

COMPARISON OF VGG16 AND VGG19 MODELS IN THE CLASSIFICATION OF DOWN SYNDROME IN THE EUROPEAN REGION WITH TRANSFER LEARNING

PERBANDINGAN MODEL VGG16 DAN VGG19 DALAM KLASIFIKASI DOWN SYNDROME DI WILAYAH EROPA DENGAN TRANSFER LEARNING

Excel Bima Evansyah¹, Christian Sri Kusuma²

^{1,2}Universitas Muhammadiyah Malang

Jl. Raya Tlogomas No.246, Babatan, Tegalondo, Kec. Lowokwaru, Kota Malang, Jawa Timur 65144

excelbimaa@webmail.umm.ac.id¹, christianskaditya@umm.ac.id²

Abstract - Down syndrome detection by utilizing facial images as the main data has been widely developed through deep learning approaches, especially Convolutional Neural Network (CNN). However, most studies only classify the disorder without paying attention to regional factors. This has limited the effectiveness of the model in the classification of Down syndrome, especially in populations in European regions that have different morphological characteristics. This study examines the performance of two pretrained CNN models, namely VGG16 and VGG19, in classifying facial images of children from Europe who are divided into 2 categories of Down Syndrome and Healthy. The dataset used in the study consists of 1,543 images from the Down syndrome class 671 images and the Healthy class 872 images. It was then expanded to 1570 images to balance the data between both Down syndrome and Healthy classes. The evaluation results of this research by applying augmentation show that the VGG16 model has superior performance compared to VGG19, with accuracy reaching 94%. Meanwhile, the VGG19 model obtained an accuracy of 90%. This difference shows that the VGG16 model has a more stable performance in detecting both categories with a better balance between precision and recall. This research is limited to European children's image data and still does not exist for ethnic teenagers or the elderly. This provides a basis for the development of facial image-based early detection systems, particularly for clinical applications or early screening in areas with similar populations.

Keywords - Augmentation, Classification Down Syndrome, European Region, VGG16, VGG19.

Abstrak - Deteksi Down syndrome dengan memanfaatkan citra wajah sebagai data utamanya telah banyak dikembangkan melalui pendekatan *deep learning*, khususnya *Convolutional Neural Network (CNN)*. Namun, sebagian besar penelitian hanya melakukan klasifikasi kelainan tersebut tanpa memperhatikan faktor wilayahnya. Hal ini menimbulkan keterbatasan efektivitas model dalam klasifikasi *Down syndrome*, khususnya pada populasi di wilayah Eropa yang memiliki karakteristik morfologis berbeda. Penelitian ini mengkaji performa dua model *pretrained CNN*, yaitu VGG16 dan VGG19, dalam mengklasifikasikan citra wajah anak-anak asal Eropa yang terbagi menjadi 2 kategori *Down Syndrome* dan *Healthy*. Dataset yang digunakan dalam penelitian terdiri 1.543 gambar dari kelas *Down syndrome* 671 gambar dan kelas *Healthy* 872 gambar. Kemudian diperluas menjadi 1570 gambar untuk menyeimbangkan data antara kedua kelas *Down syndrome* dan *Healthy*. Hasil evaluasi penelitian ini dengan menerapkan augmentasi menunjukkan bahwa model VGG16 memiliki performa yang lebih unggul dibandingkan VGG19, dengan akurasi mencapai 94%. Sementara itu, model VGG19 memperoleh akurasi sebesar 90%. Perbedaan ini menunjukkan bahwa model VGG16 memiliki performa yang lebih stabil dalam mendeteksi kedua kategori dengan nilai keseimbangan antara presisi dan recall yang lebih baik. Penelitian ini terbatas pada data citra anak-anak Eropa dan masih belum ada terhadap etnis remaja atau lanjut usia. Sehingga menjadi dasar pengembangan sistem deteksi dini berbasis citra wajah, khususnya untuk aplikasi klinis atau skrining awal di wilayah dengan populasi serupa.

Kata Kunci - Augmentasi, Klasifikasi Down Syndrome, Wilayah Eropa, VGG16, VGG19.

I. PENDAHULUAN

Down syndrome merupakan kelainan genetik yang dapat diidentifikasi sejak dalam kandungan yang terjadi akibat dari abnormalitas kromosom, yakni terdapat salinan tambahan terhadap kromosom 21 sehingga terjadi total 47 kromosom. Penambahan jumlah salinan kromosom pada janin disebabkan oleh kegagalan pembelahan sel yang terjadi dalam fase perkembangan embrio atau nondisjunction embrio[1]. Untuk mengetahui jumlah salinan kromosom pada janin, seorang wanita yang sedang mengandung melakukan pengecekan kromosom yang disebut kariotipe. Fungsi dari pengecekan tersebut agar dapat mengetahui diagnosis secara dini hipotiroid kongenital yang melalau skrining neonatal sebagai pencegah kecacatan pengaruh gangguan saraf[2]. Selain itu, penyandang *Down syndrome* mempunyai kelemahan tonus otot sehingga menimbulkan efek dalam perkembangan fisik, perkembangan motorik kasar, tumbuh kembang anak dan abnormalitas pada kerangka tubuh[3]. Dengan demikian, kelainan *Down syndrome* juga dapat diidentifikasi pasca-lahir dengan memperhatikan struktur dan karakteristik wajah yang dapat dilihat perbedaannya secara langsung, seperti mata yang miring ke atas dari sudut dalam ke sudut luar, telinga kecil yang memiliki posisi lebih rendah dari posisi normal, bentuk kepala yang datar serta memiliki ukuran mulut yang kecil, dan bentuk hidung yang kecil dengan jembatan hidung yang datar[4].

