

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Penelitian

Perkembangan teknologi kecerdasan buatan telah mendorong kemajuan pesat dalam bidang pengolahan citra digital, khususnya pada ranah computer vision. Penerapan deep learning memungkinkan sistem komputasi untuk mempelajari dan mengenali pola visual secara otomatis dari data berukuran besar, sehingga meningkatkan akurasi serta keandalan sistem pengenalan citra [1][2]. Salah satu pendekatan yang paling dominan dalam bidang ini adalah Convolutional Neural Network (CNN), yang memiliki kemampuan mengekstraksi fitur visual secara hierarkis dan mendalam [3][4]. Karakteristik tersebut menjadikan CNN sebagai fondasi utama dalam berbagai aplikasi, termasuk pengenalan wajah dan analisis atribut demografis berbasis citra [1].

Kemampuan untuk mengenali etnisitas secara otomatis dari citra wajah memiliki relevansi yang luas dalam berbagai domain aplikasi dunia nyata. Wajah manusia menyimpan atribut demografis kunci seperti kelamin, usia, dan etnisitas yang menjadi komponen integral dalam aplikasi keamanan, biometrik, pengawasan, interaksi manusia-komputer, pemasaran daring, dan layanan kesehatan [5]. Dalam ranah penegakan hukum, teknologi pengenalan wajah beserta komponen klasifikasi demografisnya telah digunakan oleh berbagai lembaga penegak hukum di tingkat federal, negara bagian, dan lokal sebagai salah satu komponen dalam investigasi kriminal [6]. Selain itu, klasifikasi etnisitas telah diterapkan secara luas dalam aplikasi video pengawasan, iklan daring, interaksi manusia-komputer, serta penelitian demografis dan biometrik [7]. Hal ini menunjukkan bahwa sistem klasifikasi etnisitas yang akurat dan adil merupakan kebutuhan teknologi yang mendesak, sekaligus membawa tanggung jawab etis yang besar dalam pengembangannya.

Pemilihan kategori etnisitas dalam penelitian ini didasarkan pada representasi demografis populasi dunia yang signifikan. Berdasarkan estimasi *United Nations World Population Prospects* 2024, kelompok Asia Selatan merepresentasikan kelas India merupakan kelompok terbesar dengan populasi mencapai 2,1 miliar jiwa atau sekitar 26% dari total populasi global. Kelompok Asia Timur mencakup sekitar 1,66 miliar jiwa ( $\pm 20\%$ ), sementara populasi Sub-Sahara Afrika yang merepresentasikan kelompok Negroid berjumlah sekitar 1,2 miliar jiwa ( $\pm 14,8\%$ ). Kawasan Eropa, Amerika Utara, dan Oseania yang merepresentasikan kelompok Caucasian berjumlah sekitar 1,15 miliar jiwa ( $\pm 14\%$ ), Asia Tenggara sekitar 700 ( $\pm 8,6\%$ ) juta jiwa, dan kawasan Timur Tengah sekitar 515 juta jiwa ( $\pm 6,3\%$ ) [8]. Secara kumulatif, keenam kelompok etnis ini merepresentasikan hampir 90% dari total populasi dunia sebesar 8,1 miliar jiwa, sehingga klasifikasi yang mencakup keenam kelas ini dinilai mampu memberikan cakupan yang representatif terhadap variasi karakteristik wajah secara global [8][9].

Dalam konteks analisis citra wajah, klasifikasi etnisitas merupakan tugas yang memiliki tingkat kompleksitas tinggi karena perbedaan visual antar kelompok etnis sering kali bersifat halus dan saling tumpang tindih. Sistem pengenalan wajah

dalam aplikasi dunia nyata harus menghadapi tidak hanya masalah penyelarasan dasar, tetapi juga variasi intra-kelas yang signifikan seperti perubahan ekspresi wajah, penuaan, atau kondisi pencahayaan serta kemiripan inter-kelas, yaitu individu dengan penampilan yang serupa, yang semuanya dapat berdampak besar pada performa pengenalan [10]. Dalam konteks klasifikasi etnisitas secara khusus, kesalahan klasifikasi banyak terjadi pada citra dengan pencahayaan yang tidak merata, bayangan yang dalam, oklusi parsial akibat rambut atau aksesoris, serta rotasi wajah yang menghalangi model untuk mengekstrak relasi spasial yang stabil [11]. Tantangan ini semakin kompleks mengingat adanya kemiripan morfologis yang tinggi antar kelompok etnis yang berdekatan secara geografis seperti *East Asian* dan *Southeast Asian* sehingga batas keputusan antar kelas bersifat halus dan sulit dipisahkan secara komputasional. Selain itu, terdapat pula variasi yang besar di dalam setiap kelas citra wajah dengan asal etnis yang sama, sehingga metode yang digunakan harus mampu mengabaikan variasi yang tidak relevan seperti pencahayaan, pose, gaya rambut, usia, dan jenis kelamin [12]. Keberhasilan sistem dalam mengenali karakteristik tersebut sangat dipengaruhi oleh arsitektur CNN yang digunakan [1][12][13].

