Krea-TIF: Jurnal Teknik Informatika
http://ejournal.uika-bogor.ac.id/index.php/krea-tif
p-ISSN: 2338-2910, e-ISSN: 2658-5836
Vol. 12, No. 1, 2024, pp. 12 – 23
DOI: 10.32832/krea-tif.v12i1.16287

Strategi *Ensemble Deep Learning* pada *Global Multi-Scale* dan *Local Attention Features* pada Pengenalan Ekspresi Wajah

Mayanda Mega Santoni*, Nurul Chamidah, Desta Sandya Prasvita
Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta, Indonesia
*E-mail koresponden: megasantoni@upnvj.ac.id

Diserahkan *23 Februari 2024;* ***Direview*** *25 Mei 2024;* ***Dipublikasikan*** *30 Mei 2024*

Abstrak

Social signal processing (SSP) merupakan bidang riset dan teknologi yang bertujuan untuk memberikan kemampuan kepada komputer untuk merasakan dan memahami sinyal sosial manusia, termasuk ekspresi wajah sebagai salah satu sinyal sosial yang kuat dalam komunikasi manusia. Dataset RAF-DB (the Real-world Affective Faces Database). Dataset ini terdiri dari tujuh kelas emosi dasar yakni anger, disgust, fear, happiness, neutral, sadness, dan surprise. Metode Global Multi-Scale and Local Attention Network (MA-Net) merupakan salah satu metode yang digunakan untuk pengenalan ekspresi wajah secara otomatis. Performa metode MA-Net pada dataset RAF-DB menghasilkan akurasi tertinggi dibandingkan penelitian sebelumnya yakni sebesar 88.40%. Walaupun telah menghasilkan akurasi yang cukup tinggi, namun metode MA-Net masih memiliki beberapa keterbatasan dalam memprediksi ekspresi wajah. Metode MA-Net kurang bisa mengenali data gambar yang memiliki masalah noise. Penelitian ini dilakukan untuk mengatasi permasalahan tersebut dengan mengusulkan strategi ensemble deep learning untuk meningkatkan performa dari metode MA-Net. Pada ensemble learning terdapat beberapa jenis fungsi agregasi, yaitu voting dan meta-learning. Hasil temuan dari penelitian ini bahwa penggunaan strategi ensemble learning khususnya pada penggunaan fungsi agregasi meta-learner atau stacking ensemble learning dapat meningkatkan performa evaluasi klasifikasi secara keseluruhan maupun pada masing-masing kelas. Penelitian lanjutan dari hasil ialah dapat mengeksplorasi teknik-teknik machine learning yang lainnya seperti transfer learning untuk meningkatkan akurasi dan generalisasi dalam pengenalan ekspresi wajah.

Kata kunci: *Bagging, Boosting, Deep Learning, Ekspresi Wajah, Ensemble Learning, Stacking*

Abstract

Social signal processing (SSP) is a field of research and technology that aims to give computers the ability to sense and understand human social signals, including facial expressions as one of the most powerful social signals in human communication. RAF-DB dataset (the Real-world Affective Faces Database). This dataset consists of seven basic emotion classes: anger, disgust, fear, happiness, neutral, sadness, and surprise. The Global Multi-Scale and Local Attention Network (MA-Net) method is one of the methods used for automatic facial expression recognition. The performance of the MA-Net method on the RAF-DB dataset produces the highest accuracy compared to previous studies, which is 88.40%. Although it has produced

quite high accuracy, the MA-Net method still has some limitations in predicting facial expressions. The MA-Net method is less able to recognize image data that has noise problems. This research aims to address the issue by proposing an ensemble deep learning strategy to enhance the performance of the MA-Net method. In ensemble learning, there are several strategies namely bagging, boosting, and stacking. The findings of this study are that the use of ensemble learning strategies, especially in the use of meta-learner aggregation functions or stacking ensemble learning, can improve the performance of classification evaluation as a whole and in each class. Further research expected from the results of this study is to explore more advanced machine learning techniques such as transfer learning to improve accuracy and generalization in facial expression recognition.

Keywords: *Bagging, Boosting, Deep Learning, Ensemble Learning, Facial Expression, Stacking*

PENDAHULUAN

Social signal processing (SSP) merupakan sebuah domain riset dan teknologi yang bertujuan untuk memberikan kemampuan kepada komputer untuk merasakan dan memahami sinyal sosial manusia. Di dalam SSP terdapat istilah *social intelligence* yang berarti kecerdasan manusia untuk mengekspresikan dan mengenali *nonverbal social signal* atau komunikasi [1]. Ekspresi wajah adalah salah satu sinyal sosial yang paling kuat dan alami bagi manusia untuk menyampaikan emosi mereka, dan memiliki peran penting dalam komunikasi. Pada bidang *computer vision*, permasalahan pengenalan ekspresi wajah secara otomatis menjadi topik yang hangat dibahas sampai saat ini. Pengenalan ekspresi wajah secara otomatis banyak diterapkan pada berbagai bidang, seperti aplikasi monitoring tingkat kelelahan pengendara mobil yang mana dapat diketahui dari proses pengenalan ekspresi wajah [2]. Selain itu, pengenalan ekspresi wajah juga dapat digunakan untuk prediksi tingkat keterlibatan siswa dalam pembelajaran daring menggunakan fitur ekspresi wajah dan *mouse behaviour* [3], dan lain-lain.

Salah satu tujuan dari pengenalan ekspresi wajah adalah mengenali atau memprediksi sekumpulan gambar atau video ke dalam beberapa jenis emosi dasar. Menurut Ekman [4] terdapat enam jenis emosi dasar yaitu senang, sedih, takut, marah, jijik, dan kaget. Namun, dalam mengenali ekspresi wajah, terdapat beberapa masalah yang dihadapi. Salah satu masalah umum yang sering dihadapi yakni oklusi wajah (*occlusion*) dan variasi pose (*pose variation*) pada wajah. Permasalahan oklusi wajah seperti wajah yang terhalangi oleh aksesoris seperti topi, kacamata, atau benda lainnya yang menutupi wajah. Di awal penelitian untuk menyelesaikan permasalahan oklusi ini yakni dengan melakukan rekonstruksi fitur-fitur geometris atau tekstur yang hilang [5], [6]. Kekurangan dari solusi ini adalah sulit untuk melakukan rekonstruksi daerah oklusi dengan baik pada data real. Sementara itu, untuk permasalahan variasi pose, beberapa metode normalisasi pose di lakukan sebelum proses pengenalan ekspresi wajah. Seperti penelitian Zhang *et al.* [7] yang mengusulkan sebuah metode untuk melatih *single* FER *classifier* dengan beberapa contoh pose. Mempelajari fitur wajah dari berbagai sudut pandang mungkin dapat mencapai kinerja yang lebih baik dalam kondisi *occlusion* dan variasi pose. Hasil ini selaras dengan studi di bidang psikologi menunjukkan bahwa mekanisme persepsi wajah manusia mengekstrak informasi pada wajah baik itu dalam bentuk holistik maupun sebagian [8].

