

Implementasi Transfer Learning pada Convolutional Neural Network dengan Arsitektur VGG dalam Klasifikasi Down Syndrome di Asia

Tarissa Rizky Salsabiila Uwar*, Christian Sri Kusuma Aditya

Fakultas Teknik, Program Studi Informatika, Universitas Muhammadiyah Malang, Kota Malang, Indonesia

Email: ^{1,*}tarissaurus@webmail.umm.ac.id, ²christianskaditya@umm.ac.id

Email Penulis Korespondensi: tarissaurus@webmail.umm.ac.id

Submitted: 21/03/2025; Accepted: 31/05/2025; Published: 01/06/2025

Abstrak—Deteksi dini *Down syndrome* merupakan langkah awal untuk mendukung intervensi dini dan memberikan edukasi dalam penanganan kesehatan bagi anak. Terlebih, *Down syndrome* memiliki karakteristik wajah tertentu, seperti mata, hidung, bibir, bentuk wajah, rambut, dan warna kulit, yang mampu dianalisis menggunakan *computer vision*. Penelitian ini bertujuan untuk klasifikasi *Down syndrome*, terkhusus di wilayah Asia. Pemilihan wilayah Asia dikarenakan wilayah tersebut termasuk dalam wilayah dengan SDI sedang/rendah. Model yang digunakan dalam penelitian ini adalah CNN berbasis arsitektur VGG16 dan VGG19 dengan mengimplementasikan *transfer learning* dan augmentasi. Augmentasi dilakukan untuk menyeimbangkan jumlah gambar antar kelas, sementara *transfer learning* digunakan untuk melatih model terlebih dahulu pada data ImageNet. Dataset yang digunakan terdiri dari dua kategori, *Down syndrome* dan *Healthy*. Hasil penelitian ini menunjukkan bahwa model VGG16 memiliki sensitivitas yang lebih tinggi dan mampu mengklasifikasikan lebih banyak kasus *Down syndrome*, namun memiliki kesalahan prediksi yang cukup besar. Sebaliknya, model VGG19 memiliki nilai spesifisitas yang lebih baik dan memiliki potensi kesalahan prediksi yang lebih kecil. Model terbaik dalam penelitian ini dipilih berdasarkan nilai *validation accuracy* tertinggi, di mana VGG19 mencapai akurasi sebesar 93% dalam iterasi terbaiknya, dan VGG16 memperoleh akurasi sebesar 91%. Dari hasil akurasi tersebut, model yang dirancang memiliki kemampuan yang optimal dalam melakukan klasifikasi *Down syndrome*, terkhusus di wilayah Asia, dengan tingkat kesalahan prediksi yang lebih rendah.

Kata Kunci: Augmentasi; Klasifikasi *Down Syndrome*; VGG16; VGG19; Wilayah Asia

Abstract—Early detection of Down syndrome is crucial for enabling early intervention and providing healthcare education for children. Down syndrome is associated with specific facial features, such as distinct characteristics of the eyes, nose, lips, face shape, hair, and skin color, which can be analyzed using computer vision techniques. This study aims to classify Down syndrome, especially in the Asian Region, which includes countries with medium/low SDI. The study proposed a CNN based on the VGG16 and VGG19 architectures by implementing transfer learning and augmentation. Augmentation is performed to balance the number of images between classes, while transfer learning is used to train the model first on ImageNet data. The dataset used consists of two categories, Down syndrome and Healthy. The results indicate that the VGG16 model has higher sensitivity and is able to classify more cases of Down syndrome, but has a fairly large prediction error. However, VGG19 model has a better specificity value and has a smaller potential for prediction error. The best model in this study was selected based on the highest validation accuracy value, where VGG19 achieved an accuracy of 93% in its best iteration, and VGG16 achieved an accuracy of 91%. These findings suggest that the proposed models, particularly VGG19, exhibit optimal performance in classifying Down syndrome, especially in the Asian region, with a lower prediction error rate.

Keywords: Asian Region, Augmentation; Down Syndrome Classification; VGG16; VGG19

1. PENDAHULUAN

Down syndrome merupakan kelainan genetika yang mampu diidentifikasi sejak dalam kandungan melalui tes prenatal skrining atau pasca kelahiran dengan menganalisa jumlah kromosom yang disebut dengan kariotipe [1]. Kelainan tersebut dapat terjadi dikarenakan adanya abnormalitas pada jumlah kromosom seseorang. Pada kondisi normal, seseorang memiliki jumlah kromosom sebanyak 46 kromosom, yang diwarisi dari sepasang ayah dan ibu sejumlah 23 kromosom. Sedangkan, seseorang yang mengidap *Down syndrome* memiliki satu salinan kromosom 21 tambahan, sehingga jumlah kromosom yang dimiliki sejumlah 47 kromosom [2]. Penambahan jumlah salinan kromosom 21 dapat terjadi dikarenakan adanya kegagalan dalam pembelahan sel yang terjadi dalam fase perkembangan embrio [3]. Kondisi tersebut mampu mempengaruhi perkembangan fisik dan intelektual pengidapnya. Oleh karena itu, seseorang dengan kelainan *Down syndrome* dapat dikenali perbedaannya secara langsung melalui fitur wajahnya yang khas. Ciri-ciri fisik seseorang dengan kelainan *Down syndrome* seperti, mata yang miring ke atas dari sudut dalam ke sudut luar, telinga kecil dan memiliki posisi yang lebih rendah dari posisi normal, bentuk kepala yang datar, memiliki ukuran mulut yang kecil, dan hidung kecil dengan jembatan hidung yang datar [4].

Dilansir oleh *World Health Organization* (WHO), prevalensi *Down syndrome* dapat diperkirakan berkisar antara 1 dari 1.000 hingga 1.100 kelahiran hidup di seluruh dunia [2]. Variasi angka kelahiran *Down syndrome* di setiap wilayah berbeda-beda, termasuk di Asia. Berdasarkan informasi yang diberikan dalam penelitian yang dilakukan oleh Lubis dkk., beberapa wilayah di Asia seperti Korea, Taiwan, dan Thailand, tercatat kasus kelahiran *Down syndrome* secara berurutan berkisar 5,03 kasus per 10.000 kelahiran, 7,92 kasus per 10.000 kelahiran, dan 1,21 kasus per 1.000 kelahiran [5]. Prevalensi *Down syndrome* dapat dipengaruhi oleh beberapa hal, yaitu *prenatal screening* dan *socio-demographic index* (SDI). Hal tersebut dikarenakan wilayah Asia termasuk dalam wilayah dengan SDI sedang atau rendah, dimana SDI merujuk pada pendapatan, pendidikan, dan tingkat kesuburan di wilayah tersebut. Wilayah dengan SDI tinggi, cenderung lebih berkembang dan maju, sehingga mampu untuk melakukan terminasi selektif. Sedangkan, wilayah dengan SDI sedang atau rendah, menunjukkan adanya peningkatan dalam

prevalensi *Down syndrome* [6]. Oleh karena itu, faktor regional menjadi dasar penting untuk memahami lebih lanjut dalam klasifikasi *Down syndrome* pada dataset citra wajah anak, terkhusus di wilayah Asia.

