



Klasifikasi Gender Berbasis Citra Wajah Menggunakan Clustering Dan Deep Learning

Okky Prasetya, Syaeful Machfud, Perani Rosyani*, Bobi Agustian

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia
Email: ¹dosen02837@unpam.ac.id, ²dosen02836@unpam.ac.id, ^{3,*}dosen00837@unpam.ac.id, ⁴dosen00679@unpam.ac.id
Email Penulis Korespondensi: dosen00837@unpam.ac.id

Abstrak—Klasifikasi gender berbasis citra wajah merupakan tantangan penting dalam bidang *computer vision*, khususnya ketika data berasal dari sumber tidak terstruktur seperti media sosial. Penelitian ini mengusulkan pendekatan terpadu yang menggabungkan preprocessing citra wajah, metode clustering, dan deep learning untuk meningkatkan akurasi klasifikasi gender. Dataset yang digunakan berasal dari Big Data Competition dan terdiri dari gambar wajah pria dan wanita yang diperoleh dari Instagram. Tahapan preprocessing dilakukan dengan OpenCV untuk deteksi dan cropping wajah. Selanjutnya, data dikelompokkan menggunakan algoritma K-Means dan DBSCAN untuk mengurangi noise dan duplikasi. Proses klasifikasi dilakukan menggunakan model *sequential learning* berbasis Inception_v3 yang diperkuat dengan teknik Agglomerative Clustering. Evaluasi sistem menunjukkan performa yang sangat baik dengan akurasi 92.97%, F1-score 0.89556, precision 0.97727, dan recall 0.83069. Hasil ini membuktikan bahwa integrasi metode clustering dan deep learning dapat secara signifikan meningkatkan efektivitas klasifikasi gender berbasis citra wajah, terutama pada data yang berasal dari sumber terbuka.

Kata Kunci: Klasifikasi Gender; Citra Wajah; Deep Learning; OpenCV; Clustering; Inception_v3; Agglomerative Clustering

Abstract—Gender classification based on facial images is a significant challenge in the field of computer vision, especially when dealing with unstructured data sourced from social media platforms. This study proposes an integrated approach combining facial image preprocessing, clustering methods, and deep learning to enhance the accuracy of gender classification. The dataset used was obtained from a Big Data Competition and consists of male and female face images sourced from Instagram. Preprocessing was performed using OpenCV for face detection and cropping. Subsequently, the data were clustered using K-Means and DBSCAN algorithms to reduce noise and redundancy. Gender classification was then conducted using a sequential learning model based on Inception_v3, enhanced with Agglomerative Clustering for feature refinement. The evaluation of the system demonstrated strong performance with an accuracy of 92.97%, F1-score of 0.89556, precision of 0.97727, and recall of 0.83069. These results confirm that the integration of clustering techniques and deep learning significantly improves the effectiveness of gender classification based on facial images, especially for open-source and non-curated datasets.

Keywords: Gender Classification; Facial Image; Deep Learning; OpenCV; Clustering; Inception_v3; Agglomerative Clustering

1. PENDAHULUAN

Klasifikasi gender berbasis citra wajah merupakan salah satu topik yang berkembang pesat dalam bidang *computer vision* dan *machine learning*. [1], [2] Penerapan sistem klasifikasi otomatis ini semakin dibutuhkan dalam berbagai sektor, seperti sistem keamanan, survei demografi, serta rekomendasi berbasis profil pengguna. Mesin dituntut untuk mampu membedakan wajah pria dan wanita dengan tingkat akurasi tinggi, meskipun harus menghadapi tantangan berupa variabilitas pose, pencahayaan, ekspresi wajah, serta keberagaman etnis pada gambar yang dianalisis. [3][4]

Namun demikian, proses klasifikasi ini memiliki sejumlah kendala yang belum sepenuhnya teratasi. Salah satu permasalahan utama adalah kualitas dan keragaman dataset yang digunakan untuk pelatihan model. Gambar-gambar wajah yang diperoleh dari sumber terbuka seperti media sosial sering kali mengandung *noise*, pencahayaan yang tidak seragam, dan bahkan duplikasi gambar dalam satu folder. Situasi ini menyebabkan kesulitan bagi sistem dalam mengenali fitur-fitur spesifik yang membedakan gender, serta meningkatkan risiko kesalahan klasifikasi, khususnya pada wajah pria dengan karakteristik feminin atau sebaliknya. [5]

