



Dataset Gambar Wajah untuk Analisis *Personal Identification*

Syefrida Yulina¹, Hoky Nawa²

^{1,2}Sistem Informasi, Jurusan Teknologi Informasi, Politeknik Caltex Riau

¹syefrida@pcr.ac.id, ²hoky@alumni.pcr.ac.id

Abstract

In today's era, which is supported by technological advances, personal assistance through the face can be carried out by sophisticated machines and robots. One of its applications is personal identification using data mining. But before conducting data training and data classification, it is necessary to carry out the process of data collection and data cleaning or data pre-processing. Currently the face dataset for personal identification at the Caltex Riau Polytechnic in particular is still in the form of raw data, namely in the form of a collection of images that have not been pre-processed. Therefore, this research will perform image preprocessing to clean up the image data that has been collected so that the data can become a cleaner source of information and can be used at a later stage. The data used in this study are image data or photos of Caltex Riau Polytechnic students. At the facial image pre-processing stage using the OpenCV library using the Python programming language. Images collected by 500 students for 5 students. The results of this study are the personal identification dataset of Caltex Riau Polytechnic students consisting of 280 images that have successfully passed the stages of grayscaling, cropping, resizing and Normalization. This dataset is stored in the file data_norm.npz. White box testing is carried out to determine the accuracy of the application of the image pre-processing stage with the test results stating that all functional basis paths applied are in accordance with the cyclometric complexity and its independent path.

Keywords: dataset, *image pre-processing*, *personal identification*.

Abstrak

Pada zaman sekarang yang didukung kemajuan teknologi, identifikasi personal melalui wajah dapat dilakukan oleh mesin dan robot robot canggih. Salah satu penerapannya adalah *personal identification* menggunakan *data mining*. Namun sebelum melakukan *training* data dan pengklasifikasian data, maka perlu melakukan proses pengumpulan data dan pembersihan data atau *pre-processing* data. Saat ini dataset wajah untuk *personal identification* pada Politeknik Caltex Riau khususnya, masih berupa data mentah yaitu berupa kumpulan gambar yang belum di *pre-processing*. Oleh karena itu, pada penelitian ini akan melakukan *image preprocessing* untuk membersihkan data gambar yang telah terkumpul agar data tersebut dapat menjadi sumber informasi yang lebih bersih dan bisa digunakan pada tahapan selanjutnya. Data yang digunakan pada penelitian ini adalah data gambar atau foto mahasiswa Politeknik Caltex Riau. Pada tahapan *pre-processing* gambar wajah menggunakan *library* OpenCV menggunakan bahasa pemrograman Python. Gambar yang dikumpulkan sebanyak 500 mahasiswa untuk 5 orang mahasiswa. Hasil dari penelitian ini adalah dataset *personal identification* mahasiswa Politeknik Caltex Riau yang terdiri dari 280 gambar yang telah berhasil melewati tahapan *grayscale*, *cropping*, *resizing* dan *Normalization*. Dataset ini disimpan kedalam bentuk file data_norm.npz. Pengujian *whitebox testing* dilakukan untuk mengetahui ketepatan penerapan tahapan *pre-processing gambar* dengan hasil pengujian menyatakan bahwa semua *basis path* fungsional yang diterapkan telah sesuai menurut *cyclometric complexity* dan *independent path* nya.

Kata kunci: dataset, *image pre-processing*, *personal identification*.

1. Pendahuluan

Personal Identification (PI) merupakan suatu teknik dan sistem sosial identifikasi orang tak dikenal dan resolusi identitas terlepas dari apakah orang tersebut hidup atau mati [1]. Kemajuan teknologi seperti *artificial intelligence* (AI) yang cukup signifikan telah menjadi perhatian untuk penerapan PI. Sejak munculnya *deep learning* sebagai teknologi berbasis AI, banyak aplikasi PI telah diperluas ke berbagai bidang seperti ilmu forensik, biometrik, medis dan pendidikan [1][2][3][4].

