

Pembelajaran Mendalam Pengklasifikasi Ekspresi Wajah Manusia dengan Model Arsitektur Xception pada Metode Convolutional Neural Network

Purnawarman Musa^{1*}, Wahid Khairul Anam¹, Saiful Bahri Musa²,
Witari Aryunani¹, Remi Senjaya¹, Puji Sularsih¹

¹Universitas Gunadarma

Jalan Margonda Raya No. 100 Depok 16424 Jawa Barat

²Politeknik Gorontalo

Jl. Saptamarga Botu Pingge Bone Bolango 96112 Gorontalo

*p_musa@staff.gunadarma.ac.id

DOI: <https://doi.org/10.21107/rekayasa.v16i1.16974>

Abstrak

Pembelajaran mendalam secara jaringan saraf tiruan dalam pengembangan inovasi yang memberikan pengetahuan tertanam pada komputer tentang suatu masalah untuk mempelajari dan memberikan solusi. Sistem pendeteksian dengan *vision framework* merupakan salah satu konsep visi komputer yang dapat mengenali objek yang diamati layaknya sistem penglihatan manusia. Penelitian kami mengusulkan sistem pengenalan ekspresi wajah memanfaatkan model berbasis *Convolutional Neural Network (CNN)* berbasis kecerdasan buatan dengan metode pembelajaran mendalam. Hasil model dari proses pelatihan data dengan CNN akan digunakan untuk sistem pendeteksi wajah berdasarkan klasifikasi ekspresi wajah. Aplikasi dirancang memanfaatkan library OpenCV, Keras, dan TensorFlow sebagai backend. Mengacu dari penelitian penggunaan model arsitektur xception pada sistem pendeteksi ekspresi wajah dengan hasil yang optimal. Berdasarkan hasil pengujian tersebut terhadap hasil penelitian mendapatkan nilai yang peningkatan akurasi dalam pelatihan dan pengujian data pada model arsitektur xception yang terlatih untuk ekspresi wajah menggunakan dataset FER-2013, dimana menghasilkan nilai akurasi 66%, serta nilai masing-masing rerata pada presisi 76%, recall 65%, dan score F1 63%.

Kata Kunci : deteksi wajah, kecerdasan buatan, pembelajaran mendalam, ekspresi wajah

Abstract

Deep learning is a neural network that creates innovations that give computer-implemented problem-solving expertise. One of the principles of computer vision is a detection system with a vision framework that can identify things encountered in the same manner as a human vision system. Using an artificial intelligence-based Convolutional Neural Network (CNN) model with deep learning techniques, we present a face emotion identification system. The categorization of facial expressions will be utilized as the basis for a face recognition system trained using CNN. The applications are intended to use the OpenCV, Keras, and TensorFlow libraries as the backend. We were discussing the study on the best use of xception architectural models in facial expression recognition systems. Based on the results of these tests, the study obtained an increased accuracy value in training and data testing on an xception architecture model trained for facial expressions using the FER-2013 dataset, resulting in an accuracy value of 66% as well as the value of each average for precision (76%), recall (65%), and F1 score (63%).

Keywords : face detection, artificial intelligence, deep learning, facial expressions

PENDAHULUAN

Deep Learning adalah subbidang Pembelajaran Mesin berdasarkan jaringan saraf tiruan (JST) memberikan pengetahuan pada komputer melakukan suatu keputusan didasari suatu kegiatan yang natural pada manusia. Komputer belajar mengkategorikan langsung dari gambar,

Article History:

Received: September 20th, 2022; **Accepted:** Feb 25th, 2023

Cite this as:

Musa et al. (2023). *Pembelajaran mendalam pengklasifikasian ekspresi wajah manusia dengan model arsitektur Xception pada metode convolutional neural network*. *Rekayasa*. Vol 16(3) 66-73.

teks, atau suara dalam pembelajaran mendalam (Xie et al., 2017). Komputer dilatih menggunakan sejumlah kumpulan data berlabel, dan nilai piksel gambar diubah menjadi representasi internal atau vektor fitur sehingga pengklasifikasi dapat mengenali atau mengkategorikan pola pada input (Tang, 2015). Tren teknologi kecerdasan buatan atau *Artificial Intelligence (AI)*, dimana cara kerja pada komputer adalah menyelidiki dan menyerupai yang dilakukan oleh orang-orang dalam hal kompleksitas dan akurasi (Vadinský, 2018) serta menyimulasikan proses dan meniru perilaku manusia (Xia et al., 2013).

