

# PENGGUNAAN METODE LINEAR DISCRIMINANT ANALYSIS UNTUK PENGENALAN WAJAH DENGAN MEMBANDINGKAN BANYAKNYA DATA LATIH

**Rifki Kosasih**

<sup>1</sup> Program Studi Komputasi Matematika, Fakultas Teknologi Industri Universitas Gunadarma

<sup>1</sup>Jl. Margonda Raya No. 100, Depok, Indonesia 16424

<sup>1</sup>rifki\_kosasih@staff.gunadarma.ac.id

## **Abstrak**

*Pengenalan wajah sangat dibutuhkan dalam sistem keamanan rumah karena dapat membantu mengetahui siapa saja yang sudah memasuki area rumah. Salah satu metode yang dapat digunakan dalam pengenalan wajah adalah metode Principle Component Analysis (PCA). Akan tetapi, metode PCA kurang optimal dalam melakukan pemisahan antar kelas. Oleh karena itu pada penelitian ini digunakan metode lain yang dapat melakukan pemisahan antar kelas secara optimal seperti metode Linear Discriminant Analysis (LDA). Data yang digunakan sebanyak 400 data citra wajah dengan komposisi 40 orang dengan tiap orang memiliki 10 citra wajah dengan berbagai ekspresi. Pada penelitian ini diusulkan untuk memperhatikan banyaknya data latih yang digunakan. Banyaknya citra wajah tiap orang yang digunakan untuk data latih adalah 5, 6, 7, 8 dan 9 citra wajah per orang. Selanjutnya dilakukan ekstraksi fitur dengan menggunakan metode LDA. Selanjutnya dilakukan klasifikasi terhadap fitur-fitur yang telah diperoleh dengan menggunakan metode K Nearest Neighbor (KNN). Berdasarkan hasil penelitian diperoleh bahwa tingkat akurasi terbesar yaitu sebesar 97,5% yang terjadi saat banyaknya citra data latih tiap orang adalah 9 dan banyaknya tetangga (K) adalah 1.*

**Kata Kunci:** Ekstraksi Fitur, KNN, LDA, pengenalan wajah

## **Abstract**

*Face recognition is very much needed in a home security system because it can help find out who has that is occupied by the house. One of the methods commonly used in face recognition is the Principle Component Analysis (PCA) method. However, the PCA method is less than optimal in separating between classes. Therefore, in this study, another method is used that can perform separation between classes optimally, such as the Linear Discriminant Analysis (LDA) method. The data used are 400 facial image data with a composition of 40 people with each person having 10 facial images with various expressions. In this study it is proposed to pay attention to the amount of training data used. The number of face images of each person used for training data are 5, 6, 7, 8 and 9 face images per person. Furthermore, feature extraction is carried out using the LDA method. Furthermore, the features that have been obtained are classified using the K Nearest Neighbor (KNN) method. Based on the results of the study, it was found that the greatest accuracy rate was 97.5% which occurred when the number of training data images per person was 9 and the number of neighbors (K) was 1.*

**Keywords:** Feature Extraction, face recognition, KNN, LDA

## **PENDAHULUAN**

Wajah seseorang memiliki ciri khusus (fitur) yang dapat membedakan orang yang satu dengan orang yang lainnya sehingga

dapat di-gunakan untuk mengenali seseorang. Pengenalan wajah merupakan salah satu topik di *computer vision* yang digunakan untuk sistem keamanan, absensi dan lain lain [1].

Pola citra pada wajah dikenali dengan cara mendapatkan ciri khusus (fitur) dari wajah, biasanya dilakukan dengan cara mengenali pola citra pada wajah. Metode yang sering dipakai dalam pengenalan pola adalah metode *manifold learning* [2],[3]. *Manifold learning* merupakan suatu metode yang digunakan untuk mereduksi dimensi dari dimensi besar menjadi dimensi yang lebih kecil tanpa menghilangkan banyak informasi [4]. Pola citra wajah seseorang akan terlihat dengan melakukan reduksi dimensi.