Untuk menjawab permasalahan tersebut, penelitian ini mengusulkan pendekatan terpadu dalam proses klasifikasi gender berbasis wajah. [6], [7] Pendekatan ini terdiri dari tiga tahap utama. Pertama, tahap *preprocessing* citra dilakukan dengan menggunakan OpenCV untuk mendeteksi dan memotong bagian wajah dari gambar, sehingga mengurangi pengaruh latar belakang dan fokus pada fitur wajah. Kedua, dilakukan *face clustering* dengan menggabungkan algoritma K-Means [8] dan DBSCAN [9] untuk menyusun data ke dalam kelompok visual yang lebih bersih dan relevan. Ketiga, proses klasifikasi gender dilakukan dengan membangun model *sequential learning* berbasis arsitektur Inception_v3, yang diperkuat dengan teknik Agglomerative Clustering untuk mengelola fitur-fitur wajah yang terlalu mirip. [10]

Agglomerative Clustering ini berfungsi untuk mengelompokkan representasi fitur wajah hasil ekstraksi Inception_v3 ke dalam hirarki kemiripan, sehingga fitur-fitur yang terlalu mirip dapat diorganisasi dan direduksi secara semantik. Tujuan dari teknik ini adalah untuk mengurangi redundansi fitur, memperjelas batas antar kelas, dan meningkatkan robustitas terhadap variasi intra-class, seperti wajah pria yang memiliki karakteristik feminin atau sebaliknya. Dengan pendekatan ini, sistem diharapkan dapat menghasilkan representasi fitur yang lebih terstruktur dan diskriminatif sebelum tahap akhir klasifikasi dilakukan.

Tujuan dari penelitian ini adalah untuk meningkatkan akurasi klasifikasi gender berbasis citra wajah, sekaligus mengurangi pengaruh noise dan duplikasi citra yang umum ditemukan dalam dataset tidak terstruktur. Selain itu,



penelitian ini juga bertujuan untuk menguji efektivitas integrasi beberapa teknik clustering dalam meningkatkan keakuratan hasil klasifikasi dan memperbaiki proses ekstraksi fitur visual dari citra wajah.[11], [12]

Penelitian ini mengisi gap yang ada dalam studi-studi sebelumnya. Sebagian besar riset terdahulu lebih menekankan pada peningkatan akurasi model klasifikasi dengan pendekatan CNN atau *deep learning*, namun belum mengadopsi tahapan *data cleansing* dan *clustering* secara menyeluruh, terutama ketika bekerja dengan data sosial media yang bersifat bebas dan penuh noise. Oleh karena itu, integrasi preprocessing, clustering, dan classification dalam satu pipeline yang terstruktur menjadi nilai tambah dalam penelitian ini.[13], [14]

Beberapa penelitian terdahulu dapat dijadikan rujukan perbandingan. Sachi et al. [15] mengembangkan model CNN untuk klasifikasi gender namun terhadap data training yang menggunakan citra abu-abu. Salsabila Et al [16] menggunakan citra wajah sebagai dasar klasifikasi, tetapi tidak mengintegrasikan metode LBP dan KNN. Sementara itu, Rosyani et al. [14] membuktikan efektivitas algoritma K-Means dalam pengelompokan fitur visual, namun penerapannya masih terbatas pada citra non-wajah seperti bunga.

Meskipun berbagai penelitian telah dilakukan dalam bidang klasifikasi gender berbasis wajah, masih terdapat sejumlah celah (*research gap*) yang belum banyak disentuh secara komprehensif. Sebagian besar studi terdahulu fokus pada penggunaan arsitektur deep learning, khususnya CNN, untuk meningkatkan akurasi klasifikasi. Namun, pendekatan-pendekatan tersebut sering kali mengabaikan tahapan penting dalam pra-pemrosesan data, seperti pembersihan dataset dari noise, duplikasi, atau ketidakteraturan distribusi pose dan ekspresi wajah, yang umum ditemukan dalam data tidak terstruktur dari sumber seperti media sosial.

Selain itu, teknik clustering seperti K-Means atau DBSCAN memang telah diterapkan dalam berbagai studi, namun umumnya hanya terbatas untuk analisis visual atau segmentasi awal, bukan sebagai bagian dari pipeline yang terintegrasi langsung dengan proses klasifikasi end-to-end. Belum banyak penelitian yang secara eksplisit menggabungkan metode preprocessing, clustering, dan deep learning dalam satu alur sistematis dan saling terhubung, seperti yang diusulkan dalam studi ini.

Dengan demikian, penelitian ini hadir untuk mengisi kekosongan tersebut dengan menawarkan pendekatan terpadu yang tidak hanya membangun model klasifikasi berbasis Inception_v3, tetapi juga mengintegrasikan tahapan clustering berbasis K-Means, DBSCAN, dan Agglomerative Clustering untuk menghasilkan dataset yang lebih bersih, representatif, dan informatif, sehingga mendukung performa klasifikasi yang lebih robust terhadap variasi intra-class dan noise data.