AI didasari tiga alasan terhadap kelebihan dari komputer: inovatif, paham konsep kecerdasan, dan perangkat yang berguna. AI memperoleh pengetahuan atau pemahaman melalui pengalaman, memahami informasi yang membingungkan atau bertentangan, beradaptasi dengan cepat dan kompeten terhadap keadaan baru, menerapkan logika dalam pemecahan masalah, dan menyelesaikan masalah (Zhang *et al.*, 2014). Menurut (Xia *et al.*, 2013), dua komponen utama pada aplikasi AI; (a) Kumpulan fakta, hipotesis, dan konsep, serta keterkaitan, dapat ditemukan di basis pengetahuan dan (b) Kapasitas mesin untuk menarik kesimpulan berdasarkan pengalaman disebut sebagai *motor inferensi* (mesin *inferensi*).

Kemampuan menginterpretasi data visual seperti halnya sistem penglihatan manusia, dimana mengubah data dari kamera video atau foto/gambar menjadi keputusan atau presentasi baru dikenal sebagai visi komputer. Proses aktivitas transformasi memungkinkan mempunyai informasi kontekstual, seperti foto atau gambar di mana sejumlah objek-objek yang berbeda dapat ditemukan. Hasil penelitian oleh (Musa *et al.*, 2022), menggunakan algoritma FaceNet dan AgeNet diukur mampu secara akurat mengidentifikasi usia pejalan kaki. Area tubuh manusia yang berfungsi sebagai titik fokus kontak sosial dan melakukan interaksi secara gestur atau mimic adalah bagian wajah (Henderson & Rothe, 2017; Janku *et al.*, 2016; Xia *et al.*, 2017).

Wajah adalah salah satu aspek khas tubuh manusia memiliki kualitas individu yang memberikan informasi dan hal-hal menarik yang dapat diperhatikan dan dipelajari (Musa *et al.*, 2009; Ramadhani *et al.*, 2017). Wajah memiliki peran penting dengan menunjukkan emosi/ekspresi. Aplikasi prediksi ekspresi memiliki dua tahap penting: representasi fitur wajah dan desain klasifikasi dan beberapa sistem untuk bidang keamanan dan bidang identifikasi pelaku kejahatan. Metode deteksi wajah oleh komputer tidak secepat dan semudah yang digunakan orang (Mustafa *et al.*, 2014).

Selama ini, peneliti-peneliti mengembangkan ekspresi wajah manusia dengan beberapa kelemahan, sebagai berikut: prediksi ekspresi wajah dari pengumpulan data (file), pra-pemrosesan manual dan tidak real-time atau membutuhkan banyak waktu. Ada dua pendekatan

deteksi wajah (Dargham *et al.*, 2012), yaitu metode deteksi berbasis fitur pada wajah dan Eigenfaces metode untuk mendapatkan eigenvector (*eigenface*), dari kovarians matriks distribusi, yang dibuat dari serangkaian pelatihan gambar wajah orang.

Studi sebelumnya oleh Hubel dan Wiesel, yang menyelidiki otak visualisasi dan persepsi kucing tentang penglihatan (Simonyan & Zisserman, 2014b, 2014a). Korteks visual hewan sebelumnya merupakan bagian integral dari sistem pemrosesan visual, tetapi operasinya telah memacu pengembangan model baru seperti Neocognitron, HMAX, LeNet-5, dan AlexNet. CNN memiliki tingkatan dengan struktur *neuron* tiga bagian (lebar, tinggi, kedalaman). Pada struktur *neuron* memiliki tingkat lebar dan tinggi adalah dimensi lapisan, sedangkan kedalaman berdasarkan hasil jumlahnya. Lapisan CNN menghasilkan lusinan atau ratusan bahkan jutaan lapisan, dimana setiap lapisan dilatih mengenali masukan citra berbeda. Setiap gambar pelatih mengalami pemrosesan gambar dari ukuran pixel yang beragam, dan setiap hasil keluaran citra diproses dan dimanfaatkan sebagai masukan ke lapisan selanjutnya.

Modul residual memodifikasi pemetaan yang dipilih diantara dua lapisan berikutnya, sehingga fitur diekstraksi dengan mekanisme pembandingan antara fitur asli dengan fitur yang dipilih (He *et al.*, 2016). Penelitian oleh (Musa & Tjandrasa, 2017), dengan penyaringan gelombang EEG menggunakan dua algoritma yaitu *WaveletPacket Decomposition (WPD)* dan *Empirical Mode Decomposition (EMD)* merepresitasi fitur ekspresi manusia. Terapan metode konvolusi 2-D yang melakukan tapis terhadap informasi digital pada bagian wajah yang nilai citranya menjadi keabuan akan menemukan ekspresi wajah (Kurniawan *et al.*, 2015). Penggabungan model fitur lokal dan global yang disarankan pada penelitian (Jain *et al.*, 2019) menghasilkan kenaikan akurasi menjadi 93,52 persen, dibandingkan dengan akurasi 74,11 persen dengan pola biner lokal dan 90,64 persen dengan fitur global. Penelitian oleh (Zahara *et al.*, 2020) dengan pendekatan CNN untuk pengenalan emosi wajah (FER-2013) menghasilkan temuan prediksi emosi wajah dengan tingkat keberhasilan 65,97 persen. Temuan simulasi yang dilakukan oleh (Vandana & Marriwala, 2022) menunjukkan bahwa keakuratan pengenalan ekspresi wajah

ditingkatkan oleh model CNN dibandingkan model CNN Sederhana.