Penggunaan data dalam bentuk gambar atau citra telah dimanfaatkan dalam berbagai prosedur diagnostik klinis karena kemampuannya dalam menyajikan informasi visual yang mendukung proses identifikasi berbagai penyakit [7]. Adanya teknologi informasi dan komputer dapat mendukung penggunaan dalam identifikasi penyakit dengan mudah melalui data citra [8]. Beberapa tahun terakhir, penggunaan teknologi berbasis *machine learning* dan *deep learning* telah digunakan dalam klasifikasi penyakit melalui data citra, terkhusus *Down syndrome*. Dalam kasus klasifikasi *Down syndrome*, beberapa penelitian terdahulu telah berhasil menunjukkan performa yang baik dengan menggunakan model seperti SVM, mini-Xception, ResNet34, dan CNN. Namun, dalam penelitian sebelumnya hanya berfokus pada klasifikasi *Down syndrome* melalui pengenalan wajah tanpa mempertimbangkan faktor regional. Hal tersebut mampu mempengaruhi tingkat akurasi dari klasifikasi *Down syndrome*, terlebih beberapa wilayah memiliki karakteristik fisik yang berbeda-beda, termasuk wilayah Asia. Berdasarkan hasil penelitian sebelumnya, CNN memiliki tingkat akurasi tertinggi diantara SVM, mini-Xception, dan ResNet34, dengan nilai akurasi sebesar 91,48% [9]. Penelitian lain yang menggunakan model serupa juga berhasil mencapai akurasi tinggi dengan nilai sebesar 90% [10] dan 91% [11]. Namun, kedua penelitian tersebut tidak hanya berfokus pada klasifikasi *Down syndrome* saja, melainkan mencakup berbagai sindrom lainnya, sehingga belum ada analisis *Down syndrome* secara mendalam. Selanjutnya, ResNet34 memiliki akurasi yang lebih unggul dibandingkan dengan mini-Xception dan SVM, yaitu sebesar 87,88% [12] disusul dengan mini-Xception sebesar 74,8% dan SVM sebesar 66,20% [13]. Berdasarkan hasil dari penelitian sebelumnya, penggunaan model *deep learning*, terlebih CNN, telah menunjukkan potensi besar dalam klasifikasi *Down syndrome* melalui analisis gambar. Potensi tersebut menunjukkan adanya kemampuan yang sangat berharga dalam pengaturan klinis untuk deteksi dan penanganan dini [14]. Oleh karena itu, penulis merekomendasikan untuk menggunakan model CNN dalam klasifikasi *Down syndrome* di wilayah Asia.

Convolutional Neural Network atau CNN merupakan salah satu jenis arsitektur *deep learning* yang sangat efektif dalam memproses informasi visual, sehingga cocok digunakan untuk tugas-tugas pengenalan wajah. Selanjutnya, CNN memiliki beberapa lapisan yang mampu secara otomatis belajar mendeteksi berbagai fitur dalam gambar, mulai dari tepi dasar hingga struktur yang lebih kompleks [15]. Dengan memanfaatkan penggunaan lapisan yang ada di dalamnya, CNN mampu memproses input pada gambar dan membantu dalam menemukan pola visual seperti tepi, tekstur, dan bentuk [16]. Pada kasus diagnosis *Down syndrome*, CNN dapat menganalisis fitur wajah untuk mendeteksi penanda fenotipe yang terkait dengan kondisi tersebut. Kemampuan CNN dalam memproses data visual yang kompleks mampu menjadikannya sebagai alat yang ampuh dalam pengenalan wajah yang objektif dan andal dalam aplikasi klinis [15]. CNN memiliki beberapa arsitektur yang cukup sering digunakan dalam analisis citra medis, yaitu VGG16 dan VGG19. Kedua arsitektur tersebut memiliki kemampuan dalam mengekstraksi fitur penting dengan tingkat akurasi yang cukup tinggi dan optimal. Dalam penelitian terdahulu, arsitektur VGG16 dan VGG19 telah berhasil mencapai akurasi terbaik dalam klasifikasi penyakit melalui citra medis, dengan akurasi secara berurut sebesar 92,31% dan 89,3% [17].

Berdasarkan penelitian terdahulu, kemampuan VGG16 dan VGG19 telah berhasil unggul dibandingkan arsitektur lain, seperti AlexNet, ResNet50, ResNet101, GoogleNet, Inception-V3, InceptionResNetV2, dan Squeezenet, dalam klasifikasi penyakit melalui citra medis. Meskipun kedua arsitektur tersebut menunjukkan akurasi tinggi, terdapat keterbatasan dari penelitian yang telah dilakukan. Penelitian terdahulu belum menerapkan teknik *transfer learning* dan data augmentasi, yang mana kedua teknik tersebut mampu meningkatkan performa dan akurasi dari model yang digunakan [18], [19]. Merujuk pada temuan yang ditemukan dalam penelitian sebelumnya, penelitian ini bertujuan untuk penerapan model *transfer learning* VGG16 dan VGG19, serta memanfaatkan teknik data augmentasi untuk meningkatkan performa dan akurasi dalam klasifikasi *Down syndrome*, terkhusus di wilayah Asia. Dalam penelitian ini, diharapkan penerapan model yang diusulkan mampu meningkatkan akurasi dan performa dalam klasifikasi *Down syndrome* dibandingkan dengan penelitian sebelumnya, terkhusus di wilayah Asia.

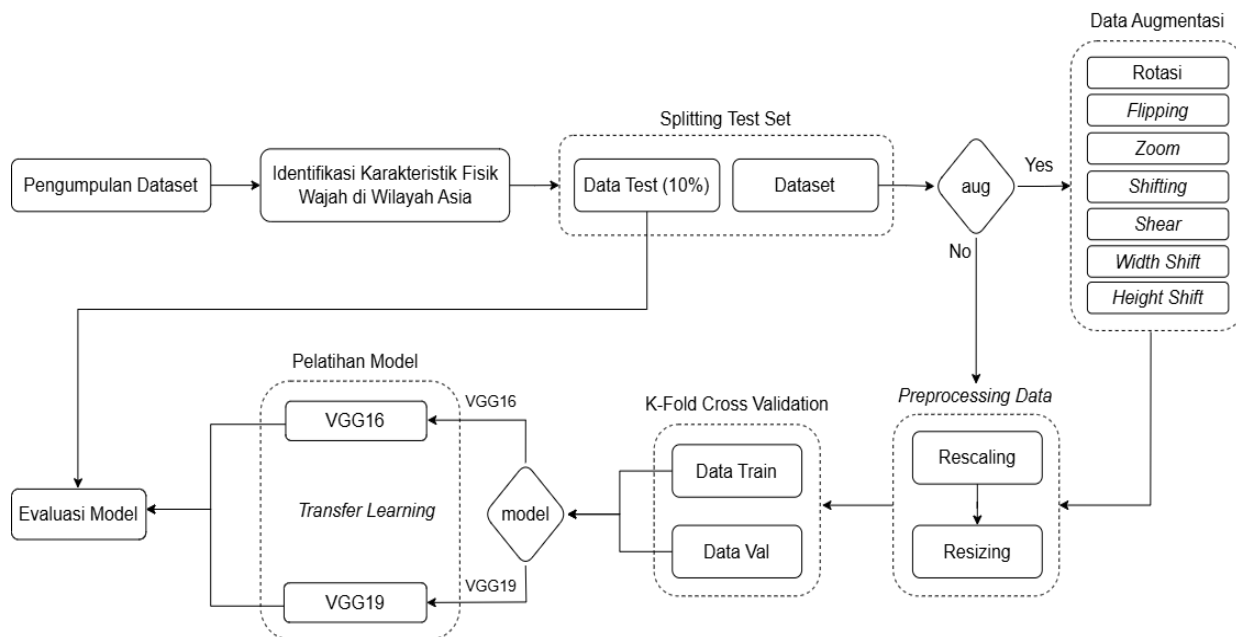
2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Serangkaian tahapan yang dilakukan dalam penelitian ini disusun secara sistematis dan terstruktur. Setiap tahapan disusun untuk melancarkan proses klasifikasi data, sehingga mampu berjalan dengan optimal dan mendukung pencapaian tujuan penelitian secara efektif.

Berdasarkan Gambar 1 dari tahapan penelitian, proses penelitian ini diawali dengan pengumpulan data dalam bentuk citra wajah anak. Tahapan berikutnya adalah identifikasi karakteristik fisik wajah anak berdasarkan wilayah di Asia, seperti memperhatikan dari bentuk mata, hidung, bibir, wajah, rambut, dan warna kulit. Tahap selanjutnya adalah memisahkan 10% data yang digunakan sebagai *test set* untuk menjaga keaslian data. Penelitian ini melakukan dua pendekatan untuk dianalisa lebih lanjut, yaitu dengan menggunakan augmentasi dan tanpa augmentasi. Setelah memilih salah satu diantara teknik tersebut, proses berikutnya *pre-processing data* dengan melakukan *rescaling* dan *resizing* agar model mampu bekerja secara optimal. Selanjutnya, *k-fold cross validation* dilakukan untuk membagi dataset menjadi *train* dan *validation set*. Tahapan berikutnya adalah pelatihan model CNN berbasis VGG16 dan

VGG19 menggunakan teknik *transfer learning*. Tahapan terakhir adalah evaluasi model untuk melihat seberapa baik model mampu melakukan tugasnya dalam klasifikasi *Down syndrome* di wilayah Asia.