Menurut data yang dilansir oleh *World Health Organization* (WHO), sekitar 3.000 hingga 5.000 bayi lahir dengan keadaan *Down syndrome* di setiap tahunnya[5]. Kasus *Down syndrome* memiliki berbagai varian, dengan persentase terbesar ditemukan pada varian *trisomy 21* dengan jumlah kasus sebesar 94%, diikuti dengan jumlah varian translokasi sebesar 3-4% dan varian mozaik atau *mosaicism* sebesar 2-4%[6]. Secara umum, ketiga variasi tersebut memiliki kemiripan dalam karakteristik wajah yang dapat dikenali secara visual, namun tetap memiliki variasi antarindividu. Selain itu, karakteristik wajah berdasarkan regionalnya juga memiliki perbedaan tiap wilayah, terutama Eropa. Keterbatasan ini memunculkan tantangan dalam mengenali morfologi wajah yang berbeda, seperti anak-anak Eropa, yang memiliki karakteristik etnis dan struktur wajah yang khas. Sekitar 8.031 kelahiran anak dengan *Down syndrome* setiap tahun pada periode 2011-2015. Tingkat kelahiran hidup mencapai 10,1 per 10.000 kelahiran, namun angka tersebut mampu meningkat sebesar 21,7 per 10.000 kelahiran apabila tidak dilakukan terminasi selektif pasca deteksi prenatal [7]. Oleh karena itu, wilayah Eropa dikategorikan sebagai *High Socio-Demographic Index* (SDI), yang mana wilayah tersebut cenderung maju dan berkembang sehingga mampu melakukan terminasi selektif[8].

Modernisasi di bidang teknologi informasi dan komputer memudahkan untuk penggunaan identifikasi penyakit secara komprehensif [9]. Model *deep learning* mampu melakukan ekstraksi fitur secara optimal dari citra wajah tanpa memerlukan *feature engineering* manual, sehingga mempercepat proses diagnosis. Penggunaan teknologi dalam beberapa tahun terakhir telah digunakan pada klasifikasi citra klinis, terutama *Down syndrome*. Berdasarkan penelitian terdahulu yang dilakukan oleh Parades dkk., penerapan model CNN telah berhasil menunjukkan akurasi tertinggi dibandingkan dengan model SVM, mini-Xception, dan Resnet34 dalam kasus klasifikasi *Down syndrome*, yakni sebesar 91,48% [10]. Kasus klasifikasi *Down syndrome* yang menggunakan model serupa juga ditemukan dalam penelitian yang dilakukan oleh Sherif dkk., dengan akurasi tertinggi mencapai 90% [11]. Berdasarkan penelitian sebelumnya, implementasi model *deep learning*, terutama CNN, menunjukkan adanya kemungkinan besar dalam melakukan tugas klasifikasi *Down syndrome* menggunakan data dalam bentuk citra. Oleh karena itu, penelitian ini bertujuan untuk mengetahui dan menganalisa efektifitas model dalam klasifikasi *Down syndrome* berdasarkan wilayah Eropa. Selanjutnya kebaruan dalam penelitian ini terletak pada integrasi arsitektur CNN yang dioptimalkan dengan model *transfer learning*, yaitu VGG16 dan VGG19. Implementasi pemilihan model tersebut dikarenakan memiliki tingkat akurasi yang baik dan jumlah layer yang sesuai, sehingga model berlatih untuk memahami dan mempelajari fitur-fitur yang kompleks [12].