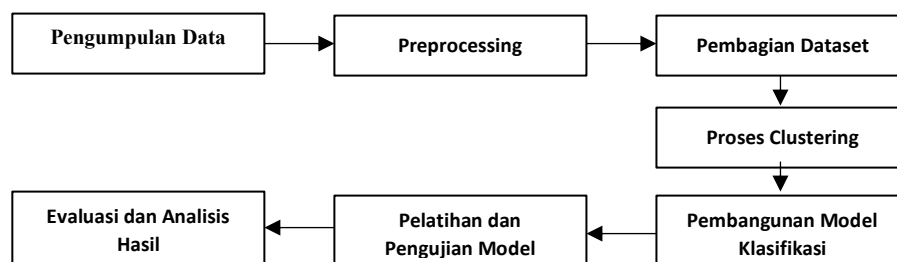
Dengan pendekatan menyeluruh yang ditawarkan, penelitian ini diharapkan dapat memberikan kontribusi praktis dalam membangun sistem klasifikasi gender yang lebih robust dan adaptif terhadap keragaman data visual di dunia nyata.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan memanfaatkan algoritma machine learning dan pengolahan citra digital. Fokus utama penelitian adalah membangun sistem klasifikasi gender berbasis wajah dengan tiga komponen utama: *preprocessing citra*, *clustering wajah*, dan *model klasifikasi*. Dataset yang digunakan berasal dari kompetisi Big Data, di mana gambar wajah diambil dari platform media sosial Instagram dan tidak melalui proses kurasi manual, sehingga menuntut proses pemurnian data yang cermat sebelum proses pelatihan model.

2.1 Tahapan Penelitian

Tahapan penelitian dilakukan secara bertingkat dan dapat dilihat pada Gambar 1. Setiap tahap berperan penting dalam mempersiapkan dan memproses data agar sesuai dengan kebutuhan klasifikasi. Tahapan ini mencerminkan alur sistematis dari input data mentah hingga diperoleh hasil klasifikasi akhir.



Gambar 1. Tahapan Penelitian

Dari gambar 1 akan dijelaskan tahapan yang akan dilalu untuk menyelesaikan penelitian ini:

2.2 Pengumpulan Data

Mengambil dataset dari sumber Big Data Competition dan Mengidentifikasi struktur folder dan jumlah data.

2.3 Preprocessing Data



Pemeriksaan kemiripan gambar dalam folder, Deteksi wajah dan cropping dengan OpenCV, Penyimpanan hasil preprocessing ke dalam basis data. Pada tahap ini dilakukan beberapa proses pembersihan data:

- Deteksi wajah dengan OpenCV menggunakan CascadeClassifier
- Cropping bagian wajah untuk menghilangkan latar belakang
- Identifikasi gambar duplikat dengan teknik perbandingan histogram
- Normalisasi ukuran citra agar seragam

2.4 Pembagian Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari gambar wajah pria dan wanita yang disusun dalam folder sesuai label kategorinya. Sebelum proses pelatihan dimulai, dilakukan tahapan **pembagian data** dan **penyeimbangan data** untuk memastikan bahwa model tidak mengalami bias terhadap salah satu kelas.

a. Pembagian Data

Dataset awal dibagi menjadi dua subset, yaitu:

- Data latih (training set): 942 gambar pria dan 1.354 gambar wanita.
- Data uji (testing set): 990 gambar yang terdiri dari gambar pria dan wanita secara acak.

Pembagian ini dilakukan dengan proporsi sekitar 70% untuk pelatihan dan 30% untuk pengujian, mempertahankan distribusi asli dataset.

b. Penyeimbangan Data

Karena jumlah gambar wanita lebih banyak dibandingkan pria, terjadi ketidakseimbangan kelas pada data latih. Untuk mengatasi hal ini, diterapkan teknik oversampling sederhana terhadap kelas minoritas (pria), yaitu dengan menggandakan (duplicating) sebagian data gambar pria secara acak hingga jumlahnya seimbang dengan kelas wanita. Setelah proses penyeimbangan:

- Jumlah data latih menjadi: 1.354 gambar pria dan 1.354 gambar wanita.
- Data uji tidak diubah dan tetap merepresentasikan distribusi alami, agar evaluasi model tetap mencerminkan kondisi nyata.

Penyeimbangan hanya dilakukan pada data latih, karena tujuan utama adalah meningkatkan kemampuan generalisasi model tanpa bias kelas selama proses pelatihan. Sementara itu, data uji tetap dibiarkan tidak seimbang untuk menjaga validitas pengujian dan dilakukan evaluasi dengan metrik tambahan seperti precision, recall, dan F1-score, guna menangani efek dari distribusi kelas yang tidak seimbang.

