

Klasifikasi K-NN dalam Identifikasi Penyakit COVID-19 Menggunakan Ekstraksi Fitur GLCM

Nisa Nafisah^{1*}, Riza Ibnu Adam^{2*}, Carudin^{3*}

* Teknik Informatika, Universitas Singaperbangsa Karawang

nisa.nafisah17162@student.unsika.ac.id¹, riza.adam@staff.unsika.ac.id², carudin@staff.unsika.ac.id³

Article Info

Article history:

Received 2021-08-12

Revised 2021-10-20

Accepted 2021-10-22

Keyword:

Covid-19,

GLCM,

Data mining,

K-NN

ABSTRACT

Covid-19 merupakan penyakit yang sedang mewabah di berbagai belahan dunia termasuk Indonesia. Penyakit ini menginfeksi saluran pernapasan yang disebabkan oleh jenis virus corona baru. Untuk mengetahui adanya virus covid-19 di dalam tubuh dapat dilakukan pemeriksaan medis seperti cek darah, pemeriksaan radiologi rontgen (x-ray) dan swab. Penelitian ini melakukan identifikasi penyakit covid-19 berdasarkan citra rontgen yang mana citra tersebut diekstraksi menggunakan metode fitur ekstraksi GLCM yaitu *contrast*, *correlation*, *energy*, dan *homogeneity*. Setelah didapat nilai dari ekstraksi lalu diklasifikasikan menggunakan metode klasifikasi data mining yaitu k-nearest neighbor dengan melakukan 3 pemodelan nilai inputan k. Hasil penelitian yang diperoleh dari pengklasifikasian didapat akurasi sebesar 80% pada model 3 dengan nilai k = 5 dan pada model 1 dan 2 diperoleh hasil akurasi sebesar 90% dengan nilai k = 1 dan k = 3.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Pandemi yang sedang melanda seluruh dunia belakangan ini merupakan penyakit Coronavirus Disease 2019 atau COVID-19 yang mana penyakit ini menginfeksi saluran pernafasan yang disebabkan oleh jenis virus corona baru. Virus corona merupakan sekelompok besar virus yang memiliki duri seperti mahkota di permukaannya, dalam kata latin untuk mahkota berarti *coronam* [1]. Sebelumnya, pada tahun 2002 telah terjadi pandemic yang disebabkan oleh virus corona yang menginfeksi saluran pernapasan yang bernama SARS-CoV (*Severe Acture Respiratory Syndrome*) [2] dan juga pada tahun 2012 terjadi infeksi virus MERS-CoV yang pertama kali teridentifikasi di Saudi Arabia [3]. Sehingga organisasi Kesehatan dunia atau WHO menetapkan SARS-CoV-2 sebagai virus yang menyebabkan penyakit COVID-19. Untuk mengetahui adanya COVID-19 di dalam tubuh, diperlukan pemeriksaan medis seperti rapid test, swab, dan PCR. Orang yang memiliki gejala-gejala terindikasi COVID-19 akan di pinta melakukan cek darah, pemeriksaan radiologi rontgen menggunakan sinar-x (X-ray), lalu dilanjutkan swab [4]. Gejala-gejala orang terindikasi penyakit COVID-19 meliputi demam atau menggigil, batuk, sesak napas, sulit bernapas, kelelahan, pegal-pegal, sakit kepala, hilangnya

indra perasa atau penciuman, sakit tenggorokan, hidung tersumbat atau meler, mual atau muntah, dan diare [5]. Pemeriksaan radiologi berperan penting dalam melawan COVID-19 yaitu dapat menilai dan melihat perkembangan kondisi penyakit pada pasien karena pada beberapa kasus menyatakan negative pada tes virologi sedangkan saat dilakukan pemeriksaan radiologi dapat menunjukkan adanya infeksi virus pada tahap awal [6]. Radiologi itu sendiri merupakan cabang ilmu kedokteran yang bertujuan untuk mengetahui bagian dalam tubuh manusia menggunakan teknologi pencitraan yang berperan untuk diagnos penyakit, salah satu kondisinya yaitu untuk mengetahui adanya kelainan pada paru-paru [7]. Salah satu pemeriksaan radiologi yaitu pemeriksaan foto rontgen. Rontgen merupakan tes diagnostic berbasis pencitraan dengan sinar-x (x-ray) yang bertujuan untuk melihat bagian dalam tubuh seperti paru-paru dan saluran pernapasan dalam bentuk citra [8]. Oleh karena permasalahan tersebut, maka pada penelitian akan dilakukan pengidentifikasian penyakit COVID-19 berdasarkan citra x-ray.

