

IMPLEMENTASI CONVOLUTIONAL NEURAL NETWORKS (CNN) UNTUK KLASIFIKASI EKSPRESI CITRA WAJAH PADA FER-2013 DATASET

Derry Alamsyah, Dicky Pratama

Program Studi Teknik Informatika, STMIK GI MDP, Program Studi Sistem Informasi, STMIK GI MDP

Jl. Rajawali No. 14 Palembang

derry@mdp.ac.id, dqpratama@mdp.ac.id

Abstract - session recognition is an interesting topic, where facial expressions in today's technological advances can support several fields such as health, business, and so on. Facial expression recognition can be done using the extraction of certain features. Meanwhile, Convolutional Neural Network (CNN) can recognize an object in the image through the features found by itself in the convolution process. By using CNN's advantages, this study aims to see CNN's performance in facial expressions of happiness and sadness in unideal data conditions. So based on this research, on the FER2013 dataset, CNN using the Adamax optimizer produced a fairly good performance where the value is given is 66% compared to Adam, N-Adam, and SGD.

Keywords - CNN, Facial Expression, FER-2013

Abstrak - Pengenalan ekspresi merupakan topik penelitian yang menarik, dimana peran ekspresi wajah dalam kemajuan teknologi saat ini dapat mendukung beberapa bidang seperti kesehatan, bisnis, dan sebagainya. Pengenalan wajah dapat dilakukan dengan menggunakan ekstraksi fitur-fitur tertentu. Sementara itu, Convolutional Neural Network (CNN) dapat mengenali objek pada citra melalui fitur yang ditemukannya sendiri dalam proses konvolusinya. Dengan menggunakan keunggulan CNN, maka penelitian ini bertujuan untuk mengetahui performa CNN dalam mengenali ekspresi wajah bahagia (happy) dan sedih (sad) pada kondisi data tidak ideal. Maka berdasarkan hasil penelitian ini, pada dataset FER2013, CNN dengan menggunakan Adamax optimizer menghasilkan performa yang cukup baik dimana akurasi yang diberikan adalah sebesar 66% dibandingkan dengan Adam, N-Adam, dan SGD.

Kata Kunci - CNN, Ekspresi Wajah, FER-2013.

I. PENDAHULUAN

Wajah merupakan bagian tubuh yang dapat mengindikasikan emosi seseorang melalui ekspresinya, apakah ia bahagia (*happy*), sedih (*sad*), marah (*angry*), jijik (*discusting*), terkejut (*sureprise*) ataupun biasa saja (*neutral*) [1]. Selain itu wajah memegang porsi yang besar dalam mengenali ekspresi seseorang dibandingkan dengan suara dan gestur tubuh lainnya [2].

Pengenalan ekspresi wajah menjadi penting seiring dengan meningkatnya teknologi yang ada pada komputer, telepon genggam, robot dan sebagainya. Peningkatan tersebut membuat interaksi antara manusia dengan teknologi menjadi semakin tidak terelakkan. Sebagai contoh sistem belanja online saat ini. Dengan menggunakan ekspresi wajah, maka dapat dibuat sistem yang mampu mengenali kepuasan pelanggan salah satunya. Selain dari itu, masih terdapat peran lain yaitu pada bidang kesehatan psikologi, interaksi manusia dan komputer (robot), dan sebagainya.

Data wajah dapat diambil melalui media kamera yang menghasilkan citra atau serangkaian citra (video). Kemudian citra wajah dapat dikelompokkan menjadi dua kelompok yaitu citra wajah dengan kondisi ideal (latar belakang homogen, posisi standar

yang umumnya dihasilkan dari suatu proses di studio atau lab [3][4][5] dan kondisi sebaliknya (tidak ideal) [1]. Sementara itu pengenalan wajah melalui citra dapat dilakukan dengan menggunakan sejumlah metode diantaranya menggunakan fitur lokal [3] atau global [6], fitur bentuk [4], ataupun fitur piksel citra itu sendiri (*raw image feature*) [7][8][9].