2.5 Proses Clustering

K-Means digunakan untuk segmentasi awal data wajah dan DBSCAN [17] digunakan untuk menghilangkan noise dan mengelompokkan wajah dengan bentuk cluster yang lebih bebas

2.6 Pembangunan Model

Tahapan clustering dalam penelitian ini dilakukan sebagai bagian dari proses pembersihan dan pengorganisasian data wajah sebelum tahap pelatihan model klasifikasi gender. Clustering diterapkan bukan pada piksel gambar secara langsung, melainkan pada representasi fitur vektor yang diekstraksi dari citra wajah menggunakan teknik feature descriptor dasar (*HOG* dan *LBP*) sebelum dimasukkan ke dalam model deep learning.

a. K-Means Clustering: Segmentasi Awal

Pada tahap pertama, dilakukan segmentasi awal data menggunakan algoritma K-Means clustering, yang bertujuan untuk mengelompokkan gambar-gambar wajah ke dalam cluster berdasarkan kemiripan fitur dasar (embedding awal) seperti bentuk wajah umum, posisi wajah (pose), dan ekspresi dominan. Ini membantu mengidentifikasi struktur kasar dari distribusi data dan membagi dataset besar menjadi kelompok visual yang lebih terstruktur.

b. DBSCAN: Identifikasi Noise dan Cluster Fleksibel

Setelah segmentasi awal, diterapkan algoritma DBSCAN (Density-Based Spatial Clustering of Applications with Noise) untuk memfilter noise dalam data dan mendeteksi outlier berupa gambar yang terlalu buram, duplikat, atau memiliki kemiringan ekstrim. DBSCAN bekerja dengan mengelompokkan data berdasarkan kedekatan densitas fitur, sehingga lebih fleksibel dalam mengenali kelompok data yang tidak berbentuk bulat sempurna seperti pada K-Means.

c. Hasil dan Tujuan Clustering

Hasil dari proses clustering digunakan untuk:

Menghasilkan subset data latih yang lebih bersih dan relevan, dengan cara mengeliminasi gambar wajah yang terdeteksi sebagai noise atau outlier oleh DBSCAN.

Meningkatkan kualitas dataset latih sebelum masuk ke tahap pelatihan model Inception_v3, sehingga model belajar dari data yang lebih representatif dan tidak bias oleh noise atau data ekstrim.

Tidak digunakan sebagai fitur langsung dalam klasifikasi, tetapi sebagai tahap data cleansing berbasis clustering.

d. Integrasi ke dalam Pipeline Klasifikasi

Dengan integrasi proses clustering ini, pipeline klasifikasi menjadi lebih robust karena:

Mengurangi *data redundancy* dan *noisy samples* pada dataset tidak terstruktur.

Meningkatkan kualitas generalisasi model pada data uji.



Memberikan pondasi bagi pembelajaran fitur yang lebih stabil di arsitektur CNN (Inception_v3).

2.7 Pelatihan dan Pengujian Model

Model dilatih menggunakan data latih dan Model diuji dengan data uji dan dievaluasi metrik performanya

2.8 Evaluasi dan Analisis Hasil

Mengukur performa model berdasarkan nilai akurasi, F1-Score, precision, recall, dan loss, Visualisasi grafik akurasi dan loss

3. HASIL DAN PEMBAHASAN

Pada bagian ini menyajikan hasil yang diperoleh dari serangkaian tahapan penelitian yang telah dilakukan, mulai dari preprocessing citra wajah, penerapan metode clustering,[18] hingga pembangunan dan evaluasi model klasifikasi gender. Setiap hasil dianalisis secara kualitatif dan kuantitatif untuk mengevaluasi efektivitas pendekatan yang diusulkan dalam meningkatkan akurasi dan kestabilan sistem klasifikasi berbasis citra. Pembahasan dilakukan untuk menafsirkan temuan penelitian dalam konteks studi terdahulu serta mengidentifikasi kontribusi dan implikasi dari pendekatan yang diterapkan.[19]

3.1 Hasil Preprocessing Data

Tahap awal dalam penelitian ini adalah preprocessing citra, yang bertujuan untuk memurnikan dataset dari noise visual, latar belakang yang tidak relevan, serta gambar-gambar ganda. Proses deteksi wajah dilakukan dengan menggunakan OpenCV CascadeClassifier, dan dilanjutkan dengan cropping bagian wajah. Citra wajah yang telah diproses kemudian diklasifikasikan ke dalam dua kategori utama: pria dan wanita. Hasil preprocessing menghasilkan 942 citra pria dan 1.354 citra wanita untuk data latih, serta 990 citra campuran untuk data uji.