Pada proses pengenalan, citra wajah diproyeksikan dari ruang dimensi yang tinggi ke ruang dimensi yang lebih rendah tanpa menghilangkan banyak informasi dan di-klasifikasikan dengan menggunakan metode jarak terdekat seperti jarak Euclid

Salah satu metode *manifold learning* yang digunakan untuk pengenalan wajah adalah *Principal Component Analysis* (PCA). Metode PCA merupakan metode reduksi dimensi dengan melakukan transformasi linier dari suatu ruang berdimensi tinggi ke dalam ruang berdimensi rendah [5],[6]. Akan tetapi, metode PCA kurang baik dalam melakukan pemisahan antar kelas sehingga tingkat akurasi pengenalan wajah menjadi turun [7]. Oleh karena itu, untuk mengatasi hal tersebut dalam penelitian ini di-gunakan metode *Linear Discriminant Analysis* (LDA) untuk

pengenalan wajah. Metode ini bekerja dengan menemukan subruang linier yang memaksimalkan pemisahan dua kelas berdasarkan kriteria *Fisher*. Untuk memaksimalkan pemisahan dua kelas tersebut dapat dilakukan dengan cara meminimalkan jarak matriks sebaran *within-class*  $S_w$  dan memaksimalkan jarak matriks sebaran *between class*  $S_b$  secara simultan sehingga menghasilkan kriteria *Fisher* yang maksimal. LDA akan menemukan sub-ruang linear di mana kelas-kelas saling terpisah dengan memaksimalkan kriteria *Fisher* [8],[9]. Setelah mendapatkan kelas-kelas, selanjutnya dilakukan klasifikasi dengan menggunakan metode K-Nearest Neighbor (KNN).

Beberapa penelitian telah dilakukan di bidang pengenalan wajah seperti kosasih, melakukan pengklasteran wajah manusia dengan menggunakan metode isomap. Data yang digunakan adalah data citra wajah yang terdiri dari 5 orang dengan tiap orang memiliki 4 citra dengan berbagai macam ekspresi wajah. Dari hasil penelitiannya terlihat bahwa citra-citra wajah yang berasal dari orang yang sama dengan ekspresi wajah yang berbeda-beda mengelompok ke dalam satu area akan tetapi dalam penelitian-nya belum dilakukan pengklasifikasian wajah [10].

Selanjutnya kosasih melanjutkan penelitian-nya melakukan klasifikasi wajah berdasarkan pengklasteran hasil isomap dengan menggunakan metode K-Nearest Neighbor (KNN). Dalam penelitiannya diperoleh tingkat akurasi sebesar 83,33% [1].

Fahrurozi menggunakan metode PCA dan *Local Binary Pattern* (LBP) untuk melakukan pengenalan wajah. Dalam penelitiannya Fahrurozi menggunakan data sample 8 orang dengan tiap orang mempunyai 5 citra wajah yang berbeda-beda. Tingkat akurasi yang diperoleh untuk mengenali wajah seseorang adalah sebesar 75% [11]. Oleh karena itu dalam penelitian ini digunakan data citra yang lebih banyak dari penelitian sebelumnya.

Selanjutnya Pratiwi menggunakan metode *Principle Component Analysis* (PCA) untuk melakukan pengenalan wajah. Dalam penelitiannya tentang pengenalan wajah diperoleh bahwa tingkat akurasi dengan metode tersebut sebesar 82,81 % [12].

