

Klasifikasi Gender dan Usia Berdasarkan Citra Wajah Manusia Menggunakan Convolutional Neural Network

Ri Munarto¹, Rian Fahrizal¹, Ardian Darma¹

¹Jurusan Teknik Elektro, Fakultas Teknik, Universitas Sultan Ageng Tirtayasa, Cilegon, Banten.

Informasi Artikel

Naskah Diterima : 8 Nopember 2021

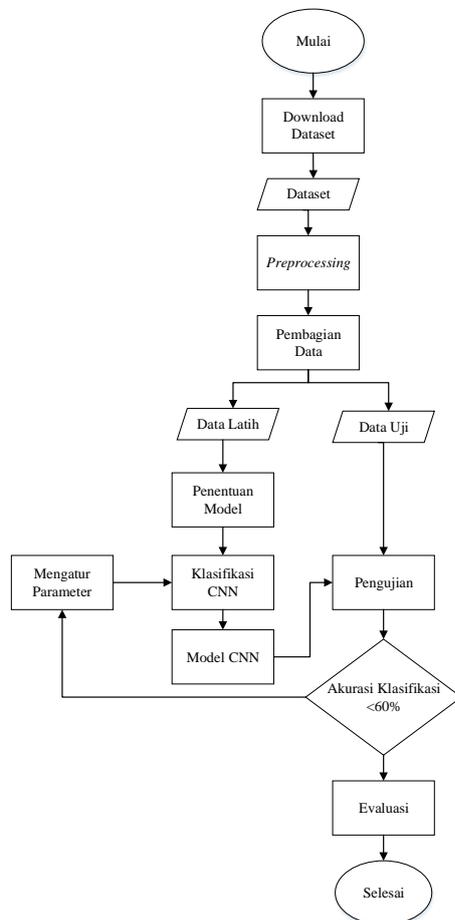
Direvisi : 9 Nopember 2021

Disetujui : 16 Nopember 2021

doi: 10.36055/setrum.v10i2.12991

*Korespodensi Penulis:
rim_munarto@yahoo.com

Graphical abstract



Abstract

The focus of this study is to create a classification model to predict gender and age using human facial images. The classification method of this study uses a convolutional neural network. The purpose of this study is to build a classification model that can predict gender and age from existing human facial image data. To do this, there are several processes carried out, starting from data collection, preprocessing, data sharing, training and data testing. The data used comes from UTKFace which has 23,708 human face image data with two labels, namely gender and age. The gender label consists of male and female, and the age label consists of ages 0-20, 21-40, and 41+ years. This study concludes that the convolutional neural network method can classify gender with 89.18% accuracy performance, 89.28% precision, and 89.17% sensitivity. And in classifying age, it can produce 74.14% accuracy performance, 78.07% precision, and 70.65% sensitivity.

Keywords: Gender, Age, CNN

Abstrak

Fokus pada studi ini adalah untuk membuat sebuah model klasifikasi untuk memprediksi gender dan usia menggunakan citra wajah manusia. Metode klasifikasi studi ini menggunakan convolutional neural network. Tujuan dari studi ini adalah membangun model klasifikasi yang dapat memprediksi gender dan usia dari data citra wajah manusia yang sudah ada. Untuk melakukannya, terdapat beberapa proses yang dilakukan, mulai dari pengumpulan data, pra proses, pembagian data, pelatihan dan pengujian data. Data yang digunakan berasal dari UTKFace yang memiliki 23.708 data citra wajah manusia dengan dua label, yaitu gender dan usia. Label gender terdiri dari laki- laki dan perempuan, serta label usia terdiri dari usia 0-20, 21-40, dan 41+ tahun. Penelitian ini menyimpulkan bahwa metode convolutional neural network dapat mengklasifikasi gender dengan performa akurasi 89,18%, presisi 89,28%, dan sensitivitas 89,17%. Dan dalam mengklasifikasi usia dapat menghasilkan performa akurasi 74,14%, presisi 78,07%, dan sensitivitas 70,65%.

Kata kunci: Gender, Usia, CNN

© 2021 Penerbit Jurusan Teknik Elektro UNTIRTA Press. All rights reserved

1. PENDAHULUAN

Data dan otomasi sangat dibutuhkan secara luas pengaplikasiannya. Demikian pula, perkiraan usia dan jenis kelamin juga merupakan kumpulan data yang dapat diklasifikasi secara otomatis yang sangat dibutuhkan. Teknik klasifikasi usia dan jenis kelamin dapat meningkatkan kemampuan persepsi dan interaksi komputer. Berbagai aplikasi untuk klasifikasi usia dan jenis kelamin termasuk *human computer interaction*(HCI) dapat memberikan layanan yang sesuai dan disesuaikan kepada pengguna berdasarkan jenis kelamin dan usia. Data yang dikumpulkan hanya gambar wajah dan tidak

perlu mengumpulkan usia dan gender seseorang karena sifatnya yang sangat pribadi. Ini juga dapat digunakan untuk akurasi verifikasi atau otentikasi berdasarkan teknik biometrik. Klasifikasi usia dapat mengontrol masuknya konten yang tidak diinginkan di televisi dan internet dari anak – anak. Dalam pekerjaan seperti polisi, militer, dan pekerjaan pemerintah memerlukan perkiraan usia selama perekrutan dan saat usia pensiun. Dalam keamanan dan pengawasan membatasi akses anak – anak dari mesin penjual otomatis dewasa seperti alkohol, rokok dan film situs dewasa.

Klasifikasi gender dan usia pada wajah memiliki banyak tantangan. Klasifikasi gender terbagi menjadi dua kelas yaitu laki – laki dan perempuan. Mudah bagi manusia untuk membedakannya, tidak mudah bagi mesin untuk mengklasifikasikannya. Untuk klasifikasi gender dapat mengenali pria atau wanita berdasarkan informasi tambahan seperti gaya rambut, bentuk wajah, aksesoris, dan fitur wajah. Sedangkan dalam perkiraan usia, belum memungkinkan untuk memprediksi usia yang sebenarnya sampai saat ini. Manusia juga masih sering salah dalam memprediksi usia secara aktual yang sebenarnya karena setiap wajah manusia memiliki sangat banyak keunikan. Jadi klasifikasi usia masih menggunakan pengelompokan rentang usia untuk memprediksi usia menggunakan citra wajah.[1]

Penelitian – penelitian sebelumnya terutama didasarkan pada fitur ekstraksi ciri yang terpisah dengan proses klasifikasi. Belakang ini deep model seperti convolutional neural network telah terbukti efektif untuk mengekstraksi fitur visual tingkat tinggi dan digunakan untuk verifikasi wajah. Penelitian ini memiliki tujuan mengikuti contoh sukses yang diberikan oleh sistem pengenalan wajah yang dijelaskan dalam beberapa tahun terakhir menunjukkan kemajuan luar biasa dapat dibuat dengan menggunakan convolutional neural network (CNN).[2] Dengan kesuksesan deep learning model dalam berbagai masalah computer vision dalam beberapa tahun terakhir,[3] penelitian terbaru tentang prediksi usia dan jenis kelamin sebagian besar bergeser ke model berbasis deep neural network. Dengan membuat arsitektur jaringan sederhana, yang dirancang dengan dataset gambar wajah manusia yang dilabeli usia dan gender yang akurat.

