

Klasifikasi Ras berdasarkan Citra Wajah menggunakan Convolutional Neural Network

Rizza Aulia Rahman¹, Anditya Arifianto², Kurniawan Nur Ramdhani³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹rizzauliahman@students.telkomuniversity.ac.id, ²anditya@telkomuniversity.ac.id,

³kurniawannr@telkomuniversity.ac.id

Abstrak

Salah satu fungsi penting identifikasi ras dalam kehidupan sehari-hari adalah untuk keperluan kontrol kesehatan yang dilakukan oleh pemerintah. Namun, tidak mudah untuk menentukan ras dari masing-masing orang secara langsung, terlebih apabila jumlah masyarakat yang banyak sehingga juga akan memakan waktu lama.

Salah satu cara yang dapat dilakukan untuk menyelesaikan masalah tersebut adalah dengan menggunakan sistem otomatis. Sistem otomatis disini adalah berupa perangkat lunak yang dapat digunakan untuk mengidentifikasi ras berdasarkan citra wajah seseorang. Dalam penelitian ini metode yang digunakan adalah Convolutional Neural Network (CNN).

Melalui penelitian ini, didapatkan sebuah struktur CNN yang optimal untuk mengklasifikasikan ras. Struktur CNN tersebut memberikan akurasi training sebesar 91.63% setelah dilatih dengan sekitar 5400 citra wajah dan akurasi testing sebesar 89.739% untuk 600 citra wajah. Selain itu, ketika diuji dengan dataset citra baru, CNN tersebut menghasilkan akurasi klasifikasi sebesar 76.67% untuk 30 citra wajah.

Kata kunci : ras, klasifikasi, citra wajah, convolutional neural network, CNN

Abstract

One of the main purpose of race identification in daily life is for the government health control program. But, it is not easy to determine the race of each person directly, it will take a long time, especially because high number of people in the society. One of the technique that could be implemented to solve this problem is by utilizing an automated software system that could identify the person's race based on their face image. For this research, the Convolutional Neural Network (CNN) is used for the software.

The result of this research is an optimal structure of CNN for race classification. The best CNN structure, give 91.63% as the training accuracy value after trained using around 5400 facial images and 89.739% as the testing accuracy value on 600 facial images. Also, when the CNN is tested to classify a new face dataset, it gives 76.67% classification accuracy on 30 facial images.

Keywords: race, classification, face image, convolutional neural network, CNN

1. Pendahuluan

Latar Belakang

Ras merupakan salah satu identitas penting yang dimiliki setiap orang. Salah satu fungsi penting ras dalam kehidupan sehari-hari adalah untuk keperluan kontrol kesehatan yang dilakukan oleh pemerintah [10]. Selain itu, ras juga merupakan salah satu identitas penting yang pertama kali diidentifikasi pada korban suatu bencana. Pada beberapa negara di dunia, ras juga merupakan salah satu identitas yang diperlukan dalam sensus penduduk.

Menurut Louis Agassiz, terdapat tiga ras utama yang ada di dunia ini, yaitu caucasoid, mongoloid, dan negroid [1]. Sehingga, dalam penelitian ini, ras yang akan diklasifikasikan adalah ketiga ras tersebut. Setiap ras yang ada pasti memiliki berbagai ciri khusus yang membedakannya dengan ras lain. Ciri tersebut dapat berupa hal yang mudah dilihat, misalkan warna kulit, pola rambut, atau bentuk muka, maupun hal yang hanya dapat diketahui melalui pemeriksaan dokter, contohnya adalah bentuk tulang pinggul, bentuk tulang tengkorak, ataupun struktur gigi [3, 6]. Untuk identifikasi menggunakan ciri-ciri yang dapat dilihat tersebut dapat dilakukan hanya melalui pengamatan langsung. Namun berdasarkan own-race bias yang timbul, akurasi dari pengamatan langsung ini sangatlah kecil [11]. Selain itu, masalah yang ditimbulkan dari identifikasi dengan pengamatan langsung adalah waktu yang dibutuhkan menjadi sangat lama apabila data yang harus diidentifikasi berjumlah besar.

Untuk mengatasi masalah-masalah yang muncul tersebut, dibutuhkan sebuah sistem yang dapat mengenali ciri-ciri dari masing-masing ras. Fitur yang akan digunakan dalam sistem pengenalan adalah berdasarkan wajah.

Hal ini dikarenakan wajah merupakan fitur utama yang dapat digunakan untuk mengidentifikasi 3 ciri utama dari seseorang, yaitu ras, jenis kelamin, dan umur [7]. Dengan menggunakan sistem komputer, diharapkan proses ini dapat dilakukan secara otomatis dan dengan akurasi yang tinggi. Sedangkan, metode yang akan dipakai dalam penelitian ini adalah Convolutional Neural Network (CNN) yang mana digunakan untuk mengklasifikasikan ras Caucasoid, Negroid, dan Mongoloid. Metode ini dipakai karena berdasarkan penelitian yang telah dilakukan sebelumnya, contohnya adalah oleh Amr Ahmed dan Kai Yu pada tahun 2008, CNN memberikan akurasi yang cukup tinggi dalam mengklasifikasikan ras Asian dan Caucasoid [2].

Perumusan Masalah

Dalam penelitian ini, ada beberapa permasalahan yang ingin diselesaikan. Permasalahan pertama adalah bagaimana caranya membangun sistem Convolutional Neural Network (CNN) yang dapat digunakan untuk mengklasifikasikan ras melalui citra wajah. Selain itu, perlu didapatkan kombinasi nilai-nilai parameter pada CNN yang memberikan performansi terbaik untuk proses klasifikasi.