II. SIGNIFIKASI STUDI

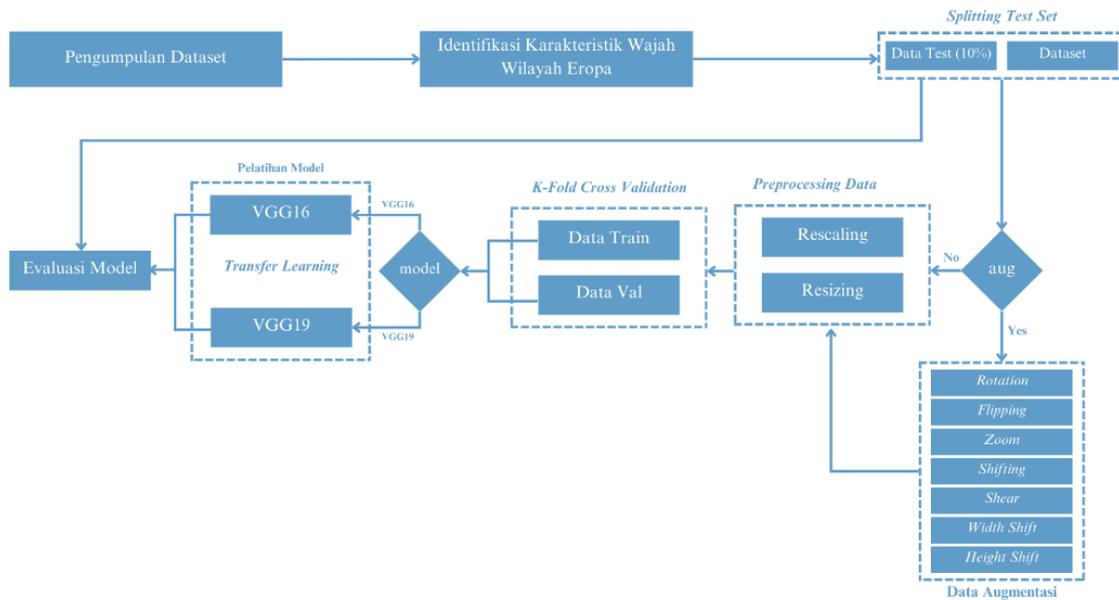
A. Penelitian Terdahulu

Penelitian ini berkontribusi dalam ranah klasifikasi citra medis dengan menyoroti perbandingan performa dua arsitektur *Convolutional Neural Network* menggunakan *transfer learning* dengan model VGG16 dan VGG19 dalam mengklasifikasikan wajah anak-anak *Downsyndrome* wilayah Eropa. Beberapa penelitian sebelumnya yang umumnya hanya menggunakan data umum atau tidak menyebutkan asal populasi, penelitian ini memanfaatkan dataset khusus anak-anak *Down syndrome* di wilayah Eropa, yang relevan mengingat wilayah Eropa memiliki karakteristik High-SDI dengan kecenderungan angka kelahiran hidup *Down syndrome* yang lebih rendah akibat intervensi prenatal [8]. Meskipun demikian, keberadaan citra wajah dari anak-anak *Downs syndrome* sangat penting untuk pengembangan sistem klasifikasi berbasis visual dalam konteks teknologi pengenalan pola. Signifikan praktis dari penelitian ini terletak pada mengenai eektivitas arsitektur VGG16 dan VGG19 dalam mengenali karakteristik wajah *Down syndrome*, yang dapat menjadi dasar pengembangan sistem klasifikasi otomatis dalam skenario non-klinis seperti sistem pendukung edukasi, pelabelan data atau basis untuk penelitian biometrik. Dengan menerapkan pendekatan *transfer learning* dan augmentasi data, penelitian ini juga menunjukkan bagaimana performa model tetap dapat dioptimalkan.

Dalam beberapa penelitian sebelumnya, CNN telah terbukti berhasil mengidentifikasi kelainan *Down syndrome* menggunakan data citra wajah. Selanjutnya, terdapat penelitian yang dilakukan oleh Sherif dkk. [11], ditemukan bahwa penggunaan CNN menggunakan model *transfer learning* dengan model VGG-Face terbukti mampu menghasilkan nilai akurasi tertinggi dibandingkan dengan model lainnya, yaitu VGG16, ResNet-50, dan ResNet-152. Model yang dibangun menunjukkan eektivitas model transfer learning dan augmentasi data untuk peningkatan akurasi dalam klasifikasi wajah sindrom multi-kelas dengan ditemukan accuracy sebesar 90%. Berdasarkan penelitian yang dilakukan oleh Pranatha et al, model CNN dibangun menggunakan arsitektur ResNet. Metode yang digunakan pada penelitian tersebut dengan menggunakan metode *transfer learning*. Penelitian tersebut berfokus pada klasifikasi wajah penderita *Down syndrome* dengan memperoleh hasil akurasi yang dihasilkan sebesar 87,88% [13]. Kontribusi utama dari penelitian ini adalah evaluasi kuantitatif dan komparatif terhadap dua model CNN pretrained dalam konteks data Down Syndrome regional, yang hingga kini masih jarang dikaji secara spesifik. Penggunaan validasi 5-fold dan analisis *confusion matrix* menambah validasi dalam menilai stabilitas dan akurasi model, sehingga hasilnya dapat menjadi referensi bagi peneliti lain yang ingin membangun atau menyempurnakan sistem klasifikasi berbasis visual untuk kasus medis serupa.

B. Metode Penelitian

Penelitian ini mengusulkan untuk mengembangkan metode pendekatan *deep learning* dengan menggunakan dua arsitektur VGG16 dan VGG19, dengan beberapa tahapan yang disajikan pada Gambar 1. Mulai dari pengumpulan dataset, identifikasi karakteristik fisik wajah, pembagian dataset, pemodelan metode dan evaluasi model.



Gambar 1. Diagram Alur Metode Penelitian

1. Pengumpulan Dataset

Tahap awal penelitian ini dimulai dengan pengumpulan data citra wajah anak dengan kondisi *Down syndrome* dan *Healthy*. Dataset diperoleh dari situs *Roboflow* dengan judul “*ds Computer Vision Project*”, terdiri dari dua kelas utama: *Down syndrome* 1.499 gambar dan *Healthy* 1.498 gambar, dengan total 2.997 gambar. Contoh visualisasi masing-masing kelas ditampilkan pada Gambar 2.



Gambar 2. Visualisasi dataset dari *Down syndrome* dan *Healthy*

2. Identifikasi Karakteristik Wajah di Wilayah Eropa

Penelitian ini melaksanakan identifikasi karakteristik wajah anak pada wilayah Eropa untuk mengoptimalkan efektifitas model dalam klasifikasi *Down syndrome*. Proses identifikasi dilakukan secara manual dengan memperhatikan beberapa kriteria berdasarkan karakteristik wajah dari wilayah Eropa [14]. Beberapa kriteria yang diamati lebih lanjut untuk proses identifikasi, disajikan pada Tabel 1.

TABEL I
KRITERIA KARAKTERISTIK WAJAH DOWNSYNDROME WILAYAH EROPA

<i>Karakteristik Fisik</i>	<i>Eropa</i>
Mata	Kelopak Mata ganda, palung air mata terbuka, besar
Hidung	Mancung
Bibir	Tipis, kencang
Bentuk Wajah	Pusat wajah menonjol ke luar, berbentuk baji, wajah panjang
Rambut	Tebal, bergelombang atau keriting
Kulit	Terang atau coklat

Setelah melakukan identifikasi berdasarkan Tabel 1, dataset yang digunakan dalam penelitian ini adalah gambar yang sudah diidentifikasi pada seperti bentuk mata, hidung, wajah, rambut dan kulit. Hasil distribusi data dari masing-masing kelas menunjukkan angka yang cukup seimbang antara lain kelas *Down syndrome* 671 gambar dan kelas *Healthy* 872 gambar, sehingga model yang dirancang dapat melakukan tugas klasifikasi secara optimal.