Gambar 1. Diagram Alur Metodologi Penelitian

2.2 Dataset

Data yang digunakan dalam penelitian ini merupakan data dalam bentuk citra wajah anak yang diperoleh dari situs Roboflow. Dataset tersebut terdiri dari dua kelas utama, yaitu *Down syndrome* dan *Healthy* dengan jumlah total sebesar 2.997 data.

2.3 Identifikasi Karakteristik Fisik Wajah di Wilayah Asia

Dalam penelitian ini, karakteristik wajah anak-anak di wilayah Asia diidentifikasi secara manual melalui beberapa aspek tertentu. Proses ini dilakukan untuk memahami lebih lanjut apakah terdapat pola tertentu yang mungkin ditemukan dan melihat seberapa baik model mampu melakukan klasifikasi *Down syndrome* di wilayah Asia. Beberapa aspek yang diamati dalam penelitian ini, disajikan pada Tabel 1 [20]

Tabel 1. Karakteristik Fisik Wajah di Wilayah Asia

Karakteristik Fisik	Asia
Mata	Lipatan Sempit
Hidung	Rendah, lebar rata-rata
Bibir	Rata-rata penuh
Bentuk Wajah	Pendek dengan pipi pipih
Rambut	Tipis, lurus atau bergelombang sedikit
Kulit	Kekuningan

Pada Tabel 1, menunjukkan beberapa aspek tertentu dari karakteristik fisik diamati lebih lanjut pada individu di wilayah Asia, seperti bentuk mata, hidung, wajah, mata, hidung, bibir, wajah, rambut, dan warna kulit. Setiap aspek menunjukkan ciri khas, yang mampu membantu dalam proses klasifikasi lebih lanjut.

2.4 Splitting Test Set

Tahapan berikutnya adalah memisahkan data sebesar 10% dari dataset untuk digunakan sebagai *test set*. Hal tersebut dilakukan untuk menjaga keaslian dari data uji yang digunakan pada saat evaluasi model. Pemisahan *test set* sebelum dilakukan *preprocessing* untuk menghindari kebocoran data, seperti data terkena augmentasi atau *balancing*, yang dapat mempengaruhi efektivitas model dalam memberikan hasil uji yang maksimal.

2.5 Data Augmentasi

Pemanfaatan teknik data augmentasi dilakukan sebagai strategi yang memungkinkan mampu meningkatkan variasi data yang tersedia saat digunakan pada proses pelatihan model [19]. Selain itu, penerapan augmentasi juga digunakan untuk menyeimbangkan jumlah gambar pada kelas dengan jumlah gambar lebih sedikit agar distribusi data menjadi lebih seimbang. Tabel 2 menyajikan informasi dari nilai-nilai yang digunakan untuk augmentasi data.

Tabel 2. Nilai Data Augmentasi

Data Augmentasi	Nilai
<i>Rotation</i>	30°
<i>Width Shift</i>	0.2
<i>Height Shift</i>	0.2
<i>Shear Transformation</i>	0.2
<i>Zoom</i>	0.2
<i>Horizontal Flip</i>	<i>True</i>
<i>Fill Mode</i>	<i>Nearest</i>

Tabel 2 menunjukkan berbagai teknik augmentasi dan nilai parameter yang diterapkan untuk transformasi gambar pada dataset yang digunakan penelitian ini. Setiap teknik augmentasi memiliki parameter tertentu yang telah disesuaikan untuk mempertahankan fitur-fitur penting pada gambar.

2.6 Pre-processing Data

Tahapan berikutnya adalah *pre-processing data*, yang mana data dipersiapkan terlebih dahulu sebelum dilatih pada model yang telah dirancang. *Pre-processing* yang digunakan dalam penelitian ini meliputi *rescaling* dan *resizing*. *Rescaling* dilakukan untuk menormalisasikan gambar dengan mengubah nilai skala atau rentang piksel gambar dari 0-255 pada tiap kanal warna (RGB) menjadi 0-1 yang dibagi pada tiap nilai piksel dengan nilai 255. Berikutnya adalah *resizing*, dilakukan untuk menyesuaikan ukuran dari input dimensi gambar sesuai dengan input model dengan ukuran 224×224 piksel.

2.7 K-Fold Cross Validation

Penelitian ini mengimplementasikan *k-fold cross validation* untuk mengevaluasi performa dari model yang dirancang. Pada penelitian ini, dalam mengevaluasi performa dari model, digunakan *5-fold cross validation* dikarenakan kemampuannya dalam generalisasi data [21]. Proses ini dilakukan dengan membagi dataset menjadi 5 bagian atau *folds*, di mana setiap iterasi model dilatih menggunakan 4 bagian sebagai *training set* (80%) dan diuji dengan 1 bagian sebagai *validation set* (20%). Proses tersebut diulang sebanyak lima kali, kemudian hasil evaluasi dihitung sebagai rata-rata dari seluruh *fold*. Implementasi *k-fold cross validation* dalam penelitian ini untuk mengoptimalkan penggunaan data, mengurangi kemungkinan adanya *overfitting*, dan menghasilkan evaluasi yang stabil.

2.8 Arsitektur Model CNN

Penelitian ini melakukan proses pelatihan dan pengujian model dengan mengimplementasikan arsitektur CNN dalam klasifikasi *Down syndrome* di wilayah Asia. Model ini memiliki puluhan hingga ratusan lapisan yang memproses data citra melalui lapisan jaringan dan menghasilkan output pada kelas tertentu. Pada model CNN, lapisan dibagi menjadi dua bagian dengan fungsi masing-masing, yaitu lapisan ekstraksi dan lapisan klasifikasi. Lapisan ekstraksi digunakan untuk melakukan tugas ekstraksi fitur dari input data. Berikutnya, lapisan klasifikasi digunakan untuk melakukan klasifikasi data berdasarkan fitur-fitur yang dihasilkan dari lapisan ekstraksi [22]. Lapisan ekstraksi meliputi lapisan konvolusi, fungsi aktivasi, dan *pooling layer*. Sedangkan, lapisan klasifikasi meliputi *fully connected layer*, dan dilengkapi dengan fungsi aktivasi seperti *dropout* dan *softmax* [22], [23]. Dalam penelitian ini, beberapa arsitektur model CNN diterapkan, yaitu VGG16 dan VGG19, yang selanjutnya akan dibandingkan melalui hasil dari kinerja setiap arsitektur yang digunakan.

a. VGG16

VGG16 memiliki jumlah lapisan sebanyak 16 lapisan, dengan 13 *convolutional layer* dan 3 *fully connected layer* [24]. Meskipun memiliki struktur arsitektur yang sederhana, VGG16 mampu mempelajari fitur secara mendalam, sehingga memungkinkan ekstraksi fitur secara menyeluruh. Selain itu, dengan struktur yang lebih sederhana, model yang dirancang menggunakan VGG16 mampu meningkatkan akurasi dan risiko *overfitting* [25]. Namun, dikarenakan memiliki jumlah parameter yang besar, dibutuhkan waktu yang lebih lama dan memiliki kebutuhan komputasi yang tinggi [26]. Arsitektur VGG16 dalam penelitian ini dirancang menggunakan input data berukuran $224 \times 224 \times 3$, yang mewakili panjang 224 piksel, lebar 224 piksel, dan *depth* 3 (RGB).

b. VGG19

VGG19 memiliki jumlah lapisan yang lebih banyak dibandingkan dengan VGG16, yaitu 19 lapisan, dengan 16 *convolutional layer* dan 3 *fully connected layer*. Perbedaan jumlah lapisan konvolusi yang dimiliki oleh VGG19, memungkinkan model yang digunakan mampu mengekstraksi fitur dengan kompleksitas yang lebih tinggi. Sama halnya dengan VGG16, lapisan konvolusi memiliki peran penting dalam mendeteksi pola-pola kompleks pada berbagai tingkat abstraksi [24]. Selanjutnya, input data yang digunakan pada arsitektur VGG19 serupa dengan sebelumnya, yaitu berukuran $224 \times 224 \times 3$.