Tahap ini berperan penting dalam memastikan bahwa model hanya menerima fitur-fitur visual yang relevan, sehingga dapat mengurangi potensi kesalahan klasifikasi akibat variasi latar belakang dan pose. Gambar 1 memperlihatkan contoh gambar wajah mentah sebelum dilakukan proses cropping



Gambar 1. Contoh Gambar Wajah Sebelum Cropping



Gambar 2. Hasil Deteksi dan Cropping Wajah

Gambar 1 merupakan sample dari deteksi wajah yang kemudian dicropping dibagian wajah saja. Sedangkan gambar 2 merupakan hasil facedetection dan cropping wajah. Gambar – gambar tersebut disimpan didalam database untuk diproses kedalam clustering gambar. Proses ini menghasilkan 942 citra pria dan 1.354 citra wanita untuk data latih, serta 990 citra campuran untuk data uji. Hasil preprocessing ini sangat krusial dalam menyiapkan data yang bersih dan relevan, serta mengurangi gangguan dari variasi latar belakang dan pose wajah.

3.2 Proses Clustering

Setelah preprocessing, tahap selanjutnya adalah pengelompokan citra berdasarkan kemiripan visual menggunakan teknik clustering. Proses clustering untuk mengelompokkan wajah berdasarkan kemiripan visual.

- K-Means Clustering dengan nilai $k=2$ berhasil membentuk dua klaster besar yang mencerminkan distribusi wajah pria dan wanita. Metode ini dipilih karena efisiensinya dalam memproses data besar secara cepat.
- DBSCAN digunakan untuk menyaring outlier dan noise. Metode ini lebih fleksibel dalam membentuk klaster dengan bentuk tidak beraturan dan sangat berguna dalam menghadapi ekspresi wajah yang tidak konsisten.

Hasil dari proses ini ditampilkan pada Gambar 3, di mana wajah yang sama dengan ekspresi berbeda berhasil dikelompokkan dalam klaster yang sama, namun menunjukkan potensi kesalahan klasifikasi akibat fitur visual yang ambigu.



Gambar 3. hasil klustering K-means dan DBScan

Namun demikian, pada beberapa kasus seperti ditunjukkan pada Gambar 4, wajah pria dengan karakteristik feminin masih cenderung salah diklasifikasikan sebagai wanita. Untuk mengatasi hal ini, diterapkan Agglomerative Clustering guna memisahkan representasi fitur yang terlalu mirip.

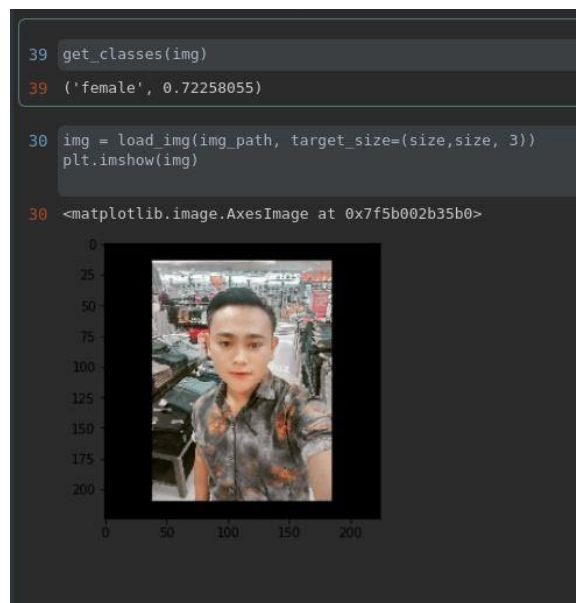
```

39 get_classes(img)
39 ('female', 0.72258055)

30 img = load_img(img_path, target_size=(size,size, 3))
    plt.imshow(img)

30 <matplotlib.image.AxesImage at 0x7f5b002b35b0>
0
25
50
75
100
125
150
175
200
0 50 100 150 200

```

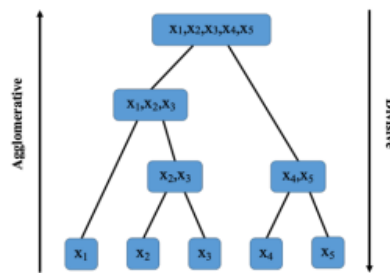


Gambar 4. Wajah Pria dengan karakteristik feminin

Gambar 4 merupakan gambar yang menjadi permasalahan dalam mengkluster wajah berdasarkan kelasnya. Untuk mengurangi masalah kami pada saat proses klasifikasi gender maka kami menggunakan Agglomerative Clustering. Agglomerative Clustering dapat mengurangi fitur-fitur yang berlebihan pada saat pengelompokan data. *Klasifikasi Gambar*

Model klasifikasi gender dibangun menggunakan pendekatan *sequential learning* dengan arsitektur Inception_v3 sebagai feature extractor. Lapisan tambahan berupa dense layer dan batch normalization digunakan untuk memperdalam proses pembelajaran. Fungsi aktivasi sigmoid dipilih untuk menghasilkan prediksi biner (pria atau wanita).