Arsitektur populer seperti VGG19 dan ResNet101 telah menunjukkan performa unggul dalam berbagai tugas klasifikasi visual berskala besar. Pemilihan kedua arsitektur ini didasarkan pada pertimbangan komplementaritas karakteristik teknisnya. VGG19, dengan arsitektur yang lebih dalam dan kernel konvolusi berukuran 3x3 yang kecil, dikenal memiliki kemampuan yang kuat dalam menangkap detail halus pada citra yang kompleks [14]. Kemampuan ekstraksi fitur hirarki VGG19 terbukti kompetitif dalam berbagai tugas klasifikasi citra. Di sisi lain, ResNet101 menawarkan keunggulan arsitektural yang berbeda, mengatasi masalah *vanishing gradient* dengan memperkenalkan *skip connections* yang memungkinkan informasi mengalir langsung dari satu lapisan ke lapisan lainnya, mengatasi masalah degradasi yang biasa terjadi saat jaringan semakin dalam. Dengan demikian, VGG19 mewakili paradigma jaringan dalam dengan ekstraksi fitur bertahap yang sederhana namun efektif, sementara ResNet101 mewakili paradigma *residual learning* yang memungkinkan informasi mengalir langsung dari satu lapisan ke lapisan lainnya, mengatasi masalah degradasi yang biasanya terjadi saat jaringan dalam dengan ekstraksi fitur bertahap yang sederhana namun efektif, sementara ResNet101 mewakili paradigma *residual learning* yang memungkinkan kedalaman jaringan yang lebih besar dengan efisiensi komputasi yang lebih baik. Meskipun demikian, performa kedua arsitektur tersebut belum tentu optimal ketika diterapkan pada tugas spesifik seperti klasifikasi etnisitas [1][12][15].

Salah satu tantangan utama dalam klasifikasi etnisitas adalah kemampuan model dalam membedakan karakteristik wajah yang memiliki tingkat kemiripan tinggi antar kelompok etnis. Setiap arsitektur CNN memiliki strategi ekstraksi fitur yang berbeda, sehingga dapat menghasilkan respons yang bervariasi terhadap fitur-fitur wajah tertentu. Perbedaan ini berpotensi memengaruhi tingkat akurasi serta konsistensi model dalam mengklasifikasikan etnisitas, sehingga diperlukan analisis yang lebih mendalam untuk memahami perilaku masing-masing arsitektur pada domain citra wajah.

Meskipun berbagai dataset pengenalan wajah telah tersedia untuk penelitian analisis demografis, pengujian komparatif yang secara khusus mengevaluasi performa arsitektur CNN berukuran besar seperti VGG19 dan ResNet101 pada tugas klasifikasi etnisitas dengan penggunaan dataset terpisah untuk setiap kategori etnis masih sangat terbatas [12][16]. Sebagian besar penelitian sebelumnya lebih berfokus pada penggunaan satu arsitektur tertentu atau membandingkan model CNN yang lebih ringan dengan dataset yang telah teranotasi secara terpusat, sehingga belum memberikan bukti empiris yang memadai mengenai bagaimana karakteristik arsitektur yang lebih dalam dan kompleks memengaruhi kemampuan model dalam menangani variasi fitur wajah lintas kelompok etnis ketika data dikumpulkan dan dikurasi secara independen per kategori. Dalam penelitian ini, klasifikasi etnisitas dilakukan dengan menggunakan enam kategori etnis utama yang merepresentasikan kelompok populasi global, yaitu *East Asian*, *Southeast Asian*, *India*, *Caucasian*, *Negroid*, dan *Middle Eastern*, sehingga diharapkan mampu memberikan gambaran yang lebih representatif terhadap variasi karakteristik wajah secara global [12]. Perbedaan tersebut dinilai cukup signifikan dalam konteks computer vision sehingga kedua kelompok tersebut diperlakukan sebagai kategori yang terpisah untuk meningkatkan kemampuan model dalam mempelajari pola visual yang lebih spesifik.