2.9 Pelatihan Model

Penelitian ini menerapkan teknik *transfer learning*, dengan melatih model terlebih dahulu di dataset dalam jumlah besar, yaitu ImageNet. Hal tersebut dilakukan untuk meningkatkan performa model dalam mempelajari data dengan

memanfaatkan pengetahuan yang telah dipelajari dari dataset besar. Selanjutnya, sebagian besar lapisan awal dari model yang bertugas untuk menangkap fitur secara umum, seperti tepi, pola, tekstur, dan bentuk dasar, dibekukan. Dengan membekukan lapisan tersebut, model hanya perlu memanfaatkan pengetahuan yang diperoleh dari data yang telah dilatih sebelumnya. Sementara, beberapa lapisan terakhir dilatih ulang untuk menangani tugas klasifikasi yang lebih spesifik, yaitu klasifikasi *Down syndrome*. Selanjutnya, ditambahkan beberapa *fully connected layer* untuk mengolah fitur yang telah diekstrak dari model sebelumnya. Lapisan tambahan tersebut terdiri dari *Flatten layer* dan *Dense layer*. Lapisan *flatten* berfungsi untuk mengubah fitur-fitur menjadi vektor satu dimensi [27]. Selanjutnya, vektor fitur melalui lapisan-lapisan *dense* yang meliputi neuron-neuron yang saling terhubung sepenuhnya, di mana fitur-fitur tersebut diproses untuk membentuk representasi abstrak [28]. Selain itu, untuk menghindari adanya risiko *overfitting*, diterapkan *Dropout layer* diantara kedua lapisan *dense*. Pada lapisan *output*, fungsi aktivasi yang digunakan adalah *softmax*, di mana berfungsi untuk mengklasifikasikan gambar multi-kelas karena mampu menentukan probabilitas pada masing-masing kelas [12]. Nilai parameter lapisan tambahan yang digunakan, disajikan pada Tabel 3.

Tabel 3. Nilai Parameter Lapisan Tambahan

Lapisan	Jumlah Neuron	Fungsi Aktivasi
<i>Flatten</i>	-	-
<i>Dense</i>	128	ReLU
<i>Dropout</i>	0.5	-
<i>Dense</i>	64	ReLU
<i>Dropout</i>	0.5	-
<i>Dense (output)</i>	2	Softmax

Selanjutnya, model dikompilasikan menggunakan *Adam optimizer*, dengan tujuan untuk mempercepat proses optimasi tanpa mengalami perubahan bobot yang terlalu drastis. Dalam penelitian ini, fungsi loss yang digunakan adalah *Categorical Crossentropy* yang cocok digunakan untuk tugas klasifikasi dengan lebih dari satu kelas. Selain menggunakan parameter-parameter yang telah disebutkan, penelitian ini juga mengusulkan untuk menggunakan *callback function* yang digunakan untuk menghentikan proses pelatihan model secara otomatis saat membalikkan nilai parameter yang telah ditentukan. Fungsi *callback* yang diterapkan adalah *EarlyStopping* yang digunakan untuk memantau metrik akurasi validasi apabila tidak ada peningkatan dari akurasi validasi dalam beberapa *epoch*. Tabel 4 menyajikan tabel nilai-nilai parameter yang digunakan dalam pelatihan model.

Tabel 4. Nilai Parameter Pelatihan Model

Parameter	Nilai
<i>Optimizer</i>	Adam (lr = 0.0001)
<i>Loss Function</i>	<i>Categorical Crossentropy</i>
Metrik Evaluasi	<i>Accuracy</i>
<i>Batch Size</i>	32
<i>Epoch Max.</i>	15
<i>Monitor (EarlyStopping)</i>	<i>val_accuracy</i>
<i>Patience (EarlyStopping)</i>	3

Tabel 4 menyajikan parameter dan nilai yang digunakan pada proses pelatihan model, meliputi *optimizer*, *loss function*, metrik evaluasi, *batch size*, *epoch max.*, *monitor*, dan *patience*. Penerapan parameter berserta nilai yang digunakan, disesuaikan untuk mengoptimalkan kinerja model dalam melakukan tugas klasifikasi.

2.10 Evaluasi Model

Dari percobaan-percobaan yang telah dilakukan, model akan dievaluasi dengan tujuan untuk menilai seberapa baik model yang telah dilatih dalam mengklasifikasikan *Down syndrome*. Dalam penelitian ini, evaluasi model yang digunakan mencakup dua komponen, yaitu *k-fold cross validation* dan *confusion matrix*. *K-fold cross validation* diterapkan sebagai metode evaluasi performa model dengan cara membagi dataset *training* dan *validation* ke dalam beberapa bagian yang berbeda. Dalam proses ini, evaluasi dilakukan dengan mencatat hasil dari nilai akurasi dan loss dari tiap fold, kemudian diambil nilai rata-ratanya untuk mendapatkan gambaran dari kemampuan model dalam mengenali pola dari dataset. Selanjutnya, dilakukan pengujian menggunakan *test set* yang telah dipisahkan sebesar 10% sejak awal. Evaluasi tersebut dilakukan untuk mengetahui kemampuan model dalam membuat keputusan dengan membandingkan hasil tersebut dengan label asli. Untuk memahami distribusi kesalahan prediksi, digunakan *confusion matrix* dengan memperhatikan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, dan *f1-score*. Untuk memrepresentasikan visual *confusion matrix*, terdapat empat komponen utama, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN).

3. HASIL DAN PEMBAHASAN

3.1 Pemahaman Dataset

Data yang digunakan terdiri dari 2.997 gambar yang telah diberikan label dengan dua kategori, yaitu *Down syndrome* dan *Healthy*. Dataset tersedia di situs Roboflow sebagai dataset yang dapat digunakan dalam project penelitian. Dalam penelitian ini, yang menjadi fokus penting adalah mengklasifikasikan *Down syndrome* di wilayah Asia. Dari total dataset yang dimiliki, diolah terlebih dahulu dengan melakukan identifikasi karakteristik fisik wajah di wilayah Asia secara manual. Maka, gambar yang digunakan dalam proses pelatihan dan evaluasi model pada penelitian ini hanya gambar yang memenuhi kriteria karakteristik fisik wajah di wilayah Asia. Total distribusi dari dataset yang digunakan, disajikan pada Tabel 5, serta visualisasi dari dataset yang digunakan tertera pada Gambar 2.

Tabel 5. Distribusi Dataset yang digunakan

Kelas	Jumlah Data
<i>Down syndrome</i>	753
<i>Healthy</i>	619



Gambar 2. Visualisasi Dataset dari Kelas *Down syndrome* dan *Healthy*

Berdasarkan Tabel 5, dataset yang digunakan dalam penelitian ini terdiri dari 735 gambar pada kelas *Down syndrome* dan 619 gambar pada kelas *Healthy*, sehingga total data yang digunakan sebanyak 1.372 gambar. Jumlah gambar dari masing-masing kelas menunjukkan angka yang cukup seimbang. Dengan distribusi data yang cukup seimbang, memungkinkan model untuk melakukan tugas klasifikasi secara optimal dalam mengenali kedua kategori.

3.2 Splitting Test Set

Dalam penelitian ini, dilakukan pemisahan test set sebesar 10% terlebih dahulu dari masing-masing kelas pada dataset yang digunakan. Hal tersebut dilakukan untuk menjaga keaslian dari data saat diuji menggunakan model yang telah dilatih. Distribusi data yang digunakan dalam *test set*, disajikan pada Tabel 6.

Tabel 6. Distribusi Data setelah *Splitting Test Set*

Kelas	Jumlah Data
<i>Down syndrome</i>	75
<i>Healthy</i>	61

Dataset yang tersisa setelah dilakukan pemisahan *test set*, digunakan untuk pelatihan model. Data ini akan dibagi menggunakan *k-fold cross validation* dengan membaginya menjadi beberapa bagian sebagai *training set* dan *validation set*. Dengan pendekatan tersebut, setiap data akan terbagi secara merata dan memiliki kesempatan untuk digunakan sebagai data latih maupun data validasi.

3.3 Data Augmentasi

Penelitian ini menggunakan dua pendekatan, yaitu dengan augmentasi dan tanpa augmentasi. Augmentasi yang digunakan dalam penelitian ini bertujuan untuk meningkatkan variasi dataset dan menyeimbangkan jumlah gambar pada kelas dengan jumlah yang lebih sedikit. Tabel 7 merupakan distribusi jumlah gambar dari masing-masing kelas sebelum proses augmentasi dilakukan.