Tujuan

Setelah menjalankan penelitian ini, diharapkan dapat dihasilkan sebuah sistem Convolutional Neural Network (CNN) yang dapat mengklasifikasikan ras Mongoloid, Negroid, dan Caucasoid dengan baik, dimana sistem CNN tersebut telah dioptimisasi dengan menggunakan kombinasi nilai-nilai parameter yang tepat.

2. Studi Terkait

Convolutional Neural Network

Convolutional Neural Network (CNN) merupakan sebuah struktur pengembangan dari Artificial Neural Network (ANN) biasa, dimana kedua jaringan tersebut masih tersusun oleh neuron-neuron yang memiliki bobot (weight) dan juga nilai bias yang didapatkan dari hasil learning. Perbedaan yang ada pada CNN adalah adanya convolution layer dan juga pooling layer yang muncul sebelum masuk ke pemrosesan oleh ANN. Selain itu, CNN juga secara eksplisit merupakan suatu sistem yang digunakan sebagai alat klasifikasi citra digital (image).

Secara umum, CNN disusun oleh beberapa layer, yaitu :

- Input Layer, merupakan sebuah layer yang berisi sebuah citra 3 dimensi yang memiliki kedalaman 3 layer, sebagai representasi layer R, G, dan B
- Convolution Layer, merupakan layer yang berfungsi untuk menghitung output dari neuron yang terhubung di dalam suatu local region. Layer ini berbentuk 3 dimensi dengan kedalaman sesuai dengan jumlah filter yang dipakai
- Activation Layer, layer ini berfungsi untuk melakukan perhitungan fungsi aktivasi terhadap hasil output dari convolution layer
- Pooling Layer, berfungsi untuk melakukan downsampling terhadap citra hasil konvolusi. Contohnya adalah melakukan downsampling dari citra berukuran [64x64x10] menjadi citra berukuran [32x32x10]
- Fully-Connected Layer, layer ini berfungsi untuk menghitung skor class yang ada terhadap citra yang telah diolah. Layer ini merupakan salah satu persamaan dari CNN dengan ANN biasa.

Dengan ini dapat ditarik suatu kesimpulan bahwa CNN masih tetap memiliki prinsip yang sama dengan ANN, yaitu mengubah suatu layer input menjadi skor class-class yang ada. Namun, proses yang ada di dalam CNN sendiri berbeda dimana citra input yang disediakan terlebih dahulu dilakukan proses konvolusi dan juga pooling sebelum diklasifikasikan. Metode klasifikasi yang digunakan secara umum adalah Fully-Connected Layer, namun tidak menutup kemungkinan digunakannya metode klasifikasi lain.

Klasifikasi ras menggunakan sistem komputer

Beberapa penelitian pernah dilakukan sebelumnya untuk melakukan proses klasifikasi ras menggunakan sistem komputer [5]. Salah satunya pada tahun 2007, Huchuan Lu dan Hui Lin telah membangun suatu sistem klasifikasi ras yang menggunakan Gabor filter sebagai metode ekstraksi fitur yang ada pada citra, dilanjutkan dengan Support Vector Machine (SVM) yang dioptimasi menggunakan Adaboost dalam proses klasifikasi [9]. Hasil yang didapat dari penelitian ini adalah akurasi klasifikasi ras African American, Caucasian, dan Eastern Asian sebesar 93%. Sedangkan pada tahun 2010, Meltem Demirkus dan Kshitiz Grag mengembangkan suatu sistem yang menggunakan Skin Color filter sebagai feature extractor [4]. Sistem ini juga menggunakan SVM dalam proses klasifikasinya,

dan akurasi yang didapatkan dari sistem ini menjadi 94%, sedikit lebih tinggi dari sistem milik Huchuan Lu. Namun, kedua sistem ini memiliki kelemahan, yaitu kompleksitas algoritma yang tinggi dan juga kebutuhan memori yang besar disebabkan oleh banyaknya perhitungan kuadrat dalam algoritma yang ada pada SVM. Selain itu, kedua sistem tersebut juga memiliki kemungkinan lebih besar untuk terjadi overfit apabila data yang diklasifikasikan menjadi lebih banyak.

Sedangkan pada tahun 2008, Amr Ahmed dan Kai Yu telah menerapkan Convolutional Neural Network (CNN) yang dioptimasi menggunakan metode Transfer Learning sebagai sistem klasifikasi ras [2]. CNN yang digunakan pada sistem ini merupakan CNN yang memiliki kedalaman 7 layer, dimana 4 layer merupakan layer konvolusi, 2 layer sebagai layer pooling, dan 1 layer fully-connected. Dalam sistem ini, ditunjukkan bahwa performansi CNN akan menjadi lebih baik apabila data citra yang digunakan pada fase training lebih banyak. Dengan sistem ini, didapatkan akurasi klasifikasi ras Asian dan Caucasian sebesar 95%. Pada sistem ini, kemungkinan overfit menjadi lebih kecil seiring dengan semakin banyaknya data yang akan diklasifikasikan. Selain itu, beban perhitungan komputasional yang ada pada CNN lebih ringan daripada beban perhitungan pada SVM. Namun, karena pada layer konvolusi sistem ini menggunakan transformasi sigmoid dimana terdapat perhitungan eksponensial, beban perhitungan komputasional sistem ini tetap tinggi. Disebutkan juga bahwa pada layer terakhir sistem ini menggunakan fully-connected layer, dimana pada layer ini perhitungan parameter yang dilakukan menjadi sangat banyak.

