



Pengenalan Wajah menggunakan Principle Component Analysis (PCA) dengan Model Algoritma Machine Learning untuk Mengidentifikasi Jenis Kelamin pada Kartu Identitas Mahasiswa

Face Recognition using Principle Component Analysis (PCA) with Machine Learning Algorithm Model to Identify Gender on Student Identity Cards

Dian Ade Kurnia¹, Umi Hayati², Tuti Hartati³, Salsa Loni Manikari⁴, Fahmi Afandi⁵

¹Manajemen Informatika, STMIK IKMI Cirebon

^{2,3,4}Teknik Informatika, STMIK IKMI Cirebon

¹Dianade2014@gmail.com, ²umi.haya41@gmail.com, ³toohart2013@gmail.com, ⁴Lonimanikari06@gmail.com,

⁵brengestextreme.pkfr@gmail.com

Abstract

Numerous individuals are still investigating the field of facial recognition research. This is evident from the advancement of computer vision technologies, which are utilized in a variety of real-world applications. This study's objective was to identify a face based on the traits or gender presentation on a college student's identification card. SEMMA (Sample, Explore, Modify, Model, and Assessment) is a data science or machine learning technique that utilizes the Support Vector Machine (SVM) and Artificial Neural Network (ANN) methods (ANN). Nevertheless, the modeling is also helped by preprocessing with the Principle Component Analysis (PCA) technique, which seeks to decrease the dimensions of current picture characteristics to selected features. According to the findings of this study, the accuracy performance of the SVM algorithm improved by 77.50% and 78.10%, respectively. The performance improvement was greater than in prior experiments that did not use dimension reduction approaches with PCA.

Keywords: face recognition, computer vision, SVM, ANN, PCA

Abstrak

Kajian tentang pengenalan wajah sampai saat ini masih banyak orang yang melakukan eksplorasi, hal ini dapat dilihat dari perkembangan teknologi *Computer Vision* yang diterapkan diberbagai aplikasi kehidupan. Tujuan penelitian ini adalah untuk mengidentifikasi wajah seseorang berdasarkan ciri atau featur jenis kelamin pada kartu identitas mahasiswa di sebuah perguruan tinggi. Metode yang digunakan melalui pendekatan data sains atau machine learning yaitu SEMMA (Sample, Explore, Modify, Model dan Asses) dengan penerapan pemodelan 2 (dua) algoritma yakni Support Vector Machine (SVM) dan Artificial Neural Network (ANN). Namun pemodelan tersebut juga didukung dengan pre-processing dengan teknik Principle Component Analysis (PCA) yang tujuannya mereduksi dimensi dari berbagai fitur gambar yang ada menjadi fitur yang terpilih. Hasil yang diperoleh dari penelitian ini bahwa adanya peningkatan performance pada aspek akurasi 77.50% untuk algoritma SVM dan 78.10%. Perolehan kinerja tersebut lebih baik dari penelitian sebelumnya yang tidak melibatkan teknik dimensi reduksi menggunakan PCA.

Kata kunci : pengenalan wajah, computer vision, SVM, ANN, PCA

1. Pendahuluan

Pengenalan wajah merupakan proses identifikasi seseorang berdasarkan apa yang mereka lihat. Saat ini dengan teknologi digital dapat dengan mudah untuk mengakses objek gambar maupun video[1]. Mengenali wajah manusia menggunakan teknik pengenalan wajah telah menjadi bagian penting dalam aktivitas penelitian para ilmuwan dibidang *computer vision*[2]. Proses

pengenalan wajah terdiri dari dua bagian, pertama adalah proses ekstraksi fitur dan kedua adalah membandingkan fitur yang diekstraksi dengan beberapa metode[3].