Beberapa penelitian yang telah mengimplementasikan pengetahuan teknik pada citra medis yakni Afriyana, Purnamasari, & Patmasari [9] yang mendeteksi kelainan tulang belakang dengan menggunakan *gray level co-*

occurrence matrix (GLCM) dan *k-nearest neighbor* (KNN) dimana dalam proses pengujiannya dilakukan dengan menggunakan 128 citra tulang belakang, dengan kelas citra normal 53 citra, 36 citra dekstroskoliosis, dan 39 citra levoskoliosis. Hasil yang diperoleh dengan akurasi sebesar 84,84% pada sistem dengan parameter GLCM orde dua di keempat parameter, dengan arah 0° , 45° , 90° , 135° dan level kuantitas 8, sedangkan pada parameter KNN dengan nilai $k = 1$ di keempat parameter yang digunakan.

Selanjutnya [10] penelitian yang telah dilakukan mendeteksi penyakit COVID-19 berdasarkan citra X-Ray yang menggunakan pendekatan deep learning berbasis residual deep network dengan melakukan evaluasi untuk mengetahui performa metode berupa precision, recall, F1, dan accuracy. Hasil yang diperoleh dengan menggunakan metode ini memberikan nilai precision 0.98, recall 0.95, F1 0.97, dan accuracy sebesar 99%.

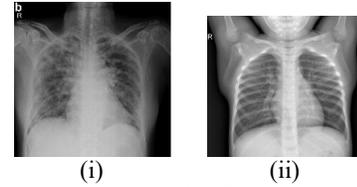
Penelitian selanjutnya [11] mengklasifikasikan pneumonia menggunakan metode K-Nearest Neighbor (K-NN) dan ekstraksi fitur *gray level co-occurrence matrix* (GLCM). Terdapat beberapa tahapan sebelum dilakukan pengklasifikasian yaitu cropping, resizing, contrast stretching, dan thresholding. Hasil yang diperoleh dari penelitian ini menunjukkan bahwa akurasi terbaik per kelas adalah 66,20% untuk $k = 5$.

Lalu pada penelitian ini akan dilakukan pengklasifikasian penyakit COVID-19 dengan mengimplementasikan teknik data mining, data mining merupakan penguraian pengetahuan yang terdapat pada basis data untuk memperoleh informasi yang pada prosesnya menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* [12], salah satu teknik data mining yang akan digunakan yaitu metode *k-nearest neighbor* (K-NN) yang mana parameter untuk pengklasifikasian diperoleh dari nilai-nilai fitur ekstraksi metode *gray level co-occurrence matrix* (GLCM) yakni *contrast*, *correlation*, *energy*, dan *homogeneity* agar dapat mengidentifikasi penyakit COVID-19 berdasarkan citra radiologi x-ray thorax. Thorax itu sendiri merupakan bagian tubuh yang terletak di antara leher dan abdomen [13] yang berbentuk kerucut yaitu lebih besar pada bagian bawah daripada bagian atasnya dan lebih panjang pada bagian belakang daripada bagian depan, yang tersusun atas rangka tulang dada (*Sternum*), 12 ruas tulang belakang (*Thoracalis*), dan 12 pasang tulang rusuk [14]

II. METODE PENELITIAN

A. Data Preparation

Tahap awal pada penelitian ini dilakukan persiapan data, dataset yang digunakan diperoleh dari situs terbuka Kaggle.com, pada dataset terdapat 69 citra terinfeksi COVID-19 dan 25 citra normal dengan total 94 citra dengan format .JPG, lalu citra yang telah disiapkan akan diklasifikasikan berdasarkan hasil ekstraksi fitur GLCM dengan 4 fitur nilai yang akan menjadi parameter yakni *contrast*, *correlation*, *energy*, dan *homogeneity*.



Gambar 1. (i) Citra yang terindikasi COVID-19 dan (ii) Citra yang tidak terindikasi COVID-19

B. Preprocessing

Sebelum dilakukan proses pengklasifikasian data mining, dilakukan *preprocessing* pada data citra yang akan digunakan.