Menanggapi tantangan ini, sebuah teknologi *computer vision* telah menjadi fitur untuk membangun sistem cerdas yang dapat menganalisis dan mengenali identitas seseorang secara otomatis yang terintegrasi pada suatu perangkat dan dataset tertentu [5]. Dalam mengembangkan sebuah sistem berbasis *computer vision*, perlu dilakukan *training* data sebelum data diklasifikasikan sehingga dapat membantu menyelesaikan tugas tersebut [6]. Data *training* harus dibersihkan terlebih dahulu melalui tahapan *image pre-*



Lisensi

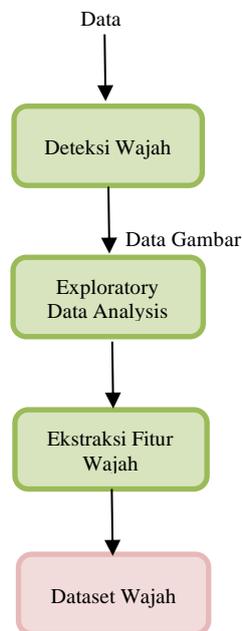
Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

processing agar meningkatkan kualitas gambar sehingga dapat dianalisis secara efektif.

Penelitian ini mengumpulkan sejumlah gambar/foto dari mahasiswa Politeknik Caltex Riau untuk *personal identification* yang akan melalui tahapan *image pre-processing*, dimana proses diawali dengan pendeteksian area wajah, menganalisis data menggunakan *Exploratory Data Analysis* (EDA), dan menentukan fitur-fitur wajah [7][8]. Dengan adanya dataset *personal identification* diharapkan dapat memudahkan tahapan klasifikasi dengan meningkatkan nilai akurasi.

2. Metode Penelitian

Dalam menghasilkan dataset *personal identification*, dilakukan tahapan *image pre-processing* dan analisis terhadap gambar wajah yang telah dikumpulkan, seperti: tahapan pendeteksian area wajah, *Exploratory Data Analysis* (EDA), dan ekstraksi fitur wajah. Metodologi penelitian dapat dilihat pada Gambar 1.



Gambar 1 Tahapan Proses Dataset Wajah

2.1 Deteksi wajah

Wajah akan dideteksi dari gambar statis menggunakan algoritma Haar Cascade, dimana dalam penerapannya digunakan *library* OpenCV berupa file xml dari algoritma tersebut. Setelah wajah terdeteksi maka dilanjutkan ke tahapan *pre-processing* gambar [9]. *Pre-processing data* merupakan teknik awal data mining untuk mengubah data mentah atau bisa dikenal dengan *raw data* yang dikumpulkan dari berbagai sumber menjadi informasi yang bersih dan dapat digunakan pada tahap selanjutnya [10]. Pada proses *pre-processing* dilakukan *grayscale* untuk konversi warna gambar asli ke bentuk warna *grayscale*, *resizing* untuk merubah

ukuran gambar, normalisasi data gambar. Tahapan ini dilakukan untuk mengambil informasi dengan cara memfilter dan membersihkan data gambar.

2.2 Exploratory Data Analysis (EDA)

Tahapan EDA dilakukan untuk analisis terhadap ukuran gambar dan analisis fitur wajah pada gambar. Pada penentuan ukuran gambar dianalisis berdasarkan nilai statistik dari keseluruhan ukuran gambar dengan mengambil nilai rata-rata ukuran gambar, sedangkan penentuan fitur wajah berdasarkan jumlah komponen wajah menurut varian maksimumnya. EDA ini menghasilkan visualisasi data melalui representasi grafis [10].

2.3 Ekstraksi Fitur Wajah

Proses reduksi dimensi gambar seperti mata, hidung, bibir, pipi dari wajah yang terdeteksi menggunakan *Principal Component Analysis* (PCA). PCA ini mengekstraksi fitur wajah optimal menurut nilai varian maksimum dengan menerapkan algoritma *Eigenface* untuk pemilihan komponen utama yang memiliki kontribusi [11][12].

Dalam hal reduksi dimensi dan penentuan komponen pada sebuah gambar, maka PCA digunakan untuk analisis utama. Biasanya ini dikenal dengan proyeksi ruang *Eigen* atau transformasi karhunen-loeve [7]. Gambar proyeksi ke sub-ruang yang ukuran simetris utama menangkap perubahan yang paling tinggi diantara gambar. Tujuan dari PCA adalah untuk mengurangi dimensi gambar.

Proses-proses yang dilakukan oleh PCA dalam menemukan bagian-bagian terpenting pada dimensi gambar, adalah seperti [13][9]:

- Setiap set dari data gambar mengurangi titik data kemudian mempertimbangkan titik fokus dari dimensi yang dipakai
- Menghitung kovarian matriks. Kovarian matriks adalah matriks persegi dari dimensi gambar yang mengumpulkan varians dalam diagonal utama dan kovarian dalam unsur-unsur luar diagonal utama.
- Menghitung nilai *Eigen*: vektor *Eigen*
- Mengurutkan nilai *Eigen*
- Memilih vektor *Eigen* dan menghasilkan data baru
- Memproyeksikan serta membandingkan data latih dengan data uji menggunakan ukuran kedekatan.