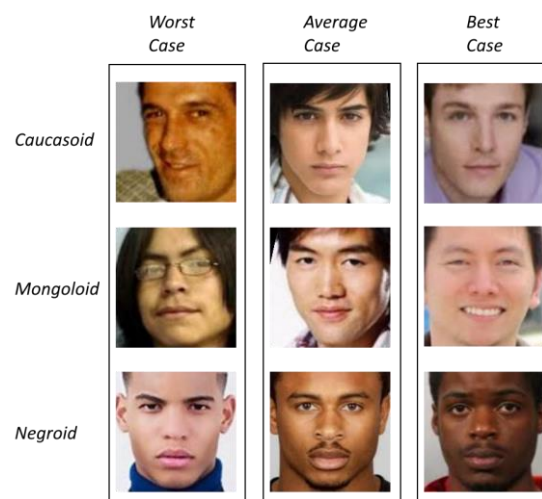
3. Sistem yang Dibangun

Dalam pengujian ini, penulis akan melakukan klasifikasi ras manusia menggunakan Convolutional Neural Network (CNN). CNN ini selanjutnya dilatih menggunakan dataset wajah UTKFace yang telah dipisah secara acak menjadi training dataset dan testing dataset.

3.1 Dataset UTKFace

Dataset UTKFace adalah sebuah dataset citra wajah yang disediakan secara gratis untuk keperluan penelitian. Dataset ini terdiri dari lebih dari 20000 citra wajah, dimana masing-masing citra dianotasi dengan umur, jenis kelamin, dan etnis yang sesuai. Masing-masing citra wajah juga sudah disesuaikan dan dipotong pada bagian sekitar wajah.

Dataset ini sendiri terdiri dari tiga ras utama sesuai dengan yang disebutkan oleh Louis Agassiz dalam bukunya, yaitu Caucasoid, Mongoloid, dan Negroid [1]. Selain itu, dataset ini juga mengandung citra wajah yang menggambarkan worst case, average case, dan best case berdasarkan ciri-ciri yang disebutkan pada buku yang sama. Gambar 1 menunjukkan beberapa contoh dari masing-masing citra tersebut.



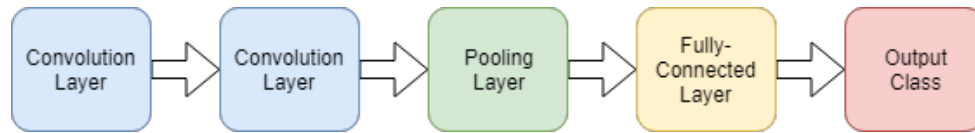
Gambar 1. Contoh worst case, average case, dan best case dari masing-masing ras

Dalam penelitian ini, dilakukan seleksi terhadap dataset terlebih dahulu sehingga hanya sebagian dari dataset yang akan digunakan. Metode seleksi dataset yang digunakan adalah dengan mengambil citra wajah yang berumur antara 20-50 tahun, dan merupakan etnis kulit putih, hitam, atau india. Lalu, didapatkan sekitar 10000 citra wajah

dari hasil seleksi yang telah dilakukan. Dari 10000 citra wajah tersebut, diambil masing-masing 2000 citra dari setiap ras, lalu dibagi lagi menjadi 90% sebagai dataset training dan 10% sisanya sebagai dataset testing.

3.2 Convolutional Neural Network

Sistem Convolutional Neural Network (CNN) secara umum terdiri dari layer konvolusi, layer pooling, dan juga sebuah layer fully-connected. Secara umum, struktur CNN ditunjukkan pada gambar 2.



Gambar 2. Skema CNN secara umum

Dalam CNN ini, berdasarkan referensi dari LeCun [8] dan juga telah digambarkan pada Gambar 2, Ekstraksi fitur yang ada pada CNN dilakukan oleh layer konvolusi dan layer pooling. Dalam pengujian ini, CNN yang digunakan dibangun dalam 9 struktur yang berbeda, dimana variasi untuk masing-masing struktur terletak pada ukuran filter pada layer konvolusi, jumlah layer konvolusi, dan juga jumlah filter pada setiap layer konvolusi. Selain itu, perlu ditentukan metode apakah perlu dilakukan suatu metode pre-processing, dalam hal ini mengubah citra menjadi citra grayscale, terhadap citra yang akan digunakan dalam proses training.

Setelah melalui proses ekstraksi fitur, citra hasil ekstraksi akan diklasifikasikan menggunakan Fully-Connected Layer. Dalam layer ini akan terdapat suatu input dan output layer, dan juga 2 hidden layer. Fungsi aktivasi yang akan digunakan pada layer ini adalah ReLU. Output yang dihasilkan dari layer ini merupakan nilai yang menggambarkan kecocokan citra dengan kelas ras yang ada.

Tabel 1 adalah struktur dari setiap CNN yang digunakan dalam pengujian. Yang pertama, 3 CNN dibangun dengan ukuran filter yang berbeda.