Penelitian tentang convolutional neural network untuk klasifikasi gender dan usia pada citra wajah manusia telah beberapa kali dilakukan. Perbedaannya adalah dataset, parameter, raproses, arsitektur model yang digunakan, dan hasil akurasi yang didapatkan dalam menggunakan convolutional neural network untuk mengolah dan menganalisis gambar, diantaranya adalah sebagai berikut. Penelitian pertama adalah gender and age classification based on human features to detect illicit activity in suspicious sites.[4] Penelitian kedua age and gender classification using multiple convolutional neural network.[5] Penelitian ketiga automatic age and gender classification using supervised apperance model.[6] Penelitian keempat age and gender classification using convolutional neural network.[7] Penelitian kelima memprediksi usia dan jenis kelamin menggunakan convolutional neural network.[8]

Berdasarkan latar belakang diatas, penelitian ini dilakukan untuk mengimplementasikan convolutional neural network. Fokus penelitian ini adalah mengklasifikasikan gender dan usia berdasarkan citra wajah manusia. Dataset yang digunakan merupakan kumpulan citra wajah manusia yang telah dilabeli gender dan usianya untuk dilakukan proses training menggunakan convolutional neural network. Hasil akhir yang diinginkan adalah sistem dapat mengklasifikasikan citra wajah manusia menurut kelas usia dan gender dengan baik.

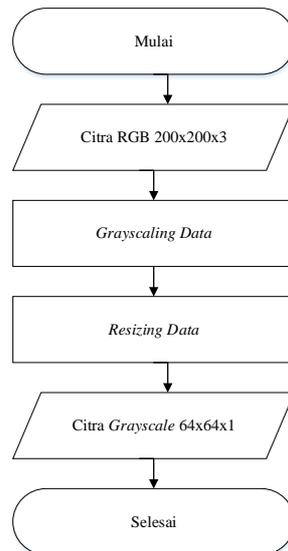
2. METODE PENELITIAN

2.1 Dataset

Dataset yang digunakan pada sistem ini adalah UTKFace dataset. Dataset ini didapatkan di website <http://susanqq.github.io/UTKFace/>. Dataset ini merupakan input data yang akan digunakan untuk data latih dan data uji. Dari data ini model CNN dapat mempelajari fitur dari gambar wajah manusia. Citra dalam dataset ini adalah citra wajah manusia yang memiliki dimensi 200x200x3.

2.2 Preprocessing

Sebelum dilakukan proses klasifikasi CNN, citra terlebih dahulu dilakukan *preprocessing*. *Preprocessing* merupakan perlakuan terhadap citra sebelum training. Citra perlu dikondisikan agar sesuai dengan tujuan dan kebutuhan sebelum training dan testing klasifikasi *convolutional neural network*.



Gambar 2.1 Tahapan Preprocessing

Tahapan preprocessing yang dilakukan terbagi menjadi dua tahap. Tahapan preprocessing yang pertama dilakukan adalah mengubah warna citra dari RGB yang memiliki 3 kanal menjadi grayscale yang memiliki 1 kanal sehingga membuat dimensi citra menjadi 200x200x1. Selanjutnya di tahap ke dua dari preprocessing adalah mengubah ukuran piksel citra dari 200x200x1 menjadi 64x64x1.

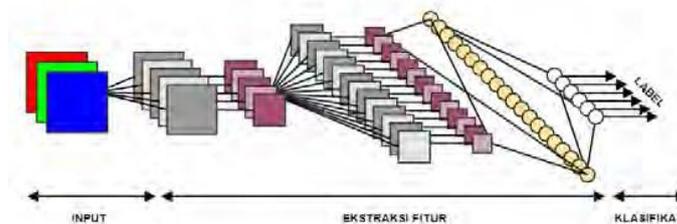
2.3 Pembagian Data

Tahapan selanjutnya setelah dilakukan preprocessing, semua dataset dibagi menjadi data latih dan data uji. Pembagian data dilakukan menjadi 80% untuk data latih dan 20% untuk data uji. Setelah didapatkan model terbaik, barulah model tersebut digunakan pada data uji kemudian dapat dilihat akurasi dari model yang dihasilkan.

2.4 Pembangunan Model

Struktur model CNN berpengaruh pada kualitas model. Pada CNN jenis layer lebih bervariasi dibandingkan dengan MLP sehingga arsitektur jaringan memiliki banyak variasi. Ukuran jaringan (kedalaman dan lebar) adalah hyperparameter jaringan yang perlu di tuning terhadap dataset yang digunakan.

Pada CNN terdapat dua bagian utama dari arsitektur model yaitu bagian ekstraksi fitur dan bagian klasifikasi. Bagian ekstraksi fitur berfungsi untuk mentransformasi data input sehingga dapat diklasifikasi dengan baik. Hal tersebut dilaksanakan dengan dengan mempropagasi data melalui berbagai transformasi linear maupun non linear. Bagian klasifikasi berfungsi untuk mereduksi data sehingga dimasukkan ke dalam classifier.



Gambar 2.2 Pembagian Arsitektur CNN

1. Input

Input yang digunakan adalah citra yang telah dilakukan grayscale dan resize pada tahap preprocessing. Setelah citra input memiliki channel grayscale yang sama dan ukuran citra yang sama, selanjutnya citra siap untuk di training dan di ekstraksi fitur.

2. Ekstraksi Fitur

Ekstraksi fitur yang berfungsi untuk mengambil fitur atau ciri – ciri dari gambar. Ekstraksi fitur pada metode convolutional neural network menggunakan feature learning yaitu ekstraksi fitur ikut di training bersama layer klasifikasi. Ekstraksi fitur pada convolutional neural network terdiri dari convolutional layer, max pooling, dan activation function(ReLU).

a. Convolutional Layer

Convolutional layer melakukan operasi konvolusi untuk mentransformasi data secara linear berdasarkan informasi data spasial, lalu melakukan transformasi non linear menggunakan fungsi aktivasi. Karena mengolah data dua dimensi, input dari layer tersebut berukuran 4 dimensi, yaitu (jumlah_data, jumlah_kernel_output, tinggi_data, lebar_data).