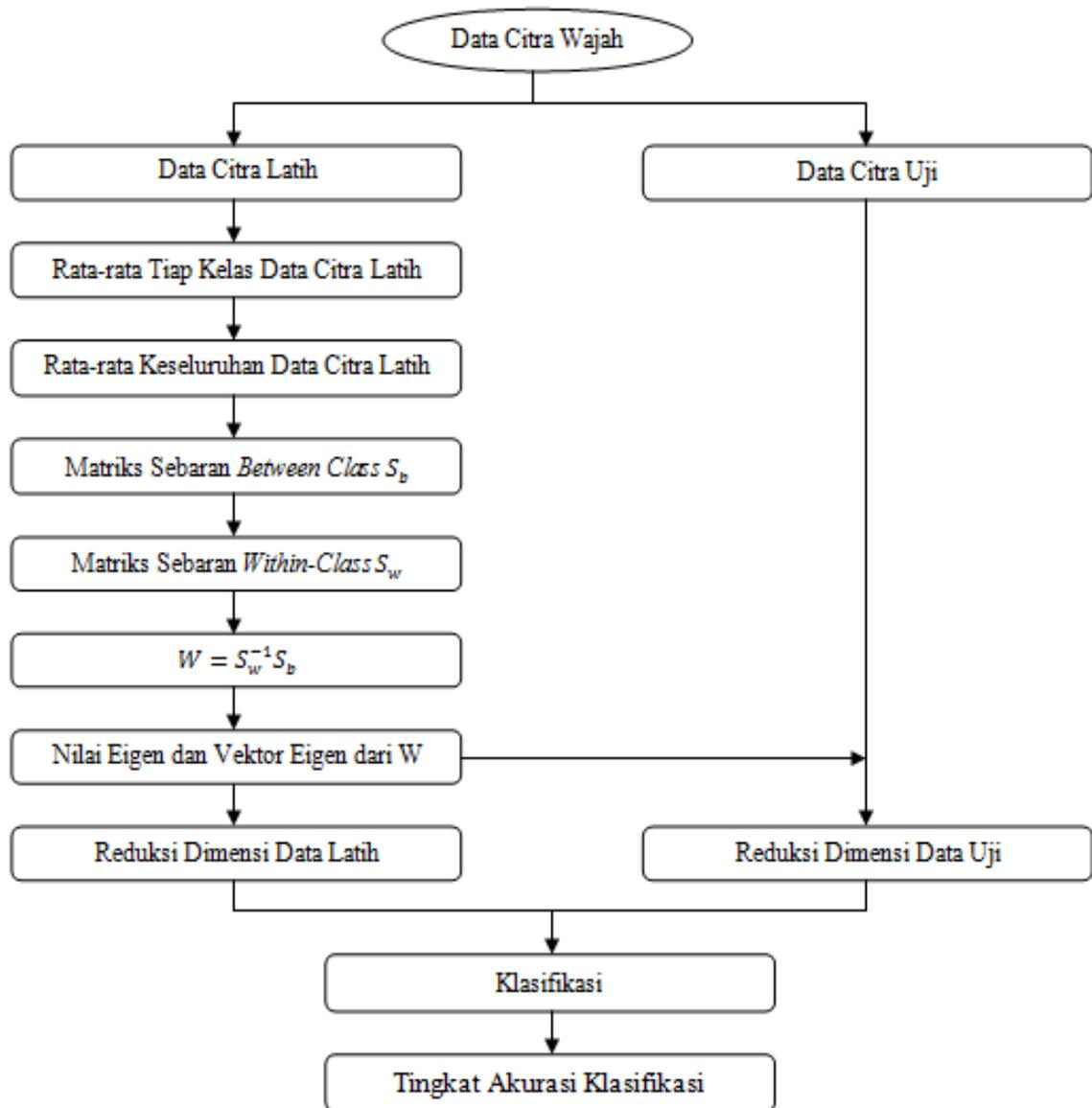
Berdasarkan beberapa penelitian sebelumnya, data yang digunakan belum memperhatikan banyaknya data latih dan data uji yang digunakan, sehingga pada penelitian ini dilakukan pengenalan wajah dengan menggunakan metode LDA dengan memperhatikan data latih dan data uji serta memperhatikan banyaknya tetangga yang

digunakan dalam proses klasifikasi dengan metode K Nearest Neighbor (KNN).

## **METODE PENELITIAN**

Pada penelitian ini diusulkan pengenalan seseorang melalui data citra wajah dengan melalui beberapa tahapan. Tahapan-tahapan penelitian tersebut dapat dilihat pada Gambar 1.

Pada Gambar 1, tahapan pertama dalam penelitian ini adalah mengumpulkan data citra wajah untuk dijadikan data latih dan data uji. Tahapan selanjutnya adalah mencari fitur wajah dengan menggunakan metode *Linear Discriminant Analysis* (LDA) seperti mencari rata-rata kelas data, sebaran data, menghitung matriks  $W$ , mencari nilai eigen dari matriks  $W$  dan melakukan reduksi dimensi pada data latih dan data uji. Hasil reduksi dimensi merupakan fitur wajah yang digunakan untuk mengenali wajah seseorang. Tahapan terakhir adalah melakukan klasifikasi wajah dengan menggunakan metode *K-Nearest Neighbor* (KNN).



