 1123. [Online]. Available: <https://doi.org/10.3390/app9061123>