Layer tersebut memiliki parameter bobot dan bias. Bobot terdiri dari banyak kernel dan memiliki ukuran empat dimensi yaitu (jumlah_kernel_output, jumlah_kernel_input, tinggi_kernel, lebar_kernel), sedangkan bias memiliki ukuran satu dimensi saja yaitu (jumlah_kernel_output). Bobot digunakan dalam melakukan konvolusi, sedangkan bias ditambahkan setelah operasi konvolusi. Hyperparameter yang digunakan dalam konvolusi adalah ukuran padding. Jika padding bernilai satu, maka citra akan diberi tambahan nilai nol berdimensi satu di sekeliling citra ketika melalui proses konvolusi. Hal tersebut berpengaruh pada ukuran citra output.

b. Max Pooling Layer

Layer tersebut adalah sebuah subsampling layer yang bertujuan mereduksi ukuran data. Citra akan dibagi menjadi beberapa pool, lalu akan dikalkulasi satu nilai dari tiap pool untuk menghasilkan data output. Dalam kasus ini, diambil nilai terbesar dari tiap pool.

Seluruh subsampling layer tidak memiliki parameter bobot sehingga tidak memerlukan inisialisasi. Yang menjadi hyperparameter dalam max pooling layer adalah pool (pool_size), serta jarak antar pool (pool_stride). Pooling dilakukan pada tiap feature maps pada data.

3. Layer Klasifikasi

Layer klasifikasi merupakan tempat dimana akan terjadinya klasifikasi yang akan menentukan kelas dari citra yang diklasifikasi. Susunan layer akan berbeda – beda sesuai dengan arsitektur model yang digunakan. Pada penelitian ini menggunakan beberapa komponen utama antara lain fully connected layer, dropout, dan activation function (sigmoid).

a. Fully Connected Layer

Fully connected layer melakukan operasi perkalian matriks untuk mentransformasi data secara linear. Namun berbeda dengan convolutional layer, layer tersebut tidak menyimpan informasi spasial dan input harus berukuran dua dimensi yaitu [jumlah_data, data]. Oleh karena hal tersebut, bobot dari fully connected layer hanya berukuran dua dimensi saja, yaitu (jumlah_kernel_output, jumlah_kernel_input).

b. Dropout Layer

Dropout layer adalah layer yang berfungsi melakukan regularisasi pada CNN dengan mematikan neuron secara acak dengan kemungkinan drop_rate pada tiap iterasi pelatihan. Ketika tahap pengujian, dropout dinonaktifkan untuk mendapatkan hasil menyeluruh, namun output dari masing – masing neuron yang menggunakan dropout dikalikan dengan drop_rate agar outputnya proporsional dengan ketika tahap pelatihan. Dalam tugas akhir ini akan diterapkan inverted dropout dimana penyesuaian skala output dilakukan pada tahap pelatihan dan bukan pada tahap pengujian. Hal tersebut dilakukan agar model dapat bekerja lebih cepat pada tahap pengujian.

c. Sigmoid Layer

Sigmoid layer berbeda dengan layer lainnya karena harus berada pada ujung jaringan. Layer tersebut menerima input sejumlah kelas pada klasifikasi, lalu menghitung cross entropy untuk tiap kelas. Prediksi kelas dilakukan dengan mengambil kelas dengan nilai cross entropy tertinggi.

Dalam fitting model, digunakan dua fungsi yang berhubungan dengan nilai cross entropy. Pada tahap pelatihan, perhitungan cross entropy loss untuk menghitung loss hasil propagasi data. Pada tahap pengujian, digunakan perhitungan error yang membandingkan label prediksi terhadap label yang benar.

2.5 Pelatihan dan Pengujian

Proses pelatihan dan pengujian menggunakan metode Convolutional Neural Network. Pada tahap ini data yang dikumpulkan dalam bentuk citra dibagi menjadi data latih dan data uji.

Training dilakukan menggunakan data latih dan testing dilakukan menggunakan data uji. Selanjutnya data latih di training dan didapatkan hasil nilai loss dan akurasi training. Apabila akurasi training masih belum mendapatkan hasil yang diinginkan maka perlu dilakukan pengaturan ulang parameter atau perubahan desain model CNN untuk meningkatkan akurasi dan mengurangi nilai loss. Apabila nilai akurasi yang diinginkan sudah tercapai, maka bisa dilakukan pengujian.

Pengujian dilakukan dengan memprediksi citra yang belum pernah di latih yaitu citra data uji menggunakan model klasifikasi CNN yang telah dibuat. Setelah dilakukan pengujian dilakukan analisis evaluasi performa pengujian. Evaluasi performa model dalam mengklasifikasi data uji dapat dianalisis dari akurasi, sensitivity, dan presisi yang didapat dari confusion matrix.

2.6 Peralatan Penelitian

Penelitian ini menggunakan perangkat keras, perangkat lunak, dan fasilitas penunjang.

1. Google Colab (GPU, RAM 12.72GB, DISK 68.40GB)
2. Tensorflow versi 2.3.0
3. Keras versi 2.4.3
4. Pandas versi 1.0.5
5. Numpy versi 1.18.5
6. Python 3
7. Matplotlib
8. Laptop Lenovo Z40

2.7 Waktu dan Tempat Penelitian

Penelitian ini dilakukan sejak bulan Desember 2019 sampai dengan bulan Januari 2021. Bertempat di Fakultas Teknik Universitas Sultan Ageng Tirtayasa, Cilegon, Banten.

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Dataset yang digunakan adalah dataset UTKFace. Dataset ini memiliki sekitar 20.000 citra wajah manusia. Setiap citra wajah manusia ini dibedakan dengan cara dilabeli gender dan usianya. Pelabelan pada setiap file gambar ditulis pada nama file dengan format usia_gender_namafile.jpg. Label pada dataset sesuai dengan usia dan gender pada setiap citra yaitu rentang usia dari 0-116 tahun dan untuk gender terdiri dari laki – laki dan perempuan. Penulisannya untuk usia ditulis sesuai angka usia dari wajah citra tersebut, untuk label gender ditulis 0 untuk pria dan 1 untuk wanita. Citra pada dataset ini memiliki resolusi sebesar 200 x 200 piksel dan memiliki 3 kanal warna yang berarti gambar pada dataset ini memiliki warna RGB (red, green, dan blue). Sampel dataset UTKFace ditampilkan pada Gambar 3.2.



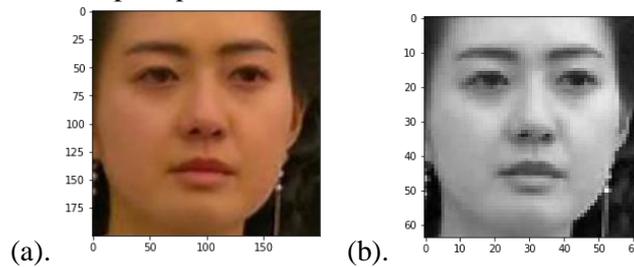
Gambar 3.2 Sampel Dataset UTKFace

Pemilihan penggunaan dataset UTKFace didasarkan pada pertimbangan jumlah banyaknya dataset yang memiliki banyak data. Dataset ini memiliki label yang cukup lengkap untuk kelas gender dan usia, yaitu label laki – laki dan perempuan untuk gender dan label 0-116 tahun. Setiap gambar yang sudah dilabeli pada dataset UTKFace diharapkan dapat mewakili setiap kelas yang diinginkan, sehingga pembuatan sistem klasifikasi gender dan usia berdasarkan citra wajah manusia dapat dilakukan dengan baik.