Pada penelitian Zhao *et al.* [9] mengusulkan sebuah metode *deep learning* yakni *Global Multi-Scale and Local Attention Network* (MA-Net). Metode MA-Net terdiri dari tiga komponen utama yakni *feature pre-extractor*, *multi-scale module*, dan *local attention module*. Pada komponen *feature pre-extractor* memiliki keunggulan pada *shallower convolution* yakni konvolusi yang lebih dangkal dibandingkan *deeper convolution*, *rich geometry features* yang

mengacu pada fitur-fitur geometri yang memiliki informasi mengenai struktur dan hubungan spasial antara objek. Dengan fitur-fitur geometri yang kaya, jaringan dapat menjadi lebih efektif dalam mengurangi kerentanan terhadap masalah seperti oklusi atau variasi pose [9].

Sementara itu pada komponen *multi-scale module* dirancang untuk mengekstrak fitur-fitur dengan *receptive field* yang berbeda, sehingga meningkatkan keberagaman dan ketahanan fitur global. Terinspirasi dari Res2Net, yang mengekstrak beberapa fitur *multi-scale* dalam sebuah *single basic block*. Selain itu pada *local attention module* juga digunakan untuk mengekstrak *local salient features*, untuk mengurangi gangguan pada keadaan *occlusion* dan *non-frontal pose*. Ekstraksi fitur lokal dengan memilih wilayah lokal berdasarkan *landmark* wajah atau *cropping* dapat mengakibatkan ketidakselarasan atau ketidakpastian. Oleh karena itu, MA-Net membagi *feature maps* yang telah diekstraksi menjadi beberapa wilayah lokal tanpa tumpang tindih langsung, yang relatif sederhana tetapi efisien [9].

Pada penelitian Zhao *et al.* [9], metode MA-Net diterapkan pada empat *dataset* ekspresi wajah yakni CAER-S, RAF-DB, AffecNet, dan SFEW. Pada penelitian ini akan difokuskan hanya pada performa MA-Net pada *dataset* RAF-DB. Performa metode MA-Net pada *dataset* RAF-DB menghasilkan akurasi sebesar 88.40%, yang lebih tinggi jika dibandingkan penelitian sebelumnya seperti penelitian Li *et al.* [10] dengan akurasi tertinggi 84.22%, penelitian Zeng *et al.* [11] dengan akurasi tertinggi 86.77% dan penelitian Wang *et al.* [12] dengan akurasi tertinggi 88.14%. Meskipun metode MA-Net telah menghasilkan akurasi yang cukup tinggi, namun metode ini masih memiliki beberapa keterbatasan dalam memprediksi ekspresi wajah. Metode MA-Net kurang bisa mengenali data citra yang memiliki masalah *noise*, seperti pada citra emosi wajah yang memiliki citra kabur yang menyebabkan ketidakjelasan ekspresi dan mengakibatkan label yang tidak konsisten dan salah [9]. Oleh karena itu, tujuan dari penelitian ini adalah meningkatkan performa metode MA-Net agar dapat mengenali ekspresi wajah pada *dataset* RAF-DB dengan lebih akurat. Manfaat yang didapatkan dari penelitian ini adalah peningkatan keandalan sistem pengenalan ekspresi wajah pada berbagai bidang aplikasi praktis.

METODE PENELITIAN

Tahapan pada penelitian ini meliputi studi literatur, metode yang diusulkan yakni strategi *ensemble deep learning*, dan implementasi dari metode yang diusulkan, mulai dari pengumpulan data, pembagian data, pembentukan model pelatihan serta pengujian.

Studi literatur

Studi literatur merupakan tahapan penting untuk memperkaya pengetahuan dan landasan teori mengenai *state of the art* penggunaan *dataset* RAF-DB dalam memprediksi ekspresi wajah. Literatur yang digunakan diambil dari tahun 2021-2023. Rangkuman penelitian terdahulu dapat dilihat pada Tabel 1.

Salah satu cara yang dapat dilakukan untuk meningkatkan performa akurasi *deep learning* yakni dengan cara melakukan *ensemble learning* [13], [14], [15]. Proses mengurangi bias dan varian model memainkan peran penting dalam menentukan keberhasilan proses pembelajaran atau pelatihan dalam *machine learning*. Penelitian yang dilakukan Mohammed dan Kora [16] telah menjabarkan bahwa menggabungkan *output* dari berbagai algoritma klasifikasi berpotensi mengurangi kesalahan generalisasi tanpa meningkatkan varians tambahan pada model. Konsep fundamental ini merupakan pendekatan dari *ensemble learning*. Penelitian Chang *et al.*[17] mengetahui *ensemble* model dapat meningkatkan hasil evaluasi dalam deteksi *engagement*. Beberapa penelitian lainnya yang membuktikan bahwa penggunaan *bagging* pada *deep learning* khususnya *Convolutional Neural Network* (CNN), dapat meningkatkan performa klasifikasi. Penelitian yang telah dihasilkan oleh Zhang *et al.* [18], [19] menyampaikan bahwa

bagging deep learning yang diusulkan dapat memprediksi diagnosa penyakit Covid 19 dengan tingkat akurasi sebesar 98.89%. Jika dibandingkan dengan metode tanpa menggunakan *ensemble learning* yakni algoritma ResNet memiliki akurasi sebesar 97.22%. Hal ini menunjukkan bahwa penggunaan *bagging deep learning* dapat meningkatkan hasil akurasi klasifikasi. Begitu juga dengan penelitian Deng *et al.*[19] menyimpulkan bahwa penggunaan strategi *bagging ensemble* dapat memberikan hasil akurasi paling tinggi jika dibandingkan dengan penggunaan *single* model klasifikasi. Model *bagging ensemble* yang diusulkan bersifat *robust* dan meningkatkan akurasi model secara keseluruhan serta sekaligus mengurangi kesalahan klasifikasi pada *single* model.