Gambar 1. Gambaran Umum Penelitian

### Pengumpulan Data Citra

Data citra yang digunakan adalah data citra wajah yang terdiri dari 40 orang (kelas) dengan tiap orang mempunyai 10 citra wajah dengan berbagai macam ekspresi, sehingga

seluruh data citra wajah yang digunakan dalam penelitian sebanyak 400 citra. Data citra wajah tersebut diperoleh dari ORL Database [13]. Sampel citra wajah dapat dilihat pada Gambar 2.



Gambar 2. Sample Data Citra Wajah Seseorang Dengan 10 ekspresi yang berbeda-beda (1 dari 40 Orang)

Tabel 1.

Nama Percobaan	Data Latih	Banyak Data Latih	Data Uji	Banyak Data Uji
1	5 citra pertama per orang	$5 \times 40 = 200$	5 citra pertama per orang	$5 \times 40 = 200$
2	6 citra pertama per orang	$6 \times 40 = 240$	4 citra pertama per orang	$4 \times 40 = 160$
3	7 citra pertama per orang	$7 \times 40 = 280$	3 citra pertama per orang	$3 \times 40 = 120$
4	8 citra pertama per orang	$8 \times 40 = 320$	2 citra pertama per orang	$2 \times 40 = 80$
5	9 citra pertama per orang	$9 \times 40 = 360$	1 citra pertama per orang	$1 \times 40 = 40$

Setelah data citra wajah diperoleh, selanjutnya data tersebut dibagi menjadi 2 yaitu data latih dan data uji. Pada penelitian ini, di-usulkan untuk memperhatikan pembagian data dengan jumlah data latih dan data uji yang berbeda-beda untuk memperoleh tingkat akurasi yang terbaik dalam pengenalan wajah seperti yang ditunjukkan pada Tabel 1.

Pada tahap pertama, data latih yang digunakan sebanyak 200 citra yang terdiri dari 5 citra wajah pertama untuk tiap orang. Data uji pada tahap ini merupakan 5 wajah berikutnya untuk tiap orang dari sebanyak 40 orang yaitu sebanyak 200 citra. Pada tahap

kedua, data latih yang digunakan adalah 6 citra wajah pertama per orang (240 citra) untuk data latih dan 4 citra wajah berikutnya per orang (160 citra) digunakan sebagai data uji. Pada tahap ketiga, data latih yang digunakan adalah 7 citra wajah pertama per orang (280 citra) untuk data latih dan 3 citra wajah berikutnya per orang (120 citra) digunakan sebagai data uji. Pada tahap keempat, data latih yang digunakan adalah 8 citra wajah pertama per orang (320 citra) untuk data latih dan 2 citra wajah berikutnya per orang (80 citra) digunakan unntuk data uji. Pada tahap terakhir, data latih yang

digunakan adalah 9 citra wajah pertama per orang (360 citra) untuk data latih dan 1 citra wajah berikutnya per orang (40 citra) digunakan pada data uji.

Setelah itu, tahapan selanjutnya melakukan proses reduksi dimensi dengan menggunakan metode *Linear Discriminant Analysis* (LDA).

### Linear Discriminant Analysis (LDA)

*Linear Discriminant Analysis* (LDA) adalah salah satu metode *manifold learning* yang merupakan suatu metode untuk mengekstrak fitur dengan cara mereduksi dimensi dari dimensi tinggi ke dimensi yang lebih rendah [14]. Berikut ini adalah tahapan-tahapan dari metode LDA.