3.2 Preprocessing

1. Grayscale Citra

Tahap ini mengubah citra berwarna atau RGB yang memiliki 3 kanal warna Red, Green, dan Blue menjadi citra grayscale yang memiliki 1 buah kanal warna sehingga yang ditampilkan hanya nilai intensitas warna keabuan seperti pada

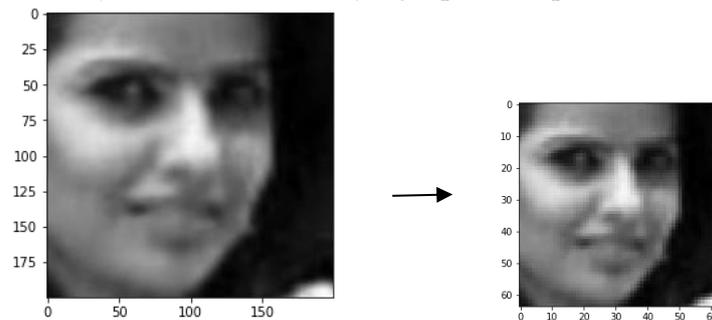


Gambar 3.2 Manipulasi Citra (a). Citra RGB, (b). Citra Grayscale

Proses grayscale menyederhanakan kanal yang sebelumnya pada RGB memiliki 3 kanal diubah menjadi 1 kanal saja. Hal ini dilakukan untuk menyederhanakan kompleksitas citra pada ekstraksi fitur di convolutional layer. Diharapkan dapat meningkatkan akurasi dan menyingkat waktu pada tahap training.

2. Resize Citra

Proses resize citra dilakukan untuk mengubah ukuran piksel pada semua citra pada dataset. Ukuran citra diubah menjadi ukuran 64x64 piksel. Proses resize ditujukan untuk menyeragamkan ukuran citra serta untuk menyesuaikan kebutuhan yang diperlukan pada klasifikasi CNN.



Gambar 3.3 Resize Citra

3.3 Membagi dataset untuk dilatih dan diuji

Tahap selanjutnya adalah dataset yang telah ada akan dibagi menjadi data latih dan data uji menggunakan model sklearn yang sudah ada pada library Python. Pada penelitian ini digunakan library sklearn karena lebih sedikit baris kode dibanding library lainnya seperti tensorflow atau numpy. Library sklearn juga biasa digunakan untuk clustering atau mengelompokkan data dengan karakteristik yang sama ke suatu wilayah yang sama dan data dengan karakteristik yang berbeda ke wilayah yang lain.

Dimasukan fungsi `train_test_split` pada program guna memisahkan array atau matriks menjadi data latih dan data test secara acak. Pada proses ini digunakan `test_size=0.2`, yang berarti pembagian data testing adalah 30% dari seluruh dataset.

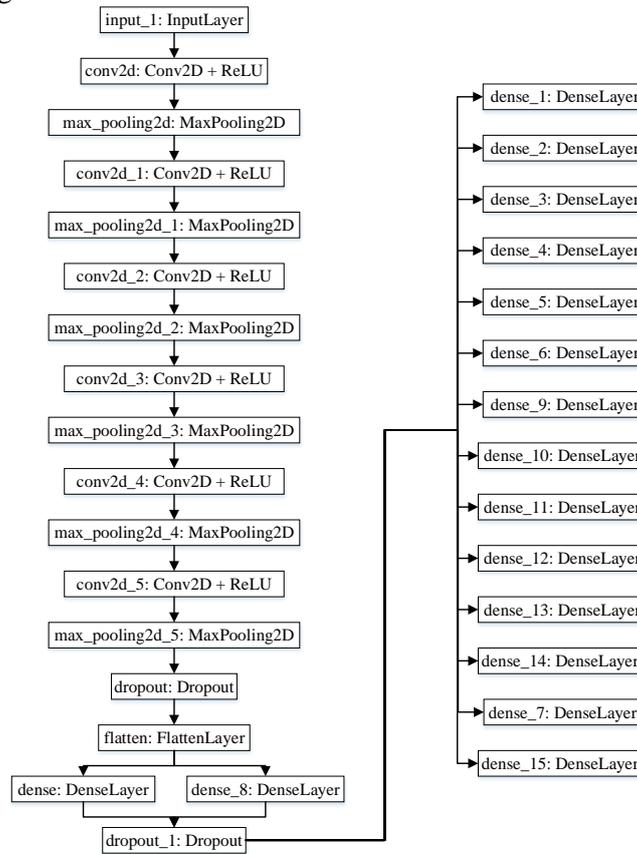
Tabel 3.1 Pembagian Shape

<code>x_train.shape</code>	<code>(18966, 64, 64, 1)</code>
<code>x_test.shape</code>	<code>(4742, 64, 64, 1)</code>
<code>y_train.shape</code>	<code>(18966, 2)</code>
<code>y_test.shape</code>	<code>(4742, 2)</code>

3.4 Arsitektur Jaringan

Setelah dilakukan preprocessing dan pembagian dataset latih dan uji, selanjutnya dataset siap untuk dilatih. Selanjutnya penentuan model yang digunakan, dimana disini menggunakan model

convolutional neural network yang didesain sendiri untuk mendapatkan hasil yang optimal. Dalam algoritma CNN pembentukan arsitektur jaringan dapat mempengaruhi hasil dan akurasi model. Berikut rancangan arsitektur jaringan CNN yang telah didesain untuk mendapatkan hasil yang baik dalam mengklasifikasi gender dan usia.



Gambar 3.4 Arsitektur Model CNN

Arsitektur model yang telah dibuat dapat dilihat pada Gambar 3.4. Dalam proses training diperlukan tahapan perlakuan sedemikian rupa untuk sistem dapat mengenali gambar sesuai labelnya dengan baik. Pembuatan arsitektur model menggunakan acuan pada metode convolutional neural network (CNN). Secara garis besar arsitektur CNN terdiri dari dua bagian besar yaitu feature extraction layer dan fully connected layer. Feature extraction layer adalah layer yang digunakan untuk mengekstraksi ciri suatu input gambar. Pada umumnya feature extraction layer memiliki dua tahap yaitu convolutional layer dan pooling. Bagian yang kedua adalah fully connected layer merupakan layer yang digunakan untuk menerima nilai dari feature layer dan mengklasifikasikan nilai tersebut pada salah satu kelas tertentu. Tahapan pada fully connected layer adalah flatten dan dense layer. Dari garis besar arsitektur yang telah disebutkan, pengguna dapat memodifikasi sedemikian rupa sesuai dengan kebutuhan.

