

ANALISIS PERBANDINGAN PENGARUH NORMALISASI DAN STANDARDISASI TERHADAP AKURASI MODEL CNN PADA KLASIFIKASI EKSPRESI WAJAH

Gilang Isnan Fiyoriirly¹; Dian Ade Kurnia²; Yudhistira Arie Wijaya³; Mulayawan⁴;

Program Studi Teknik Informatika¹
Program Studi Manajemen Informatika²
Program Studi Sistem Informasi^{3,4}

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
Gilang.isnan27@gmail.com

(*) Corresponding Author : gilang.isnan27@gmail.com
Published : 31 Desember 2025

Abstract— This study aims to compare the effect of [0,1] normalization and Z-Score standardization on the accuracy of a CNN model for classifying smiling and non-smiling facial expressions. The research uses the Smiling or Not Face Data dataset from Kaggle, consisting of 9,200 grayscale facial images sized 64x64 pixels. These images were processed through data augmentation and two different preprocessing approaches before being trained using an identical CNN architecture. The novelty of this research lies in the direct comparison of two preprocessing techniques under a controlled experimental setup to determine which method yields the most optimal classification results. The model trained with [0,1] normalization achieved an accuracy of 87.08%, precision of 87.07%, recall of 87.07%, and f1-score of 87.06%. In contrast, the model trained with Z-Score standardization achieved an accuracy of 86.30%, precision of 86.28%, recall of 86.29%, and f1-score of 86.27%. These results indicate that normalization is more effective for binary facial expression classification on medium-sized grayscale datasets. Therefore, [0,1] normalization is recommended as the preferred preprocessing method when building CNN models for similar tasks.

Keywords: CNN, normalization, standardization, preprocessing, facial expression, classification

Abstrak— Penelitian ini bertujuan membandingkan pengaruh normalisasi [0,1] dan standardisasi (Z-Score) terhadap akurasi model CNN pada klasifikasi ekspresi wajah senyum dan tidak senyum. Penelitian dilakukan menggunakan dataset Smiling or Not Face Data dari Kaggle yang terdiri dari 9.200 citra wajah grayscale berukuran 64x64 piksel. Citra-citra ini kemudian diproses melalui teknik augmentasi serta dua pendekatan preprocessing berbeda sebelum dilatih menggunakan arsitektur CNN yang sama. Keunikan penelitian ini terletak pada perbandingan langsung antara dua metode preprocessing dalam satu eksperimen yang terkontrol, guna mengidentifikasi teknik mana yang memberikan hasil klasifikasi paling optimal. Model dengan preprocessing normalisasi [0,1] menghasilkan akurasi 87,08%, precision 87,07%, recall 87,07%, dan f1-score 87,06%. Sementara itu, model dengan preprocessing standardisasi Z-Score menghasilkan akurasi 86,30%, precision 86,28%, recall 86,29%, dan f1-score 86,27%. Hasil ini menunjukkan bahwa teknik normalisasi lebih unggul dalam konteks klasifikasi biner ekspresi wajah pada dataset dengan ukuran menengah dan format grayscale. Oleh karena itu, normalisasi direkomendasikan sebagai metode preprocessing yang lebih efektif dalam membangun model CNN untuk tugas serupa. Kata kunci: CNN, normalisasi, standardisasi, preprocessig, ekspresi wajah, klasifikasi

Kata Kunci : K-Means; Pengelompokan Saham; Bank BNI; Akuntansi Investasi; KDD

INTRODUCTION

Keandalan CNN dalam klasifikasi ekspresi wajah sangat dipengaruhi oleh proses praproses data citra sebelum pelatihan model dilakukan. Salah satu tahapan paling krusial dalam praproses adalah teknik normalisasi dan standardisasi data. Langkah

ini bertujuan untuk menstabilkan distribusi input, mengurangi bias akuisisi data, dan meningkatkan kemampuan generalisasi model (Agung et al., 2024; Madusanka et al., 2023). Normalisasi data memproyeksikan nilai piksel ke dalam rentang tertentu seperti [0,1], sedangkan standardisasi

merubah distribusi nilai ke dalam z-score dengan rata-rata nol dan deviasi standar satu. Dalam banyak kasus, keberhasilan sistem klasifikasi bergantung pada seberapa optimal preprocessing diterapkan untuk menciptakan konsistensi dalam data pelatihan dan pengujian (Chen et al., 2021; Swiderska-Chadaj et al., 2020).