Eigenface merupakan hasil dari *eigen* vektor yang didapat dari kovarian matriks dengan distribusi acak pada data gambar wajah yang berdimensi tinggi. Metode ini mengubah gambar wajah kedalam kumpulan karakteristik fitur dari gambar wajah yang disebut

eigenface. Eigenface menyimpan beberapa bagian dari dimensi wajah yang tidak terlihat pada data gambar asli.

3. Hasil dan Pembahasan

Pada bagian ini membahas tentang pengumpulan data, *pre-processing*, dataset *personal identification* dan pengujian.

3.1 Pengumpulan dataset

Terdapat lima orang mahasiswa yang berpartisipasi dalam pengambilan data gambar (Gambar 2). Masing-masing mahasiswa tersebut memiliki seratus gambar wajah dengan posisi yang bervariasi. Sehingga data awal sebelum *pre-processing* berjumlah lima ratus gambar berwarna.

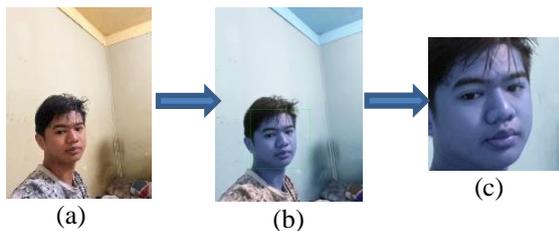


Gambar 2. Dataset Gambar Mahasiswa

Setelah data terkumpul maka dilakukan tahapan pendeteksian bagian wajah pada data gambar.

3.2 Pre-processing Data

Pada data gambar awal berupa gambar berwarna/RGB dengan ukuran gambar yang berbeda yang memiliki ekstensi file .jpg. Pada setiap gambar dilakukan *pre-processing* data yang diawali proses deteksi bagian wajah dan konversi gambar ke warna *grayscale*. Pada proses deteksi wajah dengan menggunakan library yang berekstensi xml yaitu *haarcascade_frontalface_default.xml*. Selanjutnya dilakukan konversi warna gambar dari warna RGB ke warna *grayscale*. Pada Gambar 3 memperlihatkan salah satu hasil deteksi wajah pada gambar.



Gambar 3. Proses Pendeteksian Wajah dan Hasil Wajah pada Gambar

Gambar 3 menunjukkan proses pendeteksian wajah dimana pada Gambar 3(a) memperlihatkan gambar asli dengan warna RGB, Gambar 3(b) merupakan hasil konversi gambar ke warna *grayscale*, dan Gambar 3(c) adalah gambar wajah yang telah dipotong berdasarkan posisi wajah yang telah terdeteksi.

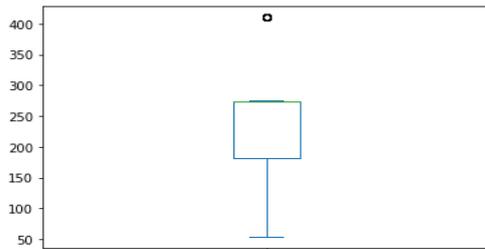
Data gambar setelah *pre-processing* berjumlah 280 gambar wajah yang berhasil dideteksi. Perbandingan data setelah *pre-processing* dengan sebelum adalah

280:500 dengan hasil 0.56 persen. Hal ini disebabkan pada tahapan deteksi wajah pada gambar, dimana posisi wajah menentukan hasil deteksi. Pada data gambar awal banyak memiliki posisi wajah yang miring sehingga banyak yang tidak berhasil terdeteksi. Tabel 1. Memperlihatkan sebagian analisis terhadap proses deteksi wajah.