3. *Splitting Test Set*

Untuk menjaga orisinalitas data uji, sebanyak 10% data dari masing-masing kelas dipisahkan sebagai *test set* sebelum dilakukan *preprocessing*. Langkah ini bertujuan mencegah penyebaran data yang dapat memengaruhi hasil evaluasi, sehingga model dapat diuji secara optimal. Distribusi data *test set* pada kelas *Down syndrome* berjumlah 67 dan *Healthy* berjumlah 50.

4. *Data Augmentasi*

Selain itu, teknik augmentasi diterapkan untuk meningkatkan proporsi data pada kelas minoritas, sehingga distribusi antar kelas menjadi lebih seimbang [9]. Hal tersebut dilakukan untuk meminimalisir *overfitting* dan meningkatkan kinerja model agar dapat mempelajari fitur data secara optimal. Rincian diterapkan secara variatif pada setiap kelas augmentasi disajikan pada Tabel 2.

TABEL II
NILAI DATA AUGMENTASI

<i>Data Augmentasi</i>	<i>Nilai</i>
<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>	True
<i>Fiil Mode</i>	Nearest

5. *Pra-pemrosesan Gambar*

Pada langkah ini, data dipersiapkan sebelum pelatihan model. Pra-pemrosesan gambar yang dilakukan meliputi *rescaling* dan *resizing*. *Rescaling* mengubah rentang nilai piksel dari 0-255 menjadi 0-1 dengan membagi setiap nilai piksel dengan 255. *Resizing* menyesuaikan ukuran gambar menjadi 244 × 244 piksel, sesuai dengan input model.

6. *Perancangan Model Convolutional Neural Network*

Rancangan penyusunan menggunakan model aritektur VGG-16 dan VGG19 dengan nilai parameter yang sama. Hal tersebut dilakukan untuk melihat performa dan kinerja dari kedua model dalam klasifikasi *Down syndrome* di wilayah Eropa. Model *Convolutional Neural Network* (CNN) memiliki puluhan hingga ratusan lapisan dalam memproses data citra melalui lapisan jaringan dan menghasilkan sebuah output tertentu. Aritektur VGG-16 memiliki jumlah lapisan sebanyak 16 lapisan, yang terdiri dari 13 *convolutional layer* dan Arisetktur VGG-19 memiliki jumlah lapisan yang sedikit lebih kompleks dibandingkan VGG-16, yaitu sebanyak 18 lapisan dengan terdiri dari 16 *convolutional layer* [15]. Perbedaan antara kedua model tersebut dapat dilihat dari jumlah lapisan konvolusinya. Pada lapisan tersebut, base model berfungsi untuk melakukan ekstraksi fitur

dengan bobot yang telah dilatih pada data dalam jumlah besar, yaitu *ImageNet*. Hal tersebut dinamakan dengan *transfer learning* yang memiliki tujuan untuk meningkatkan performa dari model dalam mempelajari data dengan memanfaatkan pengetahuan dari data yang telah dilatih sebelumnya [12]. Selanjutnya, teknik *transfer learning* lain yang digunakan dalam penelitian ini adalah *freezing layer*, di mana sebagian besar lapisan dibekukan untuk mempertahankan pengetahuan awal, kecuali lima lapisan terakhir yang dilatih ulang agar dapat menyesuaikan dengan dataset spesifik. Proses pembelajaran utama terjadi pada bagian *fully connected layer*, dengan penyesuaian bobot antar-neuron. Untuk mencegah overfitting, digunakan *dropout layer* dengan nilai 0.5. Selain itu, pelatihan model juga melibatkan parameter penting seperti *loss function*, *optimizer*, dan *evaluation metrics*, yang dirangkum pada Tabel 3.

TABEL III
NILAI PARAMETER MODEL PELATIHAN

<i>Parameter</i>	<i>Nilai</i>
<i>Batch Size</i>	32
<i>Number of Epoch</i>	15
<i>Loss Function</i>	<i>Categorical Crossentropy</i>
<i>Evaluation Metrics</i>	<i>Accuracy</i>
<i>Optimizer</i>	Adam (lr = 0.001)

Selain parameter utama, penelitian ini juga menerapkan *callback function* berupa *EarlyStopping* untuk menghentikan pelatihan secara otomatis jika tidak terjadi peningkatan performa. Fungsi ini memantau nilai *val_accuracy* dengan *patience* sebesar 15, yang berarti pelatihan akan dihentikan jika dalam 15 *epoch* berturut-turut tidak ada peningkatan signifikan. Dalam penelitian ini, evaluasi model yang digunakan meliputi dua komponen, yaitu *Cross Validation* dan *Confusion Matrix*. Penerapan metode CV atau cross-validation menggunakan *5-fold cross validation* karena mampu membantu dalam mengukur performa model dalam melakukan generalisasi data [16]. Hasil *accuracy* dari tiap *fold* disimpan dan dihitung rata-ratanya untuk menentukan performa model secara menyeluruh. Metode evaluasi model lain yang digunakan dalam penelitian adalah confusion matrix yang berfungsi untuk mengukur performa dari model yang digunakan [9]. Untuk merepresentasikan visual confusion matrix, terdapat empat komponen utama, yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, *False Negative (FN)*. Selain itu, terdapat beberapa metrik yang digunakan untuk mengukur dan mengevaluasi kinerja dari model klasifikasi yang telah dibangun, yaitu *accuracy*, *precision*, *recall*, dan *f1-score*.