Bedasarkan uraian penelitian terdahulu analisa secara visi komputer menghasilkan akurasi dan proses prediksi ekspresi wajah yang efektif, maka tujuan penelitian kami akan mengembangkan dan meninjau metode prediksi atau pengenalan ekspresi wajah dengan mengusulkan konsep Deep Learning dengan menggunakan algoritma *Convolutional Neural Networks* (ConvNet) sebagai solusi untuk masalah praktis serta meningkatkan efisiensi dan efektivitas dalam berbagai bidang, seperti bisnis, kesehatan, teknologi, dan lainnya. Kontribusi dari penelitian kami dalam pengenalan ekspresi wajah memiliki manfaat dalam dunia nyata seperti: (a) meningkatkan interaksi manusia-komputer, (b) menunjang pengembangan robot dan sistem AI yang lebih cerdas dan interaktif, (c) membantu dalam bidang keamanan, seperti verifikasi identitas dan deteksi kebohongan, bidang psikologi dan neurosains untuk memahami emosi dan persepsi wajah, bidang advertising dan marketing untuk menilai respons terhadap iklan dan produk.

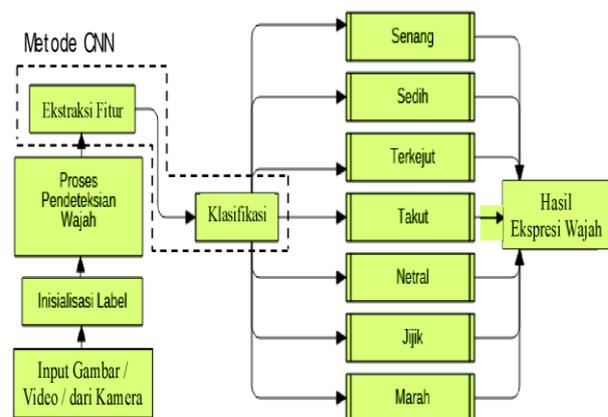
METODE PENELITIAN

Metode penelitian dalam sistem pendeteksi ekspresi wajah menggunakan algoritma *Deep Learning* dengan memanfaatkan *Convolutional Neural Networks* (CNN). *Convolutional Neural Networks* (CNN) adalah satu turunan dari konsep metode *deep learning* yang dikembangkan dari *Multi Layer Perceptron* (MLP) untuk menganalisis input dua dimensi, seperti gambar atau suara (Springenberg et al., 2014). CNN sering diperuntukan mengidentifikasi item untuk mendeteksi serta mengelompokkan berbagai hal. CNN melakukan pembelajaran langsung dari sekumpulan informasi dari gambar atau suara, merepresentasikan kebutuhan untuk ekstraksi karakteristik manual atau dataset yang telah dilabelkan memanfaatkan konsep pembelajaran yang diawasi (*supervised learning*), dimana supervised learning bekerja karena terdapat data pelatihan dan variabel yang ditargetkan, dan pendekatan ini bertujuan untuk mengelompokkan data ke dalam data yang ada.

Metode untuk mendeteksi ekspresi wajah pada riset kami memiliki dua kelompok proses utama yaitu bagian pelaksanaan pada pelatihan dataset dan pelaksanaan pada pengujian terhadap pendeteksian menggunakan sistem hasil aplikasi yang dilakukan

pada penelitian kami. Pada proses pelatihan data, pertama pemanggilan dataset yang akan dilakukan pra-proses dataset FER-2013. Kemudian model pelatihan akan dikompilasi dan pelatihan data dilakukan dengan parameter yang telah ditentukan sebelumnya. Proses pelatihan menghasilkan model terlatih sebagai parameter prediksi.