Tabel 1. Tabel Analisis Proses Deteksi Wajah

Nama File Gambar	Wajah terdeteksi	Analisis
Anti	Ya	
Anti1	Ya	
Anti2	Tidak	Terhalang masker
Anti3	Ya	
Anti4	Tidak	Ada objek boneka
Anti5	Ya	
Anti6	Ya	
Anti7	Tidak	Ada objek boneka
Anti 8	Tidak	
Anti9	Tidak	Wajah tertutup
Anti10	Ya	
Anti11	Ya	
Anti12	Tidak	Ada objek boneka
Anti13	Tidak	Objek terlalu jauh
Anti14	Tidak	Objek terlalu jauh
Anti15	Tidak	Objek terlalu jauh
Anti16	Tidak	Objek terlalu jauh
Anti17	Ya	
Anti18	Ya	
Anti19	Ya	
Anti20	Ya	
Anti21	Ya	
Anti22	Ya	
Anti23	Tidak	Wajah tertutup
Anti 24	Ya	
Anti25	Tidak	Wajah terlalu miring
Anti26	Tidak	Wajah terlalu miring
Anti27	Ya	
Anti28	Ya	
Anti29	Ya	
Anti30	Ya	
Anti31	Tidak	Objek terlalu ramai
Anti32	Tidak	Wajah terlalu miring
Anti33	Ya	
Anti34	Ya	
Anti35	Ya	
Anti36	Ya	
Anti37	Tidak	Background terlalu ramai
Anti38	Tidak	Wajah terlalu miring
Anti39	Ya	
Anti40	Ya	
Anti41	Ya	

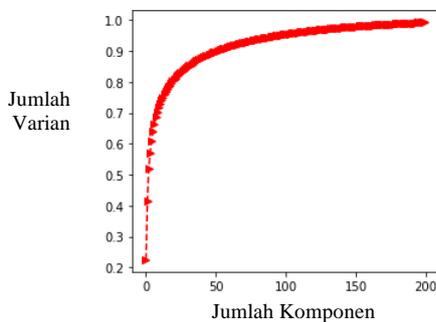
Kemudian dilakukan proses *resizing* untuk merubah ukuran gambar dengan ukuran yang sama agar mempermudah dalam proses selanjutnya. Gambar 4 memperlihatkan visualisasi plotting untuk semua ukuran gambar wajah yang bervariasi. Ukuran yang digunakan untuk semua gambar wajah adalah dengan resolusi 100x100 piksel. Hal ini berdasarkan nilai statistik ukuran gambar wajah keseluruhan dengan mempertimbangkan nilai rata-rata, nilai minimum dan nilai maksimum ukuran gambar. Jika gambar wajah berukuran kecil maka tahapan *enlarge* dilakukan untuk memperbesar ukuran gambar, sedangkan gambar berukuran besar maka dilakukan tahapan *shrink* untuk memperkecil ukuran gambar.



Gambar 4. Plotting Ukuran Gambar Wajah

Setelah tahapan *resizing* selesai, maka dilanjut ke tahapan *normalization*. Pada proses normalisasi menggunakan metode min-max, yaitu mencari nilai fitur normalisasi dengan cara membagi nilai lama dengan nilai maksimum. Sehingga didapatkan data normalisasinya kedalam bentuk tipe data array.

Selanjutnya pada setiap data gambar yang telah di *pre-processing* kemudian dilakukan ekstraksi fitur wajah dengan algoritma PCA yang menghasilkan *eigenface* dari masing-masing gambar. Analisis dilakukan dengan mengetahui jumlah komponen dengan jumlah varian yang dimiliki oleh inputan data gambar kemudian menghitung matrik kovarians. Hal ini dilakukan untuk mengetahui hubungan antar variabel dari inputan. Selanjutnya berdasarkan perhitungan nilai dan vektor eigen matriks kovarians maka didapatkan identifikasi komponen utama. Pada Gambar 5. memperlihatkan bahwa komponen fitur yang digunakan adalah sebanyak 50 komponen sehingga komponen lainnya dibuang.



Gambar 5. Analisis Penentuan Jumlah Fitur Wajah



Gambar 6. Data Gambar Wajah Asli

3.3 Hasil dataset setelah *pre-processing*

Pada Gambar 6. memperlihatkan gambar wajah asli sebelum diekstraksi fitur wajah, sedangkan Gambar 7. memperlihatkan hasil gambar *eigenface*. Tabel 2 memperlihatkan dataset berupa array. Dataset ini disimpan kedalam file dengan ekstensi npz.

