

# The Effects of Preprocessing Techniques on Nasnetmobile's Performance for Classifying Knee Osteoarthritis Based on the Kellgren-Lawrence System

Marcell Jeremy Wiradinata<sup>1\*</sup>, Daniel Martomanggolo Wonohadidjo<sup>2\*</sup>

\* School of Information Technology, Universitas Ciputra Surabaya, Indonesia

[mwiradinata@student.ciputra.ac.id](mailto:mwiradinata@student.ciputra.ac.id)<sup>1</sup>, [daniel.m.w@ciputra.ac.id](mailto:daniel.m.w@ciputra.ac.id)<sup>2</sup>

## Article Info

### Article history:

Received 2024-10-18

Revised 2024-10-31

Accepted 2024-11-13

### Keyword:

*Knee Osteoarthritis (KOA),  
Deep Learning,  
X-Ray Image Classification,  
NASNetMobile.*

## ABSTRACT

Knee osteoarthritis (KOA) is a degenerative joint disorder characterized by the progressive deterioration of protective cartilage at the ends of bones, leading to pain and limited mobility. Deep learning provides an effective approach to classify whether X-ray images indicate the presence of KOA; however, dataset preprocessing techniques can enhance the efficacy of deep learning models. This study highlights the importance of preprocessing techniques in improving image contrast, particularly in utilizing the NASNetMobile model to assess the severity of KOA through X-ray images. KOA classification based on the Kellgren-Lawrence system consists of five severity levels; however, simplifying it into two categories can improve the performance of deep learning models. By fine-tuning parts of the NASNetMobile model and using the Nadam optimizer, the model initially achieved only 59% validation accuracy. However, by applying various preprocessing techniques, the model's validation accuracy improved to 80%.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

## I. PENDAHULUAN

Osteoarthritis Lutut atau Knee Osteoarthritis (KOA) adalah salah satu dari banyak bentuk osteoarthritis dan merupakan gangguan degeneratif di mana kartilago lutut secara progresif memburuk. Data yang dikumpulkan dari tahun 2020 dari 42 negara menyimpulkan bahwa osteoarthritis lutut adalah bentuk osteoarthritis yang paling umum dan hingga tahun 2050, diproyeksikan akan terus menjadi salah satu penyebab utama tahun hidup dengan disabilitas (YLD) bagi populasi lansia [1]. Penyakit yang memengaruhi jutaan orang di seluruh dunia saat ini tidak memiliki obat yang diketahui selain untuk upaya deteksi dini, perawatan pencegahan, dan pengobatan untuk meredakan gejala bagi individu di tahap-tahap lebih lanjut [2], [3]. Seorang dokter kemungkinan besar akan menentukan tingkat keparahan osteoarthritis menggunakan pencitraan sinar-X dan kemudian menetapkan angka atau grade berdasarkan sistem penilaian Kellgren-Lawrence (KL) atau atlas Osteoarthritis Research Society International (OARSI) untuk mewakili tingkat keparahan degenerasi [4]. Memetakan tingkat keparahan OA berdasarkan pencitraan sinar-X sangat rentan terhadap kesalahan manusia dan para

penulis menyarankan untuk menerapkan penggunaan model Machine Learning (ML) untuk membantu dokter dalam mencapai grade yang lebih objektif [5], [6].

Penelitian ini berfokus pada pengaruh teknik pra-proses yang meningkatkan kontras pada implementasi model NASNetMobile untuk melakukan klasifikasi Knee Osteoarthritis (KOA) berdasarkan gambar x-ray menggunakan metode fine-tuning serta menggunakan pengoptimal Nadam. Model NASNetMobile digunakan pada penelitian ini karena studi terkait [7], menggunakan model deep learning NASNetLarge dan menunjukkan validation accuracy sebesar 98,8% dan rata-rata akurasi pelatihan untuk mengenali kehadiran atau ketidakhadiran Osteoarthritis sebesar 98,3%. Meskipun tidak ada hasil spesifik tentang NASNetMobile, namun dari keluarga model ini, seperti NASNetLarge, dapat diperkirakan bahwa NASNetMobile juga akan memberikan hasil yang baik dengan menggunakan sumber daya komputasi yang lebih rendah.

Dalam pelatihan model, pengoptimal Nadam akan digunakan. Pengoptimal Nadam menggabungkan Nesterov momentum dengan pengoptimal Adam sehingga

memperbarui parameter melalui gradien pembaruan yang diproyeksikan, bukan melalui nilai variabel saat ini. Pendekatan ini meningkatkan kecepatan dan kelancaran konvergensi, dengan memanfaatkan peluruhan gradien eksponensial untuk menstabilkan proses pengoptimalan [8], [9] sehingga cocok digunakan untuk tugas-tugas pembelajaran mendalam (deep learning) yang kompleks, seperti klasifikasi citra, pemrosesan bahasa alami, atau deteksi objek. Kecepatan dan kelancaran dari pengoptimalan Nadam terefleksikan melalui penelitian [10] yang mendapatkan tingkat akurasi tertinggi menggunakan pengoptimalan Nadam.