Berdasarkan kondisi tersebut, penelitian ini dirancang untuk melakukan pengujian dan evaluasi komparatif secara sistematis terhadap kinerja arsitektur VGG19 dan ResNet101 dalam klasifikasi etnisitas berbasis citra wajah dengan menggunakan dataset yang merepresentasikan kategori etnisitas yang dianalisis. Melalui skema eksperimen yang terkontrol dengan preprocessing, pelatihan, dan evaluasi yang seragam, penelitian ini bertujuan untuk mengukur dan membandingkan performa kedua arsitektur secara objektif, sehingga dapat mengidentifikasi kekuatan dan kelemahan masing-masing dalam konteks klasifikasi etnisitas. Pendekatan ini memungkinkan evaluasi yang lebih representatif terhadap kemampuan model dalam mengenali pola visual yang khas dari setiap kelompok etnis. Hasil penelitian ini diharapkan dapat memberikan kontribusi ilmiah berupa temuan empiris yang valid mengenai pengaruh arsitektur CNN terhadap performa klasifikasi etnisitas, serta menjadi dasar rekomendasi praktis dalam pengembangan sistem pengenalan wajah yang lebih akurat, andal, dan bertanggung jawab dalam penerapan dunia nyata.

## **1.2 Perumusan Masalah**

1. Bagaimana cara melakukan klasifikasi etnisitas berbasis citra wajah dengan menggunakan arsitektur VGG19 dan ResNet101?
2. Bagaimana pengaruh arsitektur CNN terhadap performa klasifikasi etnisitas dan rekomendasi arsitektur yang tepat dalam klasifikasi etnisitas berdasarkan hasil evaluasi performa dan potensi bias?

## **1.3 Tujuan Penelitian**

Penelitian ini bertujuan untuk menganalisis, mengevaluasi, dan membandingkan performa dua arsitektur Convolutional Neural Network, yaitu

VGG19 dan ResNet101, dalam tugas klasifikasi etnisitas berbasis citra wajah. Secara spesifik, penelitian ini bertujuan untuk:

1. Menganalisis kinerja arsitektur CNN VGG19 dan ResNet101 dalam melakukan klasifikasi citra wajah untuk mengetahui tingkat akurasi serta kemampuan masing-masing arsitektur dalam membedakan kelas etnis.
2. Menentukan arsitektur CNN yang lebih optimal untuk klasifikasi etnisitas berdasarkan citra wajah melalui evaluasi performa dan analisis potensi bias, sehingga dapat memberikan rekomendasi model yang lebih akurat, efisien, dan seimbang antar kelas etnis.