Proses pendeteksian dengan hasil model arsitektur Xception, terlebih dahulu mendapatkan gambar/video atau menangkap gambar menggunakan webcam. Model terlatih didefinisikan sebagai parameter prediksi ekspresi wajah pada sistem. Kemudian dilakukan pelabelan dari hasil citra yang telah diprediksi. Kemudian metode pendeteksian wajah berbasis fitur didefinisikan dengan metode *Haar Cascade Classifier*, untuk mendeteksi wajah serta dilakukan segmentasi citra untuk prediksi ekspresi terhadap data input wajah. Model arsitektur *Xception* merupakan model arsitektur berbasis *Convolutional Neural Networks* yang dapat menghilangkan lapisan yang terhubung sepenuhnya (*fully-connected layer*) dan jumlah parameter yang digunakan jauh berkurang pada lapisan konvolusional yang ada (Chollet, 2016). Teknik ekstraksi fitur dan mengkombinasikan dalam lapisan konvolusional terpisah (*Depthwise Separable Convolutions*), terdiri dari dua lapisan yang berbeda, yakni lapisan *depthwise* dan lapisan *pointwise convolutions*.



Gambar 1. Ilustrasi Deteksi Wajah Serta Deteksi Ekspresi

Tujuan utama adalah untuk memisahkan korelasi antar spasial pada saluran korelasi silang dengan terlebih dahulu menerapkan filter $D \times D$ pada setiap saluran input M . Kemudian menerapkan konvolusi filter $N 1 \times 1 \times M$ untuk menggabungkan saluran input M ke dalam saluran

output N. Konvolusi $1 \times 1 \times M$ diterapkan menggabungkan setiap nilai pemetaan fitur tanpa pertimbangan keterkaitan spasial pada saluran. Proses prediksi akan membutuhkan data deteksi wajah yang akan diproses kembali dengan model deteksi yang dihasilkan pada proses pelatihan. Label pada masing-masing model deteksi akan diinisialisasikan dan model prediksi terlatih yang dihasilkan oleh proses pelatihan dipanggil, seperti yang diperlihatkan pada Gambar 1. Sementara itu Identifikasi kebutuhan perangkat keras dan lunak disajikan pada Tabel 1 dan **Error! Reference source not found.**

Tabel 1. Kebutuhan dan Spesifikasi Perangkat keras

Nama Alat	Spesifikasi
Prosesor	Intel® Core i3-4030U @1.90GHz (4 CPU)
Kartu Grafis	AMD Radeon R5M230 DDR3 2GB VRAM
Memori	6 GB DDR3
Penyimpanan	500 Gb Hard Drive

Tabel 2. Kebutuhan dan Spesifikasi Perangkat Lunak

Nama Perangkat Lunak / Library	Versi
Tkinter	8.6
Keras	2.2.0
TensorFlow	1.9.0
OpenCV	3.4.2.16
Numpy	1.14.5
Pandas	0.23.3
Scipy	1.1.0

Pengumpulan data latih dalam dataset dijadikan parameter pelatihan dan mempersiapkan dataset (Carrier & Courville, 2013). Penelitian ini memanfaatkan FER-2013 dataset didapat secara sekunder dari kaggle.com yang umumnya digunakan oleh para peneliti sebagai model data pendeteksian ekspresi terdiri dari 35.887 berbentuk gambar dengan warna abu-abu (*grayscale*) berukuran 48×48 pixel tiap gambar, dan telah dikategorikan pada 7 emosi ekspresi wajah (klasifikasi pada Tabel 3).

Tabel 3. Dataset latih Ekspresi Wajah pada FER-2013

Klasifikasi	Kode Klasifikasi	Jumlah Dataset
Angry (Marah)	0	4593
Disgust (Jijik)	1	547
Fear (Takut)	2	5121
Happy (Bahagia)	3	8989
Sad (Sedih)	4	6077
Surprise (Terkejut)	5	4002

Neutral (Netral)	6	6198
------------------	---	------

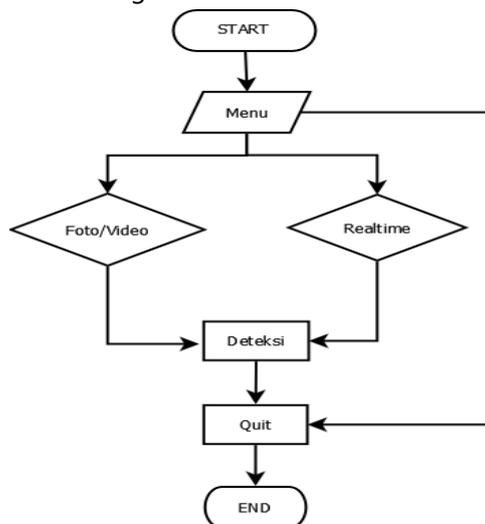
Pada tahap pengembangan model arsitektur, algoritma *Deep Learning* dirancang dengan memanfaatkan *Convolutional Neural Networks* (CNN). Selain itu, proses pelatihan dengan menggunakan parameter yang disesuaikan dari hasil perhitungan algoritma *Deep Learning* dengan memanfaatkan metode *Convolutional Neural Networks* (CNN) yang akan memperlihatkan hasil model data terlatih yang digunakan mendeteksi ekspresi wajah yang ditunjukkan pada listing program 1.