Sejumlah penelitian mengenai topik ini telah dilakukan. Pada studi [11] ini, disajikan metode diagnosis otomatis yang canggih berdasarkan neural network dan transfer learning untuk KOA. Para penulis ingin menggunakan gambar sinar-X untuk meningkatkan efisiensi dan akurasi diagnosis OA menyarankan untuk menggunakan Residual Neural Network 50 (ResNet-50), Visual Geometry Group 16 (VGG-16), dan jaringan saraf konvolusi berurutan (CNN) dalam strategi jaringan neural berbasis transfer learning. Untuk meningkatkan akurasi yang diprediksi dari model, mereka menggabungkan transfer learning dengan pendekatan data augmentation serta menggunakan berbagai macam teknik pra-proses pada dataset. Menurut para penulis, setiap model yang direkomendasikan mampu mengidentifikasi OA dengan tingkat akurasi yang diprediksi lebih baik dari 90%. Pre-trained VGG-16 adalah model dengan kinerja terbaik, dengan testing accuracy 92% dan training accuracy 99%. Para peneliti pada studi ini tidak menggunakan model NASNetMobile atau pengoptimalan Nadam, namun menggunakan sejumlah teknik pra-proses yang meningkatkan akurasi pada model deep learning mereka.

Pada suatu penelitian [10] bertujuan untuk mengembangkan sistem penilaian tingkat keparahan osteoarthritis lutut (KOA) otomatis menggunakan deep neural network. Studi tersebut membandingkan berbagai neural network dan fungsi pengoptimalan, dan menemukan bahwa kombinasi EfficientNet-B7 dengan pengoptimalan Nadam menghasilkan kinerja terbaik. Studi ini menargetkan tujuan yang sama dengan penelitian ini dan menyoroti potensi neural network dalam mengotomatiskan proses penilaian. Namun, studi ini memiliki keterbatasan, tidak mengeksplorasi keluarga NASNet dan menggunakan transfer learning daripada fine-tuning.

Terdapat studi [12] yang meneliti efektivitas fine-tuning deep learning models untuk deteksi dan analisis perkembangan osteoarthritis lutut, model DenseNet169 yang disempurnakan ini mencapai nilai presisi, perolehan, dan skor F1 yang tinggi, sehingga menyoroti potensinya untuk aplikasi dunia nyata. Namun, penelitian ini memiliki beberapa keterbatasan, termasuk tidak menjajaki penggunaan model pembelajaran mendalam lainnya, seperti NASNetMobile, dan tidak menyelidiki pengoptimalan alternatif, seperti pengoptimalan Nadam, namun menggunakan metode fine-tuning. Klasifikasi knee osteoarthritis (KOA) dari X-ray tanpa teknik pra-proses

sangat sulit karena variasi kualitas gambar, seperti perbedaan kontras dan noise, yang dapat mengaburkan detail penting [11]. Hal ini membuat model deep learning sulit mendeteksi pola yang konsisten, mengurangi akurasi klasifikasi. Teknik pra-proses penting untuk meningkatkan kualitas gambar dengan menyesuaikan kontras, menghilangkan noise, dan memperjelas area penting, sehingga model lebih mudah mengenali fitur KOA dan menghasilkan klasifikasi yang lebih akurat.

Penelitian ini terdiri dari empat bab, masing-masing disusun secara sistematis. Untuk memperjelas konteks dan tujuan penelitian, bab pertama membahas pendahuluan dari penelitian ini. Bab kedua pada metode yang digunakan, seperti desain dataset dan model. Hasil penelitian disajikan dalam bab ketiga. Ini mencakup visualisasi hasil, analisis, dan perbandingan dengan penelitian lain untuk mengevaluasi efisiensi dan efektivitas model yang diterapkan. Terakhir, Bab keempat membahas kesimpulan dari penelitian, rekomendasi untuk penelitian lebih lanjut, dan aplikasi praktis dari temuan.

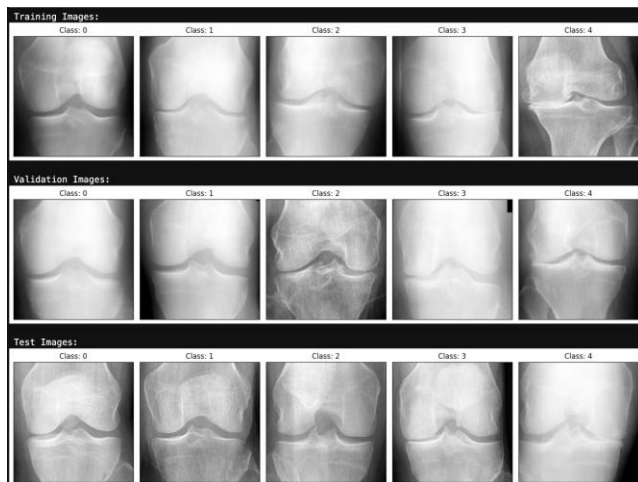
## II. METODE

Penelitian ini dibagi menjadi empat bagian. Bagian-bagian ini membahas perangkat lunak dan keras yang digunakan, dataset yang digunakan, proses pelatihan model, dan metode untuk mengevaluasi kinerja model.