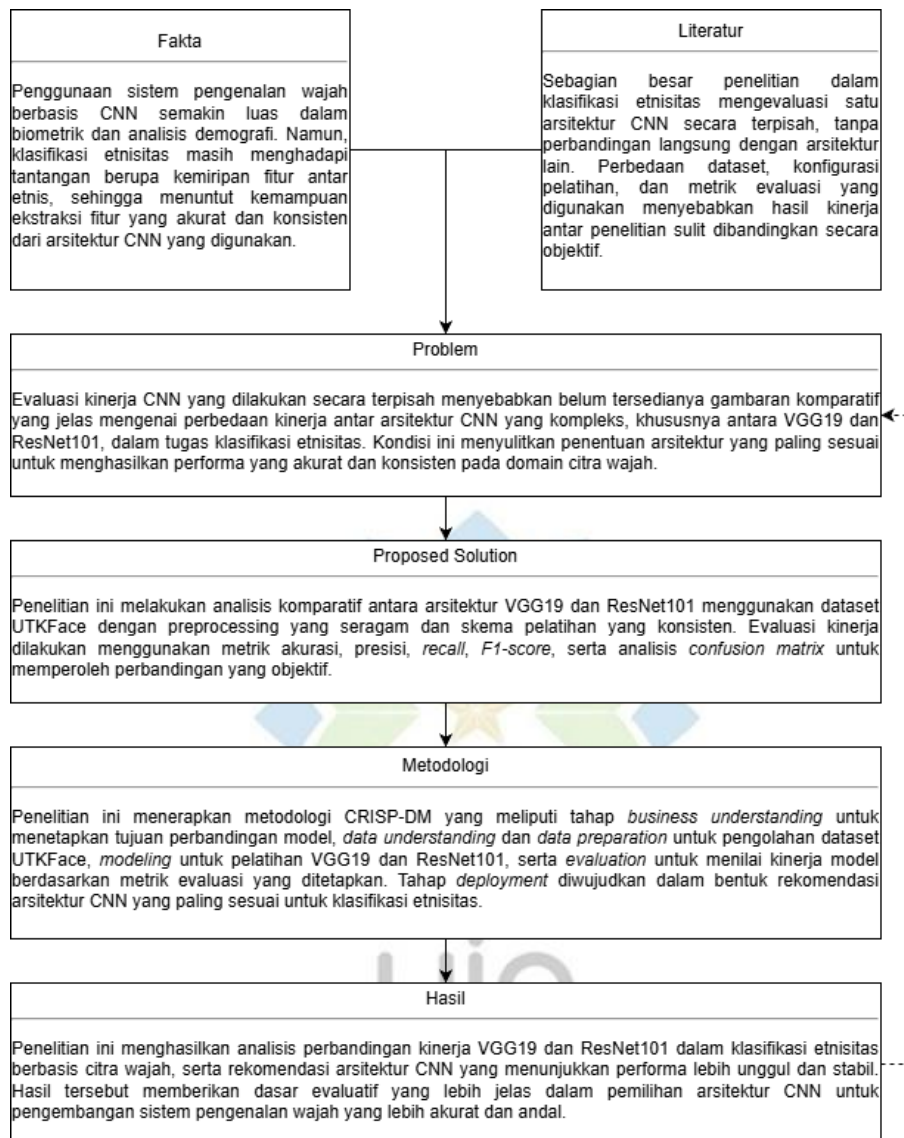
#### **1.4 Batasan Masalah Penelitian**

Agar penelitian ini lebih terfokus dan terarah, maka ruang lingkup penelitian dibatasi pada hal-hal berikut:

1. Dataset yang digunakan dalam penelitian ini dipilih berdasarkan kemampuannya dalam merepresentasikan setidaknya enam kategori etnis, yaitu *East Asian*, *Southeast Asian*, *India*, *Caucasian*, *Negroid*, dan *Middle Eastern*. Proses klasifikasi yang dianalisis hanya mencakup atribut etnisitas, tanpa mempertimbangkan atribut demografis lain seperti usia atau jenis kelamin.
2. Preprocessing data hanya mencakup tahapan dasar, seperti *resize* dan normalisasi, tanpa eksplorasi teknik preprocessing lanjutan atau metode face alignment berbasis landmark.
3. Strategi optimasi yang dieksplorasi terbatas pada teknik transfer learning dan *fine-tuning hyperparameter* (seperti *learning rate*, *batch size*).
4. Penelitian ini tidak membedakan secara konseptual antara istilah ras dan etnisitas. Seperti pada banyak penelitian klasifikasi wajah berbasis biometrik, kedua istilah tersebut digunakan secara bergantian untuk merujuk pada pengelompokan manusia berdasarkan karakteristik visual wajah yang tersedia dalam dataset.

#### **1.5 Kerangka Pemikiran Penelitian**

Kerangka pemikiran penelitian ini dirancang untuk memetakan alur logis penelitian secara sistematis, mulai dari identifikasi konteks yang lebih luas hingga luaran spesifik yang diharapkan, sebagaimana divisualisasikan pada Gambar 1.1.



Gambar 1.1 Kerangka Pemikiran

Alur kerangka pemikiran penelitian ini diawali dengan peninjauan terhadap dua domain utama yang menjadi landasan penelitian, yaitu perkembangan teknologi pengenalan wajah berbasis computer vision serta kajian penerapan Convolutional Neural Network (CNN) dalam klasifikasi demografis. Pemanfaatan sistem pengenalan wajah semakin luas pada bidang biometrik dan analisis demografi, sehingga menuntut model klasifikasi yang memiliki tingkat akurasi tinggi, stabilitas kinerja yang baik, serta mampu meminimalkan potensi bias. Namun, pada tugas klasifikasi etnisitas, masih terdapat tantangan berupa kemiripan fitur wajah antar kelompok etnis dan ketidakseimbangan distribusi data, yang dapat memengaruhi kemampuan model dalam melakukan klasifikasi secara konsisten [17][18][19][20].