- *Konversi citra RGB ke grayscale*

Setelah data citra diperoleh, tahap awal *preprocessing* yaitu memastikan data citra yang akan digunakan sudah dalam skala keabuan. Pada data terdapat beberapa citra yang masih berskala warna *RGB*. Untuk merubah citra *RGB* menjadi *grayscale*, metode umum yang digunakan yaitu dapat dilihat pada persamaan 1.

$$\frac{(R+G+B)}{3} \quad (1)$$



Gambar 2. Hasil konversi *RGB* ke *grayscale* pada citra

- *Cropping Citra*

Setelah semua data citra sudah dalam skala keabuan, langkah selanjutnya yaitu dilakukannya pemangkasan atau *cropping* pada citra, karena data citra yang diperoleh memiliki beragam pengambilan sudut gambar, maka dilakukan *cropping* untuk menyamakan tiap citra agar tepat pada objek yang akan diklasifikasikan yaitu paru-paru.



Gambar 3. Hasil pemangkasan pada citra

- *Resize Citra*

Setelah pemangkasan citra dilakukan, selanjutnya dilakukan *resize* pada citra karena ukuran tiap data citra yang telah di pangkas berbeda-beda, maka agar mempermudah proses ekstraksi fitur, dilakukan normalisasi pada ukuran tiap data citra sebesar 250×250 piksel.

Gambar 4. Hasil *resize* pada citra

C. Ekstraksi Fitur

Setelah dilakukan penyesuaian ukuran pada tiap data citra, lalu selanjutnya dilakukan ekstraksi fitur GLCM ekstraksi fitur merupakan proses pengenalan ciri juga menganalisis tekstur dalam suatu objek, tekstur tersebut dibuktikan dengan persamaan matematika, sehingga pada proses pengenalan dapat dibandingkan dengan objek lainnya menggunakan hasil analisis tekstur tersebut [15].

Metode yang digunakan dalam ekstraksi fitur yaitu GLCM, GLCM merupakan salah satu metode yang efektif yang dapat melakukan klasifikasi dan identifikasi dengan memproses ekstraksi ciri [11] dimana GLCM itu sendiri merupakan suatu matriks yang elemennya merupakan jumlah pasangan piksel yang memiliki tingkat kecerahan tertentu yang pada pikselnya terpisah oleh jarak d dan suatu sudut inklinasi θ [16]. Terdapat 4 fitur yang umumnya dipakai dalam berbagai penelitian yakni *contrast*, *correlation*, *energy*, dan *homogeneity*.

- 1) *Contrast* yaitu fitur yang melakukan perhitungan pada jumlah keberagaman intensitas keabuan citra. Nilai *contrast* akan semakin tinggi jika derajat keabuan setiap pasangan pikselnya semakin jauh.

$$Contrast = \sum_x \sum_y |x - y|^2 \cdot p(x, y) \quad (2)$$

- 2) *Correlation* yaitu fitur yang melakukan perhitungan untuk menunjukkan ukuran ketergantungan derajat keabuan pada citra untuk memberikan petunjuk adanya struktur linear dalam citra.

$$Correlation = \frac{\sum_x \sum_y (x - \mu_x) \cdot (y - \mu_y) \cdot p(x, y)}{\sigma_x \sigma_y} \quad (3)$$

- 3) *Energy* yaitu fitur yang menyatakan konsentrasi pasangan dengan intensitas keabuan pada matriks

$$Energy = \sum_x \sum_y \{p(x, y)\}^2 \quad (4)$$

- 4) *Homogeneity* yaitu jumlah level keabuan yang sejenis pada citra jika piksel semakin seragam maka nilai *homogeneity* makin tinggi.

$$Homogeneity = \sum_x \sum_y \frac{p(x, y)}{1 + |x - y|} \quad (5)$$

III. HASIL DAN EVALUASI

A. Modelling

Setelah data citra di ekstraksi, lalu membagi dataset dilakukan pemodelan menggunakan algoritma *k-nearest neighbor* dan *naïve bayes*.

K-nearest neighbor merupakan salah satu algoritma klasifikasi sederhana yang menyimpan semua kondisi dan mengklasifikasikan ke dalam kondisi baru berdasarkan jarak terdekat [17] dimana metode ini diproses berdasarkan jarak terdekat dari *query instance* ke data latih lalu mengambil k buah data terdekat [18]. *Euclidean distance* salah satu perhitungan untuk menghitung jarak pada data baru dengan data lama [19] [20].