Selain model *bagging*, terdapat juga jenis *ensemble learning* lainnya yakni *boosting* dan *stacking*. Mohammed *et al.*[24] mengusulkan sebuah *stacking ensemble deep learning* pada prediksi jenis kanker berdasarkan data TCGA yang dapat menghasilkan performa yang lebih baik dibandingkan pada model klasifikasi lainnya. Begitu juga dengan penelitian Ghasemieh *et al.*[25] yang mengusulkan sebuah model pembelajaran mesin menggunakan *stacking ensemble learning* pada prediksi penerimaan kembali pasien penyakit jantung secara darurat.

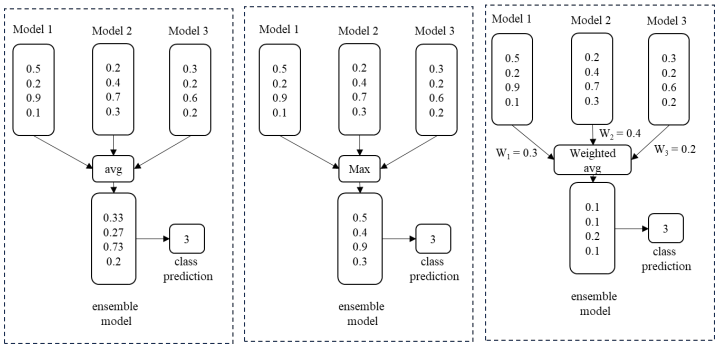
Ensemble Deep Learning

Pada *ensemble learning* menggunakan fungsi agregasi G untuk menggabungkan model $m_1, m_2, \dots, m_h$) dimana h adalah jumlah model dalam *ensemble learning*. Setiap model akan memprediksi *output* yang berbeda $\phi\,(x_i)$ untuk setiap sampel x_i dalam sebuah *dataset* D berukuran n dan fitur berdimensi m, dimana D = $\{(x_i, y_i),\ 1 \leq i \leq n$ dan $x_i \in F^m$. Fungsi agregasi G digunakan untuk menggabungkan prediksi dari model-model tunggal dan menghasilkan prediksi final yang baru untuk setiap sampel x_i. Prediksi *output* dari metode *ensemble* pada *dataset* D dapat dinotasikan pada Persamaan 1 [15].

$$y_i = \phi\,(x_i) = G(m_1, m_2, \dots, m_h) \quad (1)$$

dimana y_i adalah kelas prediksi *output* dari sampel x_i, $\phi\,(x_i)$ adalah *output ensemble learning* dari sampel x_i, G adalah fungsi agregasi, dan $m_1, m_2, \dots, m_h$ adalah model-model tunggal.

Beberapa jenis fungsi agregasi pada *ensemble learning*, yaitu *voting* dan *meta-learning* [16]. Metode *voting* digunakan sebagai fungsi agregasi dalam masalah klasifikasi atau regresi untuk meningkatkan kinerja prediksi pada model. Selain itu, metode *voting* merupakan fungsi agregasi yang umum digunakan untuk metode *bagging* dan *boosting ensemble learning*. Metode *voting* mencakup tiga pendekatan, *averaging voting*, *max-voting*, dan *weighted average voting*. Ilustrasi tiga pendekatan ini dapat dilihat pada Gambar 1.



Gambar 1 Ilustrasi *ensemble learning* menggunakan metode *voting*, kiri untuk *averaging voting*, tengah untuk *max-voting*, kanan untuk *weighted averaging voting*

Jenis fungsi agregasi *ensemble learning* selanjutnya yakni *meta-learning* yang biasa dikenal sebagai "*learning to learn*" dimana pembelajaran dilakukan berdasarkan pembelajaran

sebelumnya. Metode ini digunakan pada fungsi agregasi pada *stacking* (*stacked generalization*) *ensemble learning*. Pada metode ini, *output* dari setiap model tunggal digabungkan dan menjadi *input* baru pada *meta-learner* yang nantinya akan menghasilkan sebuah keluaran baru, seperti yang diilustrasikan pada Gambar 2.

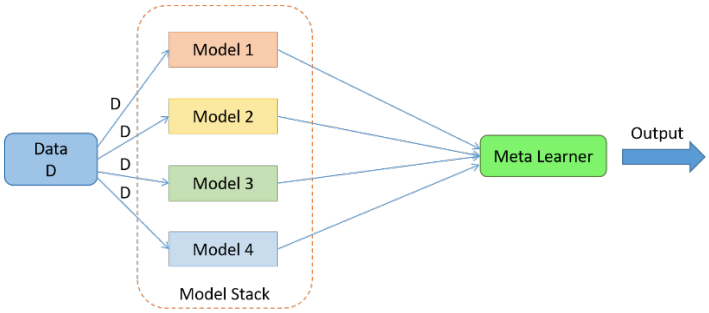
Tabel 1 Penelitian terdahulu yang menggunakan *dataset* RAF-DB