Pengenalan wajah melalui piksel dapat dilakukan dengan menggunakan metode Convolutional Neural Network (CNN). Dimana fitur wajah dapat ditemukan melalui serangkaian proses konvolusi yang dilakukan pada citra. Hal ini tentunya berbeda dengan penggunaan ekstraksi fitur lainya dimana harus menentukan bentuk ekstraksi fitur apa yang dapat menjelaskan objek (wajah). Oleh karena itu pada penelitian ini digunakan metode CNN untuk mengenali ekspresi wajah pada citra

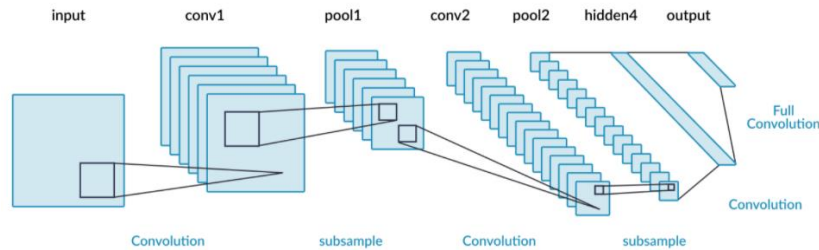
A. Convolutionan Neural Network (CNN)

Merupakan salah satu bagian dari *deep feed-forward artificial neural networks* yang banyak diaplikasi pada visi komputer (*computer vision*). Metode ini dikenal juga dengan sebutan ConvNet. Serupa dengan *artificial neural network* (ANN), CNN memiliki arsitektur yang tersusun dari *node-node* atau *neuron-neuron* yang saling terhubung pada suatu layer. Akan tetapi, berbeda dengan ANN, metode ini

menggunakan proses konvolusi dalam menemukan fitur-fitur pada citra.

Penggunaan yang meluas dari CNN dalam visi komputer ditunjukkan dengan banyaknya arsitektur yang disusun untuk sejumlah masalah. Beberapa arsitektur yang umum digunakan adalah AlexNet,

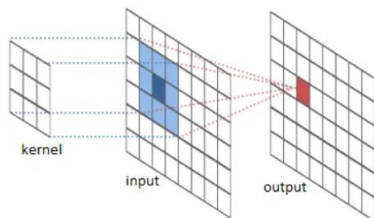
VGG, Google Inception, dsb. Sementara itu arsitektur CNN itu sendiri terdiri dari Lapis tersembunyi (*Hidden Layer*) yang berisi *convolutional layers*, *pooling layers*, *normalization layers*, *ReLU layer*, *fully connected layers* dan *loss layer* [10].



Gambar 1. Convolutional Neural Networks

B. Convolutional Layer

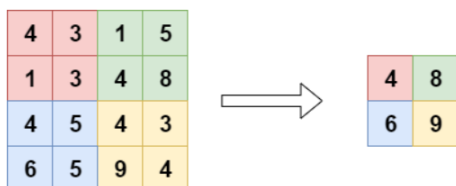
Merupakan bagian utama dari CNN, karena sebagian besar komputasi pada CNN dilakukan di lapis ini. Operasi yang dilakukan sama dengan operasi konvolusi yang biasa dilakukan pada pengolahan citra (*Image Processing*), dimana terdapat kernel dan sub citra. Kernel yang digunakan pada CNN umumnya berukuran 3x3. Kemudian untuk setiap sub citra yang berukuran sama dengan kernel dilakukan operasi konvolusi.



Gambar 2. Convolutional Layer

C. Pooling Layer

Merupakan bagian yang digunakan untuk mereduksi ukuran citra (*down sampling*) yang bertujuan untuk memudahkan operasi konvolusi pada lapis berikutnya. Proses ini dapat dilakukan dengan beberapa teknik, diantaranya adalah *max pooling* (mengambil nilai maximum dari sub citra), *average pooling* (mengambil nilai rata-rata dari sub citra), dsb. Sub citra yang digunakan pada operasi ini berukuran 8x8.