Namun pada kenyataannya, masih banyak studi yang tidak secara eksplisit mengevaluasi pengaruh perbandingan antara normalisasi dan standarisasi terhadap kinerja CNN dalam tugas klasifikasi ekspresi wajah. Sebagian besar penelitian lebih terfokus pada arsitektur model, penggunaan teknik augmentasi, atau eksplorasi dataset baru. Padahal, ketidakstabilan distribusi data antar domain atau kondisi pencahayaan dan pose yang bervariasi dalam citra ekspresi wajah menuntut adanya pendekatan preprocessing yang tepat (Liang et al., 2023; Nowozin et al., 2022). Masih jarang dijumpai penelitian yang menelusuri secara sistematis bagaimana pilihan metode preprocessing berdampak pada konvergensi, akurasi, dan kemampuan generalisasi CNN dalam konteks klasifikasi ekspresi wajah yang nyata.

Studi-studi sebelumnya cenderung mengasumsikan bahwa metode preprocessing yang digunakan bersifat universal untuk semua jenis data, tanpa mempertimbangkan variabilitas dan kompleksitas yang dimiliki oleh data ekspresi wajah. Sementara itu, pendekatan lain menggunakan teknik augmentasi data secara masif, namun tetap menghadapi keterbatasan dalam menangani pergeseran domain dan distribusi nilai intensitas piksel antar subjek (Kopalidis, T., & al., 2024; Swiderska-Chadaj et al., 2020). Hal ini memperlihatkan adanya kesenjangan dalam pemahaman dampak spesifik dari masing-masing teknik preprocessing pada struktur pembelajaran CNN.

Beberapa pendekatan yang telah diajukan dalam studi terdahulu adalah penggunaan teknik batch normalization di dalam arsitektur CNN itu sendiri atau penggunaan domain adaptation secara eksternal. Meskipun kedua teknik tersebut terbukti dapat meningkatkan stabilitas pelatihan dan generalisasi, keduanya memiliki keterbatasan dari sisi kompleksitas dan kebutuhan data pelatihan yang besar (Apicella et al., 2023; Liu et al., 2022). Dalam praktiknya, implementasi preprocessing yang sederhana seperti normalisasi atau standarisasi justru dapat memberikan dampak signifikan terhadap performa model, apabila dilakukan secara konsisten dan disiplin (Liu et al., 2022; Yu & J., 2023).

Agung et al. (2024) menunjukkan bahwa pipeline preprocessing yang cermat, termasuk normalisasi intensitas dan augmentasi data, berperan penting

dalam meningkatkan akurasi CNN pada klasifikasi multi-emotion dataset. Sementara itu, (Madusanka et al., 2023), menemukan bahwa normalisasi warna pada citra medis berdampak langsung pada peningkatan AUC model deep learning, terutama dalam menghadapi perbedaan domain antar pusat data. Penelitian oleh (Salehi et al., 2023) juga menggarisbawahi bahwa preprocessing merupakan fondasi yang tidak terpisahkan dalam klasifikasi citra medis, karena dapat meningkatkan keandalan dan reproduktibilitas model, khususnya saat digunakan pada data heterogen.

Walaupun berbagai studi telah membahas pengaruh preprocessing secara umum, masih belum ada evaluasi mendalam terhadap perbandingan dampak dari normalisasi dan standarisasi secara khusus dalam klasifikasi ekspresi wajah berbasis CNN. Gap penelitian ini menjadi penting untuk dijawab karena pemilihan preprocessing yang tepat dapat menjadi kunci dalam mengatasi tantangan-tantangan utama dalam domain ini, seperti pergeseran domain, overfitting, dan distribusi data yang tidak seragam (Salehi et al., 2023; Zhao et al., 2024). Oleh karena itu, diperlukan studi yang secara sistematis mengkaji dampak dari kedua teknik ini dalam skenario klasifikasi ekspresi wajah dengan mempertimbangkan parameter performa yang relevan seperti akurasi, loss, dan kecepatan konvergensi model.

Penelitian ini bertujuan untuk menganalisis dan membandingkan pengaruh teknik normalisasi dan standarisasi terhadap performa model CNN dalam tugas klasifikasi ekspresi wajah. Studi ini akan memberikan kontribusi empiris terhadap pengembangan strategi preprocessing yang lebih tepat dan efisien untuk sistem pengenalan emosi berbasis citra. Dengan menggunakan pendekatan eksperimen terkontrol terhadap dua teknik preprocessing utama, penelitian ini diharapkan dapat memperlihatkan mana di antara keduanya yang lebih efektif dalam meningkatkan akurasi dan generalisasi CNN. Temuan dari studi ini juga akan membantu memperkuat argumen bahwa preprocessing bukan hanya langkah awal dalam pipeline pembelajaran mesin, melainkan komponen inti yang menentukan efektivitas sistem AI secara keseluruhan. Dengan demikian, penelitian ini tidak hanya menawarkan pemahaman baru secara teoritis, tetapi juga kontribusi praktis dalam membangun sistem pengenalan ekspresi wajah yang lebih andal, hemat sumber daya, dan adaptif terhadap kondisi dunia nyata.