3.1 Pengujian

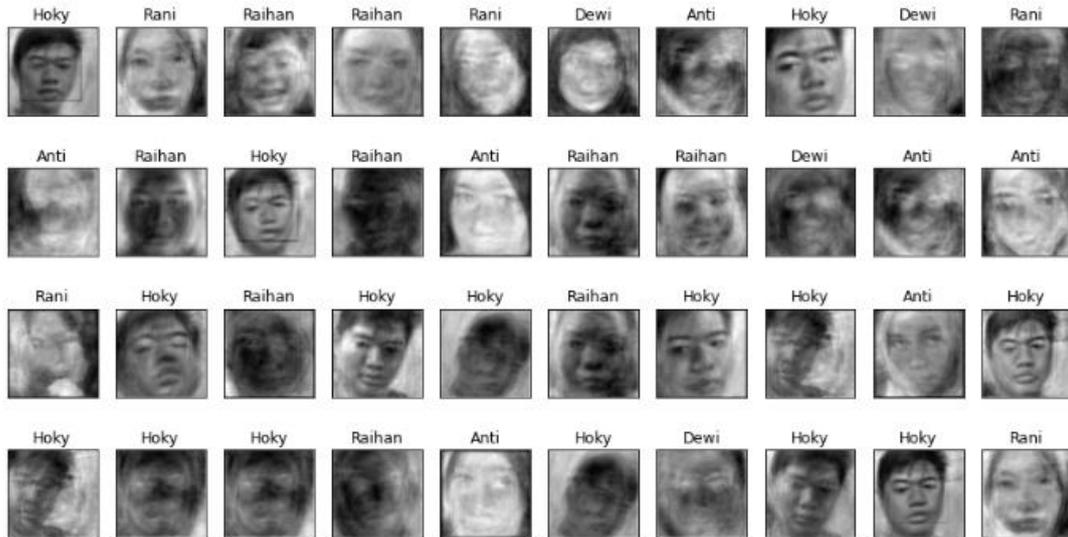
Untuk pengujian dilakukan *whitebox testing* untuk menguji penerapan algoritma dari fungsionalitas menggunakan *basis path testing* dan *respond time*. Pada pengujian fungsionalitas maka sejumlah path pada tiap fungsional ditetapkan berdasarkan barisan perintah koding pada tahapan *pre-processing*: yaitu sebanyak 120 path. Kemudian menentukan *cyclometric complexity* menggunakan rumus yaitu $(G) = E - N + 2$, dimana E merupakan jumlah edge/cabang dan N merupakan jumlah node. Maka didapatkan hasil perhitungan $(G) = 99 - 100 + 2 = 1$. Hasil dari cyclometric complexity adalah 1 sehingga terdapat 1 independent path.

Sedangkan untuk pengujian *respond time* dapat dilihat pada Tabel 3.

Tabel 3. Pengujian *Respons Time*

Jumlah Gambar	<i>Respond time (dalam detik)</i>
100	20
50	10
20	8
10	5

Hasil yang didapatkan berdasarkan *respond time* adalah bervariasi tergantung dari jumlah gambar yang diproses. Semakin banyak gambar yang diproses maka semakin besar waktu yang diperlukan untuk mengeksekusi, begitu sebaliknya.



Gambar 7. Data Gambar Wajah Eigenface

Tabel 2. Dataset *Personal Identification*

objek	0	1	2	3	4	5	6	7	8	...	9991	9992	9993	9994	9995	9996	9997	9998	9999	kode_objek	
0	Anti	0	0	0	0	0	0	0	0	...	81	89	127	140	146	136	132	147	132	:	
1	Anti	0	0	0	0	0	0	0	0	...	84	143	83	38	29	37	42	101	48	:	
2	Anti	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	:	
3	Anti	0	0	0	0	0	0	0	0	...	7	7	8	9	10	11	12	13	1	:	
4	Anti	0	0	0	0	0	0	0	0	...	64	56	62	77	79	75	79	87	76	:	
...	:	
275	Rani	0	0	0	0	0	0	0	0	...	4	5	5	5	5	5	5	5	5	1	:
276	Rani	0	0	0	0	0	0	0	0	...	48	39	40	41	45	64	71	74	69	:	
277	Rani	0	0	0	0	0	0	0	0	...	15	15	14	13	14	14	15	15	2	:	
278	Rani	0	0	0	0	0	0	0	0	...	4	4	5	6	7	6	3	2	0	:	
279	Rani	0	0	0	0	0	0	0	0	...	17	13	6	4	3	5	4	4	0	:	

4. Kesimpulan

Pada penelitian ini data awal gambar yang dikumpulkan adalah sebanyak 500 gambar wajah dari 5 orang mahasiswa. Pada tahapan *pre-processing* seperti konversi warna ke warna grayscale didapatkan bahwa semua gambar 100 persen telah berhasil dikonversikan, kemudian untuk proses deteksi wajah pada gambar didapatkan hasil 0.56 persen dari gambar yang berhasil mendeteksi bagian wajah. Dalam menentukan fitur wajah dilakukan analisis menggunakan EDA serta menerapkan algoritma PCA untuk menghasilkan gambar eigenface. Hasil dari dataset setelah *pre-processing* ini berupa file dengan ekstensi npz yang diperlukan untuk tahapan klasifikasi *personal identification* yang tidak dibahas pada tulisan ini. Hasil pengujian menggunakan *whitebox testing* menunjukkan

kesesuaian penerapan algoritma berdasarkan pada penentuan jalur logika menggunakan *basis path testing* dan *respond time*. Pada pengembangan penelitian selanjutnya dapat menerapkan algoritma lainnya serta menambahkan data yang banyak agar memiliki hasil yang sempurna pada tahapan klasifikasi.