No	Publikasi	*Dataset*, Metode dan arsitektur	Temuan dan keterbatasan penelitian
1	*Learning Deep Global Multi-Scale and Local Attention Features for Facial Expression Recognition in the Wild* [9] Tahun terbit: 2021	***Dataset*:** RAF-DB **Metode:** MA-Net (*a Global Multi-Scale and Local Attention Network*) terdiri dari tiga komponen: • *Feature pre-extractor* • *A multi-scale module* • *Local attention module* **Arsitektur:** • Resnet18 (digunakan sebagai fitur *pre-extractor*) • Modifikasi Resnet18	**Temuan penelitian:** Untuk *dataset* RAF-DB, metode MA-Net mencapai tingkat akurasi yang tinggi sebesar **88.40%** **Keterbatasan penelitian:** Metode MA-Net gagal dalam beberapa kasus tertentu, seperti gambar yang kabur.
2	*Distract Your Attention: Multi-head Cross Attention Network for Facial Expression Recognition* [20] Tahun terbit: 2021	***Dataset*:** RAF-DB **Metode dan arsitektur:** DAN (*Distract Your Attention*) terdiri dari tiga komponen: • FCN (*Fully Clustering Network*): sebagai ekstraksi ciri • MAN (*Multi-head Attention Network*): sebagai *attention map* • AFN (*Attention Fusion Network*): menggabungkan fitur dan memberikan hasil prediksi	**Temuan penelitian:** Untuk *dataset* RAF-DB, metode DAN mencapai tingkat akurasi tertinggi sebesar **89.70%** **Keterbatasan penelitian:** Perlu meningkatkan kinerja model
3	*Patch attention convolutional vision transformer for facial expression recognition with occlusion* [21] Tahun terbit: 2022	***Dataset*:** RAF-DB **Metode:** PACVT (*a Patch Attention Convolutional Vision Transformer*) **Arsitektur:** • Pre-trained ResNet-18 • *Local Feature Extraction* (*Path Attention Unit* - PAU) • *Global Feature Extraction* (*Vision Tranformer* - ViT)	**Temuan penelitian:** Untuk *dataset* RAF-DB, metode PACVT mencapai tingkat akurasi tertinggi sebesar **88.21%** **Keterbatasan penelitian:** Perlu meningkatkan kinerja model
4	*Adaptive Multilayer Perceptual Attention Network for Facial Expression Recognition* [22] Tahun terbit: 2022	***Dataset*:** RAF-DB **Metode:** AMP-NET (*Adaptive Multilayer Perceptual Attention Network*) **Architecture:** ResNet-34 yang terdiri dari 3 *convolutional layer* yang berfungsi sebagai modul *global perception*, *local perception*, dan *attention perception*	**Temuan penelitian:** Untuk *dataset* RAF-DB, metode AMP-NET mencapai tingkat akurasi tertinggi sebesar **89.25%** **Keterbatasan penelitian:** Perlu menyelidiki konstruksi model pengenalan emosi multimodal
5	PAtt-Lite*: Lightweight Patch and Attention MobileNet for Challenging Facial Expression Recognition* [23] Tahun terbit: 2023	***Dataset*:** RAF-DB **Metode:** Patt-Lite *(a lightweight patch and attention network based on MobileNetV1*) **Arsitektur:** • *Truncated* MobileNetV1 • *Patch extraction* • *Attention classifier*	**Temuan penelitian:** Untuk *dataset* RAF-DB, metode Patt-Lite mencapai tingkat akurasi tertinggi sebesar **95.05%** **Keterbatasan penelitian:** meningkatkan kinerja model, terutama pada kelas minoritas (Kelas: *disgust* & *fear*)

Implementasi Ensemble Deep Learning

Metode MA-Net pada penelitian Zhao *et al.* [9] perlu dilakukan peningkatan performa akurasinya, maka pada penelitian ini mengusulkan penggunaan *ensemble learning* pada metode MA-Net. Gambar 3 merupakan ilustrasi dari tahapan implementasi *ensemble learning* pada

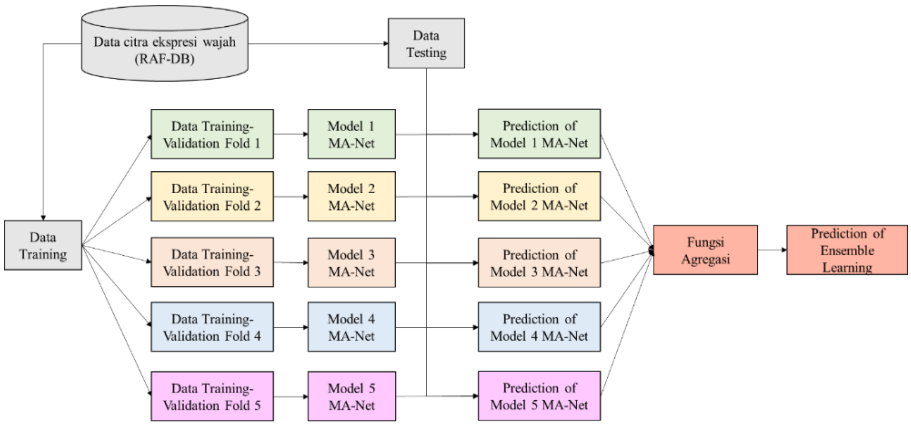
model MA-Net. Data RAF-DB dibagi menjadi data *training* dan data *testing* dengan proporsi 80% : 20%. Performa data secara keseluruhan dapat diperoleh dengan cara data *training* akan dibagi menjadi subset data baru yakni data *training* dan data *validation* dengan pembagian data menggunakan *k-fold cross validation* dengan $k = 5$. Model klasifikasi MA-Net akan dilatih pada ke-5 *fold cross validation* tersebut. Pengujian performa dari masing-masing model *fold cross validation* akan diuji menggunakan data *testing*, sehingga diperoleh prediksi dari setiap masing-masing model individu MA-Net. Selanjutnya untuk meningkatkan performa model MA-Net secara keseluruhan, maka diterapkan fungsi agregasi *voting* dan *meta-learning* untuk mendapatkan hasil prediksi dari *ensemble learning*.



Gambar 2 Ilustrasi *stacking ensemble learning* [16]

Setelah diperoleh model klasifikasi pada proses pelatihan, maka selanjutnya data pengujian diujikan pada model pelatihan. Hasil pengujian tersebut dievaluasi menggunakan nilai evaluasi akurasi (Persamaan 2). Nilai akurasi adalah nilai evaluasi yang menunjukkan perbandingan jumlah data yang terprediksi benar dengan keseluruhan data yang diujikan, dimana TP (*True Positive*) adalah jumlah data terprediksi benar di kelas positif. FP (*False Positive*) adalah jumlah data terprediksi salah di kelas positif. TN (*True Negative*) adalah jumlah data terprediksi benar di kelas negatif. dan FN (*False Negative*) adalah jumlah data terprediksi salah di kelas negatif [26].