Gambar 3. Pooling Layer (*max pooling*)

D. Normalization Layer

Lapis ini ditujukan untuk mengatasi perbedaan rentang nilai yang terdapat pada citra masukan. Beberapa teknik normalisasi dapat dilakukan, namun

yang paling umum adalah dengan mengubahnya kedalam skala 0-1 (berdasarkan rentang minimum dan maksimum). Sebagai contoh untuk citra berukuran 8 bit maka dapat dilakukan normalisasi berupa

$$N(x, y) = I(x,y)/255 \tag{1}$$

dimana nilai maksimumnya adalah 255 sementara minimumnya adalah 0.

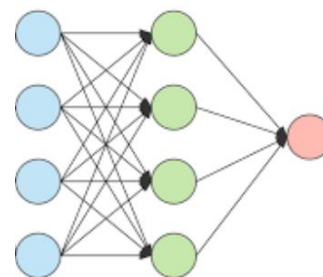
E. ReLu Layer

Lapis *Rectified Linear Unit* (ReLU) merupakan lapis yang bertujuan untuk meningkatkan nonlinearitas fungsi keputusan. Fungsi aktivasi yang digunakan adalah

$$f(x) = \max(0, x) \tag{2}$$

F. Fully Connected Layer

Pada lapis ini semua neuron dihubungkan secara keseluruhan (*fully connected*) serupa seperti yang dimiliki oleh *multi layer perceptron* (MLP). Pada bagian ini terdapat bobot dan bias yang digunakan untuk memodelkan data seperti yang dilakukan oleh MLP.



Gambar 4. Fully Connected Layer

G. Loss Layer

Lapis ini merupakan lapis akhir pada CNN yang digunakan untuk menentukan penalti terhadap hasil yang tidak sesuai dengan label (*target*) dari proses pelatihan (penentuan bobot dan bias). Terdapat sejumlah fungsi yang dapat digunakan pada lapis ini

diantaranya *sigmoid cross-entropy loss*, *softmax loss*, *euclidean loss* dsb.



Gambar 5. FER-2013 Dataset

II. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian tersusun dari beberapa tahap yaitu:

1. Mengumpulkan Data
Data yang digunakan pada penelitian ini adalah dataset Facial Expression Recognition 2013 (FER-2013). Dimana, citra wajah yang diberikan berukuran 48x48 berupa citra *grayscale*. Jumlah data yang digunakan dalam penelitian ini adalah sebanyak 100 data untuk masing-masing kelas ekspresi (yaitu *happy* dan *sad*). Sampel data ditunjukkan oleh Gambar 5.
2. Pengolahan Data
Data yang dikumpulkan dibagi kedalam tiga bagian untuk masing-masing kelas yaitu data latih (*train data*), data validasi (*validation data*), dan data uji (*test data*) yang dikumpulkan secara acak dengan komposisi 56, 19 dan 25 citra.
2. Merancang Sistem
Rancang sistem yang dibuat dengan arsitektur CNN yang terdiri dari tiga *convolutional* layer dengan masing-masing fungsi aktivasi ReLu dan *max pooling* pada *pooling layer*. Sementara itu terdapat 2 *dense layer* dengan fungsi aktivasi ReLu dan *sigmoid*. Terdapat satu luaran dari model ini dimana kelas ditentukan berdasarkan hasil dari fungsi *sigmoid*. Arsitektur ini

ditunjukkan oleh Gambar 6, dimana kernel yang digunakan berukuran 3x3.