III. HASIL DAN PEMBAHASAN

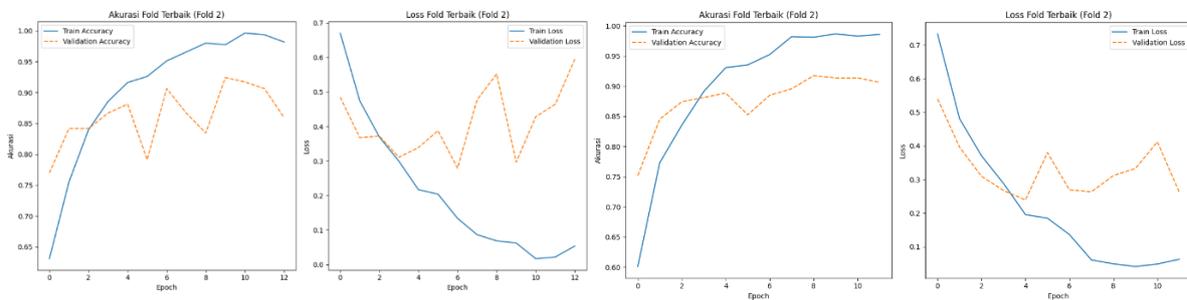
A. Hasil Pelatihan Model VGG16 dan VGG19 tanpa Augmentasi

Dataset yang telah melewati *image preprocessing*, selanjutnya diuji dengan menggunakan model yang ditentukan. Nilai-nilai parameter yang digunakan pada model untuk melakukan klasifikasi *Down syndrome* di wilayah Eropa, dapat dilihat pada Tabel 2 dan Tabel 3. Pelatihan model dilakukan menggunakan *5-Fold Cross Validation* dan hasil tiap *fold* pada kedua model disajikan pada Tabel 4.

TABEL IV
HASIL PENGARUH 5-FOLD CROSS VALIDATION TANPA AUGMENTASI

Model	Iterasi	Acc (%)	Loss (%)
VGG16	Fold – 1	0.8993	0.2610
	Fold – 2	0.9245	0.2791
	Fold – 3	0.8561	0.3665
	Fold – 4	0.8993	0.2833
	Fold – 5	0.9097	0.2261
	Mean	0.8978	0.2832
VGG19	Fold – 1	0.9137	0.2671
	Fold – 2	0.9173	0.2390
	Fold – 3	0.9101	0.2978
	Fold – 4	0.9101	0.2428
	Fold – 5	0.8953	0.271
	Mean	0.9093	0.2650

Berdasarkan Tabel 4, hasil *5-Fold Cross-Validation* menunjukkan bahwa VGG19 memiliki rata-rata akurasi lebih tinggi 90.93% dibanding VGG16 89.78%, dengan selisih 1.15%. Namun, VGG16 menunjukkan nilai loss yang lebih rendah, menandakan stabilitas model yang lebih baik. Performa masing-masing fold dapat dilihat pada Gambar 3.



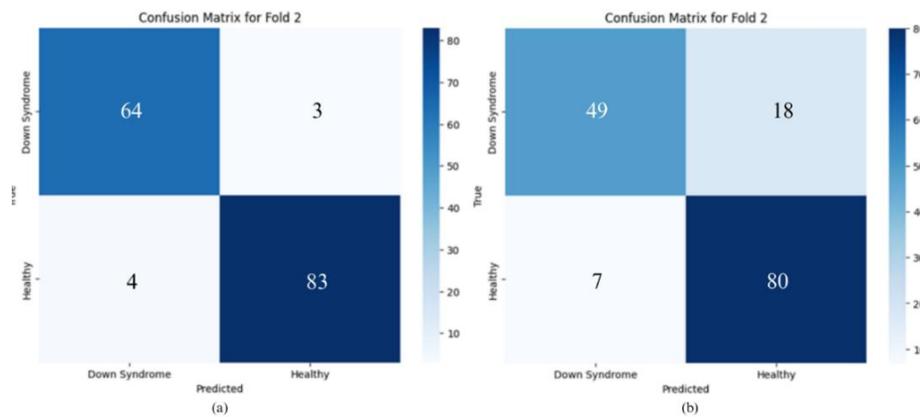
Gambar 3. (a) Grafik Akurasi dan Loss VGG16 dan (b) Grafik Akurasi dan Loss VGG19

Pada Fold-2 Gambar 3, akurasi VGG16 dan VGG19 meningkat seiring *epoch*. VGG19 menunjukkan *train accuracy* yang cepat naik hingga mendekati 1,0, namun *validation accuracy* dan *loss-nya* fluktuatif, mengindikasikan *overfitting*. Sebaliknya, VGG16 memiliki peningkatan akurasi yang lebih stabil dan *validation loss* yang lebih konsisten, menandakan generalisasi yang lebih baik. Meski akurasi VGG19 lebih tinggi, VGG16 lebih andal dalam menangani data baru.

TABEL V
HASIL EVALUASI MODEL ITERASI TERBAIK VGG16 DAN VGG19

Model	Class	Fold	Precision (%)	Recall (%)	F1-Score (%)	Acc (%)
VGG16	<i>Down Syndrome</i>	Fold-2	0.94	0.96	0.95	0.95
	<i>Healthy</i>		0.97	0.95	0.96	
VGG19	<i>Down Syndrome</i>	Fold-2	0.88	0.73	0.80	0.84
	<i>Healthy</i>		0.82	0.92	0.86	

Evaluasi pada Fold-2 Tabel 5 menunjukkan bahwa VGG16 memiliki performa lebih baik dibanding VGG19. VGG16 meraih akurasi 95% dengan skor F1 tinggi dan seimbang pada kedua kelas 0.95 untuk *Down syndrome*, 0.96 untuk *Healthy*, menandakan keseimbangan precision dan recall yang optimal. Sebaliknya, VGG19 hanya mencapai akurasi 84%, dengan F1-score yang lebih rendah khususnya pada kelas *Down syndrome* (0.80), menunjukkan kecenderungan salah klasifikasi pada kelas tersebut. Secara keseluruhan, VGG16 lebih andal dalam mendeteksi kedua kelas.