Agar klasifikasi lebih akurat dalam kasus wajah dengan kemiripan tinggi (misalnya pria dengan wajah feminin), digunakan Agglomerative Clustering untuk memisahkan fitur yang terlalu serupa. Dengan pendekatan ini, model mampu membedakan lebih jelas struktur fitur wajah yang saling tumpang tindih. Proses pengelompokan Agglomerative dijelaskan sebagai berikut. Pertama, algoritma memperlakukan setiap fitur sebagai sebuah cluster dan menginisialisasi jarak antara masing-masing dua cluster. Kemudian gabungkan dua cluster terdekat menjadi cluster baru dan hitung jarak antara cluster baru dan yang lainnya cluster. Proses penggabungan diulangi sampai standar kondisi tercapai atau semua fitur milik grup [20], [21]. Agglomerative Clustering bekerja dengan menyatukan fitur-fitur mirip ke dalam struktur hierarkis, sehingga fitur wajah yang tumpang tindih dapat dipisahkan dengan lebih baik. Proses ini digambarkan secara konseptual dalam Gambar 5.



. Gambar 5. Hierarchiel AgglomerativeClustering

Metode ini lebih unggul dalam memaksimalkan koefisien informasi antara dua fitur sehingga didapatkan kesamaan dari fitur-fitur gambar ini

3.3 Evaluasi Performa Model

Evaluasi performa model klasifikasi gender berbasis wajah dilakukan dengan menggunakan lima metrik utama, yaitu loss value, accuracy, precision, recall, dan F1-Score. Penggunaan metrik yang beragam ini bertujuan untuk memberikan gambaran yang lebih menyeluruh terhadap kualitas model, terutama ketika bekerja dengan data yang sempit mengalami ketidakseimbangan kelas. Tabel 1 menampilkan hasil evaluasi performa model:

Tabel 1. Metrik Evaluasi

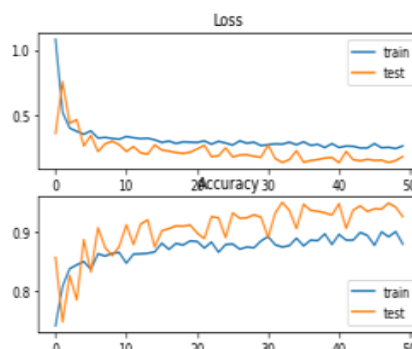
Metrik Evaluasi	Nilai
Loss Value	0.14178
Accuracy	92.97%
F1-Score	0.89556
Precision	0.97727
Recall	0.83069

Penjelasan Metrik pada tabel 1:

- Loss Value (0.14178): Menunjukkan seberapa jauh prediksi model dari nilai target sebenarnya. Nilai yang rendah mengindikasikan bahwa model telah belajar dengan baik selama proses pelatihan dan tidak mengalami overfitting atau underfitting yang signifikan.
- Accuracy (92.97%): Menggambarkan persentase total prediksi yang benar dibandingkan seluruh prediksi. Akurasi tinggi menunjukkan bahwa sebagian besar prediksi model telah sesuai dengan label sebenarnya.
- Precision (0.97727): Mengukur tingkat ketepatan prediksi positif, yaitu dari seluruh prediksi pria, sebanyak 97.7% benar-benar merupakan wajah pria. Nilai ini sangat tinggi, menunjukkan bahwa model memiliki tingkat false positive yang sangat rendah.
- Recall (0.83069): Menunjukkan kemampuan model dalam menangkap semua data positif yang sebenarnya. Artinya, dari seluruh gambar pria sebenarnya, hanya 83.0% yang berhasil dideteksi dengan benar. Nilai ini sedikit lebih rendah dibanding precision, yang mengindikasikan adanya false negatives yang lebih tinggi.
- F1-Score (0.89556): Merupakan rata-rata harmonik dari precision dan recall, yang menyeimbangkan keduanya. Nilai mendekati 0.90 ini menunjukkan bahwa model memiliki performa yang baik secara keseluruhan, meskipun ada sedikit ketidakseimbangan antara precision dan recall.