3. Mengimplementasikan Sistem
Implementasi sistem ini menggunakan 4 jenis *optimizer* yaitu Adam, Adamax, *N-adam*, dan SGD. Sementara itu data dibagi kedalam data latih dan validasi dalam proses pelatihnannya dengan menggunakan 10 *epoch*.
4. Evaluasi
Evaluasi pada penelitian ini menggunakan *confussion matrix* dan menggunakan nilai akurasi (*accuracy*), presisi (*precission*), dan *recall*. Dimana tabel *confussion matrix* yang digunakan ditunjukkan oleh Tabel 1 berikut:

Tabel 1. *Confussion Matrix*

| | | Prediksi | |
|----------|-------|----------|-----|
| | | Happy | Sad |
| Masukkan | Happy | TP | FN |
| | Sad | FP | TN |

Dengan

$$Acc = (TP + FN) / (P + N) \tag{3}$$

$$Prec = TP / (TP + FP) \tag{4}$$

$$Rec = TP / P \tag{5}$$

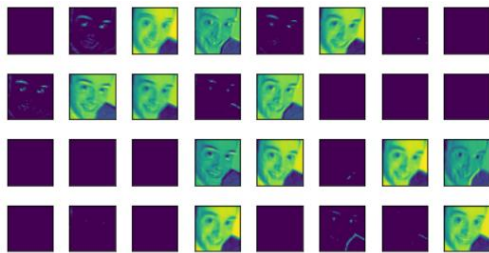
| Layer (type) | Output Shape | Param # |
|--------------------------------|--------------------|---------|
| conv2d (Conv2D) | (None, 46, 46, 32) | 320 |
| max_pooling2d (MaxPooling2D) | (None, 23, 23, 32) | 0 |
| conv2d_1 (Conv2D) | (None, 21, 21, 64) | 18496 |
| max_pooling2d_1 (MaxPooling2D) | (None, 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: 1,142,273
Trainable params: 1,142,273

Gambar 6. Arsitektur CNN

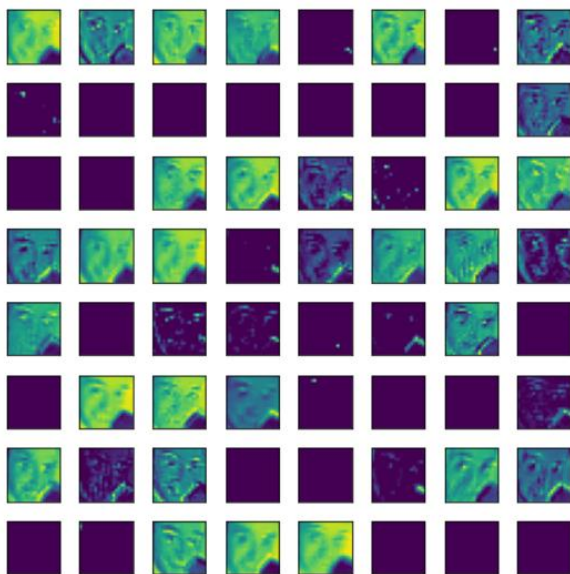
III. HASIL DAN PEMBAHASAN

Dalam melatih model CNN, data yang digunakan dibagi kedalam 2 bagian yaitu data latih dan data validasi yang digunakan untuk masing-masing *epoch*. Data latih dengan ukuran 48x48 pixel dimasukkan kedalam input layer, kemudian dilakukan proses konvolusi pada *convolutional layer* dengan menggunakan kernel berukuran 3x3. Proses ini dilakukan terhadap 32 kernel sehingga menghasilkan 32 fitur seperti yang ditunjukkan oleh Gambar 7.



Gambar 7. Hasil konvolusi pertama

Kemudian dilanjutkan reduksi ukuran fitur melalui *max pooling layer*. Sehingga menghasilkan ukuran 23x23 pixel. Selanjutnya dilakukan kembali proses konvolusi dengan kernel 3x3 sebanyak 64 kali sehingga menghasilkan 64 fitur. Hasil ini ditunjukkan oleh Gambar 8.



Gambar 8. Hasil konvolusi kedua

Selanjutnya kembali dilakukan proses reduksi fitur sehingga menghasilkan citra berukuran 10x10. Proses konvolusi yang terakhir dilakukan sebanyak 128 kali dengan ukuran kernel yang sama yaitu 3x3 dan menghasilkan 128 fitur seperti yang ditunjukkan oleh Gambar 9.