Tabel 1. Tabel struktur CNN berdasarkan ukuran filter konvolusi

Struktur CNN berdasarkan ukuran filter		
CNN 1	CNN 2	CNN 3
<ul style="list-style-type: none"> • Layer konvolusi (8 filter: 3x3) • Layer konvolusi (16 filter: 3x3) • Layer max pooling (pooling size: 2x2) • Layer konvolusi (32 filter: 3x3) • Layer konvolusi (64 filter: 3x3) • Layer max pooling (pooling size: 2x2) • Layer fully-connected (unit: 512) • Layer aktivasi Rel-U • Layer fully-connected (unit: 3) 	<ul style="list-style-type: none"> • Layer konvolusi (8 filter: 5x5) • Layer konvolusi (16 filter: 5x5) • Layer max pooling (pooling size: 2x2) • Layer konvolusi (32 filter: 5x5) • Layer konvolusi (64 filter: 5x5) • Layer max pooling (pooling size: 2x2) • Layer fully-connected (unit: 512) • Layer aktivasi Rel-U • Layer fully-connected (unit: 3) 	<ul style="list-style-type: none"> • Layer konvolusi (8 filter: 7x7) • Layer konvolusi (16 filter: 7x7) • Layer max pooling (pooling size: 2x2) • Layer konvolusi (32 filter: 7x7) • Layer konvolusi (64 filter: 7x7) • Layer max pooling (pooling size: 2x2) • Layer fully-connected (unit: 512) • Layer aktivasi Rel-U • Layer fully-connected (unit: 3)

Pertama-tama dilakukan 2 skenario pengujian, yaitu menggunakan citra RGB dan citra grayscale. Untuk pengujian ini, struktur yang digunakan adalah struktur CNN 1. Berdasarkan hasil pengujian kedua skenario tersebut, diambil skenario dengan performansi terbaik. Untuk skema-skema pengujian berikutnya, maka diasumsikan metode pre-processing yang digunakan adalah metode yang memberikan performansi lebih baik.

Skenario selanjutnya adalah melakukan pengujian terhadap setiap struktur yang ada pada tabel 1. Setelah melakukan pengujian ini, struktur dengan performansi terbaik digunakan untuk pengujian selanjutnya, dimana parameter yang diubah adalah jumlah layer konvolusi yang ada pada CNN. Untuk parameter ini, diuji 3 struktur CNN berbeda, yang ditunjukkan pada tabel 2.

Tabel 2. Tabel struktur CNN berdasarkan jumlah layer konvolusi

Struktur CNN berdasarkan jumlah layer konvolusi		
CNN 1	CNN 2	CNN 3
<ul style="list-style-type: none"> • Layer konvolusi (8 filter) • Layer konvolusi (16 filter) • Layer max pooling (pooling size: 2x2) • Layer konvolusi (32 filter) • Layer konvolusi (64 filter) • Layer max pooling (pooling size: 2x2) • Layer fully-connected (unit: 512) • Layer aktivasi Rel-U • Layer fully-connected (unit: 3) 	<ul style="list-style-type: none"> • Layer konvolusi (8 filter) • Layer konvolusi (16 filter) • Layer max pooling (pooling size: 2x2) • Layer fully-connected (unit: 512) • Layer aktivasi Rel-U • Layer fully-connected (unit: 3) 	<ul style="list-style-type: none"> • Layer konvolusi (8 filter) • Layer konvolusi (16 filter) • Layer max pooling (pooling size: 2x2) • Layer konvolusi (32 filter) • Layer konvolusi (64 filter) • Layer max pooling (pooling size: 2x2) • Layer konvolusi (128 filter) • Layer konvolusi (256 filter) • Layer max pooling (pooling size: 2x2) • Layer fully-connected (unit: 512) • Layer aktivasi Rel-U • Layer fully-connected (unit: 3)

Berdasarkan hasil pengujian CNN yang ada pada tabel 2, struktur CNN dengan performansi terbaik kembali digunakan untuk pengujian parameter berikutnya, yaitu pengujian parameter jumlah filter pada setiap layer konvolusi. Tabel 3 menunjukkan daftar skenario untuk pengujian parameter jumlah filter.

Tabel 3. Tabel struktur CNN berdasarkan jumlah filter pada setiap layer konvolusi

Skenario 1						
Layer konvolusi	Jumlah filter					
	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6
2	4	8				
4	4	8	16	32		
6	4	8	16	32	64	128
Skenario 2						
Layer konvolusi	Jumlah filter					
	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6
2	8	16				
4	8	16	32	64		
6	8	16	32	64	128	256
Skenario 3						
Layer konvolusi	Jumlah filter					
	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6
2	16	32				
4	16	32	64	128		
6	16	32	64	128	256	512

Masing-masing CNN dengan struktur yang telah disebutkan pada tabel 1, 2, dan 3 dilatih dengan menggunakan dataset training dari dataset UTKFace. Dalam proses training tersebut, metode validasi yang digunakan adalah k-fold cross validation dengan nilai k=10. Metode tersebut membagi dataset menjadi sejumlah k bagian (fold), lalu melakukan training terhadap k-1 fold yang ada dan memvalidasikannya dengan satu fold yang tersisa. Proses ini diulang hingga semua k fold yang ada pada dataset telah digunakan untuk proses validasi. Untuk akurasi pada proses training didapatkan menggunakan rumus

$$score = \frac{\sum_{n=1}^k score_n}{k} \tag{1}$$

$$score_n = \frac{N_{correct}}{N_{total}} * 100\% \tag{2}$$

Dimana score dan score_n adalah nilai akurasi total dan nilai akurasi pada fold n, N_{correct} adalah jumlah data yang benar diklasifikasikan, dan N_{total} adalah jumlah data keseluruhan yang ada pada fold yang digunakan untuk proses validasi. CNN tersebut lalu diujikan menggunakan dataset testing untuk mendapatkan nilai performansinya. Rumus untuk mendapatkan nilai performansi adalah sama dengan rumus untuk mendapatkan nilai akurasi klasifikasi, yaitu

$$score_{test} = \frac{N_{correct}}{N_{total}} * 100\% \tag{3}$$

Dimana score_{test} adalah nilai akurasi testing, N_{correct} adalah jumlah data yang diklasifikasikan dengan benar pada proses testing, sedangkan N_{total} adalah jumlah total data yang ada pada dataset testing.