Proses ekstraksi fitur dapat dilakukan dengan cara mereduksi dimensi sebuah data melalui teknik diantaranya adalah Principle Component Analysis (PCA), Independent Component Analysis (ICA), dan

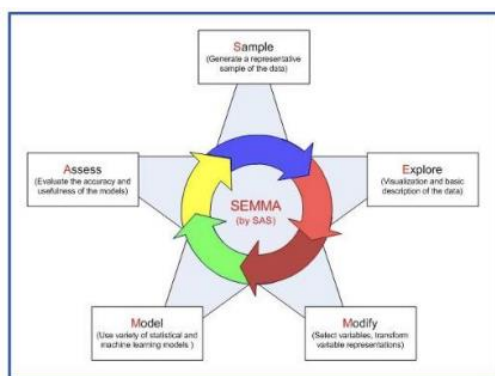
Linear Discriminant Analysis (LDA) [4]. PCA merupakan metode pengukuran statistik yang mengurangi dimensi gambar dalam ruang data menjadi dimensi yang lebih kecil [5]

Banyak penelitian yang telah dilakukan terkait dengan penggunaan teknik pengurangan dimensi gambar dengan PCA ini ketika diimplementasikan ke model algoritma machine learning seperti Support Vector Machine (SVM), Back Propagation Neural Network (BPNN), Artificial Neural Network (ANN) dan sebagainya. Seperti yang telah dilakukan oleh [6] bahwa rata-rata akurasi dalam sistem mengenal wajah dihasilkan sebesar 75.6%. Pada penelitian lainpun PCA telah digunakan terhadap algoritma Support Vector Machine (SVM) untuk mengenal wajah seseorang, dimana hasilnya ada peningkatan sebesar 15% setelah digunakan PCA sebagai fungsi untuk mereduksi dimensi gambar [7] [8]

Berdasarkan dari penelitian terkait [9] [10] dapat diyakini bahwa teknik reduksi dimensi gambar dengan PCA memiliki kontribusi terhadap model algoritma yang diusulkan. Oleh karena itu, penelitian ini bertujuan untuk menerapkan model machine learning yang menggunakan algoritma berbasis artificial neural network (ANN) dan Support Vector Machine (SVM) dengan melibatkan teknik reduksi dimensi PCA untuk mengidentifikasi Kartu Tanda Mahasiswa pada sebuah perguruan tinggi. Sehingga hasilnya jauh lebih baik dari penelitian sebelumnya.

2. Metode Penelitian

Metode yang digunakan dalam kajian ini adalah melalui pendekatan metodologi data science yang dinamakan SEMMA [11] yang terdiri dari tahapan Sample, Explore, Modify, Model, Asses yang digambarkan seperti gambar 1.

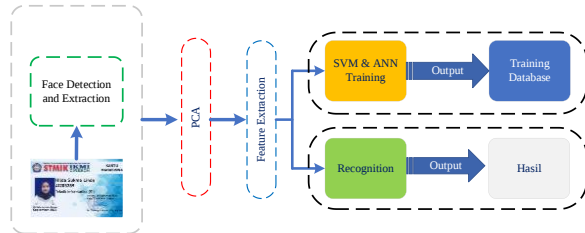


Gambar 1 Metodologi SEMMA

Metode ini sesuai dengan namanya melakukan serangkaian kegiatan yang bersifat siklus (berulang) yakni **sample** merupakan kegiatan ekstraksi data untuk mendapatkan dataset yang cukup guna mendapatkan informasi yang signifikan namun tidak terlalu besar

sehingga mudah untuk diproses selanjutnya, **explore** merupakan kegiatan mengeksplorasi data dengan mencari trend dan anomali untuk mendapatkan pemahaman tentang data, **modify** merupakan kegiatan modifikasi data dengan membuat dan memilih serta transformasi variabel untuk proses pemodelan, **model** merupakan kegiatan pemodelan dari data dengan mencari secara otomatis kombinasi data yang dapat dipakai untuk prediksi, **asses** mengevaluasi pola yang ditemukan apakah berguna dan cukup andal.