Proses berikutnya kembali mereduksi ukuran fitur menjadi 4x4 yang kemudian diubah menjadi fitur vektor berukuran 2048. Untuk selanjutnya diubah menjadi 512 fitur utama yang kemudian digunakan untuk menentukan nilai kelas pada akhir layer. Selanjutnya, hasil atau performa dari model yang dilatih sebanyak 10 *epoch* ditunjukkan oleh Tabel 2 berikut:

Tabel 2. Akurasi data latih dan validasi

| Optimizer | Data Latih | Data Validasi |
|-----------|------------|---------------|
| Adam | 0.84 | 0.66 |
| Adamax | 0.95 | 0.61 |
| N-Adam | 0.65 | 0.58 |
| SGD | 0.56 | 0.5 |

Sementara itu, pengujian dilakukan terhadap 25 data uji untuk masing-masing ekspresi, dimana hasil dari masing-masing model berdasarkan *optimizer* yang digunakan ditunjukkan oleh Tabel 3, 4, 5 dan 6.

Tabel 3. Adam Optimizer Confusion Matrix

| | Happy | Sad |
|-------|-------|-----|
| Happy | 22 | 3 |
| Sad | 15 | 10 |

Berdasarkan Tabel 3. Jumlah terbanyak yang mampu dikenali adalah milik ekspresi *happy* sebanyak 22 citra dari 25 citra uji. Sementara itu penyimpangan terbanyak dimiliki oleh ekspresi *sad*.

Tabel 4. Adamax Optimizer Confusion Matrix

| | Happy | Sad |
|-------|-------|-----|
| Happy | 17 | 8 |
| Sad | 9 | 16 |

Selanjutnya berdasarkan Tabel 4, jumlah emosi yang dikenali cukupimbang untuk *happy* maupun *sad* yaitu 16 dan 17 citra wajah. Hal tersebut serupa dengan penyimpangan yang dialami untuk keduanya.

