Penerapan *Convolutional Neural Network* Untuk Klasifikasi Citra Ekspresi Wajah Manusia Pada MMA *Facial Expression Dataset*

Tinaliah^{*1}, Triana Elizabeth²

^{1,2}Universitas Multi Data Palembang; Jl Rajawali No 14 Palembang, 0711-376400

¹Program Studi Manajemen Informatika, Fakultas Ilmu Komputer dan Rekayasa, Palembang

²Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Rekayasa, Palembang

e-mail: ^{*1}tinaliah@mdp.ac.id, ²trianaelizabeth@mdp.ac.id

Abstrak

Ekspresi wajah manusia secara umum mewakili emosi atau perasaan yang sedang dirasakannya saat itu. Klasifikasi citra ekspresi wajah dapat membantu untuk mengetahui apakah emosi yang sedang dirasakan seseorang. CNN adalah jenis neural network yang digunakan untuk mengekstrak fitur – fitur dari sebuah citra dan sangat unggul apabila diterapkan pada data citra. Pada penelitian ini akan dilakukan klasifikasi citra ekspresi wajah dengan menerapkan Convolution Neural Network pada dataset MMA Facial Expression. Dimana data akan dibagi menjadi 2 kelas, yaitu happy dan sad. Pengujian dilakukan menggunakan data testing untuk masing – masing kelas dari model CNN yang telah dibuat menggunakan optimizer yang telah ditentukan, yaitu : Adadelata, Adagrad, Adam, Adamax, Nadam, Rmsprop, dan SGD. Berdasarkan hasil pengujian dapat disimpulkan bahwa CNN dapat melakukan klasifikasi citra ekspresi wajah manusia dengan baik menggunakan optimizer SGD dengan nilai akurasi, yaitu 63%.

Kata kunci—Ekspresi Wajah, CNN, Optimizer

Abstract

Human facial expressions generally represent the emotions or feelings they are feeling at that time. Classification of facial expression images can help to find out what emotions a person is feeling. CNN is type of neural network is used to extract features from an image and is very superior if implemented on image data. In this study, facial expression image classification will be carried out by applying the Convolution Neural Network to the MMA Facial Expression dataset. Where the data will be divided into 2 classes, namely happy and sad. The test is carried out using testing data for each class of the CNN model that has been created using a predetermined optimizer, namely: Adadelata, Adagrad, Adam, Adamax, Nadam, Rmsprop, and SGD. Based on the test results, it can be concluded that CNN can classify images of human facial expressions well using the SGD optimizer with an accuracy value of 63%.

Keywords—Facial Expressions, CNN, Optimizer

1. PENDAHULUAN

Setiap manusia memiliki emosi yang diwakili oleh ekspresi wajah yang ditampilkan. Melalui ekspresi wajah, manusia dapat menyampaikan emosi atau perasaannya yang dirasakannya saat itu. Manusia memiliki 6 ekspresi dasar, yaitu: senang, sedih, terkejut, takut, marah, dan jijik [1].

Pengenalan ekspresi wajah digunakan untuk mengetahui emosi yang sedang dirasakan seseorang melalui ekspresi yang ditampilkannya. Pengenalan ekspresi wajah saat ini diterapkan pada bidang teknologi khususnya pada bidang pelayanan jasa, yaitu dalam hal mengetahui tingkat kepuasan pelanggan selama berbelanja. Melalui kamera cctv, dapat diketahui apakah pelanggan merasa senang saat berbelanja atau merasa kurang puas atas pelayanan yang diberikan.

Namun tidak semua emosi bisa terlihat dengan jelas melalui ekspresi wajahnya. Contohnya ada orang yang mempunyai ekspresi wajah datar atau dapat menyembunyikan perasaannya. Klasifikasi citra ekspresi wajah digunakan untuk mengetahui apakah emosi yang sedang dirasakan melalui ekspresi wajahnya dengan menggunakan data training yang dilatih dan data testing yang digunakan untuk dilakukan uji coba termasuk dalam kategori apa ekspresi yang muncul melalui citra gambar.