Setelah CNN dengan performansi terbaik telah didapatkan dari hasil skenario-skenario di atas, CNN tersebut diuji menggunakan sejumlah kecil citra wajah yang benar-benar merupakan data baru. Data baru ini nantinya berisi sekitar 30 citra wajah, yang dibagi menjadi 10 citra wajah untuk masing-masing ras. Data baru ini didapatkan dengan cara pencarian gambar secara manual melalui internet, sesuai dengan ras yang ingin dicari. Pengujian ini berguna untuk memastikan bahwa performansi CNN yang didapatkan sesuai dengan performansi yang dihasilkan pada skenario-skenario sebelumnya.

Nilai-nilai yang telah disebutkan di atas merupakan nilai yang dipilih setelah mempertimbangkan beberapa hal. Salah satu hal yang paling penting sebagai pertimbangan adalah keterbatasan dalam hal ketersediaan perangkat keras. Dikarenakan CNN harus melakukan perhitungan secara langsung sekaligus dalam jumlah yang banyak, maka CNN juga membutuhkan perangkat keras (dalam hal ini kartu grafis ataupun CPU) yang sangat memadai untuk sekedar melakukan proses pelatihan. Semakin banyak layer dan ukuran gambar yang ada pada CNN, maka beban sistem juga akan semakin besar dan membutuhkan perangkat keras yang lebih banyak. Oleh karena itu, jumlah layer yang digunakan pada penelitian ini hanya berada pada kisaran 2 hingga 6 layer, dan juga jumlah filter maksimal adalah 512.

4. Evaluasi

4.1 Pengujian CNN

Berdasarkan penjelasan yang telah diberikan pada bab sebelumnya, sistem Convolutional Neural Network (CNN) yang dibangun akan diujikan terhadap dataset UTKFace yang telah diseleksi dan dibagi menjadi dataset training dan testing. Untuk pengujian sistem CNN itu sendiri dibagi menjadi 4 tahap, yaitu pengujian menggunakan dataset RGB dan grayscale, pengujian berdasarkan ukuran filter konvolusi, dilanjutkan dengan pengujian berdasarkan jumlah layer konvolusi, dan yang terakhir adalah pengujian berdasarkan jumlah filter.

4.1.1 Pengujian CNN menggunakan dataset RGB dan grayscale

Pertama-tama, perlu ditentukan apakah perlu dilakukan pre-processing terhadap citra wajah yang ada. Dalam hal ini, metode yang akan dicek adalah pengubahan citra RGB menjadi citra grayscale. Skenario ini menguji struktur CNN terhadap 2 dataset, dimana salah satu dataset telah diubah citra-citra wajah yang ada menjadi citra grayscale. Tabel 4 hasil performansi training dan testing dari pengujian masing-masing model CNN.

Tabel 4. Tabel performansi CNN ketika diuji dengan citra RGB dan grayscale

Akurasi	Skenario 1 (RGB)	Skenario 2 (Grayscale)
Training	94.18%	92.64%
Testing	90.391%	90.065%

Dari tabel 4 di atas, dapat dilihat bahwa CNN dengan dataset yang menggunakan citra RGB menghasilkan performansi yang lebih baik pada fase training maupun testing. Hal ini kemungkinan dapat terjadi karena ketika citra wajah ras caucasoid dan mongoloid diubah menjadi citra grayscale, maka hampir tidak ada perbedaan di antara keduanya, karena warna kulit yang sebenarnya merupakan fitur yang paling membedakan antara keduanya menjadi terlihat sama. Namun, perbedaan performansi yang dihasilkan tidak terlalu signifikan pada kedua skenario tersebut. Sedangkan pada proses training, metode validasi yang digunakan masih sama dengan sebelumnya, yaitu metode k-fold validation, dengan fold yaitu 10. Grafik pada gambar 4 menunjukkan performansi setiap skenario ukuran filter CNN pada setiap fold.



Gambar 3. Grafik performansi setiap fold pada masing-masing skenario CNN

Berdasarkan grafik yang ditunjukkan pada gambar 4 di atas, dapat dilihat bahwa menggunakan citra RGB memberikan performansi yang sedikit lebih baik pada proses training. Oleh karena itu, maka diambil kesimpulan bahwa menggunakan citra RGB untuk klasifikasi menggunakan CNN memberikan hasil yang lebih baik dibandingkan menggunakan citra grayscale. Oleh karena itu, untuk pengujian berikutnya, dataset yang digunakan adalah dataset yang menggunakan citra RGB.

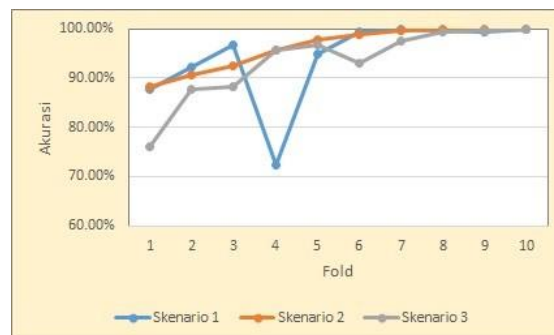
4.1.2 Pengujian CNN berdasarkan ukuran filter konvolusi

Pengujian CNN yang kedua adalah melakukan proses pelatihan model CNN yang memiliki ukuran filter konvolusi berbeda-beda terhadap dataset latih. Dalam pengujian ini, ukuran filter yang digunakan adalah 3x3, 5x5, dan 7x7, sehingga dihasilkan 3 model CNN. 3 Model CNN ini lalu di uji coba untuk mengklasifikasikan citra-citra wajah yang ada pada dataset uji, lalu dicatat hasil performansi untuk masing-masing model. Tabel 5 menunjukkan hasil performansi dari pengujian masing-masing model CNN.