Berikut ini adalah penjelasan mengenai arsitektur model CNN yang dibuat. Pada proses pertama input diproses menggunakan convolutional layer untuk dilakukan proses konvolusi yang berguna untuk mengambil ciri dari gambar. Setelah dilakukan konvolusi dilakukan max pooling untuk mengambil nilai terbesar dari konvolusi dan memperkecil ukuran piksel. Kombinasi konvolusi dan max pooling ini dilakukan selama 6 kali. Setelah dilakukan konvolusi sebanyak 6 kali dan max pooling sebanyak 6 kali, dilakukan dropout untuk mengefisiensi proses dengan cara menonaktifkan beberapa neuron. Nilai dari max pooling paling terakhir akan masuk ke flatten untuk diubah menjadi vektor untuk dapat diklasifikasi di dense layer. Dense layer dibagi menjadi dua bagian karena klasifikasi yang digunakan adalah klasifikasi multilabel. Klasifikasi multilabel adalah klasifikasi untuk satu gambar terdapat dua atau lebih label dan prediksinya, dalam hal ini gender dan usia. Jadi pembagian dense layer dari dense 0 – 7 untuk usia dan dense 8-15 untuk gender. Urutannya adalah setelah dari flatten nilai vektor untuk prediksi usia dibagi ke dense 0 dan untuk prediksi gender ke

dense 8. Setelah itu dilakukan dropout sebelum disebar ke dense layer yang lain sesuai kategorinya dense 1-7 untuk usia dan 9-15 untuk gender.

3.5 Summary Model

Setelah arsitektur model dibuat, langkah selanjutnya adalah melatih data training menggunakan model yang telah dibuat. Dengan model yang telah dibuat diharapkan proses training dapat optimal dalam mengenali data sesuai label kelasnya. Dibawah ini akan diperlihatkan proses saat terjadinya training terjadi menggunakan desain model yang telah dibuat.

Tabel 3.2 Model Summary

No	Nama	Size	Parameter
1	Input	64x64x1	0
2	conv2d	64x64x16	160
3	max_pooling2d	32x32x16	0
4	conv2d_1	32x32x32	4640
5	max_pooling2d_1	16x16x32	0
6	conv2d_2	16x16x64	18496
7	max_pooling_2	8x8x64	0
8	conv2d_3	8x8x128	73856
9	max_pooling_3	4x4x128	0
10	conv_2d_4	4x4x256	295168
11	max_pooling_4	2x2x256	0
12	conv_2d_5	2x2x512	1180160
13	max_pooling_5	1x1x512	0
14	Dropout	1x1x512	0
15	Flatten	512	0
16	Dense	512	262656
17	dropout_1	Multiple	0
18	dense_1	256	131328
19	dense_2	128	32896
20	dense_3	64	8256
21	dense_4	32	2080
22	dense_5	16	528
23	dense_6	8	136
24	dense_8	512	262656
25	dense_9	256	131328
26	dense_10	128	32896
27	dense_11	64	8256
28	dense_12	32	2080
29	dense_13	16	528
30	dense_14	8	136
31	dense_7	3	27

32	dense_15	2	18
33	Total params: 2.448.285		
34	Trainable params: 2.448.285		
35	Non-trainable params: 0		

Model yang dibentuk pada gambar diatas dapat dijelaskan sebagai berikut. Pada input layer menggunakan gambar dengan dimensi 64 x 64 x 1, dapat diartikan ukuran gambar 64 x 64 piksel dan 1 channel yang berarti grayscale. Gambar ini yang akan digunakan dalam proses training. Pada konvolusi layer pertama memiliki 16 filter channel, sehingga ukuran gambar menjadi 64x64x16. Filter berukuran 3x3 akan digerakkan dengan 1 stride (langkah) ke seluruh bagian dari input gambar, mulai dari sudut kiri atas sampai ke kanan bawah. Setiap pergeseran filter pada input gambar dilakukan operasi dot atau perhitungan matematis. Penggunaan padding pada penelitian ini adalah same, yang dimaksud adalah penambahan atau pengurangan ukuran piksel dilakukan untuk memastikan ukuran piksel tetap sama pada sebelum dan sesudah konvolusi. Sehingga ukuran output tetap 64 x 64 piksel dan 16 channel. Berikut perhitungan konvolusinya sesuai parameter yang telah ditetapkan.

$$output = \frac{W - N + 2P}{S} + 1$$

$$output = \frac{64 - 3 + 2(1)}{1} + 1 = 64$$

Keseluruhan parameternya menjadi ((3 × 3 × 1)+1 bias) × 16 = 160 parameter. Begitu juga seterusnya untuk melakukan perhitungan ukuran gambar yang terbentuk. Setelah dilakukan konvolusi pertama selanjutnya dilakukan max pooling 1. Pada max pooling pertama dilakukan downsampling yaitu mengambil nilai terbesar dari setiap pikselnya dan membagi pikselnya sebanyak setengahnya menjadi 32x32x16. Kemudian dilanjutkan dengan melakukan konvolusi pada convolution layer kedua menggunakan filter 32 lapisan dan kernel 3 x 3 menggunakan aktivasi relu sehingga output dimensi menjadi 32x32x32 channel, sehingga keseluruhan parameternya menjadi ((3×3×16)+1 bias)×32=4640 parameter.

Setelah dilakukan konvolusi pertama dan kedua, kemudian hasil operasi ini dikenakan dengan fungsi aktivasi ReLU dan pooling. ReLU adalah fungsi aktivasi yang berfungsi mempertahankan nilai sebenarnya ketika nilai itu positif dan akan memberikan angka 0 untuk nilai negatif. Sehingga nilai yang ada dalam piksel tidak ada yang bernilai negatif, hanya positif saja. Pooling layer terdiri dari sebuah filter dengan ukuran dan stride tertentu akan bergeser ke seluruh area feature map. Pooling yang digunakan yaitu max pooling. Tujuan dari penggunaan pooling layer yaitu untuk mengurangi dimensi dari feature map (downsampling). Dengan menggunakan filter 2 x 2 dan stride 1 pada operasi max pooling, proses pooling membuat piksel diperkecil setengah ukuran semulanya. Gambar yang sebelumnya memiliki dimensi 32 x 32 x 32 piksel diubah ukuran pikselnya menjadi setengahnya yaitu 16 x 16 x 32 piksel.

Konvolusi pada convolution layer yang ketiga dengan kernel 3 x 3 dan filter 64 lapisan menggunakan aktivasi relu untuk mengambil nodes paling tinggi dan meneruskannya untuk convolution berikutnya sehingga output gambar menjadi 16 x 16 x 64. Keseluruhan parameternya menjadi ((3×3×32)+1 bias)×64 filter=18.496 parameter. Setelah dilakukan convolution tiga kali kemudian dilanjutkan untuk proses pooling yang ketiga yaitu dengan mengubah input shape 16 x 16 x 64 menjadi 8 x 8 x 64 dan tidak ada parameter pada proses ini.