$$Akurasi = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (2)$$



Gambar 3 Tahapan implementasi *ensemble learning* pada model MA-Net

HASIL DAN PEMBAHASAN

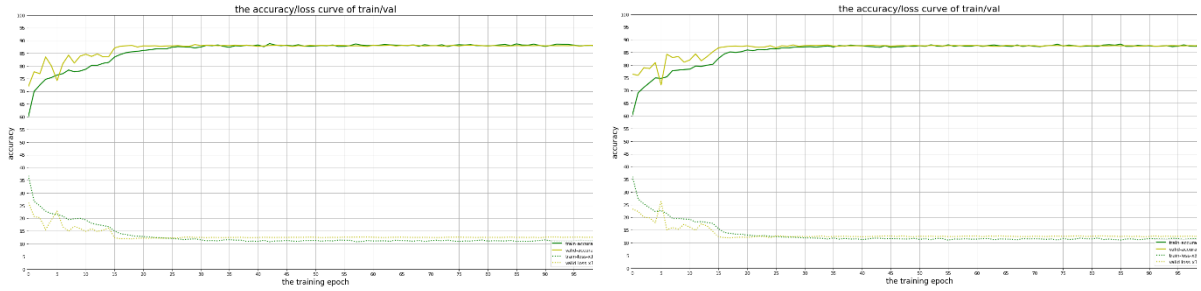
Dataset RAF-DB merupakan singkatan dari *The Real-world Affective Faces Database*. *Dataset* ini terdiri dari tujuh kelas emosi dasar yakni *anger*, *disgust*, *fear*, *happiness*, *neutral*, *sadness*,

dan *surprise*. Data ini memiliki variasi pada usia, jenis kelamin, etnis, pose kepala, kondisi pencahayaan, oklusi (misalnya, kacamata, rambut wajah, atau oklusi diri sendiri), dan lain-lain. Distribusi data latih dan data uji pada *dataset* RAF-DB dapat dilihat pada Tabel 2.

Tabel 2 Distribusi data latih dan data uji pda *dataset* RAF-DB

Emosi	Data pelatihan	Persentase data pelatihan	Data pengujian	Persentase data pengujian
Kelas = 0 (*neutral*)	2524	21%	680	22%
Kelas = 1 (*happines*)	4772	39%	1184	39%
Kelas = 2 (*sadness*)	1982	16%	477	16%
Kelas = 3 (*surprise*)	1290	11%	329	11%
Kelas = 4 (*fear*)	**281**	**2%**	**74**	**2%**
Kelas = 5 (*disgust*)	**717**	**6%**	**160**	**5%**
Kelas = 6 (*anger*)	**705**	**6%**	**162**	**5%**
Total data	12271	100%	3066	100%

Nilai akurasi tertinggi yang diperoleh metode MA-Net menggunakan *dataset* RAF-DB yakni 88.40%. Kode program yang disediakan pada penelitian Zhao *et al.* [9] ini diimplementasikan ulang, diperoleh akurasi terbaiknya sebesar 87.91%. Grafik hubungan akurasi pada setiap *epoch*-nya dan hasil implementasi ulang dapat dilihat pada Gambar 4 dengan hasil yang tidak berbeda dengan hasil terdapat pada Gambar 5. Hasil tersebut menyatakan bahwa *fear* (kelas 4) dan *disgust* (kelas 5) memiliki performa *sensitivity* (*recall*) yang paling kecil yakni 62% dan 64%. Hal ini terjadi karena dua kelas ini termasuk pada kelas minoritas (jumlah data yang sedikit), sehingga model tidak cukup akurat untuk bisa memprediksi kelas emosi ini. Ketidakseimbangan data pada setiap kelas dapat memengaruhi kinerja algoritma *machine learning* [27].



(a). Hasil *original* pada paper rujukan [9]

(b). Hasil implementasi ulang berdasarkan parameter pada paper rujukan

Gambar 4 Grafik hubungan akurasi pada setiap *epoch* yang dituliskan pada paper dan hasil implementasi ulang

	precision	recall	f1-score	support
0	0.83	0.87	0.85	680
1	0.96	0.94	0.95	1185
2	0.85	0.89	0.87	478
3	0.86	0.88	0.87	329
4	0.71	0.62	0.66	74
5	0.78	0.64	0.71	160
6	0.84	0.83	0.83	162
accuracy			0.88	3068
macro avg	0.83	0.81	0.82	3068
weighted avg	0.88	0.88	0.88	3068

Gambar 5 Evaluasi klasifikasi hasil implementasi ulang

Eksperimen selanjutnya adalah membandingkan hasil model-model individual dengan *ensemble learning* dengan berbagai fungsi agregasi. Nilai evaluasi ini dapat meningkat setelah menggunakan pendekatan *ensemble learning*. Secara keseluruhan *ensemble learning* dapat meningkatkan akurasi dari model-model individual. *Ensemble learning* dengan fungsi agregasi *averaging voting* memperoleh akurasi tertinggi sebesar 88.75%. Hasil evaluasi eksperimen

pada Gambar 6 menunjukkan bahwa walaupun ada peningkatan nilai akurasi daripada model *baseline*, namun jika dilihat dari tingkat *sensitivity* pada setiap kelas tidak menunjukkan peningkatan pada seluruh kelas. Jika kita perhatikan nilai *sensitivity* pada *fear* (kelas 4) yakni sebesar 59.46% yang mana hasil ini justru mengalami penurunan jika dibandingkan model *baseline*. Oleh karena itu, strategi *ensemble learning* menggunakan fungsi agregasi *averaging voting* belum optimal dalam meningkatkan evaluasi klasifikasi model.

Selanjutnya untuk nilai evaluasi klasifikasi eksperimen *ensemble learning* dengan fungsi agregasi *max voting* yang ditunjukan pada Gambar 7 menampilkan nilai evaluasi yang tidak jauh berbeda dengan nilai evaluasi *baseline* dengan akurasi tertinggi diperoleh sebesar 88.20%. Oleh karena itu dapat disimpulkan juga bahwa strategi *ensemble learning* menggunakan fungsi agregasi *max voting* juga belum optimal dalam meningkatkan performa evaluasi klasifikasi model. Strategi *ensemble learning* yang selanjutnya dilakukan yakni menggunakan fungsi agregasi *meta-learn* atau biasa yang disebut dengan *stacking ensemble learning*. *Meta-learner* yang dipilih pada eksperimen ini adalah *Logistic Regression*, KNN (*K-Nearest Neighbor*), dan XGBoost.