3.4 Visualisasi Performa Model

Untuk melihat tren performa model selama pelatihan, Gambar 6 berikut menunjukkan grafik akurasi dan loss terhadap jumlah epoch.:



Gambar 6. Hasil Akurasi



Grafik pada gambar 6 di atas menunjukkan tren peningkatan akurasi yang konsisten pada setiap epoch, serta penurunan nilai loss yang signifikan. Hal ini menandakan bahwa model belajar dengan stabil dan tidak mengalami overfitting. Akurasi akhir yang diperoleh sebesar 92.97% menunjukkan bahwa sistem mampu mengklasifikasikan gender dengan tingkat keandalan yang tinggi.

3.5 Pembahasan

Hasil penelitian menunjukkan bahwa integrasi preprocessing, face clustering, dan deep learning mampu meningkatkan akurasi sistem klasifikasi gender berbasis citra wajah. Teknik DBSCAN dan Agglomerative Clustering terbukti memberikan kontribusi signifikan dalam menyaring noise serta membentuk struktur fitur yang lebih representatif. Jika dibandingkan dengan penelitian sebelumnya, pendekatan ini lebih adaptif terhadap data tidak terstruktur dari media sosial dan lebih robust dalam menghadapi variasi ekspresi dan kualitas gambar. Dengan demikian, sistem ini memiliki potensi tinggi untuk diimplementasikan dalam aplikasi nyata seperti sistem keamanan, personalisasi layanan, hingga analisis demografi otomatis.

4. KESIMPULAN

Penelitian ini berhasil membangun sistem klasifikasi gender berbasis citra wajah dengan pendekatan terpadu yang menggabungkan tahapan preprocessing, face clustering, dan deep learning. Proses deteksi wajah dan pemotongan (cropping) menggunakan OpenCV terbukti efektif dalam meningkatkan fokus model pada fitur-fitur wajah yang relevan, sekaligus mengurangi gangguan dari latar belakang. Integrasi algoritma K-Means dan DBSCAN dalam proses clustering memberikan kontribusi signifikan dalam pembersihan dan pengelompokan data wajah berdasarkan kemiripan visual, sehingga menghasilkan dataset latih yang lebih bersih dan representatif. Selain itu, penggunaan Agglomerative Clustering pada tahap pasca-ekstraksi fitur berperan dalam mengurangi redundansi fitur dan memperbaiki representasi visual, khususnya untuk wajah yang memiliki kemiripan bentuk antar gender. Model klasifikasi dikembangkan menggunakan pendekatan sequential learning berbasis arsitektur Inception_v3, dan menunjukkan performa yang sangat baik. Hasil evaluasi menunjukkan bahwa sistem ini mampu melakukan klasifikasi gender secara akurat dan andal, dengan akurasi sebesar 92.97%, F1-score sebesar 0.89556, precision mencapai 0.97727, dan recall sebesar 0.83069. Nilai-nilai ini menunjukkan bahwa model memiliki tingkat ketepatan tinggi dalam memprediksi gender, meskipun masih terdapat beberapa kasus wajah pria yang tidak terdeteksi secara sempurna (false negative). Secara keseluruhan, pendekatan terpadu yang dikembangkan dalam penelitian ini meningkatkan robustness sistem terhadap noise dan keragaman visual, dan memberikan kontribusi nyata dalam pengembangan sistem klasifikasi gender berbasis citra yang aplikatif, terutama untuk data tidak terstruktur seperti citra dari media sosial atau kamera bebas di dunia nyata.