Secara garis besar blok diagram dari sistem yang akan diusulkan ini dapat ditunjukkan pada gambar 2 :



Gambar 2. Blok Diagram Sistem Pengenalan Wajah PCA + SVM & ANN

2.1. Dataset

Dataset berbentuk dataset primer sebanyak 324 gambar yang diperoleh secara langsung dari tempat penelitian yaitu di Bagian Kemahasiswaan STMIK IKMI Cirebon. Dataset berupa kumpulan file kartu mahasiswa yang berekstensi JPEG (gambar 3).



Gambar 3. Image Pra Processing

Selanjutnya gambar KTM tersebut dilakukan cropping untuk mengambil area wajahnya saja. Setelah dilakukan cropping selanjutnya dilakukan proses perubahan dari gambar warna RGB ke Black white, dan dilakukan pemrosesan ukuran dimensi dan resolusinya. Tabel 1 menjelaskan secara detail fitur dari gambar tersebut yang memiliki ukuran berikut :

Tabel 1. Fitur Gambar secara Detail

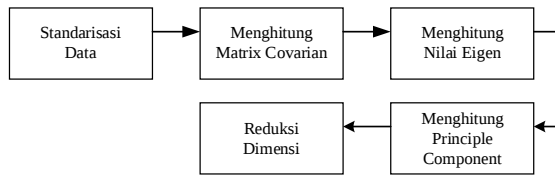
Fitur	Ukuran
Dimension	28 x 28
Weight	28 pixels
Height	28 pixels
Horizontal Resolution	64 dpi
Vertical Resolution	64 dpi
Bit Depth	24

2.2. Principle Component Analysis (PCA)

Principle Component Analysis (PCA) adalah suatu teknik analisis data yang digunakan dalam ilmu statistik

dan saat ini banyak juga digunakan pada data sains[3]. Tujuannya adalah untuk meringkas tabel data multivariat dalam skala besar hingga bisa dijadikan kumpulan variabel yang lebih kecil atau dengan istilah populernya dimensional reduksi [1].

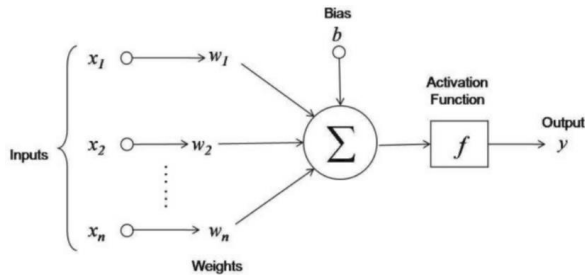
Tahapan PCA dapat dilihat pada gambar 4:



Gambar 4. Tahapan Proses PCA

2.3. Algoritma Artificial Neural Network (ANN)

ANN merupakan temuan penting dalam bidang machine learning. ANN terinspirasi oleh sistem saraf otak manusia yang disebut dengan neuron dalam melakukan klasifikasi data [12] [13].

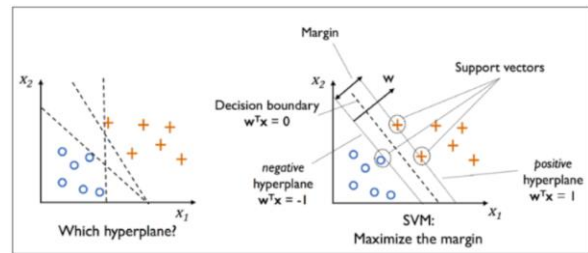


Gambar 5. Arsitektur ANN

Model terdiri dari beberapa nilai input yang dinyatakan dengan $x_1, x_2, \dots, x_n$, terhubung ke node *hidden layer* melalui faktor penimbangan. *Hidden layer* menjumlahkan nilai input menggunakan faktor penimbangan yang ditentukan selama training. Faktor penimbangan meniru kekuatan koneksi neural. Setiap *hidden node* dapat dibias oleh sinyal pengaktif sedemikian rupa sehingga menghasilkan output hanya jika jumlah sinyal input melebihi kondisi ambang batas yang telah ditentukan sebelumnya [14].