Dalam penelitian ini digunakan dua perangkat. Satu perangkat untuk melakukan pelatihan model dan satu lagi untuk melakukan teknik pra-proses dan export dataset. Untuk keperluan pelatihan model, sistem operasi yang digunakan adalah Ubuntu 20.04. Bahasa pemrograman yang digunakan untuk penelitian ini adalah Python versi 3.9.19. Framework Python dan API yang digunakan merupakan TensorFlow versi 2.14.0 dan Keras yang menyediakan berbagai fungsionalitas machine learning dan deep learning. Pelatihan model ini akan dilakukan dengan perangkat lunak JupyterLab yang merupakan bawaan dari perangkat lunak Anaconda Navigator. Perangkat lunak ini digunakan karena kemudahan dan fungsionalitas tinggi yang memenuhi keperluan penelitian ini, seperti dalam segi instalasi package atau library dan pembuatan environment baru. Dalam penelitian ini, dikarenakan digunakan dua perangkat, maka perangkat lunak AnyDesk digunakan untuk mempermudah pengiriman file dari satu komputer ke komputer lainnya. Dalam proses pelatihan model NASNetMobile digunakan komputer dengan GPU NVIDIA GeForce RTX 3060 Ti yang mendukung komputasi paralel untuk mempercepat proses pelatihan model. Selain itu, perangkat ini dilengkapi dengan penyimpanan sebesar 1 Terabyte untuk menyimpan dataset dan model yang akan dilatih. Prosesor yang digunakan adalah 13th Gen Intel(R) Core (TM) i7-13700K CPU, yang memberikan kecepatan pemrosesan yang tinggi, dan RAM sebesar 32GB. Sedangkan untuk melakukan export dan pemrosesan dataset, digunakan komputer MacBook M2 Air 2023 15" dengan spesifikasi 16GB RAM dan memiliki processor Apple Silicon M2. Hasilnya kemudian di export

dan dipindahkan ke komputer untuk melakukan pelatihan model.

Dataset yang akan digunakan [13] telah diorganisir dan berasal dari Osteoarthritis Initiative (OAI). Dataset ini mengandung sebanyak 8260 gambar x-ray daerah lutut yang telah dibagi menjadi training, validation dan testing dataset pada folder sendiri-sendiri dengan rasio pembagian 70-10-20. Gambar bersifat grayscale, dalam format gambar PNG dan terdapat dalam ukuran resolusi 224x224 atau 299x299, namun ukuran 224x224 yang akan digunakan dalam penelitian ini. Gambar-gambar telah dibagi menjadi 5 kelas, menurut sistem penilaian Kellgren-Lawrence menurut tingkat severitas OA pada lutut. Walaupun dataset ini telah dibagi menjadi lima kelas, namun untuk penelitian ini, struktur dari dataset ini dirubah sehingga terdapat dua kelas, yaitu tingkat severitas sehat dan berat. Kelas sehat terdiri dari kelas pertama dan kedua, lalu kelas dua, tiga dan empat membentuk kelas severitas berat. Pada penelitian sebelumnya [14] telah dibuktikan bahwa teknik penyederhanaan dataset dari lima kelas menjadi dua atau tiga kelas meningkatkan performa model deep learning dalam proses klasifikasi KOA. Hal ini dikarenakan sejumlah dari kelas-kelas memiliki kemiripan yang tinggi sehingga susah untuk model membedakan dan mengurangi nilai akurasi pada tahap evaluasi nantinya. Dengan menggabungkan sejumlah kelas, diharapkan tangka akurasi dapat ditingkatkan [11]. Contoh gambar-gambar pada dataset dapat dilihat pada Gambar 1.



Gambar 1. Gambar-gambar x-ray daerah lutut dari dataset KOA sebelum dilakukan pemrosesan data dan telah dibagi menjadi training, validation dan testing dataset dalam lima kelas

Setiap gambar dievaluasi untuk mengidentifikasi apakah gambar tersebut memiliki kualitas yang cukup baik, termasuk memastikan kontras, resolusi, dan isi dari gambar. Gambar 2 menunjukkan dua ilustrasi gambar x-ray lutut dengan kontras yang seimbang. Gambar-gambar tersebut menampilkan detail anatomi lutut secara jelas, sehingga memungkinkan identifikasi yang akurat terhadap struktur tulang dan persendian. Kontras yang baik membantu dalam memperjelas

perbedaan antara tulang dan jaringan lunak serta membantu melihat Asymmetric Joint Space Narrowing (AJSEN) yang merupakan salah satu gejala-gejala awal dari KOA.



Gambar 2. Dua contoh gambar x-ray daerah lutut dengan kontras yang seimbang dimana daerah renggang antara tulang atas dan bawah terlihat dengan baik.

