

Perbandingan Model Transfer Learning CNN Pada Klasifikasi Citra Penyakit Gigi dan Lidah

Mulut menjadi salah satu parameter penentu kesehatan tubuh. Di dalam mulut terdapat gigi dan lidah yang sangat rentan terhadap berbagai penyakit. Sebelumnya banyak penelitian yang melakukan klasifikasi untuk gigi maupun lidah menggunakan algoritma CNN. Namun belum ada penelitian yang melakukan klasifikasi dengan menggabungkan dataset penyakit gigi dan lidah. Oleh karena itu, penelitian ini berfokus pada klasifikasi penyakit gigi dan lidah dengan teknik transfer learning menggunakan arsitektur CNN model VGG16, VGG19, ResNet50. Tujuan utama penelitian ini untuk membandingkan ketiga model guna menemukan hasil kinerja model paling optimal dalam menangani kasus terkait. Berdasarkan hasil penelitian yang telah dilakukan, akurasi terbaik diperoleh pada skenario dengan data yang telah diaugmentasi dan model yang dilatih menggunakan 75 epoch. Dimana model VGG16 mencapai akurasi 94%, VGG19 dengan akurasi sebesar 93%, dan ResNet50 akurasi mencapai 94%.

Latar Belakang



Menurut data World Health Organization (WHO), sebanyak 80 - 90% anak berusia kurang dari 18 tahun di Eropa, Amerika, Asia, termasuk Indonesia mengidap gigi karies. Selain itu, berdasarkan hasil Survei Kesehatan Rumah Tangga (SKRT) pada tahun 2004, sebesar 90,05% penduduk Indonesia mengidap penyakit gigi karies. Tingginya masalah kesehatan mulut mendorong berbagai penelitian menggunakan kecerdasan buatan dengan berbagai model algoritma. Penelitian ini melakukan klasifikasi dengan dataset gigi dan lidah.

Metodologi

Penelitian ini melakukan klasifikasi dengan dataset gigi dan lidah yang diperoleh dari situs Kaggle dan Roboflow. Klasifikasi dilakukan dengan menggunakan 3 arsitektur CNN yaitu VGG16, VGG19, dan ResNet50.

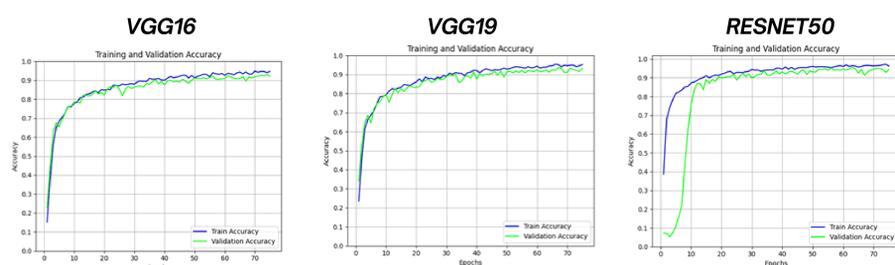
Tujuan

- Mengevaluasi dan membandingkan hasil kinerja model yang dibangun dalam klasifikasi kesehatan gigi dan lidah.
- Mengevaluasi pengaruh augmentasi dalam klasifikasi kesehatan gigi dan lidah pada model yang dibangun.
- Mengevaluasi dan membandingkan hasil kinerja setiap model yang diterapkan proses augmentasi berdasarkan jumlah epoch.

Hasil

Hasil menunjukkan bahwa teknik augmentasi dan jumlah epoch berpengaruh terhadap hasil performa semua model yang diuji. Dari semua skenario yang dilakukan dalam penelitian ini, model VGG16 menjadi yang terbaik dalam grafik saat proses pelatihan (stabil) maupun *evaluation metrics* (akurasi), sedangkan VGG19 mendapatkan nilai akurasi paling kecil dibandingkan model lain, dan ResNet50 mengalami overfitting pada awal epoch selama proses pelatihan.