Tabel 7. Distribusi Data sebelum Augmentasi

Kelas	Jumlah Data
<i>Down syndrome</i>	678
<i>Healthy</i>	558

Berdasarkan informasi yang tertera pada Tabel 7, kelas *Healthy* menunjukkan jumlah gambar yang lebih sedikit dibandingkan kelas *Down syndrome*. Ketidakseimbangan pada distribusi dari dataset yang digunakan, mampu mempengaruhi performa dari model dalam mengenali kedua kategori. Maka dari itu, dilakukan augmentasi sebagai upaya untuk meningkatkan performa model dan membantu model untuk mempelajari data secara lebih optimal.

3.4 Hasil Pelatihan Model

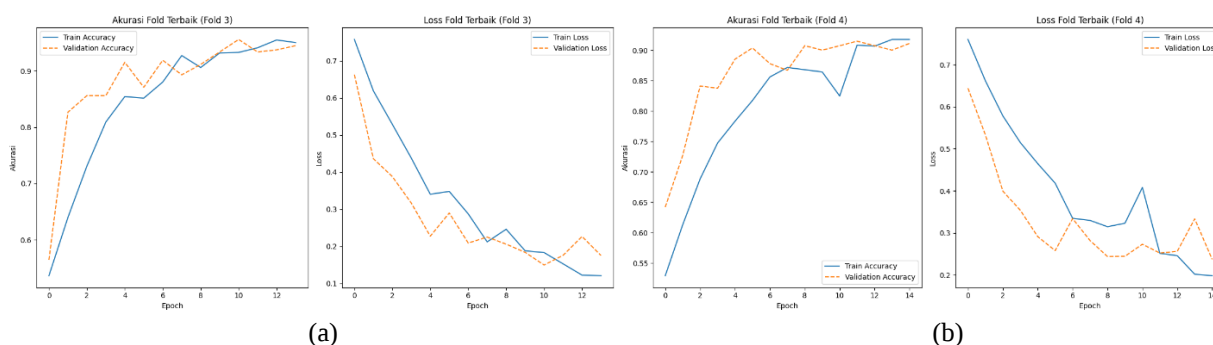
Dalam penelitian ini, pelatihan dilakukan menggunakan model CNN yang menerapkan *transfer learning* dengan dua arsitektur model, yaitu VGG16 dan VGG19. Penerapan *transfer learning* memungkinkan model untuk mempelajari fitur-fitur pada dataset dalam jumlah besar, yaitu ImageNet. Dengan memanfaatkan pengetahuan dan fitur-fitur yang telah dipelajari dari dataset besar, model tidak perlu dilatih dari awal. Selanjutnya, beberapa lapisan awal dibekukan dan beberapa lapisan akhir dilatih ulang sehingga mampu beradaptasi dengan data yang lebih spesifik. Penelitian ini menggunakan dua pendekatan untuk pelatihan model, yaitu dengan augmentasi dan tanpa augmentasi. Pendekatan tersebut dilakukan untuk mengevaluasi pengaruh dari penggunaan augmentasi terhadap performa model.

Selama proses pelatihan, evaluasi model yang dilakukan adalah dengan menganalisa hasil dari *train accuracy*, *validation accuracy*, *train loss*, dan *validation loss*. *Train accuracy* menunjukkan seberapa baik model dalam mempelajari data dari data latih, sedangkan *validation accuracy* berfungsi untuk generalisasi model pada data validasi. Selanjutnya, *train loss* memberikan hasil nilai dari seberapa besar kesalahan yang diperoleh dari model saat memprediksikan data latih, sedangkan *validation loss* menunjukkan seberapa besar kesalahan yang diperoleh dari model saat memprediksikan data validasi. Dalam penelitian ini, untuk mendapatkan hasil evaluasi yang memiliki performa terbaik, dilakukan pemantauan lebih lanjut terhadap nilai *validation accuracy* tertinggi di setiap *fold*. Model dengan nilai akurasi terbaik, akan dipilih berdasarkan hasil evaluasi dengan menggunakan *k-fold cross validation*, yang akan dibahas pada bagian selanjutnya. Tabel 8 menyajikan ringkasan hasil dari pelatihan model, di mana hasil tersebut diperoleh dari hasil dengan iterasi terbaik yang dipilih berdasarkan *validation accuracy* tertinggi menggunakan *k-fold cross validation*.

Tabel 8. Hasil Pelatihan Model dengan dan tanpa Augmentasi

Model	Augmentation	Train Acc (%)	Val Acc (%)	Acc Loss (%)	Val Loss (%)
VGG16	-	0.91	0.96	0.46	0.16
	✓	0.92	0.96	0.23	0.15
VGG19	-	0.86	0.92	0.60	0.30
	✓	0.94	0.91	0.13	0.24

Berdasarkan hasil dari Tabel 8, implementasi augmentasi di tiap arsitektur model yang digunakan berhasil menunjukkan nilai dengan akurasi yang lebih tinggi dibandingkan tanpa augmentasi. Hal tersebut menunjukkan bahwa adanya augmentasi mampu membantu model untuk melakukan tugasnya dalam mengenali pola secara lebih baik dengan meningkatkan variasi dari dataset yang digunakan. Selanjutnya, apabila dianalisa lebih lanjut dari segi loss, hasil menunjukkan bahwa model dengan augmentasi memiliki nilai loss yang lebih rendah, baik untuk *training loss* maupun *validation loss*. Dari hasil tersebut, mengindikasikan bahwa model dengan augmentasi tidak hanya meningkatkan akurasi pada model, melainkan juga membantu model untuk belajar dengan lebih stabil dan tidak mengalami *overfitting* yang signifikan. Analisa hasil selanjutnya adalah pada model VGG16 dan VGG19. Apabila ditinjau berdasarkan nilai akurasi validasi, model VGG16 memiliki nilai tertinggi di antara percobaan lainnya, baik dengan augmentasi maupun tanpa augmentasi. Sedangkan, model VGG19 memiliki penurunan sedikit pada akurasi validasinya, yaitu dari 92% menurun menjadi 91%. Berdasarkan hasil tersebut, VGG16 memiliki keunggulan dalam melakukan generalisasi data yang lebih baik dan lebih stabil dalam mengenali pola dataset dibandingkan VGG19. Gambar 3 menunjukkan grafik akurasi dan loss dari kedua model menggunakan augmentasi.



Gambar 3. (a) Grafik Akurasi dan Loss VGG16 dan (b) Grafik Akurasi dan Loss VGG19

Berdasarkan grafik diatas, Gambar 3(a) menunjukkan bahwa akurasi pelatihan meningkat secara stabil di setiap *epoch*, sementara validasi akurasi juga mengalami peningkatan namun lebih fluktuatif. Sedangkan pada pelatihan loss, grafik menunjukkan penurunan yang konsisten, dan validasi loss juga mengalami hal yang serupa walaupun terlihat lebih fluktuatif. Dari grafik pada Gambar 3(a), belum terlihat adanya indikasi bahwa model mengalami *overfitting*, dikarenakan grafik validasi loss tidak meningkat secara tajam setelah beberapa *epoch*. Pada Gambar 3(b), grafik akurasi pelatihan milik VGG19 meningkat secara cepat dibandingkan dengan model VGG16. Selain itu, grafik akurasi validasi juga meningkat dengan baik, namun terlihat adanya sedikit gap antara akurasi pelatihan dan validasi di beberapa *epoch* terakhir. Sementara pada pelatihan loss, grafik mengalami penurunan yang cepat, dengan grafik

validasi loss yang lebih fluktuatif dibandingkan milik VGG16. Dari hari temuan tersebut, model VGG19 memiliki kemungkinan mengalami *overfitting*, dikarenakan beberapa epoch terakhir menunjukkan grafik yang cukup fluktuatif. Berdasarkan kedua grafik tersebut, model VGG16 dan VGG19 memiliki keunggulan dan keterbatasan masing-masing. Hasil dari grafik VGG16, menunjukkan bahwa model tersebut memiliki kemampuan yang stabil dalam melakukan tugasnya, namun membutuhkan waktu dan jumlah *epoch* yang lebih besar untuk mencapai performa yang lebih optimal. Sedangkan, grafik VGG19 menunjukkan bahwa model tersebut mampu belajar dengan waktu yang lebih cepat dibandingkan dengan VGG16, namun terdapat indikasi adanya *overfitting*.