CNN adalah jenis *neural network* yang digunakan untuk mengekstrak fitur – fitur dari sebuah citra dan sangat unggul apabila diterapkan pada data citra [2]. CNN didesain dari pengembangan *MultilayerPerceptron* (MLP) yang digunakan untuk mengelola data dua dimensi.

Penelitian [3] melakukan analisis perbandingan klasifikasi citra cuaca dengan menggunakan algoritma CNN, SVM, dan KNN. Data yang digunakan adalah citra cuaca berjumlah 1120 citra yang terbagi menjadi 4 kelas, yaitu *cloudy*, *rain*, *shine*, dan *sunrise*. Percobaan yang dilakukan adalah membandingkan performa dari algoritma CNN, SVM, dan KNN, dengan melakukan pelatihan pada model yang dibuat. Proses pelatihan ini menggunakan 80% dari dataset, dan testing menggunakan 20% dataset. Dapat disimpulkan bahwa CNN memiliki akurasi yang baik dalam melakukan klasifikasi dataset cuaca dengan akurasi sebesar 0.942 dibandingkan dengan algoritma SVM, dan KNN.

Penelitian [4] melakukan identifikasi karakter hiragana dengan menggunakan CNN. Data yang digunakan adalah citra data hiragana yang berjumlah 1000 citra. Percobaan yang dilakukan adalah melakukan klasifikasi huruf hiragana menggunakan metode *convolutional neural network* berdasarkan data yang diuji apakah sesuai atau tidak. Dapat disimpulkan bahwa akurasi yang didapat adalah sebesar 82% dalam melakukan identifikasi karakter hiragana.

Penelitian [5] melakukan klasifikasi ekspresi citra wajah menggunakan dataset *Facial Expression Recognition* 2013 berjumlah 200 buah data yang terbagi menjadi 2 (dua) kelas, yaitu *happy* dan *sad*. Percobaan yang dilakukan adalah melakukan klasifikasi citra wajah dengan metode CNN dengan menggunakan 4 jenis *optimizer*, yaitu Adam, Adamax, N-adam, dan SGD. Dapat disimpulkan bahwa CNN dapat mengenali emosi melalui citra wajah dengan menggunakan optimizer Adamax sebesar 66%.

1.1 Convolutional Neural Network (CNN)

CNN didesain dari pengembangan *MultilayerPerceptron* (MLP) yang digunakan untuk mengelola data dua dimensi. Tahap yang dilakukan terbagi menjadi 2 (dua) bagian, yaitu ekstraksi fitur dan klasifikasi citra. Secara umum model CNN terdiri dari 4 komponen, yaitu *convolution layer*, *pooling layer*, fungsi aktivasi, dan *fully connected layer* [6]. Model CNN dapat dilihat pada Gambar 1.

Gambar 1. Model CNN [7]

1. 2 Convolutional Layer

Convolutional Layer adalah tahap yang dilakukan untuk melakukan filter terhadap citra, dimana filter itu akan bergeser ke seluruh bagian gambar, dan menghasilkan output yang disebut *feature map*. Proses *convolution layer* dapat dilihat pada Gambar 2.

Gambar 2. Convolutional Layer [8]

1. 3 Pooling Layer

Pooling layer dilakukan untuk mengurangi dimensi *feature map* sehingga mempercepat komputasi. Pada penerapannya *pooling layer* yang biasa digunakan adalah *max pooling*. *Max pooling* akan mengambil nilai terbesar untuk menyusun matriks baru berdasarkan citra yang telah direduksi.