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

Berdasarkan visualisasi confusion matrix pada Fold-2 Gambar 4 (a), VGG16 menunjukkan performa stabil dengan 64 sampel *Down syndrome* dan 83 sampel *Healthy* terklasifikasi dengan benar, serta hanya 3 dan 4 kesalahan klasifikasi. Hal ini menunjukkan VGG16 andal dalam membedakan kedua kategori. Sebaliknya, pada Gambar (b), VGG19 menunjukkan lebih banyak kesalahan: 49 sampel *Down syndrome* benar, namun 18 salah sebagai *Healthy*, dan 80 sampel *Healthy* benar, dengan 7 kesalahan sebagai *Down syndrome*. Meskipun VGG19 memiliki akurasi lebih tinggi, hasil ini menunjukkan *overfitting* dan kurangnya stabilitas, sementara VGG16 lebih konsisten dalam membedakan kedua kelas.



Gambar 5. Visualisasi Prediksi Salah Model VGG16

Melihat hasil *confusion matrix* pada model VGG16, terlihat bahwa model masih mengalami kesalahan dalam mengklasifikasikan kelas *Down syndrome* dan *Healthy*. Pada beberapa kasus, kelas *Down syndrome* diprediksi sebagai *Healthy*, yang mengindikasikan bahwa model belum sepenuhnya mampu menangkap fitur kelas *Down syndrome* secara konsisten. Sebaliknya, ada juga kelas *Healthy* yang diprediksi sebagai *Down syndrome*, yang menunjukkan bahwa model terlalu sensitif terhadap karakteristik tertentu seperti ekspresi wajah, sudut pengambilan gambar, atau pencahayaan. Kesalahan ini dapat disebabkan oleh keterbatasan dalam representasi fitur yang digunakan oleh model, sehingga menghasilkan *misclassification*.

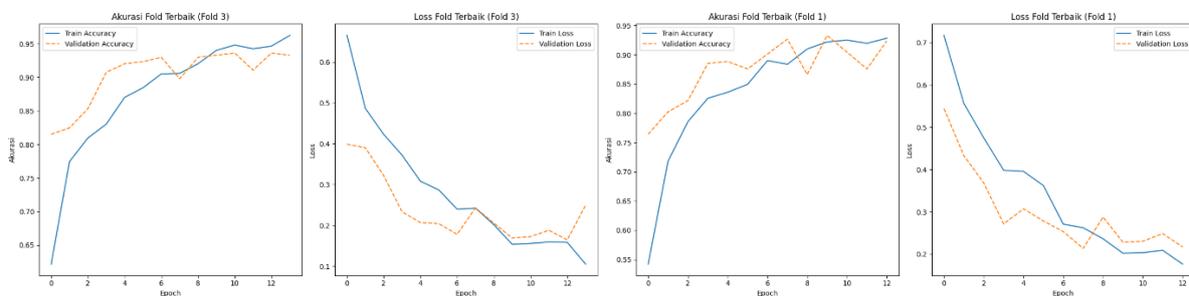
B. Hasil Pelatihan Model VGG16 dan VGG19 dengan Augmentasi

Teknik data augmentasi diterapkan untuk meningkatkan keragaman data latih guna memperbaiki generalisasi model dalam mengklasifikasikan *Down syndrome* di wilayah Eropa. Nilai-nilai parameter yang digunakan dalam model setelah menerapkan augmentasi data dapat dilihat pada Tabel 2 dan Tabel 3. Pelatihan model dilakukan menggunakan *5-Fold Cross Validation* untuk memastikan performa model yang lebih stabil dan mengurangi *overfitting*. Hasil evaluasi dari tiap fold pada kedua model disajikan pada Tabel 8.

TABEL VI
HASIL PENGARUH 5-FOLD CROSS VALIDATION DENGAN AUGMENTASI

Model	Iterasi	Acc (%)	Loss (%)
VGG16	Fold – 1	0.9140	0.2039
	Fold – 2	0.9108	0.2156
	Fold – 3	0.9363	0.1645
	Fold – 4	0.9045	0.2736
	Fold – 5	0.9045	0.2442
	Mean	0.9140	0.2204
VGG19	Fold – 1	0.9331	0.2137
	Fold – 2	0.8949	0.2723
	Fold – 3	0.8854	0.2402
	Fold – 4	0.8949	0.2515
	Fold – 5	0.9108	0.2621
	Mean	0.9038	0.2479

Hasil evaluasi pada Tabel 6 menunjukkan bahwa VGG16 memiliki performa lebih konsisten dibanding VGG19, dengan rata-rata akurasi 91,4% dan loss lebih rendah. VGG16 juga lebih efektif dalam memanfaatkan augmentasi data, sedangkan VGG19 cenderung *overfitting* dan kurang adaptif terhadap variasi data yang digunakan.



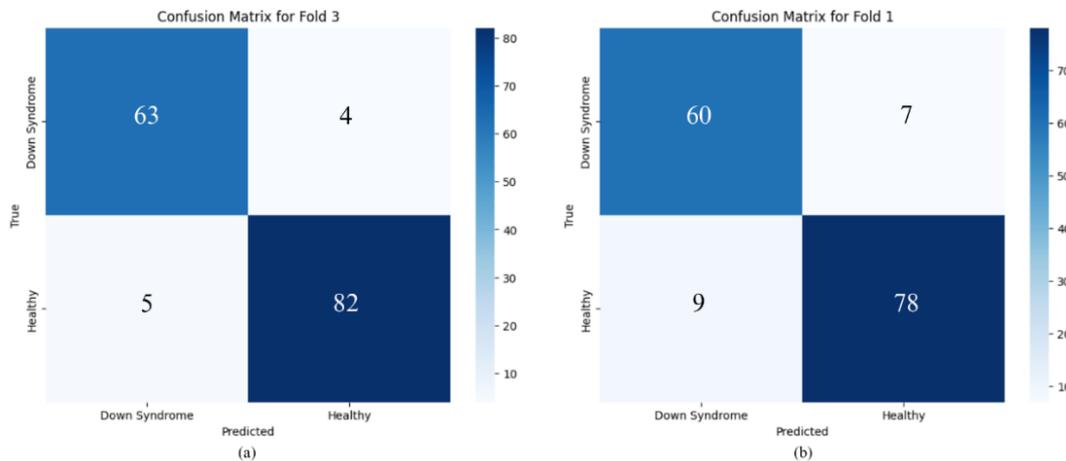
Gambar 6. (a) Grafik Akurasi dan Loss VGG16 dan (b) Grafik Akurasi dan Loss VGG19