Tabel 5. N-Adam Optimizer Confusion Matrix

| | Happy | Sad |
|-------|-------|-----|
| Happy | 3 | 22 |
| Sad | 2 | 23 |

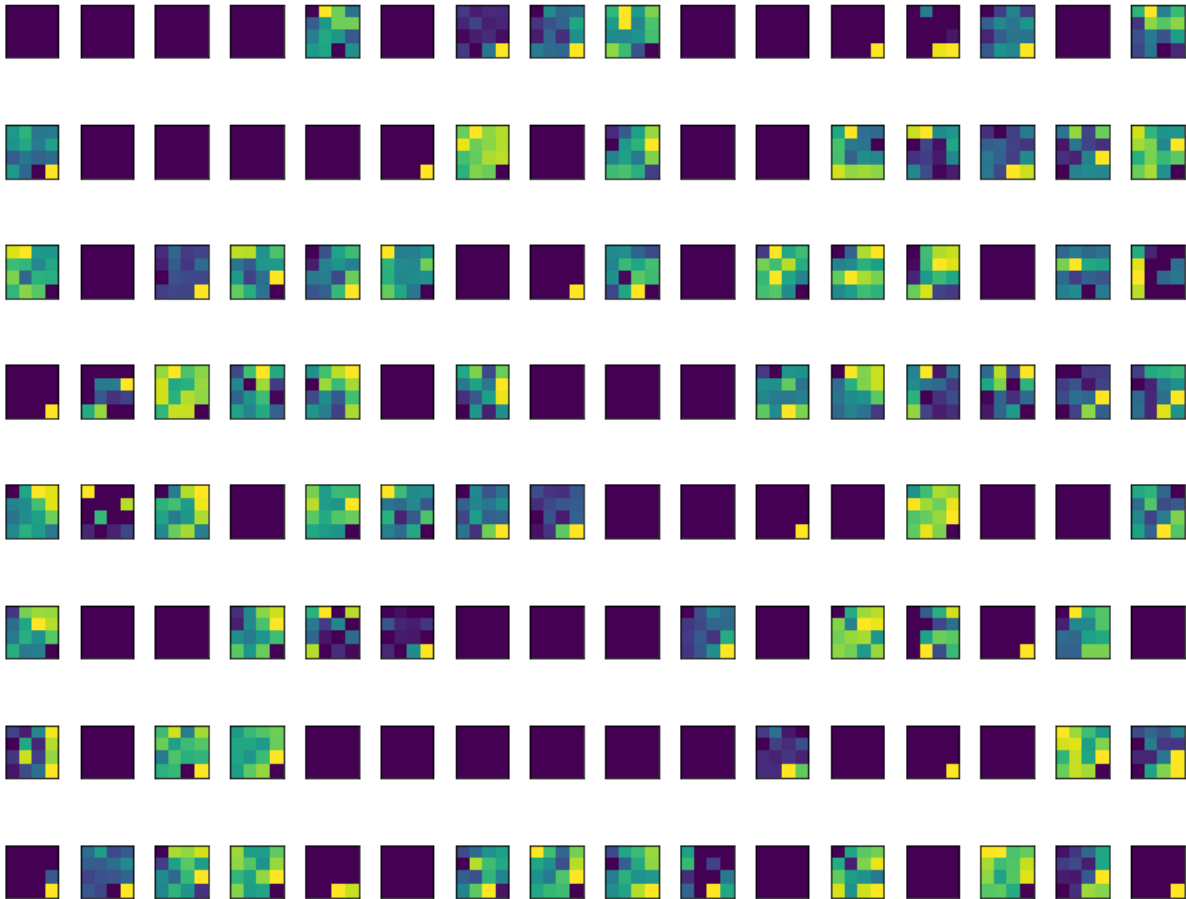
Sementara itu, terdapat kecenderungan mengenali ekspresi *sad* dimana terdapat 22 citra wajah dengan ekspresi *happy* yang dikenali sebagai *sad*. Sementara itu, hampir semua data dengan ekspresi *sad* dikenali yaitu sebanyak 23 citra dari 25 citra wajah yang ditunjukkan oleh Tabel 5.

Tabel 6. SGD Optimizer Confusion Matrix

| | Happy | Sad |
|-------|-------|-----|
| Happy | 10 | 15 |
| Sad | 11 | 14 |

Terakhir untuk berdasarkan Tabel 6, masih cukup banyak data yang belum dikenali dengan tepat yaitu 15

data citra wajah dengan ekspresi *happy* dan 11 data untuk citra wajah dengan ekspresi *sad*.



Gambar 9. Hasil Konvolusi ketiga

Performa CNN ditunjukkan oleh Tabel 7, dimana

1. Untuk Adam *optimizer*, ketepatan CNN dalam mengklasifikasi adalah sebesar 64% dengan kecenderungan mengenali ekspresi *happy* daripada *sad*. Hal tersebut ditunjukkan oleh nilai *recall* untuk ekspresi *happy* yang jauh lebih besar dibandingkan dengan *sad*, dan *precision* dari *sad* tinggi.
2. Untuk Adamax *optimizer*, ketepatan CNN dalam mengklasifikasi adalah sebesar 66% dengan porsi mengenali yangimbang ditunjukkan dengan nilai *recall* dari kedua ekspresi yang cukupimbang.
3. Untuk N-Adam *optimizer*, ketepatan CNN dalam mengklasifikasi adalah sebesar 52% dengan kecenderungan mengenali ekspresi *sad* dari pada *happy*. Hal tersebut ditunjukkan oleh nilai *recall* untuk ekspresi *sad* yang jauh lebih besar dari pada *happy* dan *precision* dari *happy* yang cukup tinggi.
4. Untuk SGD *optimizer*, ketepatan CNN dalam mengklasifikasi adalah sebesar 48%, dimana tidak ada kecenderungan dalam mengklasifikasi. Hal tersebut ditunjukkan

oleh hasil *recall* dari kedua emosi sama-sama belum cukup baik untuk dikenali.