Ucapan Terimakasih

Terima kasih kami ucapkan kepada semua pihak yang telah membantu penelitian ini, terutama kampus Politeknik Caltex Riau

Daftar Rujukan

- [1] S. Matsuda and H. Yoshimura, "Personal identification with artificial intelligence under COVID-19 crisis: a scoping review," *Syst. Rev.*, vol. 11, no. 1, p. 7, Dec. 2022, doi:

- 10.1186/s13643-021-01879-z.
- [2] J.-A. Lee and K.-C. Kwak, "Personal Identification Using an Ensemble Approach of 1D-LSTM and 2D-CNN with Electrocardiogram Signals," *Appl. Sci.*, vol. 12, no. 5, p. 2692, Mar. 2022, doi: 10.3390/app12052692.
- [3] S. N. Tkachenko, A. V. Pichugina, A. A. Shpilevoy, and A. I. Zakharov, "Development of a Face Identification System's Prototype," *J. Phys. Conf. Ser.*, vol. 2224, no. 1, p. 012003, Apr. 2022, doi: 10.1088/1742-6596/2224/1/012003.
- [4] Y. Kim, T. Soyata, and R. F. Behnagh, "Towards Emotionally Aware AI Smart Classroom: Current Issues and Directions for Engineering and Education," *IEEE Access*, vol. 6, pp. 5308–5331, 2018, doi: 10.1109/ACCESS.2018.2791861.
- [5] M. Oravec, "Feature extraction and classification by machine learning methods for biometric recognition of face and iris," *Proc. Elmar - Int. Symp. Electron. Mar.*, no. September, pp. 1–4, 2014, doi: 10.1109/ELMAR.2014.6923301.
- [6] K. Delgado *et al.*, "Student Engagement Dataset," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Oct. 2021, pp. 3621–3629, doi: 10.1109/ICCVW54120.2021.00405.
- [7] B. K. Durga and D. V. Rajesh, "Review Of Facial Emotion Recognition System," *Int. J. Pharm. Res.*, vol. 10, no. 03, Jul. 2018, doi: 10.31838/ijpr/2018.10.03.056.
- [8] F. Fadliyah, *Computer Vision dan Pengolahan Citra*. ANDI Yogyakarta, 2007.
- [9] S. Puspitodjati and T. A. Wirana, "Sistem Pengenalan Ekspresi Wajah Berdasarkan Citra Wajah Menggunakan Metode Eigenface dan Nearest Feature Line," *Komput. Dan Sist. Intelijen*, vol. 7, no. 2302–3740, pp. 311–316, 2012, [Online]. Available: <http://penelitian.gunadarma.ac.id/kommit>.
- [10] A. Riadi and R. Sulachani, "Analisis Implementasi Preprocessing Dengan Otsu-Gaussian Pada Pengenalan Wajah," *Ilk. J. Ilm.*, vol. 11, no. 3, pp. 200–205, Dec. 2019, doi: 10.33096/ilkom.v11i3.457.200-205.
- [11] G. N. Girish, C. L. Shrinivasa Naika, and P. K. Das, "Face recognition using MB-LBP and PCA: A comparative study," *2014 Int. Conf. Comput. Commun. Informatics Ushering Technol. Tomorrow, Today, ICCCI 2014*, pp. 1–6, 2014, doi: 10.1109/ICCCI.2014.6921773.
- [12] M. Agarwal, H. Agrawal, N. Jain, and M. Kumar, "Face recognition using principle component analysis, eigenface and neural network," *2010 Int. Conf. Signal Acquis. Process. ICSAP 2010*, pp. 310–314, 2010, doi: 10.1109/ICSAP.2010.51.
- [13] T. Mulyono, K. Adi, and R. Gernowo, "Sistem Pengenalan Wajah Dengan Metode Eigenface Dan Jaringan Syaraf Tiruan (Jst)," *Berk. Fis.*, vol. 15, no. 1, pp. 15–20, 2012.