Visualisasi hasil pelatihan menunjukkan perbandingan performa VGG16 dan VGG19 dalam hal akurasi dan loss. Pada grafik akurasi, VGG16 menunjukkan peningkatan stabil dengan akurasi validasi yang konsisten, menandakan model belajar dengan baik tanpa *overfitting*. Sebaliknya, VGG19 mengalami fluktuasi pada akurasi validasi, mengindikasikan ketidakstabilan dalam generalisasi. Untuk grafik loss, VGG16 menunjukkan penurunan loss yang stabil, baik pada *training* maupun validasi, sementara VGG19 menunjukkan ketidakseimbangan antara *loss training* dan *validation*, mengarah pada *overfitting*. Secara keseluruhan, VGG16 lebih stabil dan memiliki kinerja yang lebih baik setelah penerapan data augmentasi.

TABEL VII
HASIL EVALUASI AUGMENTASI ITERASI TERBAIK PADA MODEL VGG16 DAN VGG19

Model	Class	Fold	Precision (%)	Recall (%)	F1-Score (%)	Acc (%)
VGG16	<i>Down Syndrome</i>	Fold-3	0.93	0.94	0.93	0.94
	<i>Healthy</i>		0.95	0.94	0.95	
VGG19	<i>Down Syndrome</i>	Fold-1	0.87	0.90	0.88	0.90
	<i>Healthy</i>		0.92	0.90	0.91	

Setelah penerapan augmentasi data, performa VGG16 pada Fold-3 meningkat signifikan dengan akurasi 94%. *Precision* dan *recall* yang tinggi pada kedua kelas menghasilkan F1-score masing-masing 0.93 dan 0.95, menunjukkan keseimbangan prediksi yang baik. Sementara itu, VGG19 pada Fold-1 memperoleh akurasi 90%, dengan F1-score 0.91 untuk kelas *Down syndrome*. Meskipun VGG19 menunjukkan peningkatan dibandingkan sebelum augmentasi, VGG16 tetap unggul dalam stabilitas dan akurasi. Secara keseluruhan, augmentasi data memperbaiki kinerja kedua model, dengan VGG16 lebih stabil dalam memprediksi kedua kelas. Temuan ini menegaskan bahwa augmentasi meningkatkan generalisasi model terhadap variasi data.



Gambar 7. (a) Hasil Confusion Matrix VGG16 dan (b) Hasil Confusion Matrix VGG19

Pada *Confusion Matrix* Gambar 7 (a), model VGG16 berhasil mengklasifikasikan 63 sampel *Down syndrome* dan 82 sampel *Healthy* dengan benar, meskipun terdapat 4 sampel *Down syndrome* yang salah diklasifikasikan sebagai *Healthy* dan 5 sampel *Healthy* yang salah sebagai *Down syndrome*, menunjukkan akurasi yang tinggi meski masih ada kesalahan terutama dalam membedakan sampel *Down syndrome*. Di sisi lain, pada *Confusion Matrix* Gambar (b), model VGG19 menunjukkan lebih banyak kesalahan klasifikasi, dengan 60 sampel *Down syndrome* yang benar dan 7 salah sebagai *Healthy*, serta 78 sampel *Healthy* yang benar dan 9 salah sebagai *Down syndrome*. Meskipun augmentasi data meningkatkan performa, VGG16 tetap lebih stabil dibanding VGG19, dengan kedua model mengalami kesulitan dalam membedakan fitur wajah yang mirip antara kedua kelas tersebut.



Gambar 8. Hasil Visualisasi Prediksi Salah pada VGG16

Gambar di atas menunjukkan kesalahan prediksi model VGG16 dalam mengklasifikasikan sampel *Down syndrome* dan *Healthy*. Kesalahan pertama disebabkan oleh variasi ekspresi atau fitur wajah yang tidak sesuai pola khas *Down syndrome*, pencahayaan, ekspresi gambar dan pengambilan gambar. Kesalahan kedua terjadi karena beberapa sampel *Healthy* memiliki kemiripan fitur dengan *Down syndrome* atau karena model terlalu sensitif terhadap pola tertentu dalam data pelatihan.

C. Perbandingan Peforma Dua Model

Setelah hasil temuan yang telah dilakukan penerapan augmentasi data, VGG16 terbukti memiliki performa lebih unggul dibandingkan VGG19 dalam klasifikasi *Down syndrome* dan *Healthy*. Pada Fold-3, VGG16 mencapai akurasi 94% dengan F1-score 0.93 untuk *Down syndrome* dan 0.95 untuk *Healthy*, serta *precision* dan *recall* yang seimbang. Sebaliknya, VGG19 pada Fold-1 hanya mencapai akurasi 90%, dengan F1-score untuk *Down syndrome* sebesar 0.88. Perbedaan utama terlihat pada konsistensi keseimbangan antara *precision* dan *recall*, di mana VGG16 menunjukkan performa yang lebih stabil. Keunggulan ini berasal dari arsitektur VGG16 yang lebih sederhana, sehingga lebih efektif dalam mengekstraksi fitur penting tanpa mudah mengalami *overfitting*. Dengan demikian, VGG16 menjadi pilihan yang lebih baik dalam penelitian ini karena menawarkan akurasi tinggi serta keseimbangan performa antar kelas.