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (6)$$

Pada tahap pemodelan dilakukan dengan menggunakan 3 model, yaitu sebagai berikut.

1) Model 1

Pada model 1 dilakukan pemodelan *k-nearest neighbor* dengan nilai $k = 1$, hasil klasifikasi dapat dilihat pada table I.

TABEL I
HASIL KLASIFIKASI MODEL 1

Kelas	Hasil Klasifikasi
Covid 1	Benar
Covid 2	Benar
Covid 3	Benar
Covid 4	Benar
Covid 5	Benar
Covid 6	Benar
Covid 7	Benar
Normal 1	Benar
Normal 2	Salah
Normal 3	Benar

2) Model 2

Pada model 2 dilakukan pemodelan *k-nearest neighbor* dengan nilai $k = 3$, hasil klasifikasi dapat dilihat pada table II.

TABEL II
HASIL KLASIFIKASI MODEL 2

Kelas	Hasil Klasifikasi
Covid 1	Benar
Covid 2	Benar
Covid 3	Benar
Covid 4	Benar
Covid 5	Benar
Covid 6	Benar
Covid 7	Benar
Normal 1	Benar

Normal 2	Salah
Normal 3	Benar

3) Model 3

Pada model 3 dilakukan pemodelan k-nearest neighbor dengan nilai $k = 5$, hasil klasifikai dapat dilihat pada table III.

TABEL III
HASIL KLASIFIKASI MODEL 3

Kelas	Hasil Klasifikasi
Covid 1	Benar
Covid 2	Benar
Covid 3	Benar
Covid 4	Benar
Covid 5	Benar
Covid 6	Benar
Covid 7	Benar
Normal 1	Benar
Normal 2	Salah
Normal 3	Salah

B. Evaluation

Setelah dilakukan pemodelan menggunakan algoritma k-nearest neighbor dengan nilai $k = 1,3$, dan 5 diperoleh hasil akurasi dengan menggunakan persamaan berikut.

$$Akurasi = \frac{Jumlah\ benar}{Jumlah\ Data\ Uji} \times 100\%$$

Pada model 1, pemodelan algoritma k-nearet neighbor dengan nilai $k = 1$ menghasilkan 1 nilai kesalahan output yang tida sesuai dengan kelas pada data uji ke 9, maka diperoleh hasil akurasi sebesar 90% dari total 10 data uji.

$$Akurasi = \frac{9}{10} \times 100\% = 90\%$$

Pada model 2, pemodelan algoritma k-nearet neighbor dengan nilai $k = 3$ juga menghasilkan 1 nilai kesalahan output yang tidak sesuai dengan kelas pada data uji ke 9, maka diperoleh hasil akurasi sebaesar 90% dari total 10 data uji.

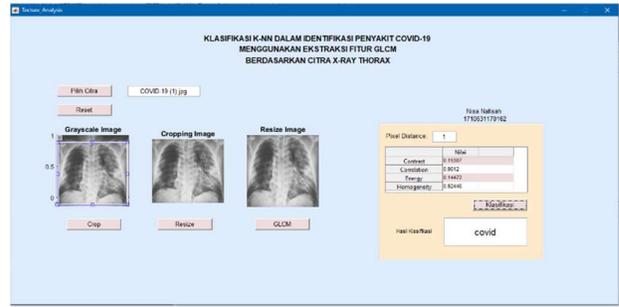
$$Akurasi = \frac{9}{10} \times 100\% = 90\%$$

Lalu pada model 3, pemodelan algoritma k-nearet neighbor dengan nilai $k = 5$ menghasilkan 2 nilai kesalahan output yang tidak sesuai dengan kelas pada data uji ke 9 dan 10, maka diperoleh hasil akurasi sebaesar 80% dari total 10 data uji.