Gambar 3. Proses Max Pooling [9]

1. 5 Fungsi Aktivasi

Fungsi aktivasi adalah fungsi non-linier yang membuat sebuah jaringan dapat mentransformasi data input ke dimensi yang lebih tinggi yang memungkinkan dilakukan klasifikasi [9]. Pada CNN terdapat fungsi aktivasi digunakan, yaitu fungsi sigmoid. Fungsi sigmoid berfungsi mentransformasi nilai dari input x menjadi antara 0 dan 1 dengan bentuk distribusi fungsi. Fungsi sigmoid memiliki bentuk, yaitu:

$$\sigma(x) = \frac{1}{(1 + e^{-x})} \quad (1)$$

1. 6 Fully Connected Layer

Fully connected layer yang bertujuan untuk menghubungkan semua neuron serupa seperti yang dimiliki oleh *MultilayerPerceptron* (MLP), yang bertujuan untuk mentransformasi dimensi data agar data dapat diklasifikasikan secara linear. Proses *fully connected layer* dapat dilihat pada Gambar 4.

Gambar 4. Proses *Fully Connected Layer*

2. METODE PENELITIAN

Pada penelitian ini metode penelitian yang digunakan adalah metode eksperimental. Dimana pengembangan sistem dilakukan dengan tahapan sebagai berikut:

1. Pengumpulan Data

Data citra yang digunakan adalah citra ekspresi wajah berjumlah 200 buah citra RGB yang diperoleh dari dataset MMA Facial Expression [10]. Citra ekspresi wajah berukuran 48 x 48 piksel yang terbagi dalam 2 (dua) kelas, yaitu *happy* dan *sad* masing – masing berjumlah 100 data citra RGB. Data citra yang dikumpulkan akan terbagi menjadi 80% data latih, dan 20% data uji. Dataset dapat dilihat pada Gambar 5.

Gambar 5. Sampel MMA Facial Expression Dataset

2. Perancangan Sistem

Model CNN yang akan dibangun terdiri dari 3 (tiga) *convolutional layer* dengan fitur berukuran 3x3 dan menggunakan fungsi aktivasi ReLU dan *max pooling* dengan kernel berukuran 2x2. Kemudian dilakukan proses *flatten*. Terdapat 2 dense layer dengan fungsi aktivasi ReLU dan sigmoid, kemudian akan didapatkan hasil output berupa 1 kelas dengan fungsi aktivasi sigmoid. Model CNN yang dibangun dapat dilihat pada Gambar 6.

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 46, 46, 32)	896
max_pooling2d (MaxPooling2D)	(None, 23, 23, 32)	0
conv2d_1 (Conv2D)	(None, 21, 21, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(None, 10, 10, 64)	0
conv2d_2 (Conv2D)	(None, 8, 8, 128)	73856
max_pooling2d_2 (MaxPooling2D)	(None, 4, 4, 128)	0
flatten (Flatten)	(None, 2048)	0
dense (Dense)	(None, 512)	1049088
dense_1 (Dense)	(None, 1)	513
Total params: 1,142,849		
Trainable params: 1,142,849		
Non-trainable params: 0		

Gambar 6. Model CNN

3. Implementasi dan uji coba

Model yang dirancang akan diimplementasikan ke dalam bahasa pemrograman python dengan menggunakan Google Colab. Optimizer yang akan digunakan ada 7 (tujuh), yaitu : Adadelta, Adagrad, Adam, Adamax, Nadam, Rmsprop, dan SGD. Kemudian dilakukan uji coba terhadap model yang telah diimplementasikan dengan menggunakan data testing yang telah disiapkan.

4. Analisis hasil uji coba

Berdasarkan uji coba yang telah dilakukan, maka akan dilakukan analisis menggunakan *confussion matrix* dengan menggunakan nilai *accuration*, *presicion*, dan *recall*. Tabel *confussion matrix* dapat dilihat pada Tabel 1.

Tabel 1. *Confussion Matrix*

Prediksi \ Sebenarnya	True	False
True	TP	FP
False	FN	TN

dimana:

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

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

3. HASIL DAN PEMBAHASAN

Citra ekspresi wajah berukuran 48 x 48 piksel akan dikonvolusi pada convolution layer menggunakan fitur berukuran 3x3 dan fungsi aktivasi ReLU. Proses ini dilakukan sehingga menghasilkan 32 fitur. Tujuan dari proses konvolusi ini adalah melakukan ekstrasi fitur dari citra ekspresi wajah. Lalu dilakukan proses pooling menggunakan metode *max pooling* berukuran 2x2. Tujuannya adalah mengurangi dimensi dengan operasi *down-sampling*.