Logistic Regression adalah proses pemodelan probabilitas hasil diskrit. *Logistic Regression* umumnya digunakan untuk memodelkan kelas biner, namun *Logistic Regression* juga dapat diterapkan pada kasus klasifikasi multikelas [28]. *Stacking ensemble learning* pada *meta-learner Logistic Regression* menghasilkan nilai evaluasi akurasi yang cukup tinggi dibandingkan dengan model *baseline*. Pada strategi ini menghasilkan akurasi tertinggi sebesar 90.19%. Hasil evaluasi klasifikasi pada eksperimen ini dapat dilihat pada Gambar 8. Jika dibandingkan dengan model *baseline*, nilai *sensitivity* (*recall*) pada masing-masing kelas meningkat untuk beberapa kelas khususnya peningkatan yang cukup tinggi pada *fear* (kelas 4) dan *disgust* (kelas 5). Pada model *baseline* nilai *sensitivity* pada *fear* (kelas 4) sebesar 62% meningkat menjadi 70.27%, sementara itu nilai *sensitivity* pada *disgust* (kelas 5) sebesar 64% pada model *baseline* dan meningkat menjadi 70.63% setelah dilakukan *stacking ensemble learning* dengan *meta-learner Logistic Regression*. Dari eksperimen ini dapat disimpulkan bahwa strategi *ensemble learning* dengan fungsi agregasi *meta-learner Logistic Regression* dapat meningkatkan performa evaluasi klasifikasi model secara optimal dengan peningkatan nilai akurasi sebesar 2.28%.

	precision	recall	f1-score	support
0	0.8450	0.8897	0.8668	680
1	0.9524	0.9460	0.9492	1185
2	0.8640	0.9038	0.8834	478
3	0.8896	0.8571	0.8731	329
4	0.7333	0.5946	0.6567	74
5	0.7338	0.6375	0.6823	160
6	0.8616	0.8457	0.8536	162
accuracy			0.8875	3068
macro avg	0.8400	0.8106	0.8236	3068
weighted avg	0.8866	0.8875	0.8865	3068

Gambar 6 Evaluasi klasifikasi eksperimen *ensemble learning* dengan *averaging voting*

	precision	recall	f1-score	support
0	0.8371	0.8765	0.8563	680
1	0.9540	0.9443	0.9491	1185
2	0.8528	0.8849	0.8686	478
3	0.8847	0.8632	0.8738	329
4	0.7231	0.6351	0.6763	74
5	0.7194	0.6250	0.6689	160
6	0.8457	0.8457	0.8457	162
accuracy			0.8820	3068
macro avg	0.8310	0.8107	0.8198	3068
weighted avg	0.8814	0.8820	0.8813	3068

Gambar 7 Evaluasi klasifikasi eksperimen *ensemble learning* dengan *max voting*

Selanjutnya dilakukan kembali *stacking ensemble learning* dengan fungsi *meta-learner* KNN (*K-Nearest Neighbor*). Algoritma KNN adalah salah satu teknik pembelajaran mesin yang populer digunakan untuk tugas klasifikasi dan regresi. Algoritma ini bergantung pada gagasan bahwa titik data yang serupa cenderung memiliki label atau nilai yang serupa [29]. Perlu dilakukan pencarian nilai k terbaik untuk menghasilkan model klasifikasi yang lebih akurat.

Pada Tabel 3 dapat dilihat bahwa nilai $k = 3$ memberikan nilai akurasi yang paling tinggi yakni 91.20%. Nilai evaluasi ini lebih tinggi dibandingkan dengan model *baseline* dan strategi

stacking ensemble learning dengan *meta-learner Logistic Regression*. Namun jika diperhatikan lebih detail pada hasil evaluasi klasifikasi pada Gambar 12, strategi ini tidak dapat bekerja secara optimal untuk meningkatkan performa dari *fear* (kelas 4) dan *disgust* (kelas 5) jika dibandingkan dengan model *baseline*. Strategi ini hanya dapat meningkatkan performa kelas yang memiliki data mayoritas seperti pada *neutral* (kelas 0), *happiness* (kelas 1) dan *sadness* (kelas 2).

	precision	recall	f1-score	support
0	0.8499	0.8824	0.8658	680
1	0.9602	0.9570	0.9586	1185
2	0.8773	0.8975	0.8873	478
3	0.9061	0.9088	0.9074	329
4	0.8525	0.7027	0.7704	74
5	0.7635	0.7063	0.7338	160
6	0.9150	0.8642	0.8889	162
accuracy			0.9019	3068
macro avg	0.8749	0.8455	0.8589	3068
weighted avg	0.9018	0.9019	0.9015	3068

Gambar 8 Evaluasi klasifikasi eksperimen *ensemble learning* dengan *Stacking* menggunakan meta learner adalah *Logistic Regression*

Tabel 3 Hasil eksperimen *stacking ensemble learning* dengan *meta-learner* KNN untuk nilai $k = 2 - 5$

Model *stacking ensemble*: KNN	**Akurasi uji**
K = 2	90.71%
K = 3	**91.20%**
K = 4	90.12%
K = 5	89.67%

Strategi *stacking ensemble learning* yang terakhir yakni menggunakan *meta-learner* XGBoost. XGBoost adalah sebuah metode optimasi *gradient boosting* yang terdistribusi yang dirancang untuk pelatihan pada *machine learning* yang lebih efisien. XGBoost adalah singkatan dari *Extreme Gradient Boosting* yang juga merupakan model pembelajaran *ensemble learning* yakni *boosting*, yang menggabungkan beberapa model lemah (*weak model*) untuk menghasilkan performa yang lebih akurat [30]. Pada Gambar 13 dapat dilihat bahwa strategi *stacking ensemble learning* menggunakan XGBoost sebagai *meta-learner* menghasilkan akurasi yang sangat akurat jika dibandingkan dengan model *baseline* dan model *ensemble learning* lainnya.

	precision	recall	f1-score	support
0	0.8554	0.9485	0.8996	680
1	0.9588	0.9612	0.9600	1185
2	0.8980	0.9205	0.9091	478
3	0.9151	0.8845	0.8995	329
4	0.8654	0.6081	0.7143	74
5	0.8793	0.6375	0.7391	160
6	0.9067	0.8395	0.8718	162
accuracy			0.9120	3068
macro avg	0.8969	0.8285	0.8562	3068
weighted avg	0.9126	0.9120	0.9101	3068

Gambar 12 Evaluasi klasifikasi eksperimen *ensemble learning* dengan *Stacking* menggunakan *meta learner* adalah KNN dengan k=3

	precision	recall	f1-score	support
0	0.9281	0.9676	0.9474	680
1	0.9855	0.9781	0.9818	1185
2	0.9503	0.9603	0.9553	478
3	0.9576	0.9605	0.9590	329
4	1.0000	0.8514	0.9197	74
5	0.9658	0.8812	0.9216	160
6	0.9689	0.9630	0.9659	162
accuracy			0.9622	3068
macro avg	0.9652	0.9374	0.9501	3068
weighted avg	0.9628	0.9622	0.9621	3068