Listing Program 1: Pembuatan Model dari Dataset Latih

```

1. hasil = Activation ('softmax', name = 'predictions' ) (x)
2. model = Model ( img input , hasil )
3. return model
    
```

Perancangan sistem mulai dari pemanggilan metode Haar Cascade Feature-based Classifier sebagai metode pendeteksian wajah, penentuan model terlatih yang dilakukan dan akan mendeteksi ekspresi wajah dengan memanfaatkan webcam sebagai media inputan data citra secara real-time atau memasukkan data berupa gambar atau video. Terdapat tiga menu utama hasil aplikasi yaitu Webcam, Quit, dan Import. Pada menu Webcam dikhususkan untuk Webcam mendeteksi ekspresi secara real-time menggunakan perangkat webcam yang telah terhubung. Menu Import, menu dialog berisi pilih file yang akan digunakan, file berupa file foto ataupun video untuk mendeteksi ekspresi wajah. Sedangkan menu Quit untuk menutup program setelah tidak digunakan.



Gambar 2. Alur Pengujian Deteksi Ekspresi Wajah

Analisa terhadap pengukuran hasil prediksi ekspresi wajah dilakukan dengan menghitung menggunakan persamaan akurasi (1), presisi (2), recall (3) dan F1 score (4) yang ditunjukkan sebagai berikut:

$$Akurasi = \frac{(TP + TN)}{(TP + FN + TN + FP)} \times 100\% \dots (1)$$

$$Presisi_{class_emosi} = \frac{TP}{TP + FP} \dots \dots \dots (2)$$

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

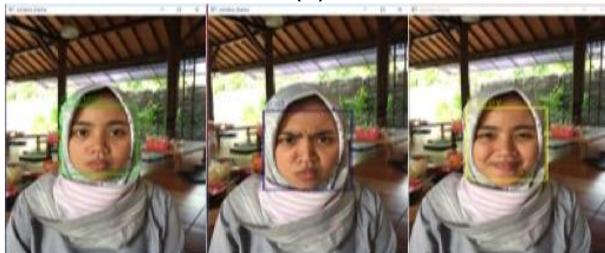
$$F1\ Score_{class_emosi} = \frac{2 \times Precision \times Recall}{Precision + Recall} \dots (4)$$

HASIL PEMBAHASAN

Hasil penelitian menunjukkan bahwa penggunaan model arsitektur CNN pada sistem pendeteksi ekspresi wajah dapat dilakukan dengan optimal secara real-time (Gambar 3), memasukkan data gambar (Gambar 4) dan video (Gambar 5). Hal ini dibuktikan dengan penggunaan lapisan konvolusi terpisah pada saat melakukan pelatihan dengan optimal.

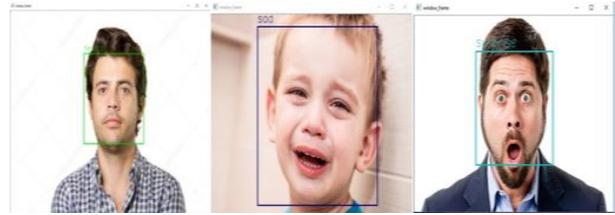


(A)



(B)

Gambar 3. Hasil uji prediksi ekspresi wajah secara real-time, (A) Peserta Pertama dengan ekspresi wajah Takut (kiri), Marah (tengah) dan Bahagia (kanan), (B) Peserta Kedua dengan ekspresi wajah Netral (kiri), Sedih (tengah) dan Bahagia (kanan)



(A)



(B)

Gambar 4. Hasil uji prediksi ekspresi wajah dengan memasukkan file gambar, (A) Uji ke-1 menggunakan gambar dari internet, hasil ekspresi wajah Netral (kiri), Sedih (tengah) dan Terkejut (kanan), (B) Uji ke-2 menggunakan file gambar dari smartphone dengan target uji sebanyak 2 orang, hasil ekspresi wajah keduanya diklasifikasikan ekspresi bahagia



Gambar 5. Hasil uji prediksi ekspresi wajah dengan memasukkan file video, hasil ekspresi wajah sebagian besar diklasifikasikan ekspresi bahagia

		PREDIKSI						
		MARAH	JUIK	TAKUT	SENANG	SEDIH	TERKEJUT	NETRAL
AKTUAL	MARAH	15	0	0	2	3	0	0
	JUIK	0	1	1	0	8	0	11
	TAKUT	0	0	17	0	0	1	3
	SENANG	0	0	0	18	0	0	2
	SEDIH	0	0	2	0	14	0	4
	TERKEJUT	0	0	2	3	3	12	0
	NETRAL	0	0	0	2	3	0	15