1. Tahapan pertama pada proses LDA adalah mengubah data citra latih dan citra uji menjadi vektor-vektor  $\{x_1, x_2, \dots, x_n\}$ .
2. Buat kelas-kelas berdasarkan banyaknya orang pada data citra latih dan citra uji.
3. Hitung rata-rata dalam kelas ( $m_i$ ) dan rata-rata keseluruhan kelas ( $m$ ) dari seluruh citra di *database*.
4. Menghitung matriks sebaran antar kelas ( $S_b$ ) menggunakan persamaan (1)
 
$$\sum_{i=1}^k n_i (m_i - m)(m_i - m)^T$$
 (1)
5. Menghitung matriks sebaran dalam kelas ( $S_w$ ) menggunakan persamaan (2)

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (x_j - m_i)(x_j - m_i)^T$$

(2)

6. Mencari matriks  $W$  dengan menggunakan persamaan (3)

$$W = S_w^{-1} S_b$$

(3)

7. Mencari vektor eigen dari  $W$  dan mengurutkan nilai eigen ( $\lambda$ ) sesuai dengan urutan nilai yang ada pada nilai eigen dari besar ke kecil.
8. Melakukan reduksi dimensi dengan cara melakukan proyeksi dengan vektor eigen yang sudah dipilih ke ruang dimensi yang lebih rendah yang disebut dengan *fisher basis* vektor [15] menggunakan persamaan (4).

$$\bar{u}_i = V^T x_i$$

(4)

9. Dimana :  $\bar{u}_i$  adalah vektor hasil proyeksi yang berdimensi lebih rendah

$V$  adalah matriks yang berisi vektor-vektor eigen dari matriks  $W$

$x_i$  adalah vektor yang mewakili citra wajah

10. Hasil reduksi dimensi tersebut merupakan fitur yang digunakan untuk

mengenali seseorang. Setelah itu, dilakukan klasifikasi berdasarkan fitur-fitur yang sudah diperoleh dengan menggunakan metode K-Nearest Neighbor (KNN).

### **K-Nearest Neighbor (KNN)**

K-Nearest Neighbor merupakan metode klasifikasi yang digunakan dengan memperhatikan objek atau tetangga terdekat [16]. Jika objek yang ingin diprediksi kelasnya dekat dengan objek di kelas A maka objek tersebut akan masuk ke kelas A. Jarak Euclid digunakan untuk mengukur jarak antara satu objek dengan objek yang lain. Algoritma KNN dapat dilihat sebagai berikut.

Algoritma KNN

input: Fitur Latih

Fitur Uji yang belum diketahui kelasnya

1. Tentukan banyaknya tetangga terdekat yang ingin digunakan untuk proses klasifikasi (nilai K).
2. Hitung jarak euclid dari fitur uji yang ingin diprediksi ke semua fitur latih.
3. Pilih K objek fitur latih yang terdekat dengan fitur uji yang ingin diprediksi.
4. Lakukan prediksi kelas dari fitur uji berdasarkan K fitur latih terdekat yang telah dipilih.

### **HASIL DAN PEMBAHASAN**

Berdasarkan uraian sebelumnya, data yang digunakan adalah data citra wajah yang

terdiri dari 40 orang (kelas) dengan tiap orang memiliki 10 citra dengan berbagai ekspresi wajah. Pada penelitian ini diusulkan bahwa banyaknya data citra latih merupakan hal penting yang harus diperhatikan. Banyaknya citra wajah tiap orang yang digunakan sebagai data latih adalah 5, 6, 7, 8 dan 9 citra wajah per orang. Setelah itu dilakukan proses LDA untuk mendapatkan fitur latih dan fitur uji. Tahapan terakhir adalah melakukan klasifikasi berdasarkan fitur yang telah diperoleh dengan menggunakan metode KNN. Pada metode klasifikasi diusulkan untuk memperhatikan banyaknya jumlah tetangga terdekat yang digunakan dalam proses klasifikasi. Pada penelitian ini digunakan jumlah tetangga dari 1 sampai 24 tetangga terdekat. Hasil klasifikasi dapat dilihat pada Tabel 2.