Selanjutnya masuk ke konvolusi pada convolution layer keempat dengan kernel 3x3 dan filter 128 lapisan menggunakan fungsi aktivasi ReLU untuk menghilangkan nilai minus pada piksel dan mempertahankan nilai positif. Sehingga output gambar menjadi 8x8x128. Keseluruhan parameternya menjadi ((3×3×64)+1 bias)×128 filter=73.856 parameter. Setelah itu dilakukan proses max pooling yang keempat yaitu mengubah input shape 8x8x128 menjadi 4x4x128 dan tidak ada parameter pada proses ini.

Selanjutnya masuk ke konvolusi pada convolution layer kelima dengan kernel 3x3 dan filter 256 lapisan menggunakan fungsi aktivasi ReLU untuk menghilangkan nilai minus pada piksel dan mempertahankan nilai positif. Sehingga output gambar menjadi 4x4x256. Keseluruhan parameternya

menjadi $((3 \times 3 \times 128) + 1 \text{ bias}) \times 256 \text{ filter} = 295.168$ parameter. Setelah itu dilakukan proses max pooling yang keempat yaitu mengubah input shape $4 \times 4 \times 256$ menjadi $2 \times 2 \times 256$ dan tidak ada parameter pada proses ini.

Selanjutnya masuk ke konvolusi pada convolution layer keenam dengan kernel 3×3 dan filter 512 lapisan menggunakan fungsi aktivasi ReLU untuk menghilangkan nilai minus pada piksel dan mempertahankan nilai positif. Sehingga output gambar menjadi $2 \times 2 \times 512$. Keseluruhan parameternya menjadi $((3 \times 3 \times 256) + 1 \text{ bias}) \times 512 = 1.180.160$ parameter. Setelah itu dilakukan proses max pooling yang keempat yaitu mengubah input shape $2 \times 2 \times 512$ menjadi $1 \times 1 \times 512$ dan tidak ada parameter pada proses ini.

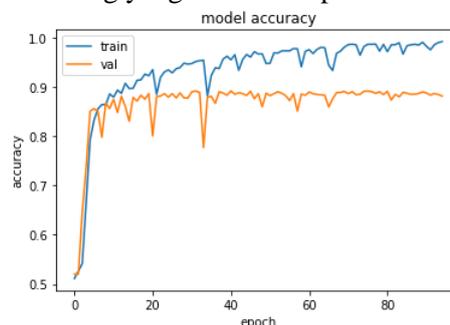
Setelah konvolusi dan pooling selesai selanjutnya dilanjutkan dengan dropout untuk menonaktifkan node yang tidak dibutuhkan untuk mengefisiensi proses learning. Setelah dropout dilakukan flatten untuk mengubah nilai menjadi nilai vektor. Setelah dilakukan flatten akan dibagi menjadi dua cabang dense layer, cabang pertama adalah Dense dan cabang kedua adalah dense_8. Dense merupakan cabang dense layer untuk klasifikasi usia dan cabang dense_8 merupakan cabang dense layer untuk klasifikasi gender. Parameter pada Dense menjadi $(512 + 1 \text{ bias}) \times 512 = 262.656$ parameter. Nilai parameter pada dense_8 juga sama dengan Dense karena berasal dari satu cabang dan input channel number yang sama yaitu 512, sehingga perhitungan parameter menjadi seperti berikut $(512 + 1 \text{ bias}) \times 512 = 262.656$ parameter. Setelah itu dilakukan dropout multiple untuk membentuk cabang baru ke setiap cabang dense layer, dense_1 – dense_7 adalah klasifikasi untuk usia dan dense_9 – dense_15 adalah klasifikasi untuk gender.

Pada dense layer bagian terakhir ini terdapat beberapa dense yang memiliki persamaan jumlah input channel number dan output channel number sehingga menghasilkan jumlah parameter yang sama. Dense_1 dan dense_9 memiliki parameter yang sama yaitu $(512 + 1 \text{ bias}) \times 256 = 131.328$ parameter. Dense_2 dan dense_10 memiliki parameter $(256 + 1 \text{ bias}) \times 128 = 32.896$ parameter. Dense_3 dan dense_11 memiliki parameter $(128 + 1 \text{ bias}) \times 64 = 8.256$ parameter. Dense_4 dan dense_12 memiliki parameter $(64 + 1 \text{ bias}) \times 32 = 2.080$ parameter. Dense_5 dan dense_13 memiliki parameter $(32 + 1 \text{ bias}) \times 16 = 528$ parameter. Dense_6 dan dense_14 memiliki parameter $(16 + 1 \text{ bias}) \times 8 = 136$ parameter. Ada perbedaan di dense layer paling terakhir yaitu pada dense_7 dan dense_15. Dense layer terakhir penentuan output channel number berdasarkan kelas yang diinginkan yaitu 3 pada dense_7 yang berarti 3 kelas pada klasifikasi usia dan 2 pada dense_15 yang berarti 2 kelas pada klasifikasi gender. Parameternya adalah $(8 + 1 \text{ bias}) \times 3 = 27$ parameter untuk dense_7. Parameter untuk dense_15 adalah $(8 + 1 \text{ bias}) \times 2 = 18$ parameter.

Total parameter yang dilatih dan berhasil dilatih adalah 2.448.285 parameter. Jumlah total parameter yang dilatih dan jumlah parameter yang berhasil dilatih adalah sama, maka tidak ada parameter yang gagal dilatih. Dengan kata lain jumlah parameter yang tidak terlatih adalah 0 perhitungan.

3.6 Hasil Training

Data training dilatih menggunakan arsitektur model desain convolutional neural network pada gambar 3.4. Setelah terbentuk model, dilakukan fit model untuk mendapatkan nilai akurasi dari data training dan data testing yang akan ditampilkan dalam bentuk grafik sebagai berikut.

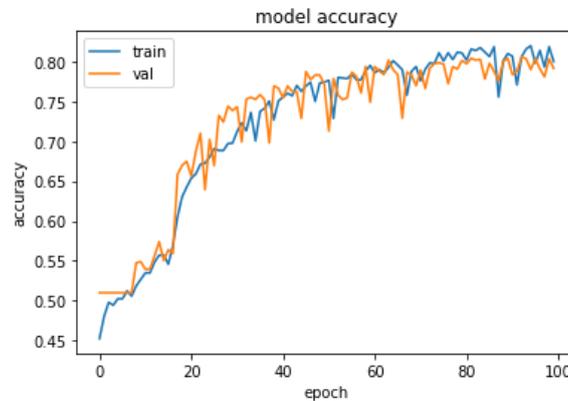


Gambar 3.5 Grafik Akurasi Training Gender

Dari Gambar 3.5 dapat terlihat grafik hasil training pada kelas gender dengan menggunakan 100 epoch yang berpengaruh pada besaran akurasi yang dihasilkan. Setelah dilakukan fit training sebanyak 100 epoch didapatkan akurasi sebesar 93,25% untuk kelas gender. Waktu yang dibutuhkan

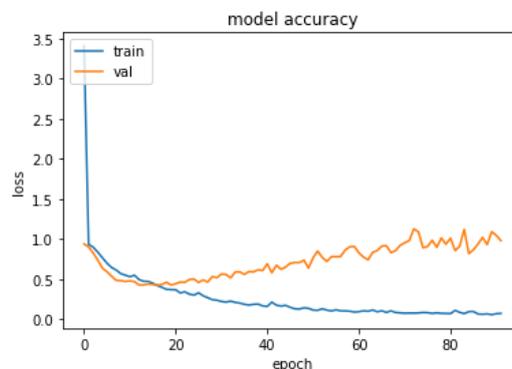
dalam menjalankan training 100 epoch adalah 288 detik. Banyaknya epoch mempengaruhi waktu training, semakin banyak jumlah epoch maka semakin lama waktu training.