Langkah selanjutnya dilakukan proses konvolusi lagi menggunakan kernel berukuran 3x3 dan fungsi aktivasi ReLU, sehingga menghasilkan 64 fitur. Lalu proses *max pooling* berukuran 2x2.

Langkah selanjutnya dilakukan proses konvolusi lagi menggunakan kernel berukuran 3x3 dan fungsi aktivasi ReLU, sehingga menghasilkan 128 fitur. Lalu proses *max pooling* berukuran 2x2.

Langkah selanjutnya adalah melakukan proses flatten sehingga hasil ekstrasi fitur yang dilakukan dapat menjadi sebuah vektor berukuran 2048, agar bisa menjadi input dari *fully connected layer*. Selanjutnya diubah menjadi 512 fitur yang kemudian akan didapatkan hasil output berupa 1 kelas dengan fungsi aktivasi sigmoid.

Proses training model CNN akan dilakukan sebanyak 20 epoch. Pengujian dilakukan menggunakan data testing untuk masing – masing kelas dari model CNN yang telah dibuat menggunakan optimizer yang telah ditentukan, yaitu : Adadeltra, Adagrad, Adam, Adamax, Nadam, Rmsprop, dan SGD. Hasil dari pengujian untuk setiap model dari optimizer dapat dilihat pada Tabel 2.

Tabel 2. Confusion Matrix Adadeltra,
Adagrad, Adam, Adamax, Nadam,
Rmsprop, SGD Optimizer Adadeltra
Optimizer

Sebenarnya Prediksi	Happy	Sad
Happy	18	2
Sad	20	0

Adagrad Optimizer

Sebenarnya Prediksi	Happy	Sad
Happy	18	2
Sad	18	2

Adam Optimizer

Prediksi \ Sebenarnya	Happy	Sad
Happy	17	3
Sad	14	6

Adamax Optimizer

Prediksi \ Sebenarnya	Happy	Sad
Happy	16	4
Sad	15	5

Nadam Optimizer

Prediksi \ Sebenarnya	Happy	Sad
Happy	11	9
Sad	14	6

Rmsprop Optimizer

Prediksi \ Sebenarnya	Happy	Sad
Happy	5	15
Sad	4	16

SGD Optimizer

Prediksi \ Sebenarnya	Happy	Sad
Happy	18	2
Sad	13	7

Tabel 3. Perbandingan Akurasi, Precision, dan Recall

Optimizer	Akurasi	Happy		Sad	
		Precision	Recall	Preission	Recall
Adadelta	0.45	0.47	0.90	0.00	0.00
Adagrad	0.50	0.50	0.90	0.50	0.10
Adam	0.58	0.55	0.85	0.67	0.30
Adamax	0.53	0.52	0.80	0.56	0.25
Nadam	0.43	0.44	0.55	0.40	0.30
Rmsprop	0.53	0.56	0.25	0.52	0.80
SGD	0.63	0.58	0.90	0.78	0.35

Berdasarkan hasil akurasi, precision, dan recall untuk setiap optimizer yang ditunjukkan pada Tabel 3, dapat dilihat bahwa:

1. SGD Optimizer merupakan optimizer yang mempunyai tingkat akurasi sebesar 63% lebih baik dibandingkan optimizer lainnya dalam melakukan klasifikasi citra ekspresi wajah. Dimana lebih cenderung berhasil mendeteksi happy daripada sad. Hal ini dapat dilihat bahwa nilai recall untuk kelas happy lebih besar dari pada kelas sad.
2. Nadam Optimizer merupakan optimizer yang mempunyai tingkat akurasi yang kurang baik dibandingkan optimizer lainnya sebesar 43% dalam melakukan klasifikasi citra ekspresi

wajah. Dimana lebih cenderung berhasil mendeteksi happy daripada sad. Hal ini dapat dilihat bahwa nilai recall untuk kelas happy lebih besar dari pada kelas sad.