Dari sisi kajian ilmiah, berbagai penelitian menunjukkan bahwa arsitektur CNN seperti VGG dan ResNet memiliki performa yang baik dalam klasifikasi visual. Meskipun demikian, evaluasi kinerja arsitektur sering dilakukan secara terpisah pada konfigurasi, dataset, dan skema pelatihan yang berbeda, sehingga hasil yang diperoleh sulit dibandingkan secara objektif. Perbedaan karakteristik arsitektur, khususnya pada tingkat kedalaman jaringan dan penggunaan residual connection, berpotensi menghasilkan kemampuan ekstraksi fitur yang berbeda, terutama pada citra wajah yang memiliki kompleksitas visual tinggi [18][21].

Kondisi tersebut mengarah pada permasalahan utama penelitian, yaitu belum tersedianya gambaran komparatif yang jelas mengenai perbedaan kinerja arsitektur VGG19 dan ResNet101 dalam tugas klasifikasi etnisitas berbasis citra wajah. Keterbatasan dasar empiris ini menyulitkan penentuan arsitektur CNN yang paling sesuai untuk menghasilkan performa yang akurat dan stabil pada domain tersebut [19].

Sebagai solusi, penelitian ini melakukan analisis komparatif terhadap arsitektur VGG19 dan ResNet101 dalam klasifikasi etnisitas berbasis citra wajah dengan memanfaatkan dataset citra wajah berlabel etnisitas. Kedua arsitektur dilatih dan dievaluasi dengan skema preprocessing yang seragam serta penerapan strategi transfer learning dan fine-tuning yang konsisten, sehingga perbedaan performa yang dihasilkan dapat mencerminkan karakteristik arsitektur masing-masing model [18][20][21]. Evaluasi kinerja dilakukan menggunakan metrik akurasi, presisi, *recall*, *F1-score*, *training time*, serta analisis *confusion matrix* untuk menilai performa model secara menyeluruh dan mengamati distribusi kesalahan klasifikasi antarkelas etnis.

Pelaksanaan penelitian ini mengikuti metodologi CRISP-DM sebagai kerangka kerja sistematis yang mencakup tahap *business understanding*, *data understanding*, *data preparation*, *evaluation*, dan *deployment* [22][23]. Tahap *deployment* diwujudkan dalam bentuk rekomendasi arsitektur CNN yang paling sesuai berdasarkan hasil evaluasi komparatif. Analisis yang dihasilkan memberikan dasar evaluative yang lebih jelas dalam pemilihan arsitektur CNN untuk pengembangan sistem klasifikasi etnisitas berbasis pengenalan wajah yang lebih akurat dan andal.

## **1.6 Sistematika Penulisan**

Sistematika penulisan penelitian ini disusun untuk memberikan gambaran umum tentang proses penulisan yang dilakukan untuk menyelesaikan penelitian ini. Sistematika penulisan penelitian ini adalah sebagai berikut:

### **BAB I PENDAHULUAN**

Bab ini terdiri dari latar belakang, perumusan masalah penelitian, tujuan penelitian, batasan masalah penelitian, kerangka pemikiran penelitian, metodologi penelitian, dan sistematika penulisan.

### **BAB II KAJIAN LITERATUR**

Bab ini terdiri dari pembahasan konsep dasar dan teori-teori yang berkaitan dengan topik masalah yang diambil dan hal-hal yang berguna dalam proses analisis permasalahan.

### **BAB III METODOLOGI PENELITIAN**

Bab ini terdiri dari pembahasan tentang alur metodologi penelitian dari tahap awal sampai akhir yang di dalamnya terdiri dari pembahasan analisa produk dan perencanaan eksekusi aplikasi.

### **BAB IV HASIL DAN PEMBAHASAN**

Bab ini terdiri dari dua hal utama, pertama pemaparan tentang temuan atau hasil penelitian berdasarkan tahapan penelitian yang dilakukan. Pemaparan hasil penelitian disesuaikan dengan urutan masalah penelitian. Kedua pembahasan hasil atau temuan penelitian untuk menjawab rumusan penelitian.

### **BAB V PENUTUP**

Bab ini terdiri dari pembahasan kesimpulan penelitian, kritik, dan saran yang penulis dapatkan selama menyelesaikan penelitian tugas akhir ini. Intisari dari bab ini dapat digunakan sebagai pertimbangan untuk pengembangan penulisan selanjutnya