2.4. Support Vectore Machine (SVM)

Support Vector Machine (SVM) merupakan salah satu metode dalam supervised learning yang biasanya digunakan untuk klasifikasi [15]. Cara kerja SVM adalah mencari hyperplane terbaik dengan memaksimalkan jarak antar kelas. Gambar 6 memperlihatkan Hyperplane yang merupakan sebuah fungsi yang dapat digunakan untuk pemisah antar kelas[16].



Gambar 6 Hyperplane yang memisahkan dua kelas positif (+1) dan negatif (-1)

Hyperplane yang ditemukan SVM diilustrasikan seperti gambar diatas posisinya berada ditengah-tengah antara dua kelas, artinya jarak antara hyperplane dengan objek-objek data berbeda dengan kelas yang berdekatan (terluar) yang diberi tanda bulat kosong dan positif. Dalam SVM objek data terluar yang paling dekat dengan hyperplane disebut support vector.

2.5. Tools yang digunakan

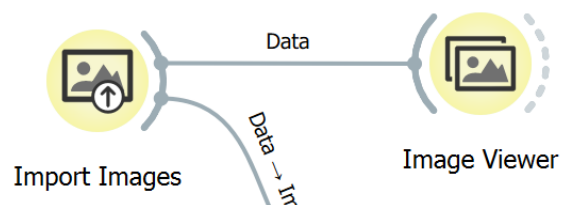
Untuk membuktikan hasil penelitian ini, digunakan tools **Orange Data Mining Versi 3.2** yang dilengkapi dengan adds on *Image Analytic* yang berfungsi sebagai library untuk mengolah data gambar.

3. Hasil dan Pembahasan

Pada bagian ini akan diimplementasikan hasil eksperimen yang telah dilakukan dari mulai tahapan data preparation, pemodelan dan evaluasi pemodelan.

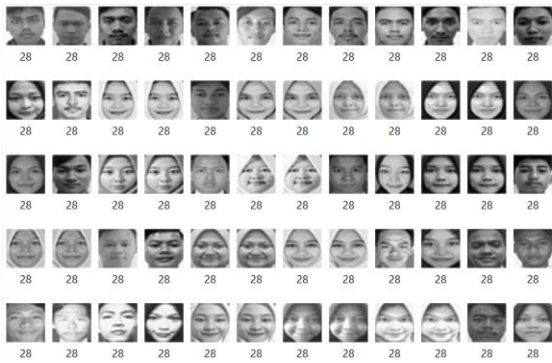
3.1. Data Preparation

Melalui fasilitas import images, orange data mining memanggil data yang telah disiapkan berdasarkan katagori jenis kelamin yaitu Laki-Laki dan Perempuan, seperti pada tampilan langkah pada gambar 7.



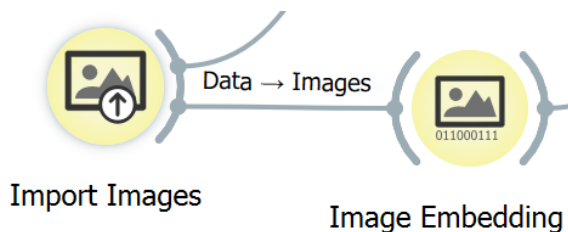
Gambar 7. Import Images untuk memanggil data

Selanjutnya data yang telah dipanggil dapat ditampilkan melalui *image viewer*, sehingga hasilnya seperti terlihat pada gambar 8.



Gambar 8. Tampilan Wajah menggunakan Image Viewer

Supaya data dapat dibaca melalui data angka, maka pada orange data mining memiliki fasilitas encoding dengan melibatkan operator *Image Embedding*. Gambar 9 memperlihatkan Teknik *Embedding* yang menggunakan arsitektur *embeded inception 3*.