Gambar 6. Confusion Matrix

Pengujian berdasarkan matriks kebingungan yang terdapat pada Gambar 6 mendapatkan hasil yang cukup baik. Hasil uji ekspresi wajah dalam pengukuran berdasarkan kategori ekspresi wajah tertinggi didapatkan pada ekspresi *happy* (Bahagia/senang) sebanyak 18 kali dideteksi benar dari 20 percobaan, dan Sedangkan jijik dengan jumlah nilainya terdeteksi paling rendah yaitu sebanyak 11 kali diprediksi benar dari 20 percobaan per klasifikasi.

Pengujian akan menghitung nilai akurasi sebagai penilaian keseluruhan yang menunjukkan seberapa efektif model dapat mengkategorikan data citra yang telah dievaluasi menggunakan persamaan (1). Hasil akurasi pada seluruh pengujian adalah 65.71%.

Analisa pada penilaian presisi akan menentukan tingkat akurasi yang digunakan model untuk setiap klasifikasi atau frekuensi yang diklasifikasikan dengan tepat menghitungnya menggunakan persamaan (2). Perhitungan presisi pada setiap klasifikasi ekspresi wajah yaitu, untuk kategori Marah adalah 1.00, kategori Jijik adalah 1.00, kategori Takut adalah 0.77, kategori Bahagia adalah 0.72, kategori Sedih adalah 0.45, kategori Terkejut adalah 0.92, dan kategori Netral adalah 0.43. Sehingga nilai presisi rerata adalah 0.76 atau memiliki presentasi untuk presisi adalah 76%.

Suata nilai mengalami penarikan (*recall*) ditentukan dengan mempertimbangkan seluruh hasil nilai yang positif dan seberapa sering hasil prediksi yang diproyeksikan memberikan nilai yang tepat dan benar dalam perhitungan. Rumus untuk menghitung *recall* ditunjukkan pada persamaan (3).

Berdasarkan hasil *recall* pada setiap klas, maka untuk kategori Marah adalah 0.75, kategori Jijik adalah 0.05, kategori Takut adalah 0.81, kategori Bahagia adalah 0.90, kategori Sedih adalah 0.70, kategori Terkejut adalah 0.60, kategori Netral adalah 0.75. Hasil rerata dari nilai *recall* adalah sebesar 0.65 atau nilai rerata *recall* dalam presentasi sebesar 65%.

Nilai Skor F1 adalah parameter pengukuran yang menyelaraskan terhadap nilai presisi dan *recall*. Perhitungan dari kedua hasilnya tersebut akan membandingkan persamaan (2) dan (3), sebagai persamaan (4) untuk menghitung Skor F1. Perhitungan pada score F1 untuk kategori Marah adalah 0.86, kategori Jijik adalah 0.10, kelas Takut adalah 0.79, kategori Bahagia adalah 0.80, kategori

Sedih adalah 0.55, kategori Terkejut adalah 0.73, kategori Netral adalah 0.57. Berdasarkan nilai-nilai dari masing-masing klasifikasi, maka rerata pada nilai score F1 sebesar 0.63 atau menghasilkan memiliki nilai rata-rata sebesar 63% untuk nilai score F1.

Dalam perbandingan akurasi yang tidak ada perbedaan dari hasil penelitian kami dengan penelitian sebelumnya (Alamsyah & Pratama, 2020; Zahara *et al.*, 2020), dimana mendapatkan nilai 66%. Namun berdasarkan hasil dari setiap klasifikasi yang diuji mendapatkan hasil yang cukup baik, kecuali untuk klafisikasi pada uji emosi jijik sangat sulit diidentifikasi dan mengalami ketidakberhasilan dalam mendeteksi ekspresi wajah dalam emosi jijik yang terdeteksi sebanyak 1 dari 20 kali percobaan yang dilakukan pengujian. Terdapat dua alasan yang dapat dari hasil tersebut, diantaranya:

1. Alasan utama dari hasil penelitian sebelumnya (Nisbett & Miyamoto, 2005; Righart & de Gelder, 2008; Righart & De Gelder, 2006, 2008) dalam bidang psikologi mengeksplorasi pengaruh konteks emosional pada pengenalan ekspresi wajah menggunakan kategori spesifik konteks emosional dan paradigma *Micro Expression Training Tool* (METT), sehingga dalam dataset FER-2013 memiliki jumlah data yang sedikit untuk klasifikasi untuk emosi jijik karena jarang dipresentasikan dalam ekspresi wajah manusia dan sering dipresentasikan melalui tindakan atau perilaku tubuh dan konteks interaksi sosial, bukan melalui ekspresi wajah saja, sehingga pada dataset FER-2013 yang mengekspresikan emosi jijik sulit dan memerlukan dataset yang lebih besar dan lebih komprehensif untuk memastikan akurasi pengenalan.
2. Dalam bersosialisasi masyarakat Indonesia pada umumnya kultur dan budaya menganut adat-istiadat ketimuran menunjukkan bahwa orang-orang dari Timur lebih dipengaruhi oleh konteks (Ko *et al.*, 2011), ekspresi wajah dalam emosi jijik merupakan suatu hal yang ekspresi dalam kategori negative, sehingga beberapa peserta pengujian sangat sulit memperagakan ekspresi wajah dalam bentuk emosi jijik.

KESIMPULAN DAN SARAN

Pembelajaran mendalam yang diterapkan telah mampu menyamai kemampuan manusia dalam

melakukan klasifikasi ekspresi wajah. Hal ini dibuktikan dengan tingkat kepekaan ketika dilakukan pengujian, dan menghitung prosentase yang didapatkan pada setiap klasifikasi prediksi ekspresi wajah, dimana hasil untuk akurasi sebesar 66%, presisi rerata sebesar 76%, recall rerata sebesar 65%, dan score F1 sebesar 63%.

Usulan dalam mengembangkan keberlanjutan dari hasil riset kami dengan memberikan saran yang memungkinkan untuk memperbaiki hasil riset ini, seperti hasil deteksi ekspresi yang lebih akurat dibutuhkan intensitas cahaya yang cukup dengan jarak yang optimal. Selain itu, diharapkan adanya penambahan dataset baru khususnya kategori jikik pada klasifikasi ekspresi wajah, serta jumlah dataset yang lebih banyak untuk mendapatkan akurasi pendeteksi ekspresi yang lebih tinggi.

DAFTAR PUSTAKA

- Alamsyah, D., & Pratama, D. (2020). Implementasi Convolutional Neural Networks (CNN) untuk Klasifikasi Ekspresi Citra Wajah pada FER-2013 Dataset. *Jurnal Teknologi Informasi*, 4(2), 350–355. <https://doi.org/10.36294/jurti.v4i2.1714>
- Carrier, P.-L., & Courville, A. (2013). The Facial Expression Recognition 2013 (FER-2013) Dataset. Retrieved from Wolfram Data Repository website: <https://datarepository.wolframcloud.com/resources/FER-2013>
- Chollet, F. (2016). Xception: Deep Learning with Depthwise Separable Convolutions. *CoRR*, abs/1610.0. Retrieved from <http://arxiv.org/abs/1610.02357>
- Dargham, J. A., Chekima, A., & Mounq, E. G. (2012). Fusing Facial Features for Face Recognition. *International Journal of Interactive Multimedia and Artificial Intelligence*, 1(5), 54. <https://doi.org/10.9781/ijimai.2012.157>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, abs/1512.0, 770–778. IEEE Computer Society. <https://doi.org/10.1109/CVPR.2016.90>
- Henderson, R., & Rothe, R. (2017). Picasso: A Modular Framework for Visualizing the Learning Process of Neural Network Image Classifiers. *Journal of Open Research Software*, 5(1), 22. <https://doi.org/10.5334/jors.178>
- Jain, V., Lamba, P. S., Singh, B., Namboothiri, N., & Dhall, S. (2019). Facial expression recognition using feature level fusion. *Journal of Discrete Mathematical Sciences and Cryptography*, 22(2), 337–350. <https://doi.org/10.1080/09720529.2019.1582866>
- Janku, P., Koplik, K., Dulik, T., & Szabo, I. (2016). Comparison of tracking algorithms implemented in OpenCV. *MATEC Web of Conferences*, 76. <https://doi.org/10.1051/mateconf/20167604031>
- Ko, S. G., Lee, T. H., Yoon, H. Y., Kwon, J. H., & Mather, M. (2011). How Does Context Affect Assessments of Facial Emotion? The Role of Culture and Age. *Psychology and Aging*, 26(1), 48–59. <https://doi.org/10.1037/a0020222>
- Kurniawan, A. L., Isnanto, R. R., Zahra, A. A., Kunci, K., Wajah, P., Komputer, V., ... Euclidean, J. (2015). Perancangan sistem Pengenalan Wajah Menggunakan Metode Ekstraksi Ciri Susunan Tapis Wavelet Gabor 2D Dengan Jarak Euclidean. *Transient: Jurnal Ilmiah Teknik Elektro*, 4(1), 39–43.
- Musa, P., Wibowo, E. P., Musa, S. B., & Baihaqi, I. (2022). Pelican Crossing System for Control a Green Man Light with Predicted Age. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 21(2), 293–306. <https://doi.org/10.30812/matrik.v21i2.1508>
- Musa, P., Yuliani, N., & Lamsani, M. (2009). Rancang Bangun Pengendali Pintu Otomatis Berdasarkan Ciri Wajah Menggunakan Metode Euclidean Distance Dan Fuzzy C-mean. *Jurnal Ilmiah Informatika Komputer Universitas Gunadarma*, 13(1), 35.
- Musa, S. B., & Tjandrasa, H. (2017). Analisis Fitur Sinyal Emosi EEG Berdasarkan Hybrid Decomposition. *ENERGY*, 7(1), 7–12.
- Mustafa, R., Min, Y., & Zhu, D. (2014). Obscenity detection using haar-like features and gentle Adaboost classifier. *Scientific World Journal*, 2014. <https://doi.org/10.1155/2014/753860>
- Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: Holistic versus analytic perception.