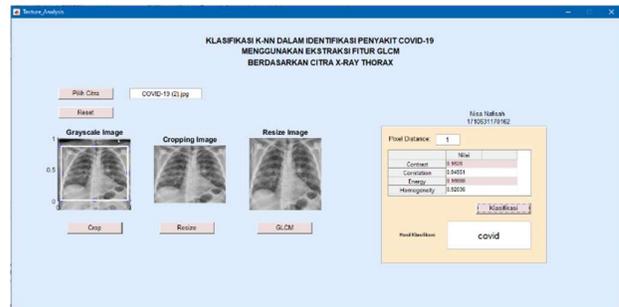
$$Akurasi = \frac{8}{10} \times 100\% = 80\%$$

C. Hasil

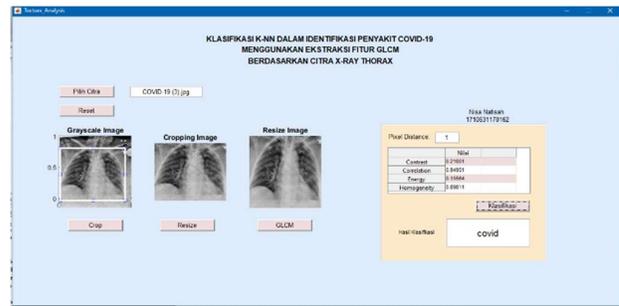
Pada tahap ini dilakukan pengimplementasian terhadap model yang telah diperoleh dengan menggunakan model yang memiliki akurasi terbaik yaitu dengan nilai $k = 3$ untuk dimasukan ke dalam program yang dibangun pada aplikasi bantu MATLAB dan menyiapkan 5 data baru untuk dilakukan pengujian. Hasil klasifikasi pada citra dapat dilihat pada gambar berikut.



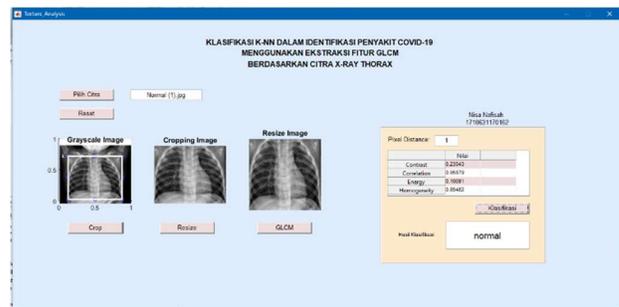
Gambar 5. Hasil klasifikasi citra covid-19 (1)



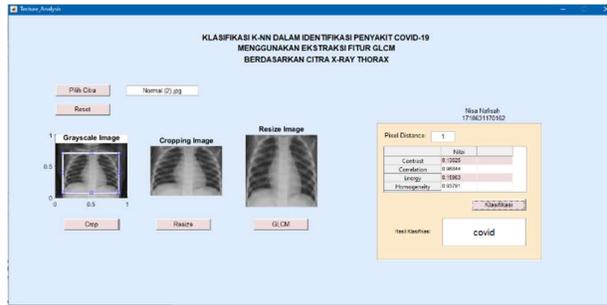
Gambar 6. Hasil klasifikasi citra covid-19 (2)



Gambar 7. Hasil klasifikasi citra covid-19 (3)



Gambar 8. Hasil klasifikasi citra normal (1)



Gambar 9. Hasil klasifikasi citra normal (2)

III. KESIMPULAN

Hasil pembahasan yang diperoleh dari penelitian yang telah dilakukan, dapat diambil beberapa kesimpulan diantaranya pengidentifikasian penyakit Covid-19 berdasarkan citra x-ray thorax dengan parameter atribut hasil dari ekstraksi fitur GLCM atau *gray level co-occurrence matrix* yaitu *contrast*, *correlation*, *energy*, dan *homogeneity* menggunakan Teknik klasifikasi data mining k-nearest neighbor atau k-nn dengan melakukan 3 pemodelan untuk mendapatkan hasil akurasi terbaik. Hasil yang diperoleh dengan menggunakan Teknik klasifikasi data mining k-nearest neighbor memperoleh hasil akurasi terbaik pada nilai $k=1$ dan $k=3$ dengan akurasi sebesar 90%.