3. Adadelta Optimizer mampu melakukan klasifikasi citra ekspresi wajah sebesar 45%. Dimana berhasil mendeteksi happy daripada sad. Hal ini dapat dilihat bahwa nilai recall untuk kelas sad = 0

4. KESIMPULAN

Berdasarkan hasil pengujian dapat disimpulkan bahwa:

1. CNN mengklasifikasikan citra ekspresi wajah manusia dengan baik menggunakan optimizer SGD dengan nilai akurasi, yaitu 63%.
2. Optimizer Nadam menghasilkan akurasi yang kurang baik dibandingkan optimizer lainnya dalam melakukan klasifikasi citra ekspresi wajah manusia.

5. SARAN

Penerapan convolutional neural network untuk klasifikasi citra ekspresi wajah manusia pada MMA Facial Expression Dataset dapat dikembangkan lebih lanjut dengan menambahkan arsitektur CNN, seperti AlexNet, VGG16, atau GoogleNet.

DAFTAR PUSTAKA

- [1] V. Triyanti, Y. Yassierli, and H. Iridiastadi, 2019, "Basic Emotion Recogniton Using Automatic Facial Expression Analysis Software," *J. Optimasi Sist. Ind.*, Vol. 18, No. 1, p. 55, doi: 10.25077/josi.v18.n1.p55-64.2019.
- [2] I. P. A. E. D. Udayana and I. K. D. G. Supartha, 2021, "Implementasi Kombinasi Metode Mean Denoising dan Convolutional Neural Network pada Facial Landmark Detection," *J. Nas. Pendidik. Tek. Inform.*, Vol. 10, No. 1, p. 1, doi: 10.23887/janapati.v10i1.29779.
- [3] M. Farid Naufal, 2021, "Perbandingan, Analisis Svm, Algoritma Untuk, dan CNN," *J. Teknol. Inf. dan Ilmu Komput.*, Vol. 8, No. 2, pp. 311–318, doi: 10.25126/jtiik.202184553.
- [4] C. Umam and L. Budi Handoko, 2020, "Convolutional Neural Network (CNN) Untuk Identifikasi Karakter Hiragana," *Pros. Semin. Nas. Lppm Ump*, Vol. 0, No. 0, pp. 527–533, [Online]. Available: <https://semnaslppm.ump.ac.id/index.php/semnaslppm/article/view/199>.
- [5] D. Alamsyah and D. Pratama, 2020, "Implementasi Convolutional Neural Networks (CNN) Untuk Klasifikasi Ekspresi Citra Wajah pada FER-2013 Dataset," *JurTI (Jurnal Teknol. Informasi)*, Vol. 4, No. 2, pp. 350–355.
- [6] S. Indolia, A. Kumar, S. P. Mishra, and P. Asopa, 2018, "Science Direct Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach," *Procedia Comput. Sci.*, vol. 132, pp. 679–688, doi: 10.1016/j.procs.2018.05.069.

- [7] “A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way | by Sumit Saha | Towards Data Science.” <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53> (accessed Sep. 26, 2021).
- [8] M. A. Hossain and M. S. Alam Sajib, 2019, “*Classification of Image Using Convolutional Neural Network (CNN)*,” *Glob. J. Comput. Sci. Technol.*, Vol. 19, No. 2, pp. 13–18, doi: 10.34257/gjcstdvol19is2pg13.
- [9] W. S. Eka Putra, “*Klasifikasi Citra Menggunakan Convolutional Neural Network (CNN) pada Caltech 101*,” *J. Tek. ITS*, Vol. 5, No. 1, 2016, doi: 10.12962/j23373539.v5i1.15696.
- [10] “MMA Facial Expression | Kaggle.” <https://www.kaggle.com/mahmoudima/mma-facial-expression> (accessed Sep. 1, 2021).