Dari Gambar 3.5 terlihat grafik akurasi training semakin bertambah seiring banyaknya jumlah epoch. Banyaknya jumlah epoch dapat meningkatkan akurasi. Terlalu banyak jumlah epoch dapat menyebabkan overfitting, yaitu pada saat training sistem terlalu terpaku pada data training sehingga akurasi training sangat besar dan akurasi testing menjadi kecil. Sistem yang baik adalah sistem yang akurasi training dan testing memiliki akurasi yang tinggi dan saling mendekati



Gambar 3.6 Grafik Akurasi Training Usia

Dari Gambar 3.6 grafik akurasi training usia terus meningkat dari setiap epoch. Dari epoch 20 sampai selesai akurasi berada diatas 70% dan tidak sampai melebihi 80%. Tidak ada peningkatan atau penurunan akurasi yang signifikan setelah epoch ke 20. Berdasarkan gambar diatas akurasi gender dari training model mencapai 93,25% dan akurasi usia mencapai 74,32%. Proses training menggunakan input gambar sebesar 64 x 64 piksel. Proses training diatur menggunakan 100 epoch, pada epoch ke 95 proses training terhenti secara otomatis karena sudah tidak ada perubahan yang signifikan.



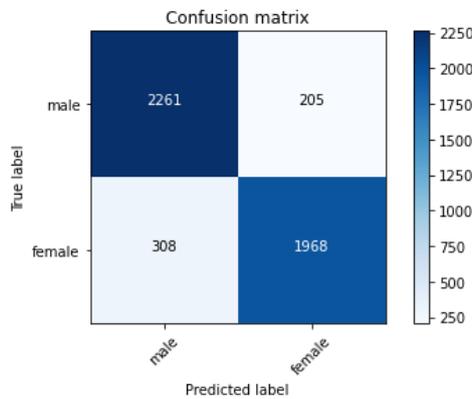
Gambar 3.7 Grafik Loss

Berdasarkan Gambar 3.7 grafik loss mengalami penurunan. Penurunan grafik loss ini merupakan hal baik karena nilai loss yang dihasilkan semakin kecil. Artinya sistem ini sudah baik karena memiliki loss sebesar 0,34%. Sistem dianggap baik apabila memiliki akurasi yang tinggi dan loss yang kecil.

Waktu yang dibutuhkan untuk training data latih tergantung pada epoch yang ditentukan. Epoch ditentukan dengan nilai 100, setelah mencapai epoch ke 92 proses training berhenti karena sistem akan berhenti training secara otomatis ketika tidak ada lagi perubahan yang signifikan pada akurasi dan loss. Waktu yang dibutuhkan untuk training adalah 288 detik dengan rincian 51 detik untuk epoch pertama dan 2 detik untuk setiap epoch selanjutnya. Waktu training pada epoch pertama lebih lama dari sebelumnya karena model baru pertama kali mengenali gambar, sedangkan pada epoch selanjutnya model hanya mengulangi mempelajari gambar yang sudah dipelajari sebelumnya dengan tujuan dapat mengenali data training lebih baik lagi untuk mencapai akurasi yang diinginkan.

3.7 Pengujian dan Evaluasi Model

Evaluasi model dilakukan untuk mengetahui kemampuan model dalam mengklasifikasi gambar. Evaluasi ini didapat dari hasil klasifikasi pada data testing. Untuk memudahkan dalam mengevaluasi model dalam mengklasifikasi data testing maka ditampilkan hasil klasifikasi dalam confusion matrix. Confusion matrix dapat memperlihatkan berapa banyak data yang terklasifikasi dengan benar dan yang terklasifikasi tidak benar.



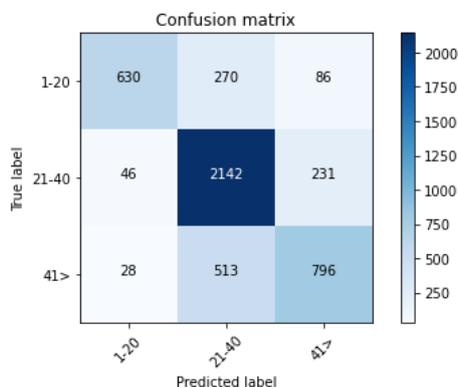
Gambar 3. 8 Confusion Matrix Gender

Dari hasil klasifikasi yang didapat pada gambar 4.8 dapat dilihat bahwa data testing yang diklasifikasi tidak seluruhnya mampu terklasifikasi sesuai dengan kategori labelnya. Pada data berlabel pria ada 2261 data yang dapat terklasifikasi sebagai pria. Sedangkan ada 205 data berlabel pria yang terklasifikasi sebagai wanita. Sama halnya pada data berlabel wanita yang terklasifikasi dengan benar ada 1968. Sedangkan ada 308 data berlabel wanita terklasifikasi sebagai pria.

Hasil dari nilai prediksi pada data testing yang benar menghasilkan nilai yang cukup besar dibandingkan dengan prediksi data testing yang salah. Dari perbandingan jumlah data yang terprediksi benar dan salah, tingkat keberhasilan sistem memprediksi data dengan benar mendapatkan akurasi sebesar 89,18%. Akurasi sangat bagus digunakan sebagai acuan performa jika jumlah data false negative dan false positive yang sangat mendekati atau simetris. Jika jumlahnya tidak mendekati maka sebaiknya menggunakan F1 Score. Evaluasi kinerja model dalam mengklasifikasi data testing mendapatkan nilai precision sebesar 89.28%, nilai recall sebesar 89.17%. Nilai precision, nilai recall, dan nilai akurasi saling mendekati. Maka dilakukan perhitungan untuk mencari nilai F1 Score.

$$F1\ Score = 2 \times (89,28\% \times 89,17\%) / (89,28\% + 89,17\%) = 89,27\%$$

Dari perhitungan diatas, F1score mendapatkan hasil 89,27%. Nilai F1score, akurasi, recall, dan precision semuanya simetris saling mendekati di angka 89%. Karena itu akurasi cukup akurat untuk menjadi acuan dalam mengklasifikasi gender di angka 89,18%. Akurasi gender yang didapatkan memiliki nilai yang lebih tinggi dari penelitian lain tentang klasifikasi usia dan gender berdasarkan citra wajah manusia menggunakan metode convolutional neural network yang telah dicantumkan pada kajian pustaka.