Tabel 5. Tabel performansi CNN ketika diuji dengan ukuran filter 3x3, 5x5, dan 7x7

Akurasi	Skenario 1 (Ukuran filter 3x3)	Skenario 2 (Ukuran filter 5x5)	Skenario 3 (Ukuran filter 7x7)
Training	94.18%	96.28%	93.40%
Testing	90.391%	89.739%	89.088%

Berdasarkan performansi yang ditunjukkan pada tabel 5, diperlihatkan bahwa pada proses training, menggunakan ukuran filter 5x5 menghasilkan akurasi yang lebih baik. Namun, hal ini tidak diikuti dengan peningkatan pada akurasi testing yang mana justru mengalami penurunan akurasi. Sedangkan, menggunakan ukuran filter 7x7 menghasilkan penurunan performansi baik pada proses training maupun testing. Hal ini dapat terjadi karena ketika menggunakan filter ukuran 5x5 dan 7x7, citra yang dihasilkan pada setiap proses konvolusi menjadi lebih kecil. Dengan berkurangnya ukuran citra, maka fitur-fitur yang ada di dalamnya pun menjadi lebih sedikit, sehingga mempengaruhi performansi CNN yang dihasilkan. Sedangkan pada proses training, metode validasi yang digunakan masih sama dengan sebelumnya, yaitu metode k-fold validation, dengan fold yaitu 10. Grafik pada gambar 4 menunjukkan performansi setiap skenario ukuran filter CNN pada setiap fold.



Gambar 4. Grafik performansi setiap fold pada masing-masing skenario CNN

Berdasarkan grafik pada gambar 4, terlihat bahwa pada proses training, CNN dengan ukuran filter 5x5 menghasilkan performansi yang lebih stabil dan lebih baik pada setiap fold, sehingga akurasi training secara keseluruhan juga menjadi lebih baik. Namun, akurasi testing yang diberikan justru lebih rendah jika dibandingkan dengan akurasi testing menggunakan ukuran filter 3x3. Oleh karena itu, dapat diambil kesimpulan bahwa menggunakan CNN dengan ukuran filter 3x3 memberikan performansi paling baik dalam skenario pengujian CNN berdasarkan ukuran filter konvolusi ini. Oleh karena itu, untuk skenario selanjutnya, CNN yang akan digunakan adalah CNN dengan struktur ukuran filter 3x3, dan tanpa menggunakan preprocessing pada citra yang akan digunakan pada sistem.

4.1.3 Pengujian CNN berdasarkan jumlah layer konvolusi

Proses pengujian selanjutnya adalah dengan mengubah jumlah layer pada sistem CNN terbaik yang didapatkan pada dua pengujian sebelumnya. Pada struktur CNN terbaik yang didapatkan, jumlah layer yang ada dalam CNN berjumlah 4. Dalam pengujian ini, struktur CNN terbaik dari pengujian sebelumnya digunakan untuk skenario pengujian pertama, lalu jumlah layer pada CNN diubah menjadi 2 untuk skenario pengujian kedua, dan 6 untuk skenario pengujian ketiga. Tabel 6 menunjukkan hasil performansi dari masing-masing skenario pengujian CNN.

Tabel 6. Tabel performansi CNN dengan jumlah layer konvolusi 4, 2, dan 6

Akurasi	Skenario 1 (Jumlah layer 4)	Skenario 2 (Jumlah layer 2)	Skenario 3 (Jumlah layer 6)
Training	94.18%	96.88%	85.28%
Testing	90.391%	87.622%	88.559%

Dari tabel 6, dapat dilihat bahwa dengan menggunakan 2 layer konvolusi meningkatkan performansi training CNN sekitar 2-3%. Namun peningkatan ini tidak diimbangi dengan peningkatan performansi testing, dimana terjadi penurunan performansi sekitar 3%. Sedangkan menambah jumlah layer konvolusi menjadi berjumlah 6 menurunkan performansi training secara drastis. Selain itu, performansi testing juga mengalami penurunan performansi. Hal ini kemungkinan dapat terjadi karena dengan menambah layer pada CNN menyebabkan semakin banyak fitur yang disimpan dari setiap citra. Namun, fitur-fitur tersebut banyak yang bukan merupakan fitur spesifik setiap ras, sehingga mempengaruhi performansi yang menjadi turun. Pada proses training, dilakukan juga proses validasi dimana metode yang digunakan adalah k-fold cross validation, dengan jumlah fold yaitu 10. Grafik pada gambar 5 menunjukkan performansi validasi pada setiap fold untuk masing-masing skenario.



Gambar 5. Grafik performansi setiap fold pada masing-masing skenario CNN

Berdasarkan grafik pada gambar 5, skenario 2 (menggunakan 2 layer konvolusi) memberikan stabilitas akurasi yang lebih tinggi pada proses validasi dibandingkan dengan skenario 1 (menggunakan 4 layer konvolusi). Hal ini sesuai dengan performansi training secara keseluruhan seperti yang telah ditunjukkan pada tabel 6. Namun, skenario tersebut menghasilkan overfit yang sedikit besar antara training dan testing. Oleh karena itu, untuk skenario pengujian selanjutnya, struktur CNN yang digunakan adalah CNN dengan ukuran filter 3x3 dan jumlah layer konvolusi 4, yang mana memberikan performansi yang lebih seimbang antara training dan testing.