3.5 Evaluasi Model dengan K-Fold Cross Validation

Dalam upaya mengurangi risiko terjadi *overfitting*, penelitian ini menggunakan *k-fold cross validation*. Dataset yang digunakan dalam proses ini dibagi menjadi 5 bagian atau *fold*, di mana pada tiap bagian akan secara bergantian menjadi *validation set*, sedangkan bagian lain digunakan sebagai *training set*. Proses dilakukan sebanyak 5 kali, sehingga tiap bagian data memiliki kesempatan untuk menjadi bagian dari *validation set*. Dalam proses ini, metrik yang digunakan untuk dipantau lebih lanjut dalam tahap evaluasi adalah *validation accuracy*. Hal tersebut dilakukan untuk menilai kemampuan model dalam melakukan generalisasi pada data validasi yang tidak termasuk dalam training. Hasil evaluasi model dengan *k-fold cross validation*, disajikan pada Tabel 9.

Tabel 9. Hasil Evaluasi Model dengan K-Fold Cross Validation

<i>Fold</i>	VGG16 (tanpa augmentasi)	VGG16 (dengan augmentasi)	VGG19 (tanpa augmentasi)	VGG19 (dengan augmentasi)
1	89%	91%	92%	85%
2	96%	91%	90%	83%
3	93%	96%	89%	83%
4	90%	89%	88%	92%
5	90%	90%	87%	89%
Mean	92%	91%	89%	87%

Berdasarkan hasil diatas, terdapat pola yang dapat ditemukan dari performa model di tiap *fold*. Model VGG16 dengan implementasi augmentasi dalam proses pelatihan, menunjukkan nilai akurasi yang paling stabil pada tiap *fold* dan memiliki rata-rata sebesar 91%. Meskipun nilai rata-rata yang dimiliki oleh model tersebut lebih kecil dibandingkan model VGG16 tanpa augmentasi, model tersebut menunjukkan nilai akurasi validasi yang lebih stabil. Berbeda dengan model sebelumnya, VGG19 menunjukkan nilai rata-rata dengan angka yang lebih rendah. Hal tersebut bisa terjadi dikarenakan adanya kemungkinan bahwa dengan jumlah arsitektur yang lebih dalam, model tidak dapat menjamin mampu melakukan generalisasi data dengan baik. Walaupun nilai rata-rata pada model VGG19 dengan augmentasi memiliki nilai yang lebih rendah di antara model lain, namun nilai akurasi validasi pada tiap *fold* menunjukkan angka yang cukup stabil. Berdasarkan temuan tersebut, baik dari model VGG16 maupun VGG19, menunjukkan bahwa implementasi augmentasi mampu membantu model dalam melakukan tugasnya untuk generalisasi data dengan lebih stabil dan optimal di tiap bagiannya. Sedangkan, model yang tidak menggunakan augmentasi, menunjukkan ada fluktuasi dan gap yang cukup besar antar *fold*.

Dari hasil temuan diatas, penelitian ini memilih untuk menggunakan hasil dari iterasi terbaik yang ditentukan berdasarkan nilai *validation accuracy* tertinggi di sebagian *fold* pada tiap model. Dalam proses ini, dipilih salah satu pendekatan dari tiap model untuk dianalisa lebih lanjut. Berdasarkan Tabel 9, model yang menggunakan augmentasi memiliki konsistensi yang lebih baik dibandingkan yang tidak menggunakan augmentasi. Maka dari itu, Tabel 10 menyajikan metrik evaluasi yang lebih rinci untuk dianalisis lebih lanjut dengan memperhatikan metrik *precision*, *recall*, *f1-score*, dan *accuracy* pada masing-masing kelas.

Tabel 10. Hasil Evaluasi Model dengan Nilai Iterasi Terbaik pada Model VGG16 dan VGG19

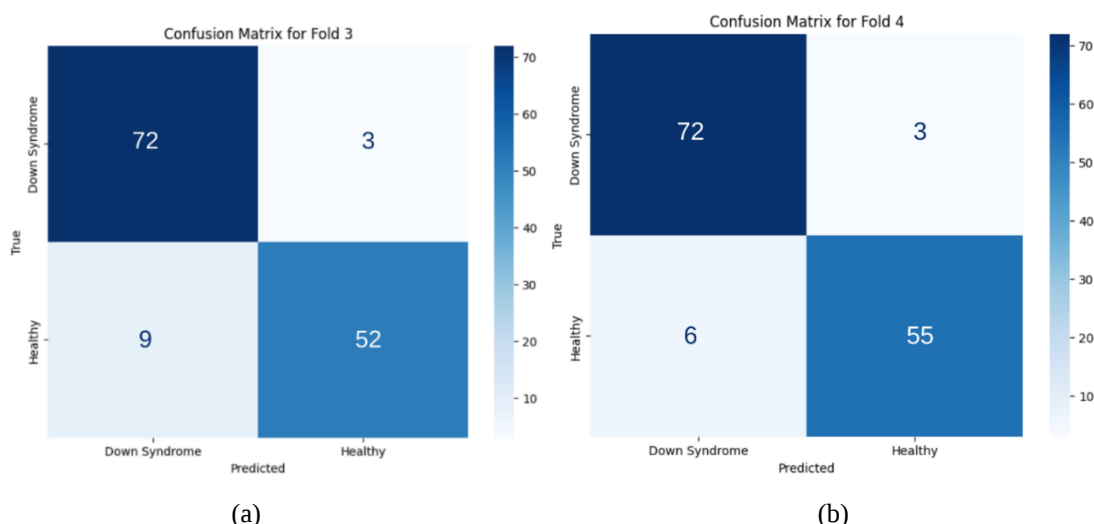
<i>Model</i>	<i>Class</i>	<i>Fold</i>	<i>Precision (%)</i>	<i>Recall (%)</i>	<i>F1-Score (%)</i>	<i>Acc (%)</i>
VGG16	<i>Down Syndrome</i>	Fold-3	0.89	0.96	0.92	0.91
	<i>Healthy</i>		0.95	0.85	0.90	
VGG19	<i>Down Syndrome</i>	Fold-4	0.92	0.96	0.94	0.93
	<i>Healthy</i>		0.95	0.90	0.92	

Berdasarkan hasil yang ditunjukkan pada Tabel 10, akurasi yang dimiliki oleh model VGG16 memiliki nilai yang lebih unggul dibandingkan dengan model VGG19. Nilai *precision* pada model VGG19 memiliki nilai yang lebih baik untuk kelas *Down syndrome*. Hal tersebut menunjukkan bahwa, model memiliki kemampuan presisi yang lebih baik dalam mengidentifikasi individu dengan *Down syndrome*. Sedangkan, kedua model menunjukkan nilai presisi yang sama untuk kelas *Healthy*. Berikutnya, *recall* antara kedua model untuk kelas *Down syndrome* memiliki nilai yang sama. Namun, untuk kelas *Healthy*, model VGG16 cenderung memiliki nilai yang lebih rendah dibandingkan VGG19. Dengan hasil tersebut, model VGG16 memiliki kekurangan dalam mengidentifikasi individu pada kelas *Healthy* secara akurat. Analisis selanjutnya pada metrik *f1-score*, yang mana model VGG19 juga memiliki nilai yang

lebih tinggi pada kedua kelas, dibandingkan dengan VGG16. Dari hasil temuan secara keseluruhan, model VGG19 memiliki kemampuan yang lebih unggul dibandingkan VGG16 pada iterasi terbaiknya. Hasil dari *accuracy*, *precision*, *recall*, dan *f1-score* yang dimiliki oleh VGG19, mampu mengindikasikan bahwa model tersebut mampu melakukan generalisasi data lebih baik dan optimal.