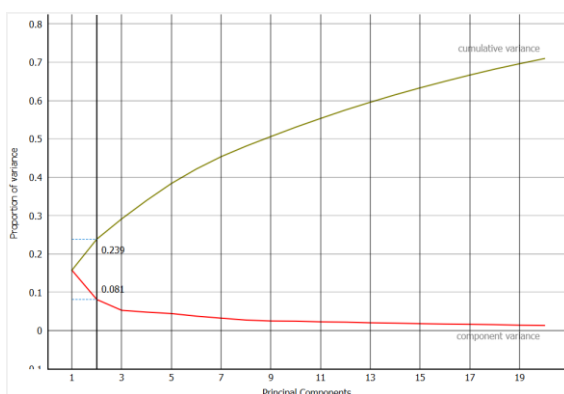


Gambar 9 Image Embedding dengan Arsitektur Inception 3

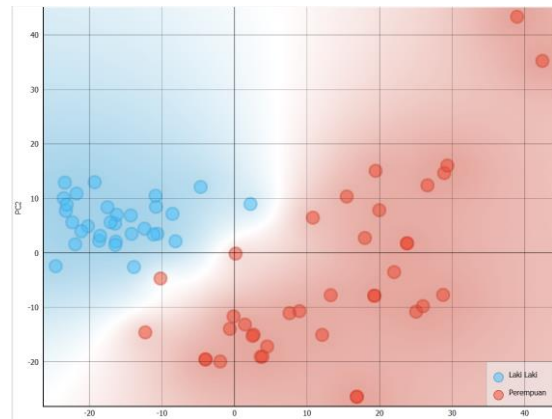
Hasil dari embeding tersebut diproses melalui operator PCA untuk mereduksi dimensi fitur dengan mengambil dua komponen yang principle, sehingga diperoleh fitur yang terseleksi seperti pada tabel 2, gambar 10 dan gambar 11.

Tabel 2. Seleksi Pitur dengan PCA

	category	Image name	Image	size	width	height	PC1	PC2
1	Laki Laki	31010140e copy	Laki Laki/31010...	9302	28	28	-14.591	-0.131903
2	Laki Laki	31190001 copy	Laki Laki/31190...	9183	28	28	-19.7074	6.27686
3	Laki Laki	31190002 copy	Laki Laki/31190...	9337	28	28	-17.2239	1.7753
4	Laki Laki	31190004 copy	Laki Laki/31190...	9138	28	28	-18.9142	-0.318818
5	Laki Laki	31190005 copy	Laki Laki/31190...	9269	28	28	-8.85618	1.61283



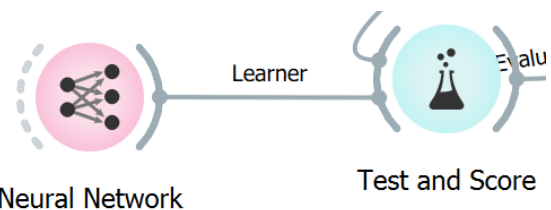
Gambar 10. Grafik PCA



Gambar 11. Grafik Sebaran Seleksi Fitur dengan PCA

3.2. Membangun Model

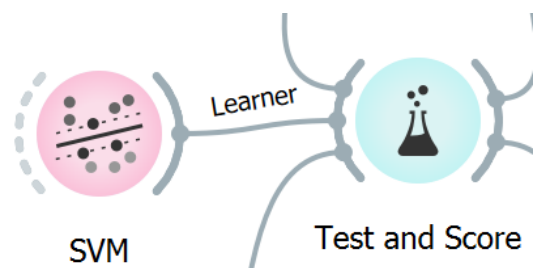
Setelah melakukan *data preparation* dan data telah siap untuk dimodelkan, maka dilakukan pemodelan dengan algoritma klasifikasi yaitu *Artificial Neural Network (ANN)* dan *Support Vector Machine (SVM)* (gambar 12).



Gambar 12. Desain Pemodelan Artificial Neural Network (ANN)

Dalam pemodelan tersebut, parameter yang digunakan adalah jumlah neuron pada *hidden layer 10*, *activation function* menggunakan *ReLU*, dan parameter optimasinya *Adam*.