Analisis Hasil Grafik



Berdasarkan hasil grafik yang didapatkan pada semua skenario pengujian, ResNet50 cenderung mengalami overfitting pada awal proses pelatihan. Sementara itu, kedua model VGG cenderung lebih stabil pada awal proses pelatihan karena memiliki struktur yang lebih sederhana dibandingkan ResNet50, sehingga tidak rentan terhadap overfitting. Namun, meskipun stabil, hasil akhir VGG ternyata masih sedikit di bawah atau setara dengan ResNet, menunjukkan bahwa meskipun mengalami overfitting pada awalnya, ResNet mampu mengejar dan bahkan melampaui kinerja VGG pada akhir proses pelatihan karena kemampuannya untuk menangkap pola yang lebih kompleks.

Pengaruh Augmentasi

	Model	Precision	Recall	F1-Score	Accuracy
Non Augmentasi	VGG16	0.92	0.92	0.92	0.92
	VGG19	0.91	0.91	0.91	0.91
	ResNet50	0.93	0.93	0.93	0.93
Augmentasi	VGG16	0.86	0.88	0.85	0.85
	VGG19	0.88	0.89	0.88	0.88
	ResNet50	0.93	0.93	0.93	0.93

ResNet50 memberikan hasil terbaik di antara ketiga model yang diujikan. Dengan nilai precision, recall, F1-Score, dan accuracy yang sama yaitu 0.93. Hal ini menunjukkan bahwa ResNet50 memiliki konsistensi performa yang tinggi dalam kedua kondisi. Di sisi lain, model VGG16 dan VGG19 mengalami penurunan performa setelah dilakukan augmentasi, ditunjukkan dengan nilai metrik yang lebih rendah dibandingkan tanpa augmentasi. Penurunan ini dapat disebabkan oleh kompleksitas tambahan dari data hasil augmentasi, dimana mungkin memerlukan lebih banyak epoch untuk mencapai performa optimal.

Pengaruh Jumlah Epoch

Epoch	Model	Precision	Recall	F1-Score	Accuracy
25	VGG16	0.86	0.88	0.85	0.85
	VGG19	0.88	0.89	0.88	0.88
	ResNet50	0.93	0.93	0.93	0.93
50	VGG16	0.90	0.91	0.90	0.90
	VGG19	0.90	0.91	0.90	0.90
	ResNet50	0.93	0.93	0.93	0.93
75	VGG16	0.94	0.94	0.94	0.94
	VGG19	0.93	0.94	0.93	0.93
	ResNet50	0.94	0.94	0.94	0.94

Tabel menunjukkan bahwa semua model mengalami peningkatan performa seiring bertambahnya epoch. Model VGG16 dan VGG19 menunjukkan peningkatan yang signifikan dalam semua metrik ketika epoch ditingkatkan. VGG16 mencapai performa terbaik pada 75 epoch dengan nilai precision, recall, F1-Score, dan accuracy sebesar 0.94 sedangkan VGG 19 memperoleh nilai precision sebesar 0.93, recall sebesar 0.94, F1-Score sebesar 0.93, dan accuracy sebesar 0.94. Hal ini menunjukkan bahwa kedua model tersebut membutuhkan lebih banyak epoch untuk mencapai performa optimal pada data yang diaugmentasi. Di sisi lain, ResNet50 mempertahankan performa stabil dengan nilai metrik yang konsisten dari 25 hingga 75 epoch.

Kesimpulan

Penelitian ini membandingkan kinerja tiga model CNN, yaitu VGG16, VGG19, dan ResNet50, dalam tugas klasifikasi gambar. Hasil penelitian menunjukkan bahwa ResNet50 secara konsisten memberikan performa terbaik, terutama setelah dilakukan augmentasi data. Meskipun ResNet50 cenderung mengalami overfitting pada awal pelatihan, model ini mampu mengatasi masalah tersebut dan mencapai akurasi yang tinggi. Model VGG16 dan VGG19, di sisi lain, membutuhkan lebih banyak epoch untuk mencapai performa optimal setelah dilakukan augmentasi data. Peningkatan jumlah epoch terbukti efektif dalam meningkatkan kinerja kedua model VGG, namun ResNet50 tetap menunjukkan keunggulannya dengan performa yang stabil dan konsisten. Hasil penelitian ini mengindikasikan bahwa pemilihan model CNN dan strategi pelatihan yang tepat sangat penting untuk mencapai kinerja yang optimal dalam tugas klasifikasi gambar.