TABEL 1.  
CONTOH GAMBAR-GAMBAR X-RAY DAERAH LUTUT DARI DATASET KOA YANG DIKELUARKAN KARENA MEMILIKI SIFAT-SIFAT YANG KURANG IDEAL SEPERTI MEMILIKI NILAI KONTRAS RENDAH, WARNANYA INVERTED, TERLALU GELAP, TERLALU TERANG ATAU TERDAPAT OBJECT YANG MENGHALANGI STRUKTUR

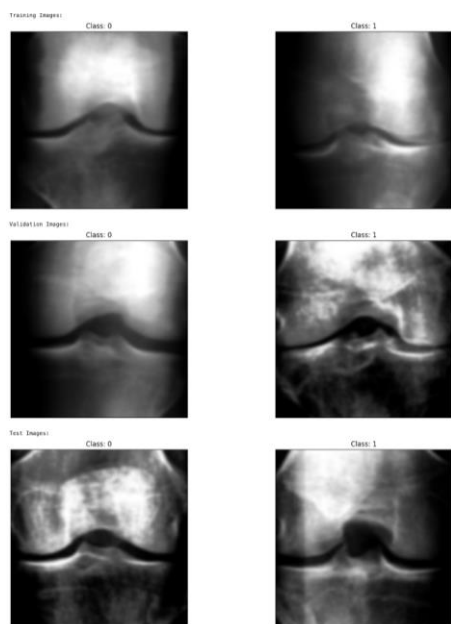
Gambar x-ray dengan warna <i>Inverted</i>	Pencahayaannya terlalu rendah	Pencahayaannya terlalu tinggi
Terdapat titik-titik pada kiri gambar x-ray	Terdapat garis putih pada gambar x-ray	Terdapat objek yang menghalangi lutut

Tabel 1 menunjukkan contoh gambar yang kurang cocok; gambar-gambar ini dikeluarkan dari dataset karena memiliki berbagai kekurangan. Misalnya, gambar yang diputar warnanya menunjukkan struktur anatomi lutut dalam warna yang berbeda dari mayoritas dataset yang lainnya. Selain itu, pencahayaan yang terlalu rendah membuat detail tulang tidak terlihat dengan jelas, sementara pencahayaan yang terlalu tinggi membuat gambar terlalu terang, mengaburkan fitur penting. Selain itu, ada gambar yang mengandung artifact, seperti titik-titik yang ditemukan pada gambar x-ray, garis putih yang menutupi area penting, atau benda asing yang menghalangi pandangan dari lutut. Untuk melatih model secara efektif, gambar-gambar seperti ini dikeluarkan agar dataset yang digunakan memiliki kualitas terbaik.

Setelah gambar dipilih, sisa dataset akan diproses lebih lanjut untuk meningkatkan kualitas kontras. Langkah pertama adalah histogram matching, yang bertujuan menyamakan

kontras pada seluruh gambar di dalam dataset dengan menggunakan gambar referensi dengan kontras optimal, seperti yang ditunjukkan pada Gambar 2. Ini dilakukan untuk menjaga konsistensi visual antara gambar, terutama jika setiap gambar memiliki kontras dan histogram yang berbeda. Kemudian, rescale intensity diterapkan untuk menyesuaikan rentang intensitas piksel, membantu menyederhanakan nilai grayscale dan memperjelas detail yang mungkin tidak terlihat. Namun penggunaan rescale intensity menyebabkan terlalu banyak noise terbentuk pada gambar, sehingga Gaussian smoothing digunakan untuk mengurangi noise tersebut, sehingga fitur penting lebih menonjol dan analisis menjadi lebih akurat. Langkah terakhir adalah cropping dengan teknik zoom untuk memperbesar area penting (ROI) dan mengurangi bagian gambar yang tidak relevan seperti area gelap dan kosong pada sisi-sisi di gambar, terutama pada gambar x-ray. Contoh gambar x-ray hasil dari pemrosesan dapat dilihat pada Gambar 3.

Dibandingkan dengan Gambar 1, dampak dari pemrosesan menyetarakan tingkat kecerahan pada gambar-gambar pada seluruh dataset, menyederhanakan nilai piksel pada tiap gambar dan memperjelas daerah joint space yang penting untuk menentukan jika lutut memiliki KOA.



Gambar 3. Gambar x-ray dari dua kelas yang telah dilakukan teknik pra-proses dan dibagi menurut dataset training (barisan paling atas), validation (barisan kedua) dan testing (barisan ketiga)

Dataset selanjutnya melalui tahap Data Augmentation, dimana berbagai transformasi diterapkan untuk memperkaya variasi gambar secara acak dan meningkatkan kemampuan generalisasi model dengan menggunakan ImageDataGenerator. Transformasi dilakukan pada training dataset, validation dan testing, namun training dataset mendapatkan augmentasi yang lebih intensif. Beberapa teknik yang diterapkan termasuk menormalkan nilai piksel dari 0-

255 menjadi 0-1, memutar gambar hingga 30 derajat, menggeser gambar secara horizontal dan vertikal hingga 20%, melakukan shear (transformasi geser), zooming secara acak, dan membalik gambar secara horizontal. Untuk area kosong yang dihasilkan oleh transformasi seperti rotasi atau pergeseran, digunakan piksel terdekat untuk mengisi ruang tersebut. Terakhir, dikarenakan jumlah dalam dataset yang tidak seimbang ini, dilakukan perhitungan class weights dengan menggunakan compute\_class\_weights dari library scikit-learn sehingga model tidak terlalu bias terhadap kelas yang memiliki jumlah gambar lebih banyak dan meningkatkan konvergensi model.