3.6 Evaluasi Model dengan Confusion Matrix

Evaluasi model berikutnya adalah menggunakan *confusion matrix*, yang mampu memberikan informasi tambahan perihal performa model dalam melakukan klasifikasi wajah anak sesuai dengan kategori yang telah ditentukan berdasarkan label yang sebenarnya. Evaluasi model menggunakan *confusion matrix* mampu membantu dalam menemukan pola kesalahan prediksi dan menganalisis seberapa baik kemampuan model dalam membedakan gambar wajah anak dengan *Down syndrome* dan *Healthy*. *Confusion matrix* memiliki empat komponen penting untuk menunjukkan distribusi dari tiap kelas, yaitu dengan *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN). Dalam evaluasi ini, hasil yang digunakan untuk dianalisis lebih lanjut adalah hasil dari evaluasi model dengan iterasi terbaik. Gambar 4 menyajikan nilai *confusion matrix* dari masing-masing model berdasarkan hasil evaluasi menggunakan iterasi terbaik.



Gambar 4. (a) *Confusion Matrix* VGG16 dan (b) *Confusion Matrix* VGG19

Berdasarkan hasil tersebut, Gambar 4 (a) dan (b) menampilkan distribusi hasil prediksi model VGG16 dan VGG19 terhadap 136 gambar wajah anak dengan kategori *Down syndrome* dan *Healthy*. Hasil dari *confusion matrix* pada model VGG16 menunjukkan bahwa, dari 75 gambar yang digunakan pada kelas *Down syndrome*, telah berhasil terprediksi benar sesuai dengan kelas yang sebenarnya sejumlah 72 gambar (*true positive*). Sementara sejumlah 3 gambar terprediksi salah, yakni terprediksi pada kelas *Healthy* (*false negative*). Selanjutnya, 61 gambar yang digunakan pada kelas *Healthy*, berhasil terprediksi benar dengan kelas yang sebenarnya sejumlah 52 gambar (*true negative*), dengan prediksi kesalahan sejumlah 9 gambar (*false positive*). Analisis *confusion matrix* berikutnya adalah pada model VGG19, gambar yang terprediksi benar sesuai dengan kelas *Down syndrome* sejumlah 72 gambar (*true positive*), sementara 3 gambar terprediksi salah (*false negative*). Berbeda dengan VGG16, model VGG19 berhasil memprediksikan gambar yang terprediksi benar sesuai dengan kelas *Healthy* sejumlah 55 gambar (*true negative*), dengan prediksi kesalahan yang lebih sedikit dibandingkan model sebelumnya, yakni 6 gambar (*false positive*).

Berdasarkan hasil dari analisis tersebut, model VGG19 memiliki jumlah prediksi kesalahan yang lebih kecil dibandingkan VGG16 dalam mengklasifikasikan individu dengan kategori *Healthy*. Oleh karena itu, model VGG19 memiliki kemampuan yang lebih optimal dalam klasifikasi *Down syndrome*. Gambar 5 menampilkan visualisasi dari hasil gambar yang terprediksi salah menggunakan model VGG19.



Gambar 5. Visualisasi Prediksi Salah pada Model VGG19

Berdasarkan hasil visualisasi diatas, terdapat beberapa faktor yang mampu mempengaruhi kesalahan prediksi. Untuk kesalahan prediksi *Down syndrome* sebagai *Healthy*, ada kemungkinan dikarenakan pencahayaan, sudut pengambilan gambar, dan ekspresi wajah. Sedangkan, kesalahan prediksi *Healthy* sebagai *Down syndrome*, dikarenakan beberapa fitur wajah, seperti bentuk mata dan struktur wajah, lebih sering ditemukan pada kelas *Down syndrome*.

3.7 Perbandingan Performa Model

Dari hasil temuan pada percobaan yang telah dilakukan, Tabel 11 menampilkan ringkasan secara keseluruhan hasil dari perbandingan antara kedua model yang telah diimplementasikan.

Tabel 11. Perbandingan Performa Model VGG16 dan VGG19

Aspek	VGG16	VGG19
Akurasi	Memiliki nilai akurasi yang lebih tinggi dalam melakukan klasifikasi <i>Down syndrome</i> , namun memiliki potensi kesalahan yang lebih banyak dalam klasifikasi <i>Healthy</i>	Memiliki kemampuan yang lebih baik dalam klasifikasi <i>Healthy</i> , namun struktur yang dirancang kurang mampu dalam menangkap pola wajah <i>Down syndrome</i>
Sensitivitas	Tinggi	Rendah
Spesifisitas	Rendah	Tinggi
Pola Kesalahan	Model gagal melakukan klasifikasi pada kelas <i>Healthy</i>	Model kurang berhasil menangkap dan klasifikasi pada kelas <i>Down syndrome</i>
<i>Overfitting</i>	Kemungkinan lebih kecil	Lebih rentan karena memiliki lapisan yang lebih kompleks

Tabel 11 menyajikan ringkasan dari perbedaan antara performa model VGG16 dan VGG19 yang dilakukan dalam klasifikasi *Down syndrome* di Asia. Perbandingan tersebut dilakukan untuk mengetahui model yang memiliki performa lebih baik dan optimal dalam melakukan tugas klasifikasi. Dari hasil tersebut, tiap model memiliki kelebihan dan keterbatasan masing-masing.

4. KESIMPULAN

Penelitian ini membandingkan performa model CNN dengan arsitektur VGG16 dan VGG19 dengan menerapkan teknik *transfer learning* dan augmentasi terhadap klasifikasi *Down syndrome* di wilayah Asia. Hasil dari penelitian ini menunjukkan bahwa implementasi *transfer learning* dan augmentasi mampu meningkatkan klasifikasi *Down syndrome*, terkhusus di wilayah Asia. Model VGG16 berhasil mencapai akurasi dengan hasil yang cukup tinggi, yaitu 91% dengan grafik akurasi dan loss yang cukup konsisten. Sementara, model VGG19 berhasil mencapai akurasi yang lebih tinggi, yaitu sebesar 93% dengan potensi prediksi kesalahan yang lebih kecil dibandingkan VGG16. Meskipun kedua model tersebut memiliki akurasi yang cukup tinggi, terdapat perbedaan yang signifikan antar keduanya. VGG16 memiliki sensitivitas yang lebih tinggi dibandingkan VGG19, yang mana model tersebut lebih mampu mengenali pola wajah anak dengan *Down syndrome*. Namun, keterbatasan yang dimilikinya adalah model cenderung mengenali wajah anak sehat atau *Healthy* sebagai *Down syndrome*. Di sisi lain, VGG19 memiliki spesifisitas yang lebih baik, yang mana model lebih akurat dalam mengklasifikasikan wajah anak dengan *Healthy*. Keterbatasan dari model tersebut adalah model kurang mengenali wajah *Down syndrome*. Berdasarkan temuan tersebut, pemilihan model tergantung dengan kebutuhan sistem yang dikembangkan. Model VGG16 cocok digunakan untuk mengklasifikasikan anak dengan *Down syndrome* yang lebih banyak. Sedangkan, model VGG19 cocok digunakan untuk mengurangi tingkat kesalahan dalam klasifikasi *Healthy* sebagai *Down syndrome*. Meskipun performa model dalam penelitian ini berhasil mencapai akurasi yang lebih unggul dibandingkan penelitian sebelumnya, terdapat beberapa keterbatasan yang dimiliki oleh penelitian ini, yaitu dataset yang digunakan kurang bervariasi, variasi regional, dan salah satu model yang dirancang memiliki potensi terjadi *overfitting*. Untuk penelitian selanjutnya, diharapkan dapat memperkaya variasi dari dataset yang digunakan, melakukan percobaan dengan arsitektur model yang lebih mampu menangkap pola untuk klasifikasi pengenalan wajah, serta menerapkan *fine-tuning* untuk mengoptimalkan performa model.

REFERENCES

- [1] M. K. S. A. Dr. Roedi Irawan, *Kelainan Genetik dan Diagnosis Sindrom Down*. Surabaya: Airlangga University Press, 2021, ISBN: 978-602-473-759-7.
- [2] T. O. Natania, R. Larasati, and E. Purwaningsih, "SYSTEMATIC LITERATURE REVIEW: PEMELIHARAAN KESEHATAN GIGI DAN MULUT PENYANDANG DOWN SYNDROME DITINJAU DARI PERAN ORANG TUA," *Jurnal Kesehatan Gigi dan Mulut (JKGM)*, vol. 3, no. 2, pp. 47–54, Nov. 2021, doi: 10.36086/jkgm.v3i2.909.
- [3] S. N. Siti and S. Sudarman, "Hubungan antara Kemampuan Mengunyah Makanan dengan Kemampuan Artikulasi pada Anak Down Syndrome di Samarinda," *Jurnal Terapi Wicara dan Bahasa*, vol. 1, no. 2, pp. 347–360, Jun. 2023, doi: 10.59686/jtwb.v1i2.39.