Tabel 7. Hasil Prediksi Ekspresi Wajah

| Opt. | Acc. | Happy | | Sad | |
|--------|------|-------|------|-------|------|
| | | Prec. | Rec. | Prec. | Rec. |
| Adam | 0.64 | 0.59 | 0.92 | 0.77 | 0.4 |
| Adamax | 0.66 | 0.65 | 0.68 | 0.67 | 0.64 |
| N-adam | 0.52 | 0.6 | 0.12 | 0.51 | 0.92 |
| SGD | 0.48 | 0.48 | 0.4 | 0.48 | 0.56 |

IV. KESIMPULAN

Berdasarkan penelitian, implementasi dan pengujian, maka dapat diambil kesimpulan sebagai berikut :

1. Kemampuan CNN dalam mengenali emosi melalui citra wajah paling baik menggunakan Adamax *optimizer* dengan nilai akurasi sebesar 66%.
2. Kemampuan CNN untuk seimbang dalam mengenali kedua ekspresi adalah dengan menggunakan Adamax *optimizer* dengan hasil *recall* untuk *happy* dan *sad* sebesar 68% dan 64%.

3. SGD *optimizer* tidak cukup baik untuk membuat CNN mengenali emosi melalui citra wajah.

Processing Symposium (IDAP), Malatya, 2017, pp. 1-5, doi: 10.1109/IDAP.2017.8090281.

DAFTAR PUSTAKA

- [1] S. S. Shafira, N. Ulfa, H. A. Wibawa and Rismiyati, "Facial Expression Recognition Using Extreme Learning Machine," 2019 3rd International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 2019, pp. 1-6, doi: 10.1109/ICICoS48119.2019.8982443.
- [2] Samadiani N, Huang G, Cai B, Luo W, Chi CH, Xiang Y, He J. A Review on Automatic Facial Expression Recognition Systems Assisted by Multimodal Sensor Data. *Sensors* (Basel). 2019 Apr 18;19(8):1863. doi: 10.3390/s19081863. PMID: 31003522; PMC-ID: PMC6514576.
- [3] S. A. Rizwan, A. Jalal and K. Kim, "An Accurate Facial Expression Detector using Multi-Landmarks Selection and Local Transform Features," 2020 3rd International Conference on Advancements in Computational Sciences (ICACS), Lahore, Pakistan, 2020, pp. 1-6, doi: 10.1109/ICACS47775.2020.9055954.
- [4] Y. He and X. He, "Facial Expression Recognition Based on Multi-Feature Fusion and HOSVD," 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China, 2019, pp. 638-643, doi: 10.1109/ITNEC.2019.8729003.
- [5] Ma, D.S., Correll, J. & Wittenbrink, B. "The Chicago face database: A free stimulus set of faces and norming data". *Behav Res* 47, 1122–1135 2015. <https://doi.org/10.3758/s13428-014-0532-5>
- [6] T. T. Kaleekal and J. Singh, "Facial Expression Recognition using Hahn Moment on Facial Patches," 2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), Bangalore, India, 2019, pp. 1338-1344, doi: 10.1109/RTEICT46194.2019.9016953.
- [7] M. Garcia Villanueva and S. Ramirez Zavala, "Deep Neural Network Architecture: Application for Facial Expression Recognition," in *IEEE Latin America Transactions*, vol. 18, no. 07, pp. 1311-1319, July 2020, doi: 10.1109/TLA.2020.9099774.
- [8] V. Tümen, Ö. F. Söylemez and B. Ergen, "Facial emotion recognition on a dataset using convolutional neural network," 2017 International Artificial Intelligence and Data Processing Symposium (IDAP), Malatya, 2017, pp. 1-5, doi: 10.1109/IDAP.2017.8090281.
- [9] Mollahosseini, A., Chan, D., and Mahoor, M. H., "Going Deeper in Facial Expression Recognition using Deep Neural Networks", arXiv e-prints, 2015/
- [10] Suyanto, "Machine Learning Tingkat Dasar dan Lanjut", 2018, Informatika.