- Trends in Cognitive Sciences*, 9(10), 467–473.
<https://doi.org/10.1016/j.tics.2005.08.004>
- Ramadhani, A. L., Musa, P., & Wibowo, E. P. (2017). Human face recognition application using PCA and eigenface approach. *2nd International Conference on Informatics and Computing (ICIC)*, ICIC 2017, 1–5.
<https://doi.org/10.1109/IAC.2017.8280652>
- Righart, R., & de Gelder, B. (2008). Rapid influence of emotional scenes on encoding of facial expressions: An ERP study. *Social Cognitive and Affective Neuroscience*, 3(3), 270–278.
<https://doi.org/10.1093/scan/nsn021>
- Righart, R., & De Gelder, B. (2006). Context influences early perceptual analysis of faces - An electrophysiological study. *Cerebral Cortex*, 16(9), 1249–1257.
<https://doi.org/10.1093/cercor/bhj066>
- Righart, R., & De Gelder, B. (2008). Recognition of facial expressions is influenced by emotional scene gist. *Cognitive, Affective and Behavioral Neuroscience*, 8(3), 264–272.
<https://doi.org/10.3758/CABN.8.3.264>
- Simonyan, K., & Zisserman, A. (2014a). Two-stream convolutional networks for action recognition in videos. *Advances in Neural Information Processing Systems*.
- Simonyan, K., & Zisserman, A. (2014b). Very Deep Convolutional Networks for Large-Scale Image Recognition. *3rd International Conference on Learning Representations, ICLR 2015, abs/1409.1*. Retrieved from <http://arxiv.org/abs/1409.1556>
- Springenberg, J. T., Dosovitskiy, A., Brox, T., & Riedmiller, M. A. (2014). Striving for Simplicity: The All Convolutional Net. *3rd International Conference on Learning Representations, ICLR 2015, abs/1412.6*. Retrieved from <http://arxiv.org/abs/1412.6806>
- Tang, Y. (2015). Deep Learning using Linear Support Vector Machines. *ArXiv:1306.0239*.
- Vadinský, O. (2018). An Overview of Approaches Evaluating Intelligence of Artificial Systems. *Acta Informatica Pragensia*, 7(1), 74–103.
<https://doi.org/10.18267/j.aip.115>
- Vandana, & Marriwala, N. (2022). Facial Expression Recognition Using Convolutional Neural Network. *Lecture Notes in Networks and Systems*, 339, 605–617.
https://doi.org/10.1007/978-981-16-7018-3_45
- Xia, J., Xie, F., Zhang, Y., & Caulfield, C. (2013). Artificial intelligence and data mining: Algorithms and applications. *Abstract and Applied Analysis*.
<https://doi.org/10.1155/2013/524720>
- Xia, X.-L., Xu, C., & Nan, B. (2017). Facial Expression Recognition Based on TensorFlow Platform. *ITM Web of Conferences*, 12, 01005.
<https://doi.org/10.1051/itmconf/20171201005>
- Xie, D., Zhang, L., & Bai, L. (2017). Deep Learning in Visual Computing and Signal Processing. *Applied Computational Intelligence and Soft Computing*, Vol. 2017.
<https://doi.org/10.1155/2017/1320780>
- Zahara, L., Musa, P., Prasetyo Wibowo, E., Karim, I., & Bahri Musa, S. (2020). The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi. *5th International Conference on Informatics and Computing (ICIC)*, 7.
<https://doi.org/10.1109/ICIC50835.2020.9288560>
- Zhang, Y., Balochian, S., Agarwal, P., Bhatnagar, V., & Housheya, O. J. (2014). Artificial intelligence and its applications. *Mathematical Problems in Engineering*.
<https://doi.org/10.1155/2014/840491>