Gambar 13 Evaluasi klasifikasi eksperimen *ensemble learning* dengan *Stacking* menggunakan meta learner adalah XGBoost

Pada strategi ini, *stacking ensemble learning* dengan XGBoost sebagai *meta-learner* menghasilkan akurasi sebesar 96.22%. Hasil ini meningkat sebesar 8.31% jika dibandingkan model *baseline*. Selain akurasi yang meningkat cukup drastis, nilai evaluasi klasifikasi yakni *sensitivity* pada setiap kelas juga mengalami peningkatan yang signifikan. Pada model *baseline* nilai *sensitivity* pada *fear* (kelas 4) sebesar 62% meningkat menjadi 85.14%, sementara itu nilai *sensitivity* pada *disgust* (kelas 5) sebesar 64% pada model *baseline* dan meningkat menjadi 88.12% setelah dilakukan *stacking ensemble learning* dengan *meta-learner* XGBoost. Hasil ini dapat meningkat secara optimal disebabkan karena algoritma XGBoost sendiri juga merupakan strategi *ensemble learning boosting* yang mana disampaikan pada bagian metode penelitian bahwa pendekatan *ensemble learning* dapat meningkatkan performa klasifikasi, sehingga hasil yang diperoleh menjadi *extra boosting*.

Nilai evaluasi klasifikasi *stacking ensemble learning* menggunakan XGBoost sebagai *meta-learner* juga menandingi model Patt-Lite [23], yang merupakan model terbaik berdasarkan *state-of the art* dari *dataset* RAF-DB. Model Patt-Lite menghasilkan akurasi sebesar 95.05%, sementara itu model *stacking ensemble learning* menggunakan XGBoost sebagai *meta-learner* menghasilkan akurasi sebesar 96.22%, dengan persentase peningkatan sebesar 1.07%. Tidak hanya dari sisi nilai akurasi, dari nilai *recall* atau *sensitivity* pada setiap kelas, khususnya pada *fear* (kelas 4) dan *disgust* (kelas 5) yang masing menjadi limitasi pada model Patt-Lite dapat diatasi pada eksperimen ini. Dapat dilihat pada Gambar 13, nilai *recall* kelas 4 sebesar 85.15% meningkat sebesar 12.17% jika dibandingkan model Patt-Lite. Begitu juga dengan kelas 5 mengalami peningkatan sebesar 5.14% jika dibandingkan dengan model Patt-Lite. Perbandingan model yang diusulkan pada penelitian ini dengan model-model *deep learning* lainnya yang menggunakan *dataset* RAF-DB dapat dilihat pada Tabel 4.

Tabel 4 Perbandingan hasil akurasi *state of the art* pada *dataset* RAF-DB dari tahun 2021 - 2023

Model *deep learning*	**Akurasi (%)**
PACVT [21]	88.21%
MA-Net (re-run)	87.91%
MA-Net [9]	88.40%
AMP-Net [22]	89.25%
DAN [20]	89.70%
Patt-Lite [23]	95.05%
MA-Net *Ensemble learning: averaging voting**	88.75%
MA-Net *Ensemble learning: max voting**	88.20%
MA-Net *Stacking ensemble*: XGBoost*	96.22%

*hasil yang diperoleh pada penelitian ini

KESIMPULAN

Penelitian ini telah berhasil meningkatkan performa metode MA-Net dalam mengenali ekspresi wajah pada *dataset* RAF-DB. Performa metode MA-Net pada *dataset* RAF-DB menghasilkan akurasi tertinggi digunakan sebagai model *baseline*. Penelitian ini menerapkan dua jenis strategi *ensemble deep learning*. Strategi *ensemble deep learning* pertama yakni *voting* dengan menggunakan dua jenis fungsi agregasi yakni *averaging voting* dan *max voting* dengan akurasi tertinggi disekitar 88%. Strategi *ensemble deep learning* kedua yakni *stacking ensemble learning* dengan menggunakan fungsi agregasi *meta-learner* yakni *Logistic Regression*, KNN (*K-Nearest Neighbor*), dan XGBoost mendapatkan akurasi lebih besar dari 90%. Strategi *stacking ensemble learning* ini dapat mengungguli model dari penelitian Patt-Lite yang merupakan model terbaik dari *state of the art* pada *dataset* RAF-DB. Oleh karena itu, pada penelitian ini dapat disimpulkan bahwa penggunaan strategi *ensemble learning* khususnya pada penggunaan fungsi agregasi *meta-learner* atau *stacking ensemble learning* dapat meningkatkan performa evaluasi klasifikasi secara keseluruhan maupun pada masing-masing kelas.

DAFTAR PUSTAKA

[1] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," *Image Vis Comput*, vol. 27, no. 12, pp. 1743–1759, 2009, doi: 10.1016/j.imavis.2008.11.007.

[2] I. Saadi, D. W. cunningham, T.-A. Abdelmalik, A. Hadid, and Y. El Hillali, "Driver's facial expression recognition: A comprehensive survey," *Expert Syst Appl*, p. 122784, Dec. 2023, doi: 10.1016/j.eswa.2023.122784.

[3] Z. Zhang, Z. Li, H. Liu, T. Cao, and S. Liu, "Data-driven Online Learning Engagement Detection via Facial Expression and Mouse Behavior Recognition Technology," *Journal*

of Educational Computing Research, vol. 58, no. 1, pp. 63–86, Mar. 2020, doi: 10.1177/0735633119825575.

[4] P. Ekman *et al.*, "Are There Basic Emotions?," 1992.

[5] F. Bourel, C. C. Chibelushi, and A. A. Low, "Recognition of Facial Expressions in the Presence of Occlusion," British Machine Vision Association and Society for Pattern Recognition, Feb. 2013, pp. 23.1-23.10. doi: 10.5244/c.15.23.

[6] Z. Hammal, M. Arguin, and F. Gosselin, "Comparing a novel model based on the transferable belief model with humans during the recognition of partially occluded facial expressions," *J Vis*, vol. 9, no. 2, 2009, doi: 10.1167/9.2.22.