DAFTAR PUSTAKA

- [1] J. Huizen, "Why is it called coronavirus?," 2020. <https://www.medicalnewstoday.com> (accessed Mar. 08, 2021).
- [2] M. D. C. Pane, "SARS," 2020. <https://www.alodokter.com/sars>.
- [3] N. R. Aida, "7 Wabah Virus yang pernah Mengguncang Dunia Selain Corona," 2020. https://www.kompas.com/tren/read/2020/02/06/171200465/7-wabah-virus-yang-pernah-mengguncang-dunia-selain-corona?amp=1&page=3&jxconn=1*la89al*other_jxampid*NmhETI FUTDVGbmJDaWdlUXZ4OW9FX2U3LVIVOTRnOVhQS21QSDk4YWpHZGxoZWNDM0hiTURoVG5iNXNjS0c2TW.
- [4] Tim, "Berita Kesehatan: Kegunaan Tes Darah, Rontgen, dan Swab," 2020. <https://m.cnnindonesia.com> (accessed Mar. 14, 2021).
- [5] Cdc.gov, "Symptoms," 2021.
- [6] B. Yanti and U. Hayatun, "Peran pemeriksaan radiologi pada diagnosis Coronavirus disease 2019," *J. Kedokt. Syiah Kuala*, pp. 53–57, 2020.
- [7] H. R., "Artikel: Perlu Tahu, Ini Jenis-Jenis Pemeriksaan Radiologi," Halodoc.com, 2019. <https://www.halodoc.com> (accessed Apr. 14, 2021).
- [8] M. Helmi, "Rontgen," 2020. <https://www.sehatq.com> (accessed Apr. 14, 2021).
- [9] Y. Afriyana, R. Purnamasari, and R. Patmasari, "Deteksi Kelainan Tulang Belakang Berdasarkan Citra Medis Digital Dengan Menggunakan Gray Level Co-occurrence Matrix (GLCM) dan K-nearest Neighbor (KNN)," in *e-Proceeding of Engineering*, 2018, pp. 4657–4682.
- [10] Y. . Hariyani, S. Hadiyoso, and T. SuprptoSiadari, "Deteksi Penyakit Covid-19 Berdasarkan Citra X-Ray Menggunakan Deep Residual Network," *ELKOMIKA J. Tek. Energi Elektr. Tek. Inform. Tek. Elektron.*, pp. 443–453, 2020.
- [11] C. Wijaya, H. Irsyad, and W. Widhiarso, "Klasifikasi Pneumonia Menggunakan KNearest Neighbor Dengan Ekstraksi GLCM," *J. Algoritma.*, pp. 33–44, 2020.
- [12] G. Gunadi and D. I. Sensus, "Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori Dan Frequent Pattern Growth (Fp-Growth) : Studi Kasus Percetakan Pt.Gramedia," *J. Telemat. MKOM*, vol. 4, pp. 118–132, 2012.
- [13] R. N. Hayati, "Segmentasi Paru-Paru Pada Citra Digital Hasil X-Ray Thorax Menggunakan Metode Level Set Untuk Menghitung Diameer Maksimal Paru-Paru," 2013.
- [14] N. Surani, "Berita: Mengenal Rontgen Thorax (Dada)," *RS Betha Medika*, 2017. <https://www.kasih-group.com/id/berita/3836/mengenal-rontgen-thorax-dada-> (accessed Apr. 19, 2021).
- [15] H. A. Pitoyo, "Implementasi Metode Support Machine Untuk Klasifikasi Daun Mangga Berdasarkan Tekstur Daun," 2020.
- [16] I. Purnamasari and T. Stujo, "Pengenalan Ciri Garis Telapak Tangan Menggunakan Ekstraksi Fitur (Glem) Dan Metode K-Nn," *J. VOI(Voice Informatics)*, pp. 32–41, 2017.
- [17] D. S. Sayad, "datamining," 2010. <https://chem-eng.utoronto.ca/~datamining/>.
- [18] F. Agustina and Z. A. Ardiansyah, "Identifikasi Citra Daging Ayam Kampung dan Mroiler Menggunakan Metode GLCM dan Klasifikasi-NN," *J. INFOKAM*, 2020.
- [19] D. E. Kurniawan and A. Dzikri, "Pengenalan Personal Berdasarkan Pengukuran Jarak Citra Wajah Menggunakan Pendekatan Linear dan Nonlinear," in *Seminar Nasional Teknik Informatika dan Komputer*, 2015, pp. 1–4.
- [20] D. E. Kurniawan, "Identifikasi Citra Wajah Menggunakan Gabor-based Kernel Principal Component Analysis," in *Applied Business and Engineering Conference (ABEC 2014)*, 2014, pp. 1–5.