Berikutnya, dataset tersebut dimasukkan ke dalam model NASNetMobile yang dilatih melalui fine tuning dimana sepertiga dari feature-extraction layer akan di unfreeze sehingga dapat lebih baik mempelajari pola pada dataset ini. NASNetMobile tanpa classification layer-nya memiliki sebanyak 769 lapisan, sehingga sebanyak 256 lapisan akan dilakukan unfreezing dan dilatih ulang dengan dataset KOA. Setelah hasil dari pelatihan didapatkan, model akan update weights-nya dengan menghitung loss berdasarkan categorical crossentropy dan melalui pengoptimal Nadam dimana penggunaan Learning Rate Scheduler dengan mengimplementasikan Exponential Decay untuk mengurangi learning rate secara bertahap selama proses pelatihan bertujuan untuk mencegah model terjebak dalam minimum lokal dan memastikan konvergensi yang lebih baik. Initial learning rate yang digunakan adalah 0.01, dan dibagi 0.1 setiap 5 epoch dalam proses pelatihan dengan learning rate minimum  $1e-5$ .

Pelatihan model dilakukan selama 100 epoch, namun EarlyStopping digunakan untuk memberhentikan pelatihan model setelah 10 epoch apabila tidak ada peningkatan pada nilai validation accuracy. Selama proses pelatihan, Modal Checkpoint akan menyimpan weights terbaik apabila terdapat peningkatan pada nilai validation accuracy pada epoch tersebut. Selama proses pelatihan, hubungan antara validation accuracy dan training accuracy menjelaskan proses belajar dari model deep learning, terutama validation accuracy, dimana nilai yang tinggi menunjukkan potensi kemampuan model dalam mengklasifikasikan data baru.

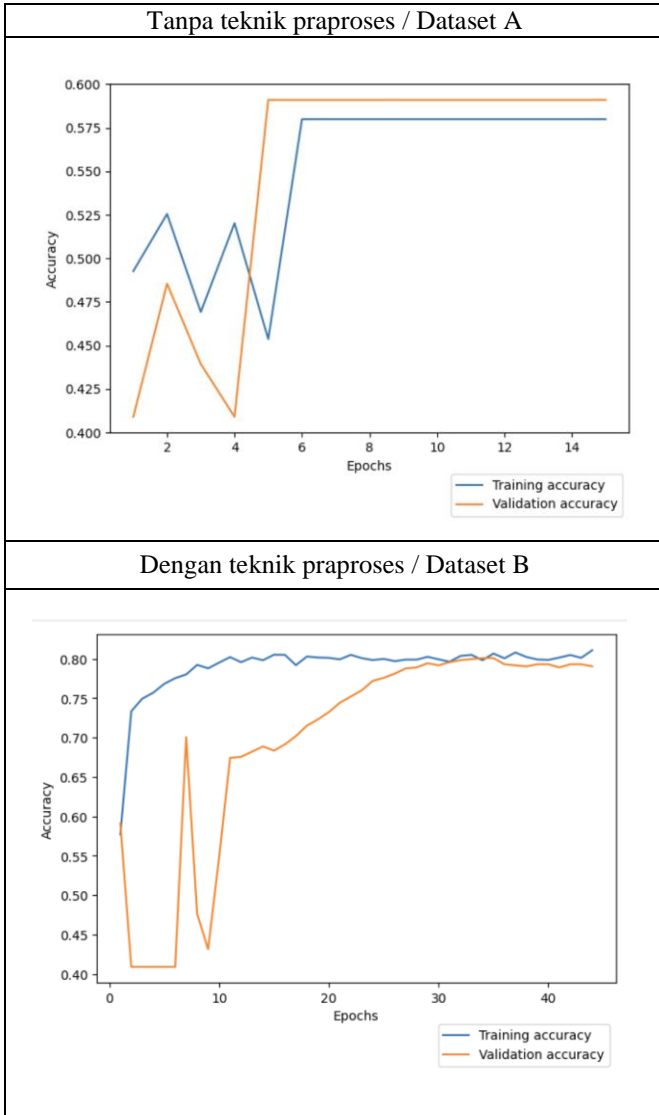
Selama proses pelatihan model, pergerakan akurasi serta loss dari sisi training dan validation dapat dipantau melalui line graph untuk memahami jika model sedang overfitting atau underfitting. Model yang telah dilatih ini akan lanjut dievaluasi dengan dataset testing dan hasilnya adalah dalam bentuk testing accuracy, precision, recall dan F1-score. Selain itu, confusion matrix dalam bentuk heatmap akan digunakan untuk memahami kecenderungan model deep learning selama tahap testing dan memperjelas asal-usul dari nilai precision, recall dan F1-score.

**III. HASIL DAN PEMBAHASAN**

Berdasarkan grafik pada TABEL 2, terlihat bahwa model NasNetMobile yang telah di fine-tune sebesar 30% pada dataset KOA ini menunjukkan kinerja yang cukup baik.

TABEL 2.