4.1.4 Pengujian CNN berdasarkan jumlah filter pada masing-masing layer konvolusi

Proses pengujian CNN yang terakhir adalah dengan menggunakan variasi jumlah filter pada masing-masing layer konvolusi yang ada pada CNN. Dalam pengujian ini, sesuai dengan yang telah dijabarkan pada bab sebelumnya, ada 3 skenario variasi jumlah filter yang digunakan. Selain itu, berdasarkan hasil pengujian sebelum-

nya, maka jumlah layer konvolusi yang digunakan berjumlah 4 layer. Ketiga skenario variasi jumlah filter untuk masing-masing layer konvolusi yang digunakan dalam pengujian ini adalah sebagai berikut

1. Skenario 1: 8, 16, 32, 64
2. Skenario 2: 4, 8, 16, 32
3. Skenario 3: 16, 32, 64, 128

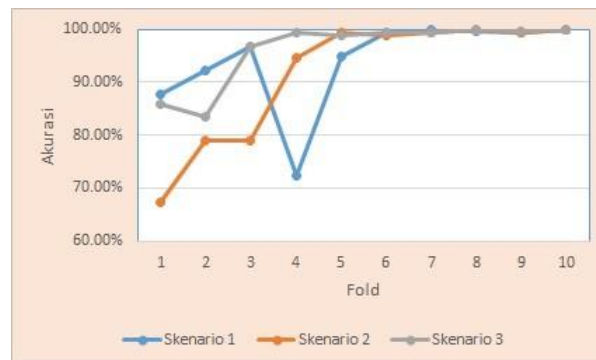
Masing-masing CNN dari skenario-skenario tersebut dilatih dan diuji menggunakan dataset training dan testing yang sama dengan yang digunakan pada pengujian-pengujian sebelumnya. Tabel 7 menunjukkan hasil dari pengujian masing-masing CNN.

Tabel 7. Tabel performansi CNN dari masing-masing skenario jumlah filter

Akurasi	Skenario 1	Skenario 2	Skenario 3
Training	94.18%	91.63%	96.21%
Testing	90.391%	89.739%	89.414%

Dari tabel 7, menggunakan jumlah filter yang lebih sedikit, dalam hal ini ditunjukkan oleh skenario 2, memberikan performansi training yang lebih rendah dibandingkan skenario 1. Sedangkan menggunakan jumlah filter yang lebih banyak seperti dalam skenario 3, memberikan performansi training yang lebih tinggi dibandingkan dengan skenario 1. Namun, performansi testing dari ketiga skenario tersebut secara keseluruhan dapat dikatakan seimbang, dimana performansi testing tertinggi masih dihasilkan oleh CNN skenario 1. Sedangkan CNN dengan performansi di bawahnya adalah skenario 2, dan terakhir adalah skenario 3. Namun, perbedaan akurasi testing antara skenario 1 dan skenario 3 hanya sekitar $\pm 0.9\%$, yang mana tidak menunjukkan perbedaan yang cukup besar sehingga dapat dikatakan performansi antara ketiga skenario tersebut cukup seimbang.

Selain itu, pada proses training juga masih dilakukan proses validasi, menggunakan metode yang sama dengan pengujian sebelumnya, yaitu k-fold cross validation, dengan jumlah fold = 10. Gambar 6 menunjukkan grafik perbandingan akurasi tiap fold pada proses training untuk masing-masing skenario.



Gambar 6. Grafik performansi setiap fold pada masing-masing skenario CNN

Jika hanya melihat dari hasil grafik 6, maka yang seharusnya memberikan performansi terbaik adalah CNN skenario 3. Namun, seperti yang telah dijelaskan pada tabel 7, CNN skenario 3 menghasilkan overfit yang sedikit besar antara performansi training dan testing. Sedangkan CNN yang memberikan overfit yang paling kecil antara training dan testing adalah CNN skenario 2. Sehingga, dapat diambil kesimpulan bahwa CNN terbaik pada pengujian ini adalah CNN dengan variasi jumlah filter 4, 8, 16, dan 32 untuk masing-masing layer konvolusi.

4.2 Pengujian Struktur CNN Terbaik

Berdasarkan hasil-hasil yang telah dijabarkan pada pengujian-pengujian sebelumnya, maka struktur CNN terbaik yang dapat digunakan untuk mengklasifikasikan ras mongoloid, caucasoid, dan negroid ditunjukkan pada tabel 8.

Tabel 8. Struktur CNN terbaik dari hasil pengujian

Layer	Nama Layer
1	Layer konvolusi
	Ukuran filter = 3x3
	Jumlah filter = 4
2	Layer konvolusi
	Ukuran filter = 3x3
	Jumlah filter = 8
3	Layer max pooling
	Ukuran pool = 2x2
4	Layer konvolusi
	Ukuran filter = 3x3
	Jumlah filter = 16
5	Layer konvolusi
	Ukuran filter = 3x3
	Jumlah filter = 32
6	Layer max pooling
	Ukuran pool = 2x2
7	Layer fully-connected
	Unit = 512
8	Layer aktivasi Rel-U
9	Layer fully-connected
	Unit = 3

Meskipun struktur CNN yang ditunjukkan pada tabel 8 tidak memberikan performansi terbaik pada pengujian terakhir, namun CNN ini memiliki tingkat overfit (perbedaan performansi antara training dan testing) yang paling rendah di antara skenario-skenario yang lain. Selain itu, model CNN ini juga membutuhkan waktu yang paling cepat dalam proses training, dan juga karena jumlah filter yang lebih sedikit pada masing-masing layer konvolusi, maka model ini juga membutuhkan resource perangkat keras yang lebih sedikit.