Berdasarkan Tabel 2, jika banyaknya citra wajah per orang adalah 5 maka tingkat akurasi terbesar terjadi saat jumlah tetangga terdekat (K) pada metode KNN adalah 1 yaitu sebesar 96 %. Jika banyaknya citra wajah per orang adalah 6 maka tingkat akurasi terbesar terjadi saat jumlah tetangga terdekat (K) adalah 1 yaitu sebesar 92,5 %. Jika banyaknya citra wajah per orang adalah 7 maka tingkat akurasi terbesar terjadi saat jumlah tetangga terdekat (K) adalah 1 yaitu sebesar 91,67 %.

Tabel 2. Hasil Klasifikasi Pengenalan Wajah Dengan Metode KNN

Jumlah Tetangga Terdekat (K)	Banyaknya Citra Wajah Tiap Orang				
	5 (%)	6 (%)	7 (%)	8 (%)	9 (%)
1	96,0	92,5	91,67	93,75	97,5
2	86,0	87,5	91,67	90,0	95,0
3	84,5	85,0	85,0	88,75	92,5
4	80,0	82,5	84,17	88,75	92,5
5	80,0	72,5	84,17	85,0	90,0
6	79,0	75,0	82,5	85,0	82,5
7	79,5	72,5	79,17	82,5	85,0
8	75,5	67,5	80,0	77,5	82,5
9	75,0	65,0	79,17	75,0	77,5
10	72,0	67,5	75,83	73,75	75,0
11	73,0	65,0	75,0	75,0	75,0
12	73,0	70,0	75,0	73,75	72,5
13	71,0	70,0	73,33	71,25	70,0
14	69,0	67,5	71,67	68,75	70,0
15	67,0	65,0	71,67	70,0	62,5
16	64,5	60,0	69,17	67,5	62,5
17	62,0	57,5	69,17	67,5	62,5
18	59,0	50,0	68,33	67,5	62,5
19	57,5	50,0	65,83	67,5	60,0
20	54,5	50,0	65,0	67,5	57,5
21	54,0	50,0	64,17	63,75	57,5
22	53,5	50,0	63,33	62,5	57,5
23	53,5	52,5	60,83	65,0	57,5
24	52,5	52,5	59,17	62,5	60,0

Jika banyaknya citra wajah per orang adalah 8 maka tingkat akurasi terbesar terjadi saat jumlah tetangga terdekat (K) adalah 1 yaitu sebesar 93,75 %. Jika banyaknya citra wajah per orang adalah 9 maka tingkat akurasi ter-besar terjadi saat jumlah tetangga terdekat (K) adalah 1 yaitu sebesar 97,5 %.

Secara keseluruhan tingkat akurasi terbesar adalah sebesar 97,5% yaitu terjadi saat banyaknya data latih per orang adalah 9 (360 data latih) dengan banyaknya tetangga terdekat adalah 1.

**SIMPULAN DAN SARAN**

Pada sistem keamanan, sistem pengenalan wajah sangat dibutuhkan untuk mengetahui siapa yang sudah masuk ke area yang diamati. Pengenalan wajah merupakan salah satu cara yang digunakan untuk mengenali seseorang karena wajah seseorang memiliki ciri-ciri khusus yang dapat membedakan orang yang satu dengan orang yang lainnya. Pada penelitian ini digunakan metode *Linear Discriminant Analysis* (LDA) untuk mengenali wajah seseorang.

Data yang digunakan adalah data citra wajah yang terdiri dari 40 orang dengan tiap orang memiliki 10 citra wajah dengan berbagai ekspresi. Data tersebut dibagi

menjadi dua bagian yaitu data latih dan data uji. Pada penelitian ini pembagian data dilakukan beberapa kali berdasarkan banyaknya citra wajah tiap orang. Perbandingan citra wajah tiap orang yang digunakan sebagai data latih dan data uji adalah 5 : 5, 6 : 4, 7 : 3, 8 : 2 dan 9 : 1 citra wajah per orang.

Setelah itu digunakan metode LDA untuk mendapatkan fitur wajah seseorang. Fitur-fitur tersebut selanjutnya diklasifikasikan dengan menggunakan metode *K Nearest Neighbor* (KNN). Berdasarkan percobaan yang dilakukan diperoleh bahwa tingkat akurasi terbesar yaitu sebesar 97,5% yang terjadi saat perbandingan jumlah data latih dan data uji adalah 9 : 1.