GRAFIK GARIS YANG MENUNJUKKAN NILAI TRAINING ACCURACY DAN VALIDATION ACCURACY PROSES PELATIHAN MODEL NASNETMOBILE DENGAN DATASET TANPA TEKNIK PRAPROSES (ATAS) DAN DATASET DENGAN TEKNIK PRAPROSES (BAWAH)



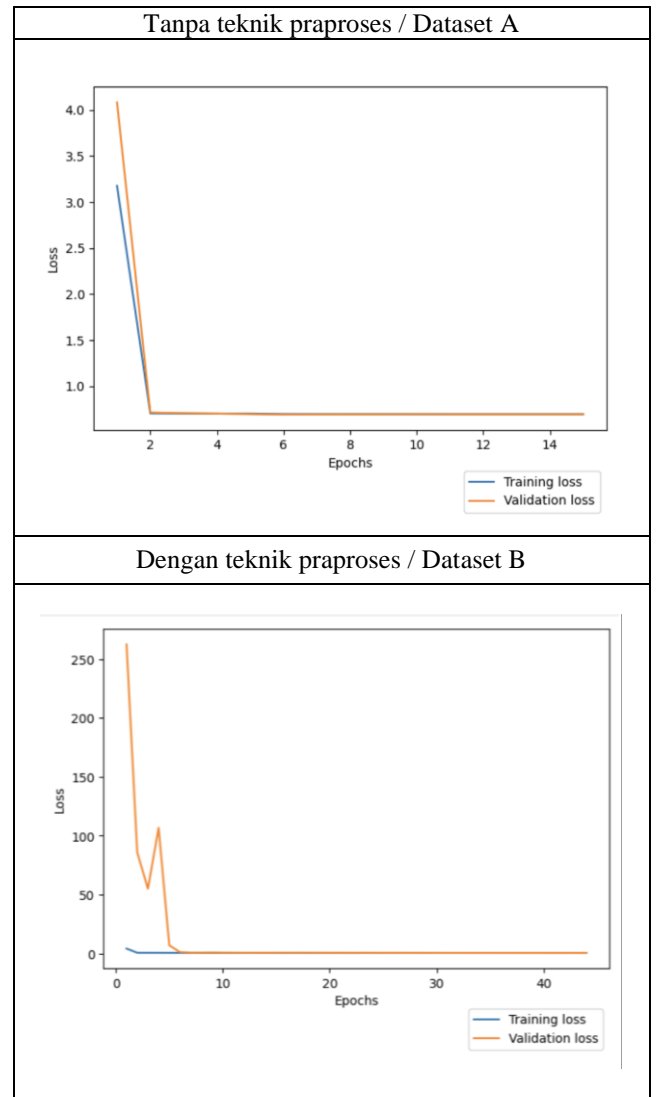
Hal ini dapat dilihat dari peningkatan akurasi pelatihan yang cukup signifikan pada beberapa epoch pertama pada kedua skenario. Namun, setelah mencapai titik tertentu, kurva akurasi pelatihan mulai stagnan pada nilai 0.575 hingga 0.6 saat dilatih dengan dataset A (tanpa teknik praproses) sedangkan untuk model yang dilatih dengan dataset B (yang telah di teknik praproses), kurva menjadi stagnan pada nilai sekitar 0.8 dan cenderung sedikit fluktuatif.

Sementara itu, kurva validation accuracy cenderung lebih stabil dengan dataset A dan mencapai nilai 0.59. Sedangkan

dataset B menunjukkan peningkatan yang lebih lambat dibandingkan dengan kurva training accuracy. Ini mengindikasikan bahwa model telah mulai overfitting pada training data, sehingga kemampuannya dalam menggeneralisir pada data yang belum pernah dilihat sebelumnya menjadi terbatas, namun hal ini teratasi dengan berjalannya pelatihan dimana model berhasil meningkatkan nilai validasi menjadi sekitar 0.8.

TABEL 3

GRAFIK GARIS YANG MENUNJUKKAN NILAI TRAINING LOSS DAN VALIDATION LOSS DARI PROSES PELATIHAN MODEL NASNETMOBILE DENGAN DATASET TANPA TEKNIK PRAPROSES (ATAS) DAN DATASET DENGAN TEKNIK PRAPROSES (BAWAH)



Grafik loss yang disajikan pada TABEL 3 memberikan perbandingan yang menarik antara kinerja model NASNetMobile ketika dilatih pada dua dataset yang berbeda. Model yang dilatih pada dataset A (tanpa teknik praproses) menunjukkan nilai loss yang berkurang secara tajam sedangkan pada dataset B (dengan teknik praproses) menunjukkan nilai validation loss yang sangat tinggi pada

tahap awal-awal pelatihan namun berhasil turun dan stabil hingga akhir pelatihan. Hasil yang ditunjukkan pada TABEL 2 dan TABEL 3 mengindikasikan bahwa teknik praproses data sangat efektif dalam membantu model belajar fitur-fitur yang lebih relevan sehingga meningkatkan kemampuan generalisasi model pada tahap pelatihan.

Meskipun menggunakan dataset yang berbeda, menurut Gambar 4, kedua model NASNetMobile menghasilkan kinerja yang hampir sama. Ini menunjukkan bahwa komponen lain yang lebih dominan mempengaruhi hasil akhir, seperti, arsitektur model yang kuat, dan pengaturan hyperparameter yang optimal. Dengan kata lain, perbedaan yang ada di antara dataset dalam hal ini tidak memiliki pengaruh yang signifikan terhadap kinerja model saat melakukan klasifikasi pada gambar-gambar x-ray daerah lutut yang baru.