Struktur CNN tersebut lalu diuji untuk mengklasifikasikan sebuah dataset wajah baru yang terdiri dari 10 citra wajah untuk masing-masing ras. Semua citra wajah yang ada pada dataset tersebut adalah citra yang didapatkan secara manual melalui pencarian gambar dan juga dipotong dan diubah ukurannya secara manual. Tabel 9 menunjukkan hasil akurasi klasifikasi secara keseluruhan dan juga masing-masing ras.

Tabel 9. Tabel hasil klasifikasi struktur CNN terbaik

	Total	Ras Caucaso- id	Ras Negroid	Ras Mongo- loid
Akurasi klasifikasi	76.67%	80.00%	90.00%	60.00%
Jumlah klasifikasi be- nar	23	8	9	6

Hasil yang ditunjukkan pada tabel 9 memperlihatkan bahwa akurasi klasifikasi untuk masing-masing ras tidak menunjukkan perbedaan yang signifikan dengan hasil klasifikasi yang didapatkan pada proses-proses sebelumnya. Namun, hal ini tidak berlaku bagi ras mongoloid, dimana akurasi yang dihasilkan hanya sekitar 60%. Hal ini kemungkinan disebabkan karena fitur-fitur yang ada pada beberapa citra wajah ras mongoloid tidak berbeda jauh dengan ras caucasoid, salah satunya adalah warna kulit, dimana beberapa citra wajah pada ras mongoloid memiliki warna kulit yang lebih terang sehingga lebih mirip dengan warna kulit ras caucasoid. Oleh karena itu, pada tabel 9 hasil klasifikasi total juga ikut terpengaruh oleh akurasi rendah yang dihasilkan pada ras mongoloid.

Dari penjelasan di atas, dapat diambil kesimpulan bahwa CNN terbaik yang dihasilkan memang merupakan struktur CNN yang optimal. Struktur CNN tersebut dapat digunakan untuk mengklasifikasikan ras dari citra wajah apapun yang disediakan, selama citra wajah tersebut telah diubah sesuai dengan ketentuan, yaitu dipotong hanya pada bagian wajah saja.

5. Kesimpulan

Berdasarkan penelitian yang telah dilakukan dan sesuai dengan tujuan yang telah disebutkan pada bab pendahuluan, ada beberapa kesimpulan yang dapat diambil

1. Untuk mengklasifikasikan dataset yang terdiri ras mongoloid, caucasoid, dan negroid, lebih baik apabila tidak dilakukan pengubahan warna menjadi grayscale pada citra, dengan perbandingan akurasi training 94.18% berbanding 92.64%, dan akurasi testing 90.391% berbanding 90.065%

2. Struktur CNN beserta kombinasi parameter terbaik berdasarkan pengujian beberapa skenario yang telah ditentukan adalah struktur CNN sesuai dengan yang ditunjukkan pada tabel 8, dengan akurasi training sebesar 91.63% dan akurasi testing sebesar 89.739%, dan akurasi terhadap dataset baru sebesar 76.67%

Selain itu, ada beberapa hal yang dapat dilakukan untuk meningkatkan performansi CNN yang didapatkan. Salah satunya adalah dengan menambah ras yang ada pada dataset beserta kumpulan citra wajah untuk ras tersebut lalu melatih CNN dengan dataset yang telah dimodifikasi. Dengan menambah ras pada dataset, maka CNN tersebut juga dapat digunakan untuk mengklasifikasikan ras-ras baru yang ditambahkan.

Selain dengan menambahkan ras baru, peningkatan pada CNN juga dapat dilakukan dengan menambah skenario pada tahap pengujian. Dengan menambahkan skenario pada tahap pengujian, maka variasi CNN yang didapatkan juga akan semakin banyak, dan memungkinkan untuk didapatkan struktur CNN yang memberikan performansi lebih baik.

Daftar Pustaka

- [1] L. Agassiz. *The Diversity of Origin of the Human Races*. Boston Christian Examiner, 1850.
- [2] A. Ahmed, W. Xu, K. Yu, and Y. Gong. Training hierarchical feed-forward visual recognition models using transfer learning from pseudo-tasks. *Computer Vision – ECCV 2008: 10th European Conference on Computer Vision*, 3:69–82, 2008.
- [3] L. M. Alcoff. *Visible Identities : Race, Gender and the Self*. Oxford University Press, 2006.
- [4] M. Demirkus, S. Guler, and K. Garg. Automated person categorization for video surveillance using soft biometrics. *Proceedings of SPIE - The International Society for Optical Engineering*, 2010.
- [5] S. Fu, H. He, and Z.-G. Hou. Learning race from face : A survey. *IEEE*, 2014.
- [6] S. Haslanger. *Gender and race: (what) are they? (what) do we want them to be?* Massachusetts Institute of Technology, 2000.
- [7] J. T. Kubota, M. R. Banaji, and E. A. Phelps. *The neuroscience of race*. National Center for Biotechnology Information, 2012.
- [8] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *IEEE*, 1998.
- [9] H. Lin, H. Lu, and L. Zhang. A new automatic recognition system of gender, age and ethnicity. 2006 6th World Congress on Intelligent Control and Automation, 2:9988–9991, 2006.
- [10] V. M. Mays, N. A. Ponce, D. L. Washington, and S. Cochran. Classification of race and ethnicity: Implications for public health. *Annu Rev Public Health*, 24:83–110, 2003.
- [11] B. S. Ryan. Alleviating own-race bias in cross-racial identifications. *Washington University Jurisprudence Review*, 8:124–128, 2015.