[7] F. Zhang, T. Zhang, Q. Mao, and C. Xu, "Joint Pose and Expression Modeling for Facial Expression Recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Dec. 2018, pp. 3359–3368. doi: 10.1109/CVPR.2018.00354.

[8] G. Yovel and B. Duchaine, "Specialized Face Perception Mechanisms Extract Both Part and Spacing Information: Evidence from Developmental Prosopagnosia." [Online]. Available: http://direct.mit.edu/jocn/article-pdf/18/4/580/1935781/jocn.2006.18.4.580.pdf

[9] Z. Zhao, Q. Liu, and S. Wang, "Learning Deep Global Multi-Scale and Local Attention Features for Facial Expression Recognition in the Wild," *IEEE Transactions on Image Processing*, vol. 30, pp. 6544–6556, 2021, doi: 10.1109/TIP.2021.3093397.

[10] S. Li and W. Deng, "Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 356–370, Jan. 2019, doi: 10.1109/TIP.2018.2868382.

[11] J. Zeng, S. Shan, and X. Chen, "Facial expression recognition with inconsistently annotated datasets," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 2018, pp. 227–243. doi: 10.1007/978-3-030-01261-8_14.

[12] K. Wang, X. Peng, J. Yang, S. Lu, and Y. Qiao, "Suppressing uncertainties for large-scale facial expression recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, 2020, pp. 6896–6905. doi: 10.1109/CVPR42600.2020.00693.

[13] J. Yu, Z. Cai, P. He, G. Xie, and Q. Ling, "Multi-model Ensemble Learning Method for Human Expression Recognition," Mar. 2022, [Online]. Available: http://arxiv.org/abs/2203.14466

[14] X. Zheng, W. Chen, Y. You, Y. Jiang, M. Li, and T. Zhang, "Ensemble deep learning for automated visual classification using EEG signals," *Pattern Recognit*, vol. 102, Jun. 2020, doi: 10.1016/j.patcog.2019.107147.

[15] W. Jiang, Y. Wu, F. Qiao, L. Meng, Y. Deng, and C. Liu, "Model Level Ensemble for Facial Action Unit Recognition at the 3rd ABAW Challenge," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, IEEE Computer Society, 2022, pp. 2336–2343. doi: 10.1109/CVPRW56347.2022.00260.

[16] A. Mohammed and R. Kora, "A comprehensive review on ensemble deep learning: Opportunities and challenges," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 2. King Saud bin Abdulaziz University, pp. 757–774, Feb. 01, 2023. doi: 10.1016/j.jksuci.2023.01.014.

[17] C. Chang, L. Chen, C. Zhang, and Y. Liu, "An ensemble model using face and body tracking for engagement detection," in *ICMI 2018 - Proceedings of the 2018*

International Conference on Multimodal Interaction, Association for Computing Machinery, Inc, Oct. 2018, pp. 616–622. doi: 10.1145/3242969.3264986.

[18] Z. Zhang, B. Chen, J. Sun, and Y. Luo, "A bagging dynamic deep learning network for diagnosing COVID-19," *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-95537-y.

[19] W. Deng, Q. Xu, J. Liu, Y. Lu, M. Fan, and X. Liu, "Image Classification Method of Longhorn Beetles of Yunnan Based on Bagging and CNN," in *2022 5th International Conference on Pattern Recognition and Artificial Intelligence, PRAI 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 198–203. doi: 10.1109/PRAI55851.2022.9904155.

[20] Z. Wen, W. Lin, T. Wang, and G. Xu, "Distract Your Attention: Multi-Head Cross Attention Network for Facial Expression Recognition," *Biomimetics*, vol. 8, no. 2, Jun. 2023, doi: 10.3390/biomimetics8020199.

[21] C. Liu, K. Hirota, and Y. Dai, "Patch attention convolutional vision transformer for facial expression recognition with occlusion," *Inf Sci (N Y)*, vol. 619, pp. 781–794, Jan. 2023, doi: 10.1016/j.ins.2022.11.068.

[22] H. Liu, H. Cai, Q. Lin, X. Li, and H. Xiao, "Adaptive Multilayer Perceptual Attention Network for Facial Expression Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 9, pp. 6253–6266, Sep. 2022, doi: 10.1109/TCSVT.2022.3165321.

[23] J. Le Ngwe, K. M. Lim, C. P. Lee, and T. S. Ong, "PAtt-Lite: Lightweight Patch and Attention MobileNet for Challenging Facial Expression Recognition," Jun. 2023, [Online]. Available: http://arxiv.org/abs/2306.09626

[24] M. Mohammed, H. Mwambi, I. B. Mboya, M. K. Elbashir, and B. Omolo, "A stacking ensemble deep learning approach to cancer type classification based on TCGA data," *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-95128-x.

[25] A. Ghasemieh, A. Lloyed, P. Bahrami, P. Vajar, and R. Kashef, "A novel machine learning model with Stacking Ensemble Learner for predicting emergency readmission of heart-disease patients," *Decision Analytics Journal*, vol. 7, Jun. 2023, doi: 10.1016/j.dajour.2023.100242.

[26] M. Grandini, E. Bagli, and G. Visani, "Metrics for Multi-Class Classification: an Overview," Aug. 2020, [Online]. Available: http://arxiv.org/abs/2008.05756

[27] J. Tanha, Y. Abdi, N. Samadi, N. Razzaghi, and M. Asadpour, "Boosting methods for multi-class imbalanced data classification: an experimental review," *J Big Data*, vol. 7, no. 1, Dec. 2020, doi: 10.1186/s40537-020-00349-y.

[28] S. Moulya and T. R. Pragathi, "Mental Health Assist and Diagnosis Conversational Interface using Logistic Regression Model for Emotion and Sentiment Analysis," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Jan. 2022. doi: 10.1088/1742-6596/2161/1/012039.

[29] L. S. Ihzaniah, A. Setiawan, and R. W. N. Wijaya, "Perbandingan Kinerja Metode Regresi K-Nearest Neighbor dan Metode Regresi Linear Berganda pada Data Boston Housing," *Jambura Journal of Probability and Statistics*, vol. 4, no. 1, pp. 17–29, May 2023, doi: 10.34312/jjps.v4i1.18948.

[30] G. Abdurrahman, H. Oktavianto, and M. Sintawati, "Optimasi Algoritma XGBoost Classifier Menggunakan Hyperparameter Gridesearch dan Random Search Pada Klasifikasi Penyakit Diabetes," 2022.