IV. KESIMPULAN

Penelitian ini bertujuan untuk membandingkan dua performa model CNN berbasis VGG16 dan VGG19, yang menerapkan dua pendekatan utama, yaitu transfer learning dan augmentasi data, dalam melakukan tugas klasifikasi *Down syndrome* pada populasi di wilayah Eropa. Berdasarkan hasil yang didapatkan, model VGG16 memiliki performa yang lebih baik dibandingkan VGG19, baik menggunakan augmentasi data maupun tidak menggunakan augmentasi data. Hasil yang didapatkan dalam klasifikasi *Down syndrome* tanpa augmentasi mencapai akurasi sebesar 95%, sedangkan yang menggunakan augmentasi mencapai akurasi sebesar 90%. Namun, hasil yang ditunjukkan pada model VGG16 tanpa augmentasi menunjukkan adanya fluktuatif pada beberapa epoch. Sementara, model VGG16 yang menggunakan augmentasi menunjukkan grafik yang lebih stabil. Terdapat sejumlah keterbatasan metodologis yang perlu dicermati. Dataset yang digunakan masih terbatas dalam representasi etnis dan variasi wajah anak-anak Eropa secara menyeluruh, yang dapat memengaruhi kemampuan generalisasi model. Selain itu, proses pelabelan data dilakukan secara manual, yang berpotensi menimbulkan bias subjektif terhadap karakteristik fenotip wajah anak dengan *Down syndrome*. Temuan *overfitting* pada salah satu model juga menandakan perlunya evaluasi lebih lanjut terhadap kompleksitas arsitektur dan strategi pelatihan. Sebagai pengembangan penelitian berikutnya, disarankan untuk memperluas cakupan dan keberagaman dataset yang digunakan, dengan mempertimbangkan dimensi variasi etnis, rentang usia, serta ekspresi wajah anak-anak di wilayah Eropa guna meningkatkan kemampuan generalisasi model. Selain itu, penerapan teknik fine-tuning lanjutan dan strategi augmentasi data yang lebih adaptif perlu dioptimalkan untuk mengurangi potensi *overfitting* yang teridentifikasi pada arsitektur antara lain EfficientNet maupun Vision Transformer, dikarenakan kemampuannya dalam merepresentasikan fitur yang lebih baik dalam memodelkan pola visual kompleks pada citra medis.

REFERENSI

- [1] 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.
- [2] S. Apriliani' and S. A. Saroh, "SEORANG ANAK LAKI-LAKI USIA 1 TAHUN SINDROM DOWN DENGAN HIPOTIROID KONGENITAL SUSPEK BRONKITIS : LAPORAN KASUS A 1 Year Boy Down Syndrome With Suspected Congenital Hypothyroid Bronchitis : Case Report," Sep. 2022.
- [3] D. G. Arvanda and A. N. Izzah, "Analisis Studi Kasus Mengenai Usia Ibu dan Gambaran Fisik Penderita Sindrom Down," *JURNAL BIOSENSE*, vol. 6, no. 02, pp. 206–218, Dec. 2023, doi: 10.36526/biosense.v6i02.3312.

- [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] A. R. Mardhia and N. Pransista, “Pendidikan Anak Berkebutuhan Khusus Pada Down Syndrome,” *PPSDP Undergraduate Journal of Educational Sciences*, vol. 2, no. 1, pp. 1–17, Mar. 2025, doi: 10.59175/pujes.v2i1.225.
- [6] A. Fadhia Haya and A. D. S. Fitri, “Hubungan antara Oral Motor Skill dengan Kejelasan Bicara pada Anak dengan Down Syndrome di Surakarta,” *Jurnal Terapi Wicara dan Bahasa*, vol. 3, no. 1, pp. 61–74, Dec. 2024, doi: 10.59686/jtwb.v3i1.140.
- [7] G. de Graaf, F. Buckley, and B. G. Skotko, “Estimation of the number of people with Down syndrome in Europe,” *European Journal of Human Genetics*, vol. 29, no. 3, pp. 402–410, Mar. 2021, doi: 10.1038/s41431-020-00748-y.
- [8] L. Chen *et al.*, “Global, Regional, and National Burden and Trends of Down Syndrome From 1990 to 2019,” *Front Genet*, vol. 13, Jul. 2022, doi: 10.3389/fgene.2022.908482.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] D. Sumarlie, C. Lubis, and T. Handhayani, “PENGENALAN KUE TRADISIONAL INDONESIA MENGGUNAKAN ALGORITMA CONVOLUTIONAL NEURAL NETWORK,” 2022. [Online]. Available: <https://arxiv.org/abs/1409.1556>,
- [13] 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.
- [14] 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.
- [15] 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.
- [16] J. Nur Hanifa and M. Yunita Wijaya, “PENERAPAN CROSS-VALIDATION PADA MODEL EFEK CAMPURAN: DAMPAK FAKTOR EKONOMI DAN KESEHATAN TERHADAP INDEKS KERENTANAN NEGARA-NEGARA DI ASIA TENGGARA TAHUN 2010-2021,” vol. 5, no. 1, 2024, doi: 10.46306/lb.v5i1.

UCAPAN TERIMA KASIH

Ucapan rasa terima kasih kepada Bapak Christian Sri Kusuma Aditya, S.Kom., M.Kom. yang telah bersedia dalam membimbing saya untuk terselenggaranya penelitian ini. Semoga penelitian ini dapat memberikan manfaat bagi pengembangan dalam bidang klasifikasi data citra serta menjadi landasan penelitian lebih lanjut.