Gambar 3.9 Confusion Matrix Usia

Dari hasil tampilan Gambar 4.9 dapat dilihat bahwa data testing kategori usia tidak seluruhnya mampu terklasifikasi dengan benar sesuai labelnya. Pada data berlabel usia 1-20 tahun ada 630 data yang terklasifikasi dengan benar, 270 data terklasifikasi di kelas 21-40 tahun, dan 86 data terklasifikasi di kelas 41+. Pada data yang berlabel usia 21-40 tahun ada 2142 data terklasifikasi dengan benar, 46 data terklasifikasi sebagai kelas usia 1-20 tahun, dan 231 data terklasifikasi pada kelas usia 41+. Pada data yang berlabel kelas usia 41+ tahun ada 796 data yang terklasifikasi dengan benar, 28 data terklasifikasi pada kelas usia 1-20 tahun, dan 513 data terklasifikasi pada kelas usia 21-40 tahun.

Hasil dari nilai prediksi pada data testing yang benar menghasilkan nilai yang cukup besar dibandingkan dengan prediksi data testing yang salah. Dari perbandingan jumlah data yang terprediksi benar dan salah, tingkat keberhasilan sistem memprediksi data dengan benar mendapatkan akurasi sebesar 74,14%. Evaluasi kinerja model dalam mengklasifikasi data testing mendapatkan nilai precision sebesar 78,07%, nilai recall sebesar 70,65%. Nilai precision dan recall memiliki selisih sebesar 7,42%. Perlu dilakukan perhitungan F1score untuk dipertimbangkan menjadi acuan dalam tingkat keberhasilan model dalam mengklasifikasi usia.

$$F1\ Score = 2 \times (78,07\% \times 70,65\%) / (78,07\% + 70,65\%) = 74,17\%$$

Nilai yang didapat dari perhitungan F1 Score diatas adalah 74,17%. Nilai F1 Score dapat dipertimbangkan menjadi acuan performa model klasifikasi. Hal ini bisa dilakukan apabila false negative dan false positive data klasifikasi nilainya tidak saling mendekati. Nilai F1 Score adalah 74,17% dan nilai akurasi adalah 74,14%, keduanya saling mendekati dan dapat menjadi acuan performa klasifikasi usia.

Nilai yang didapatkan dari klasifikasi ini masih jauh dari angka 100%. Penelitian ini dilakukan semaksimal mungkin oleh peneliti untuk mencapai nilai yang paling optimal. Ada beberapa faktor yang dapat diperbaiki lagi untuk mendapatkan hasil yang lebih optimal. Pemilihan gambar pada dataset harus dapat mewakili setiap kelasnya dengan akurat. Gambar pada kelas gender harus dapat mewakili wajah wanita dan pria seutuhnya. Begitu juga pada kelas usia, gambar wajah harus dapat mewakili gambar wajah muda, dewasa, dan tua seutuhnya. Penggunaan parameter yang digunakan dapat lebih variatif dari yang dilakukan pada penelitian ini bahkan dapat menggunakan fitur yang tidak dilakukan dalam penelitian ini untuk meningkatkan akurasi yang lebih baik.

4. KESIMPULAN

4.1 Kesimpulan

Dari hasil penelitian yang sudah dilakukan untuk membuat sistem klasifikasi gender dan usia dengan metode convolutional neural network terhadap gambar wajah manusia dapat diterapkan dan didapatkan hasil sebagai berikut.

- a) Penggunaan metode convolutional neural network untuk klasifikasi gender dan usia berdasarkan gambar wajah manusia dapat dilakukan hingga proses pelatihan dan pengujian. Penggunaan CNN bisa mengklasifikasi gambar wajah manusia yang sangat rumit dengan akurasi yang cukup tinggi.
- b) Hasil evaluasi pengukuran kinerja untuk menguji tingkat keberhasilan model dalam mengklasifikasi gender dan usia telah dilakukan menggunakan confusion matrix dan telah didapat akurasi dari klasifikasi gender sebesar 89,18% dan akurasi klasifikasi usia sebesar 74,14%.

4.2 Saran

Rekomendasi dan saran yang dapat diberikan terkait pengembangan dan penggunaan sistem klasifikasi ini untuk penelitian yang akan datang adalah sebagai berikut.

- a) Melakukan pengembangan sistem dengan menambahkan kelas baru seperti ekspresi, ras maupun informasi lain yang dapat diambil dari gambar wajah manusia.
- b) Untuk memperoleh kinerja sistem yang lebih baik dapat menggunakan model pengembangan dari convolutional neural network seperti transfer learning atau recurrent neural network untuk memperoleh keakuratan hasil yang lebih optimal.

REFERENSI

- [1] G. Trivedi and N. N., “Gender Classification and Age Estimation using Neural Networks: A Survey,” *Int. J. Comput. Appl.*, vol. 176, no. 23, pp. 34–41, 2020, doi: 10.5120/ijca2020920251.
- [2] M. Zufar and B. Setiyono, “Convolutional Neural Networks Untuk Pengenalan Wajah Secara Real-Time,” *J. Sains dan Seni ITS*, vol. 5, no. 2, p. 128862, 2016, doi: 10.12962/j23373520.v5i2.18854.
- [3] H. G. Krizhevsky A, Sutskever I, “ImageNet Classification with Deep Convolutional Neural Networks,” *Proc. Twenty-Sixth Annu. Conf. Neural Inf. Process. Syst.*, pp. 1097–1105, 2012, doi: 10.1201/9781420010749.
- [4] E. A. Torres and M. Hernández-álvarez, “Gender and age classification based on human features to detect illicit activity in suspicious sites,” *2019 Int. Conf. Comput. Sci. Comput. Intell.*, pp. 416–419, 2019, doi: 10.1109/CSCI49370.2019.00081.
- [5] K. R. Hassan and I. H. Ali, “Age and Gender Classification using Multiple Convolutional Neural Network,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 928, no. 3, 2020, doi: 10.1088/1757-899X/928/3/032039.
- [6] A. M. Bukar, H. Ugail, and D. Connah, “Automatic age and gender classification using supervised appearance model,” *J. Electron. Imaging*, vol. 25, no. 6, p. 061605, 2016, doi: 10.1117/1.jei.25.6.061605.
- [7] M. K. Benkaddour, S. Lahlali, and M. Trabelsi, “Human Age and Gender Classification using Convolutional Neural Network,” *2020 2nd Int. Work. Human-Centric Smart Environ. Heal. Well-Being, IHSH 2020*, pp. 215–220, 2021, doi: 10.1109/IHSH51661.2021.9378708.
- [8] A. Zein, “Memprediksi Usia Dan Jenis Kelamin Menggunakan Convolutional Neural Networks,” *Sainstech J. Penelit. dan Pengkaj. Sains dan Teknol.*, vol. 30, no. 1, pp. 1–7, 2020, doi: 10.37277/stch.v30i1.727.