REFERENCES

- [1] S. Tilki, H. B. Dogru, A. A. Hameed, A. Jamil, and J. Rasheed, "Gender Classification using Deep Learning Techniques," no. May, 2021.
- [2] M. U. Khan, M. Saad, S. Aziz, and J. M. Ch, "Electrocardiogram based Gender Classification," no. January 2021, 2020, doi: 10.1109/ICECCE49384.2020.9179305.
- [3] H. Fauzi, C. Erika, S. Sa'adiyah, and F. Oscandar, "Classification of Gender Individual Identification Using Local Binary Pattern on Palatine Rugae Image," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 8, no. 3, p. 422, 2022, doi: 10.26555/jiteki.v8i3.23636.
- [4] D. T. Adherda, M. Hikmatyar, and Ruuhwan, "Gender Classification Based on Voice Using Recurrent Neural Network (Rnn)," *Antivirus : Jurnal Ilmiah Teknik Informatika*, vol. 17, no. 1, pp. 111–122, 2023, doi: 10.35457/antivirus.v17i1.3049.
- [5] D. I. Mulyana and V. V. Pramansah, "Gender Classification for Anime Character Face Image Using Random Forest Classifier Method and GLCM Feature Extraction," *JUITA: Jurnal Informatika*, vol. 10, no. 2, p. 243, 2022, doi: 10.30595/juita.v10i2.13833.
- [6] H. Salsabila, E. Rachmawati, and F. Sthevanie, "Klasifikasi Gender Berdasarkan Citra Wajah Menggunakan Metode Local Binary Pattern dan K-Nearest Neighbor," vol. 8, no. 2, pp. 3137–3146, 2021.
- [7] R. Tamisier *et al.*, "FACE Cluster phenotyping predicting outcomes in a prospective multicenter cohort study of chronic heart failure patients with central sleep disorder breathing indicated for adaptive servo ventilation," *Archives of Cardiovascular Diseases Supplements*, vol. 12, no. 2–4, p. 246, Oct. 2020, doi: 10.1016/j.acvdsp.2020.03.112.
- [8] P. Rosyani, A. Suhendi, D. H. Apriyanti, and A. A. Waskita, "Color Features Based Flower Image Segmentation Using K-Means and Fuzzy C-Means," *Building of Informatics, Technology and Science (BITS)*, vol. 3, no. 3, pp. 253–259, 2021, doi: 10.47065/bits.v3i3.1060.
- [9] R. Adha, N. Nurhaliza, U. Sholeha, and M. Mustakim, "Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia," *SITEKIN: Jurnal Sains, Teknologi dan Industri*, vol. 18, no. 2, pp. 206–211, 2021.
- [10] A. Shahcheraghian, A. Ilinca, and N. Sommerfeldt, "K-means and agglomerative clustering for source-load mapping in distributed district heating planning," *Energy Conversion and Management: X*, vol. 25, no. October 2024, p. 100860, 2025, doi: 10.1016/j.ecmx.2024.100860.
- [11] B. O. Siahaan, "Implementasi Pengenalan Wajah Untuk Absensi Karyawan Dengan Metode Eigenface," *Comasie*, vol. 5, pp. 19–28, 2021.



- [12] A. Valentino, A. Ranga Sn, F. Wijoyo, I. J. Lestari, T. P. Andari, and P. Rosyani, "Studi literatur review: Alat Identifikasi Wajah untuk Absensi Mahasiswa Dengan Algoritma YOLO pada Flatform Android," *Logic : Jurnal Ilmu Komputer dan Pendidikan*, vol. 1, no. 2, pp. 233–238, 2023, [Online]. Available: <https://journal.mediapublikasi.id/index.php/logic>
- [13] P. Rosyani, S. Saprudin, and R. Amalia, "Klasifikasi Citra Menggunakan Metode Random Forest dan Sequential Minimal Optimization (SMO)," *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 9, no. 2, p. 132, 2021, doi: 10.26418/justin.v9i2.44120.
- [14] I. H. Iksari, R. Amalia, and P. Rosyani, "Segmentasi Citra Bunga Menggunakan Blob Analisis," *Building of Informatics, Technology and Science (BITS)*, vol. 3, no. 3, pp. 228–234, 2021, doi: 10.47065/bits.v3i3.1050.
- [15] L. Wijayanti, H. Sachi, and S. Octaviani, "Klasifikasi Gender Berdasarkan Gambar Menggunakan Metode Deep Learning Pada MATLAB," 2023.
- [16] H. Salsabila, E. Rachmawati, and F. Stevanie, "Klasifikasi Gender Berdasarkan Citra Wajah Menggunakan Metode Local Binary Pattern dan K-Nearest Neighbor," 2021.
- [17] D. Deng, "DBScan Clustering Algorithm Based on Density," pp. 949–953, 2020, doi: 10.1109/IFEEA51475.2020.00199.
- [18] Perani Rosyani and Fabian Syawali, "Application of Advanced Class Determination System Using K-Means Clustering Method (Case Study: SMK Al-Badar Balaraja)," *International Journal of Integrative Sciences*, vol. 2, no. 10, pp. 1557–1570, 2023, doi: 10.55927/ijis.v2i10.6347.
- [19] D. Lakshmi and R. Ponnusamy, "Facial emotion recognition using modified HOG and LBP features with deep stacked autoencoders," *Microprocess Microsyst*, vol. 82, no. January, p. 103834, 2021, doi: 10.1016/j.micpro.2021.103834.
- [20] C. H. Park, "A feature selection method using hierarchical clustering," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8284 LNAI, pp. 1–6, 2013, doi: 10.1007/978-3-319-03844-5_1.
- [21] C. Ding and X. He, "Cluster merging and splitting in hierarchical clustering algorithms," *Proceedings - IEEE International Conference on Data Mining, ICDM*, pp. 139–146, 2002, doi: 10.1109/icdm.2002.1183896.