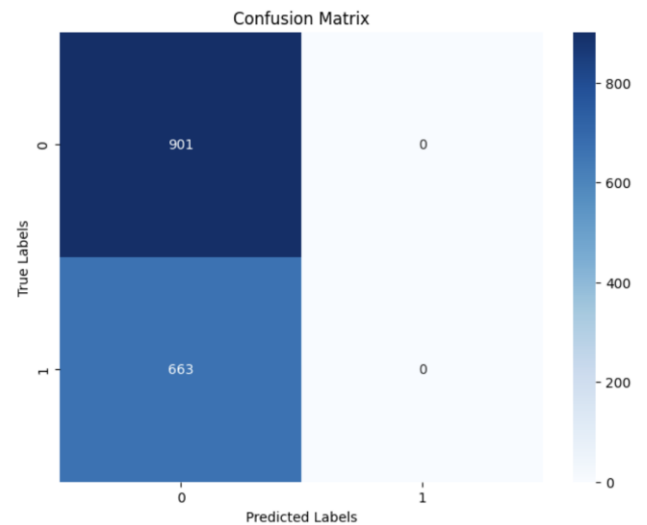
	precision	recall	f1-score
0	0.58	1.00	0.73
1	1.00	0.00	0.00
accuracy			0.58
macro avg	0.79	0.50	0.37
weighted avg	0.76	0.58	0.42

Gambar 4. Perhitungan precision, recall, f1-score serta akurasi pengujian dari model NASNetMobile dengan dataset tanpa teknik praproses dan dengan teknik praproses bersifat sama

Performa dari kedua model yang dilatih dengan dataset tanpa teknik praproses dan dengan teknik praproses dapat dievaluasi menurut tingkat precision, recall, F1-score dari tiap kelas pada dataset dan juga dari testing accuracy. Precision menunjukkan proporsi prediksi positif yang benar dari semua prediksi positif. Pada kelas 0, precision sebesar 0.58 artinya 58% dari prediksi yang dilabel sebagai kelas 0 memang benar-benar kelas 0. Namun, pada kelas 1, precision-nya 1.00 yang menunjukkan bahwa semua prediksi yang dilabel sebagai kelas 1 adalah benar. Recall menunjukkan proporsi contoh positif yang benar-benar diklasifikasikan sebagai positif dari semua contoh positif yang ada. Pada kelas 0, recall sebesar 1.00 artinya semua contoh kelas 0 berhasil diklasifikasikan dengan benar. Sebaliknya, pada kelas 1, recall-nya 0.00 yang mengindikasikan model sama sekali tidak dapat mengklasifikasikan contoh kelas 1 dengan benar. F1-score adalah rata-rata harmonik dari precision dan recall. Nilai F1-score yang tinggi menunjukkan bahwa model memiliki baik precision maupun recall yang tinggi. Pada kasus ini, F1-score kelas 0 cukup baik (0.73), tetapi F1-score kelas 1 sangat rendah (0.00) dikarenakan nilai recall yang sangat kurang. Testing Accuracy menunjukkan proporsi total prediksi yang benar. Nilai accuracy sebesar 0.58 mengindikasikan bahwa model dapat mengklasifikasikan dengan benar sekitar 58% dari seluruh data.

Confusion Matrix digunakan untuk menggambarkan kinerja dari kedua model tersebut dengan memetakan kecenderungan dari model NASNetMobile dalam melakukan

klasifikasi tiap kelas [6], [11]. Confusion Matrix dapat dilihat pada Gambar 5.



Gambar 5. Confusion Matrix yang menggambarkan kinerja klasifikasi dari kedua model NASNetMobile saat dievaluasi dengan dataset pengujian

Pada confusion matrix, terlihat bahwa model memiliki kinerja yang sangat baik dalam mengklasifikasikan kelas 0. Dari 901 sampel yang sebenarnya merupakan kelas 0, model berhasil memprediksi semuanya dengan benar. Namun, model sama sekali gagal dalam mengklasifikasikan kelas 1. Seluruh 663 sampel dari kelas 1 salah diprediksi sebagai kelas 0.

Hasil ini sejalan dengan nilai metrik evaluasi sebelumnya yang menunjukkan precision yang tinggi untuk kelas 0 tetapi recall yang sangat rendah untuk kelas 1. Hal ini mengindikasikan adanya masalah ketidakseimbangan kelas yang cukup signifikan, di mana model cenderung lebih "memilih" kelas mayoritas (kelas 0) daripada kelas minoritas (kelas 1).

## V. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, implementasi model deep learning menggunakan arsitektur NASNetMobile dalam klasifikasi gambar, khususnya untuk deteksi Knee Osteoarthritis (KOA), menunjukkan potensi yang signifikan. Dengan langkah-langkah teknik praproses yang baik seperti histogram matching, rescale intensity, gaussian smoothing, dan cropping, kualitas gambar x-ray dapat ditingkatkan sehingga meningkatkan performa model. Penggunaan teknik fine-tuning pada NASNetMobile terbukti efisien, terutama ketika dihadapkan pada dataset dengan ukuran yang terbatas. Selain itu, penerapan hyperparameter tuning dan penggunaan learning rate scheduler seperti Exponential Decay pada pengoptimalan Nadam mampu meningkatkan kinerja model secara keseluruhan, menghasilkan akurasi yang tinggi dengan jumlah epoch yang lebih sedikit. Hasil pelatihan model mendapatkan tingkat validation accuracy 60% dengan menggunakan dataset tanpa teknik praproses dan meningkat

menjadi 80% saat menggunakan teknik praproses. Penelitian selanjutnya dapat fokus pada eksperimentasi berbagai macam teknik praproses lainnya, metode full fine tuning dan atau penggunaan pengoptimal lainnya serta pembagian kelas-kelas pada dataset dengan kombinasi lainnya.

#### UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada School of Information Technology, Universitas Ciputra Surabaya, atas dukungan dan penyediaan perangkat keras yang digunakan dalam penelitian ini. Fasilitas yang disediakan sangat berperan penting dalam kelancaran pelaksanaan penelitian ini.

#### DAFTAR PUSTAKA

- [1] J. D. Steinmetz et al., "Global, regional, and national burden of osteoarthritis, 1990–2020 and projections to 2050: a systematic analysis for the Global Burden of Disease Study 2021," *The Lancet Rheumatology*, vol. 5, no. 9, pp. e508–e522, Sep. 2023, doi: 10.1016/S2665-9913(23)00163-7.
- [2] B. Kathryn, "Osteoarthritis IN OLDER ADULTS: Current Treatments," *Journal of Gerontological Nursing*, vol. 31, no. 5, pp. 11–19, May 2005, doi: 10.3928/0098-9134-20050501-05.
- [3] K. Üreten, H. Erbay, and H. H. Maraş, "Detection of hand osteoarthritis from hand radiographs using convolutional neural networks with transfer learning," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 28, no. 5, pp. 2968–2978, 2020, doi: 10.3906/ELK-1912-23.
- [4] A. Xuan et al., "The application of machine learning in early diagnosis of osteoarthritis: a narrative review," *Therapeutic Advances in Musculoskeletal Disease*, vol. 15, Jan. 2023, doi: 10.1177/1759720X231158198.
- [5] A. Tiwari, M. Poduval, and V. Bagaria, "Evaluation of artificial intelligence models for osteoarthritis of the knee using deep learning algorithms for orthopedic radiographs," *World J Orthop*, vol. 13, no. 6, pp. 603–614, Jun. 2022, doi: 10.5312/wjo.v13.i6.603.
- [6] V. K. V, V. Kalpana, and G. H. Kumar, "Evaluating the efficacy of deep learning models for knee osteoarthritis prediction based on Kellgren-Lawrence grading system," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 5, p. 100266, Sep. 2023, doi: 10.1016/j.prime.2023.100266.
- [7] V. V. Kishore, K. Sahithi, K. S. J. Reddy, K. Akash, K. S. Jyothy, and S. Yalavarthi, "Enhanced Knee Osteoarthritis Grading: Transfer Learning with Pre-Trained CNN's For Swift Diagnosis," in *Proceedings of the International Conference on Computational Innovations and Emerging Trends (ICCIET 2024)*, vol. 112, K. R. Madhavi, P. Subba Rao, J. Avanija, I. L. Manikyamba, and B. Unhelkar, Eds., in *Advances in Computer Science Research*, vol. 112., Dordrecht: Atlantis Press International BV, 2024, pp. 4–20. doi: 10.2991/978-94-6463-471-6\_2.
- [8] D. Choi, C. J. Shallue, Z. Nado, J. Lee, C. J. Maddison, and G. E. Dahl, "On Empirical Comparisons of Optimizers for Deep Learning," Jun. 16, 2020, arXiv: arXiv:1910.05446. Accessed: Oct. 22, 2024. [Online]. Available: <http://arxiv.org/abs/1910.05446>
- [9] T. Dozat, "INCORPORATING NESTEROV MOMENTUM INTO ADAM," 2016.
- [10] A. Ezgi and A. Onan, "Automatic Knee Osteoarthritis Severity Grading using Deep Neural Networks: Comparative Analysis of Network Architectures and Optimization Functions," *ICAENS*, vol. 1, no. 1, pp. 197–203, Jul. 2023, doi: 10.59287/icaens.992.
- [11] H. A. Alshamrani, M. Rashid, S. S. Alshamrani, and A. H. D. Alshehri, "Osteo-NeT: An Automated System for Predicting Knee Osteoarthritis from X-ray Images Using Transfer-Learning-Based Neural Networks Approach," *Healthcare (Switzerland)*, vol. 11, no. 9, May 2023, doi: 10.3390/healthcare11091206.
- [12] S. A. El-Ghany, M. Elmoogy, and A. A. A. El-Aziz, "A fully automatic fine tuned deep learning model for knee osteoarthritis detection and progression analysis," *Egyptian Informatics Journal*, vol. 24, no. 2, pp. 229–240, Jul. 2023, doi: 10.1016/j.eij.2023.03.005.
- [13] P. Chen, "Knee Osteoarthritis Severity Grading Dataset," vol. 1, Sep. 2018, doi: 10.17632/56rmx5bjcr.1.
- [14] A. S. Mohammed, A. A. Hasanaath, G. Latif, and A. Bashar, "Knee Osteoarthritis Detection and Severity Classification Using Residual Neural Networks on Preprocessed X-ray Images," *Diagnostics*, vol. 13, no. 8, p. 1380, Apr. 2023, doi: 10.3390/diagnostics13081380.