Pemodelan menggunakan algoritma klasifikasi *Support Vector Machine* dapat dilihat pada desain gambar 13



Gambar 13. Desain Pemodelan SVM

Parameter yang dilibatkan dalam SVM itu sendiri adalah SVM Const (C) 1.00, Regression loss epsilon 0.10 dengan parameter optimasinya adalah numerical tolerance sebesar 0.0010.

Sehingga hasil pengujian dan score dari kedua pemodelan tersebut dapat ditampilkan seperti pada tabel 3.

Tabel 3. Perbandingan Model SVM dan ANN

Model	AUC	CA	F1	Preccission	Recall
SVM	0.848	0.775	0.775	0.775	0.775
ANN	0.861	0.781	0.781	0.783	0.781

Dari tabel 3 tersebut dapat dijelaskan bahwa model algoritma Neural Network memiliki tingkat akurasi lebih baik dibandingkan dengan model algoritma SVM yaitu sebesar 78.10%, begitu pula dengan nilai *Precision* dan *Recall* serta *F1-score*.

3.3. Evaluasi Model

Evaluasi model dilakukan dengan cara mengukur performa dari model yang telah dibangun melalui tabel Confution Matrix. Confution matrik menyajikan tabel untuk mengukur kinerja dari model klasifikasi. Confution matrix akan menampilkan dan membandingkan nilai aktual yang sebenarnya dengan nilai hasil prediksi model yang dapat digunakan untuk menghitung metrik evaluasi diantaranya adalah *Accuracy*, *Precision*, *Recall* dan *F1-Score*.

Tabel 4. Confusion Matrix Model SVM

		Predicted		Σ
		Laki-Laki	Perempuan	
Actual	Laki-Laki	74.8%	20.1%	150
	Perempuan	25.2%	79.9%	174
Σ		155	169	324

Tabel 5. Confusion Matrix Model ANN

		Predicted		Σ
		Laki-Laki	Perempuan	
Actual	Laki-Laki	74.8%	18.8%	150
	Perempuan	25.2%	81.2%	174
Σ		159	165	324

Berdasarkan tabel 4 dan 5, nilai prediksi jenis kelamin laki-laki pada algoritma SVM lebih tinggi (sebesar 20.1%) dibanding dengan nilai prediksi jenis kelamin laki-laki pada algoritma ANN. Namun, nilai prediksi jenis kelamin perempuan pada algoritma SVM lebih kecil (sebesar 79.9%) dibanding dengan nilai prediksi jenis kelamin perempuan pada algoritma ANN.

4. Kesimpulan

Dari hasil uji eksperimen dataset gambar wajah pada kartu identitas mahasiswa, dengan menggunakan teknik dimensi reduksi PCA dan pemodelan SVM serta Neural Network dapat disimpulkan bahwa adanya peningkatan performance dari model yang dikembangkan antara sebelum melibatkan dimensi reduksi PCA dengan tidak melibatkannya. Hal ini membuktikan bahwa model yang dirancang sudah cukup baik untuk digunakan. Namun demikian, model ini tentunya belum maksimal, oleh karena itu penelitian berikutnya dapat

mengembangkan model klasifikasi ini dengan menggunakan teknik deeplearning. Tidak hanya saja teknik yang dikembangkan, namun tahapan pre-processing yang harus dilakukan pun menjadi sangat penting. Karena, pada objek gambar, banyak sekali fitur fitur yang dihasilkan, sehingga diharapkan, dengan teknik pra processing dan seleksi fitur yang baik akan berpengaruh kepada hasil kinerja dari model yang dibangun.

Ucapan Terimakasih

Terimakasih kepada Dirjen Pendidikan Tinggi Kemendikbud Ristek yang telah membiayai terhadap publikasi melalui Program Kompetisi Hibah Penelitian Dosen Pemula tahun anggaran 2022.