- [4] S. E. Antonarakis *et al.*, “Down syndrome,” *Nat Rev Dis Primers*, vol. 6, no. 1, p. 9, Feb. 2020, doi: 10.1038/s41572-019-0143-7.
- [5] N. D. A. Lubis, H. Sasmita, S. Amelia, and A. Yosi, “Promoting Positive Deviance from PIK POTADS North Sumatra to Improve the Quality of Life of Down Syndrome Children,” *ABDIMAS TALENTA: Jurnal Pengabdian Kepada Masyarakat*, vol. 8, no. 2, pp. 995–1000, Dec. 2023, doi: 10.32734/abdimastalenta.v8i2.11540.
- [6] L. Chen *et al.*, “Global, Regional, and National Burden and Trends of Down Syndrome From 1990 to 2019,” *Front Genet*, vol. 13, no. July, pp. 1–14, Jul. 2022, doi: 10.3389/fgene.2022.908482.
- [7] S. Hussain *et al.*, “Modern Diagnostic Imaging Technique Applications and Risk Factors in the Medical Field: A Review,” *Biomed Res Int*, vol. 2022, no. 1, p. 5164970, Jan. 2022, doi: 10.1155/2022/5164970.
- [8] N. A. Ujilast, N. S. Firdausita, C. S. K. Aditya, and Y. Azhar, “MRI Image Based Alzheimer’s Disease Classification Using Convolutional Neural Network: EfficientNet Architecture,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 1, pp. 18–25, Jan. 2024, doi: 10.29207/resti.v8i1.5457.
- [9] N. Paredes, E. Caicedo-Bravo, and B. Bacca, “Emotion Recognition in Individuals with Down Syndrome: A Convolutional Neural Network-Based Algorithm Proposal,” *Symmetry (Basel)*, vol. 15, no. 7, p. 1435, Jul. 2023, doi: 10.3390/sym15071435.
- [10] F. F. Sherif, N. Tawfik, D. Mousa, M. S. Abdallah, and Y.-I. Cho, “Automated Multi-Class Facial Syndrome Classification Using Transfer Learning Techniques,” *Bioengineering*, vol. 11, no. 8, p. 827, Aug. 2024, doi: 10.3390/bioengineering11080827.
- [11] E. Setyati, S. Az, S. P. Hudiono, and F. Kurniawan, “CNN based Face Recognition System for Patients with Down and William Syndrome,” *Knowledge Engineering and Data Science*, vol. 4, no. 2, p. 138, Dec. 2021, doi: 10.17977/um018v4i22021p138-144.
- [12] M. D. A. Pranatha, G. H. Setiawan, and M. A. Maricar, “Utilization of ResNet Architecture and Transfer Learning Method in the Classification of Faces of Individuals with Down Syndrome,” *Journal of Applied Informatics and Computing*, vol. 8, no. 2, pp. 434–442, Nov. 2024, doi: 10.30871/jaic.v8i2.8474.
- [13] N. Paredes, E. F. Caicedo-Bravo, B. Bacca, and G. Olmedo, “Emotion Recognition of Down Syndrome People Based on the Evaluation of Artificial Intelligence and Statistical Analysis Methods,” *Symmetry (Basel)*, vol. 14, no. 12, p. 2492, Nov. 2022, doi: 10.3390/sym14122492.
- [14] L. Alzubaidi *et al.*, “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,” *J Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.
- [15] K. Rezaee, “Machine learning and facial recognition for down syndrome detection: A comprehensive review,” *Computers in Human Behavior Reports*, vol. 17, p. 100600, Mar. 2025, doi: 10.1016/j.chbr.2025.100600.
- [16] Y. R., V. Raja Sarobin M., R. Panjanathan, G. J. S., and J. A. L., “Diabetic Retinopathy Classification Using CNN and Hybrid Deep Convolutional Neural Networks,” *Symmetry (Basel)*, vol. 14, no. 9, p. 1932, Sep. 2022, doi: 10.3390/sym14091932.
- [17] W. Setiawan, “PERBANDINGAN ARSITEKTUR CONVOLUTIONAL NEURAL NETWORK UNTUK KLASIFIKASI FUNDUS,” *Jurnal Simantec*, vol. 7, no. 2, pp. 48–53, Jun. 2020, doi: 10.21107/simantec.v7i2.6551.
- [18] W. M. Pradnya D and A. P. Kusumaningtyas, “Analisis Pengaruh Data Augmentasi Pada Klasifikasi Bumbu Dapur Menggunakan Convolutional Neural Network,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 4, p. 2022, Oct. 2022, doi: 10.30865/mib.v6i4.4201.
- [19] T. B. Sasongko, H. Haryoko, and A. Amrullah, “Analisis Efek Augmentasi Dataset dan Fine Tune pada Algoritma Pre-Trained Convolutional Neural Network (CNN),” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 4, pp. 763–768, Aug. 2023, doi: 10.25126/jtiik.20241046583.
- [20] E. Rachmawati, N. A. Agustina, and F. Sthevanie, “Pengenalan Ras Berdasarkan Hidung Dan Mulut Menggunakan Gray Level Co-Occurrence Matrix,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 4, pp. 729–734, Jul. 2021, doi: 10.25126/jtiik.2021844366.
- [21] Y. N. FUADAH, I. D. UBAIDULLAH, N. IBRAHIM, F. F. TALININGSING, N. K. SY, and M. A. PRAMUDITHO, “Optimasi Convolutional Neural Network dan K-Fold Cross Validation pada Sistem Klasifikasi Glaukoma,” *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, vol. 10, no. 3, p. 728, Jul. 2022, doi: 10.26760/elkomika.v10i3.728.
- [22] W. Setiawan, *Deep Learning menggunakan Convolutional Neural Network: Teori dan Aplikasi*, Cetakan I. Malang: Media Nusa Creative (MNC Publishing), 2021, ISBN: 978-602-462-446-0.
- [23] B. Nugroho and E. Y. Puspaningrum, “Kinerja Metode CNN untuk Klasifikasi Pneumonia dengan Variasi Ukuran Citra Input,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 3, pp. 533–538, Jun. 2021, doi: 10.25126/jtiik.2021834515.
- [24] Weny Indah Kusumawati and Adisaputra Zidha Noorizki, “Perbandingan Performa Algoritma VGG16 Dan VGG19 Melalui Metode CNN Untuk Klasifikasi Varietas Beras,” *Journal of Computer, Electronic, and Telecommunication*, vol. 4, no. 2, Dec. 2023, doi: 10.52435/complete.v4i2.387.
- [25] T. A. R. Ramadhani and A. R. Manga, “Analisis Perbandingan VGG-16 dan ResNet50 untuk Klasifikasi Multilabel Gambar Kerbau Toraja: Pendekatan Deep Learning,” *Jurnal Teknik*, vol. 22, no. 2, pp. 71–81, Dec. 2024, doi: 10.37031/jt.v22i2.490.
- [26] M. K. Insani and D. B. Santoso, “Perbandingan Kinerja Model Pre-Trained CNN (VGG16, RESNET, dan INCEPTIONV3) untuk Aplikasi Pengenalan Wajah pada Sistem Absensi Karyawan,” *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 5, no. 3, pp. 2612–2622, Sep. 2024, doi: 10.35870/jimik.v5i3.925.
- [27] R. A. Mas’ud and Junta Zeniarja, “Optimasi Convolutional Neural Networks untuk Deteksi Kanker Payudara menggunakan Arsitektur DenseNet,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 1, pp. 310–318, Jun. 2024, doi: 10.29408/edumatic.v8i1.25883.
- [28] M. Diwakaran and D. Surendran, “Breast Cancer Prognosis Based on Transfer Learning Techniques in Deep Neural Networks,” *Information Technology and Control*, vol. 52, no. 2, pp. 381–396, Jul. 2023, doi: 10.5755/j01.itc.52.2.33208.