Pada penelitian selanjutnya dapat digunakan metode lain sebagai pembandingan tingkat keberhasilan dalam pengenalan wajah seperti metode Local Binary Patterns Histograms (LBPH) dan metode lainnya.

## DAFTAR PUSTAKA

- [1]. R. Kosasih, "Kombinasi Metode Isomap Dan KNN Pada Image Processing Untuk Pengenalan Wajah," *CESS (Journal of Computer Engineering System and Science)*, vol. 5, no. 2, hal. 166 - 170, 2020.
- [2]. V. D. Maaten, E. Postma, dan V. D. Herik, "Dimensionality reduction a comparative review," *L. J. P. Technical Report TiCCTR*. 2009-005, 2009.
- [3]. L. Cayton, "Algorithms for manifold learning," *UCSD Technical Report CS*. 2008-0923, 2005.
- [4]. R. Kosasih, S. Madenda, C. M. Karyati dan Lussiana, "Determination the Optimal Position from T1 and T2 Weighted MR Imaging of the Abdominal Aortic Aneurysm," *Advance Science, Engineering and Medicine*, vol. 7, no. 10, hal. 915 - 919, 2015.
- [5]. R. Kosasih, "Penggunaan Metode PCA Untuk Reduksi Data Image Pembuluh Darah Vena," *Prosiding Konferensi Nasional Matematika XVII - ITS, Surabaya*, 2014, hal. 241 - 247.
- [6]. J.Y. Sari, "Pengenalan wajah pada citra digital menggunakan algoritma Eigenface dan Euclidean distance," *Skripsi, Jurusan Teknik Informatika, Universitas Halu Oleo, Kendari*, 2012.
- [7]. P. N. Belhumeur, J. P. Hespanha, dan D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, vol. 19, no.7, hal. 711 - 720.
- [8]. F. Mahmud, M. T. Khatun, S. T. Zuhori, S. Afroge, M. Aktar, dan B. Pal, "Face recognition using Principle Component Analysis and Linear

- Discriminant Analysis," *International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, 2015, hal. 1 - 4.
- [9]. I. N. T. A. Putra dan A. Harjoko, "Pengenalan Wajah Berbasis Mobile Menggunakan Fisherface dan Distance Classifier," *Jurnal Sains dan Teknologi*, vol. 7, no. 1, hal. 135 - 145, 2018.
- [10]. R. Kosasih dan A. Fahrurrozi, "Clustering of Face Images by Using Isomap method," *Proceeding on International Workshop on Academic Collaboration*, No (ISBN: 978-602-9438-86-4), 2017, hal. 52 - 56.
- [11]. A. Fahrurrozi dan R. Kosasih, "Face Recognition Using Local Binary Pattern Combined With PCA For Images Under Various Expression and Illumination," *Proceeding on International Workshop on Academic Collaboration*, No (ISBN: 978-602-9438-86-4), 2017, hal. 1 - 7.
- [12]. D. E. Pratiwi, dan A. Harjoko, "Implementasi Pengenalan Wajah Menggunakan PCA (Principle Component Analysis)," *Indonesian Journal of Electronics and Instrumentations Systems*, vol. 3, no. 2, hal. 175 - 184, 2013.
- [13]. AT&T Laboratories Cambridge, "The Database of Face, 2001. [Daring]. Available: <http://cam-orl.co.uk/facedatabase.html>. [Diakses: 14 April 2021].
- [14]. M. Anggo dan L. Arapu, "Face Recognition Using Fisherface Method," *J. Phys.: Conf. Ser.* 1028 012119, 2018.
- [15]. I. G. P. S. Wijaya, K. Uchimura, dan G. Koutaki, "Face Recognition Using Holistic Features and Linear Discriminant Analysis Simplification," *TELKOMNIKA*, vol.10, no.4, hal. 775 - 787, 2012.
- [16]. Murni, R. Kosasih, A. Fahrurrozi, T. Handhika, I. Sari dan D. P. Lestari, "Travel Time Estimation for Destination In Bali Using kNN-Regression Method with Tensorflow," *IOP Conf. Ser.: Mater. Sci. Eng.* 854 012061, 2020.