Daftar Rujukan

- [1] S. Singh and S. V. A. V. Prasad, "Techniques and challenges of face recognition: A critical review," *Procedia Comput Sci*, vol. 143, pp. 536–543, 2018, doi: 10.1016/j.procs.2018.10.427.
- [2] Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, S. Nasrin, and V. K. Asari, "Comprehensive Survey on Deep Learning Approaches," 2017, [Online]. Available: <https://arxiv.org/pdf/1803.01164.pdf>
- [3] M. A. Marjan, M. R. Islam, M. P. Uddin, M. I. Afjal, and M. al Mamun, "PCA-based dimensionality reduction for face recognition," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 19, no. 5, pp. 1622–1629, 2021, doi: 10.12928/TELKOMNIKA.v19i5.19566.
- [4] M. K. Halidu, P. Bagheri-Zadeh, A. Sheikh-Akbari, and R. Behringer, "PCA in the context of Face Recognition with the Image Enlargement Techniques," *2019 8th Mediterranean Conference on Embedded Computing, MECO 2019 - Proceedings*, no. June, pp. 10–14, 2019, doi: 10.1109/MECO.2019.8760162.
- [5] S. A. Baker, H. H. Mohammed, and H. A. Aldabagh, "Improving face recognition by artificial neural network using principal component analysis," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 6, pp. 3357–3364, 2020, doi: 10.12928/TELKOMNIKA.v18i6.16335.
- [6] N. Jamil, S. Iqbal, and N. Iqbal, "Face Recognition Using Neural Networks."
- [7] Y. Wang and Q. Wu, "Research on Face Recognition Technology Based on PCA and SVM," in *2022 7th International Conference on Big Data Analytics, ICBDA 2022*, 2022, pp. 248–252, doi: 10.1109/ICBDA55095.2022.9760320.
- [8] S. Pasricha, "Evaluation of Carcass Characteristics of Japanese Quail Using Principal Component Analysis (PCA)," *Archives of Anesthesiology and Critical Care*, vol. 4, no. 4, pp. 527–534, 2020.
- [9] Y. C. See and N. M. Noor, "FACE RECOGNITION USING COMPLETE GABOR FILTER WITH RANDOM FOREST," vol. 13, no. 13, 2018, Accessed: Aug. 10, 2021. [Online]. Available: www.arpnjournals.com
- [10] N. N. A. Mangshor, I. A. A. Majid, S. Ibrahim, and N. Sabri, "A real-time drowsiness and fatigue recognition using support vector machine," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 4, pp. 584–590, 2020, doi: 10.11591/ijai.v9.i4.pp584-590.
- [11] A. Purbasari, F. R. Rinawan, A. Zulianto, A. I. Susanti, and H. Komara, "CRISP-DM for Data Quality Improvement to Support Machine Learning of Stunting Prediction in Infants and Toddlers," *Proceedings - 2021 8th International Conference on Advanced Informatics: Concepts, Theory, and*

- Application*, *ICAICTA* 2021, no. M1, 2021, doi: 10.1109/ICAICTA53211.2021.9640294.
- [12] R. Immanuel *et al.*, "Performance Prediction of solar still using Artificial neural network," *Mater Today Proc*, Sep. 2022, doi: 10.1016/j.matpr.2022.08.311.
- [13] M. Ozcalici and M. Bumin, "Optimizing filter rule parameters with genetic algorithm and stock selection with artificial neural networks for an improved trading: The case of Borsa Istanbul," *Expert Syst Appl*, vol. 208, Dec. 2022, doi: 10.1016/j.eswa.2022.118120.
- [14] K. Rusek and P. Guzik, "Two-stage neural network regression of eye location in face images," *Multimed Tools Appl*, vol. 75, no. 17, pp. 10617–10630, 2016, doi: 10.1007/s11042-014-2114-z.
- [15] J. Estupiñan and M. Santa, "Implementation of algorithms based on support vector machine (SVM) for electric systems topic review," *Tecnura*, vol. 20, no. 48, 2016.
- [16] P. L. Jackson, "Support vector machines as Bayes' classifiers," *Operations Research Letters*, vol. 50, no. 5, pp. 423–429, Sep. 2022, doi: 10.1016/j.orl.2022.06.003.