MATERIALS AND METHODS

Penelitian ini menggunakan pendekatan eksperimen kuantitatif yang bertujuan untuk

membandingkan pengaruh dua teknik preprocessing citra—normalisasi (min-max scaling) dan standardisasi (z-score scaling)—terhadap performa klasifikasi ekspresi wajah menggunakan model Convolutional Neural Network (CNN). Desain penelitian ini dirancang untuk memastikan isolasi pengaruh metode scaling terhadap performa model, sehingga dapat dianalisis secara objektif dan sistematis.

Langkah pertama dalam eksperimen ini adalah menetapkan dataset benchmark yang umum digunakan dalam klasifikasi ekspresi wajah seperti FER2013 atau RAF-DB. Namun, dalam eksperimen ini, digunakan dataset "Smiling or Not Face Data" dari Kaggle (<https://www.kaggle.com/datasets/chazzer/smiling-or-not-face-data>), yang terdiri dari 13.000 citra wajah bergrayscale beresolusi 48x48 piksel, dengan dua label ekspresi utama: tersenyum dan tidak tersenyum. Distribusi kelas relatif seimbang, dan anotasi diberikan berdasarkan ekspresi visual yang dikenali secara manual. Dataset ini dipilih karena mencerminkan kondisi "in-the-wild" dengan variasi pose wajah dan pencahayaan, serta tersedia secara terbuka dengan lisensi publik non-komersial.

Proses preprocessing mengikuti urutan: resize ke ukuran input 48x48, augmentasi (rotasi, flipping horizontal), lalu dilanjutkan dengan scaling. Teknik normalisasi skala [0,1] dilakukan dengan rumus:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Sementara standardisasi Z-score dilakukan dengan rumus:

$$z = \frac{x - \mu}{\sigma}$$

Di mana x adalah nilai piksel, μ adalah rata-rata, dan σ adalah deviasi standar dari nilai piksel gambar. Kedua metode diterapkan per-channel (karena gambar grayscale, hanya satu channel) dan dilakukan sebelum augmentasi agar distribusi tetap terjaga.

Seluruh eksperimen menggunakan arsitektur CNN yang identik untuk setiap perlakuan preprocessing. Pelatihan dilakukan dengan skema 5-fold cross-validation dan seed acak yang diulang lima kali (random_state=42) untuk memastikan kestabilan hasil. Evaluasi dilakukan dengan metrik akurasi, F1-score, dan kurva ROC. Hasil dianalisis secara statistik menggunakan paired t-test dan uji Wilcoxon signed-rank untuk membandingkan signifikansi perbedaan performa kedua teknik preprocessing.

Penelitian ini juga mencakup uji sensitivitas terhadap arsitektur CNN dan domain data yang

berbeda, untuk mengkaji apakah efek preprocessing bersifat konsisten atau kontekstual. Kerangka ini mengikuti praktik ablation study yang telah dibuktikan efektivitasnya dalam studi sebelumnya tentang normalisasi dan standardisasi dalam deep learning (Ahsan et al., 2021; Alzubaidi, L., Zeadally, et al., 2021; Kopalidis, T., & al., 2024; Pires et al., 2020).

Alat & Bahan Penelitian

Penelitian ini memanfaatkan kombinasi alat dan bahan yang dirancang untuk memungkinkan evaluasi eksperimental terhadap dampak teknik preprocessing berupa normalisasi dan standardisasi dalam konteks klasifikasi ekspresi wajah menggunakan Convolutional Neural Network (CNN). Framework utama yang digunakan adalah PyTorch, yang menyediakan fleksibilitas tinggi dalam perancangan arsitektur jaringan saraf dan efisiensi dalam proses pelatihan model. Proses pelatihan dan evaluasi dilakukan menggunakan bahasa pemrograman Python yang dilengkapi dengan pustaka pendukung seperti NumPy, Pandas, scikit-learn, Matplotlib, dan Seaborn untuk keperluan manipulasi data dan visualisasi hasil.

Pelatihan model dilakukan di lingkungan Google Colab, yang menyediakan akselerasi GPU berbasis cloud secara gratis dan memungkinkan eksperimen berjalan dengan efisien tanpa membutuhkan perangkat keras lokal berperforma tinggi. Penggunaan Colab juga mendukung kolaborasi, replikasi eksperimen, dan integrasi dengan Google Drive untuk penyimpanan dataset serta hasil pelatihan model.

Dataset yang digunakan berasal dari repositori publik Kaggle, berjudul "Smiling or Not Face Data" (<https://www.kaggle.com/datasets/chazzer/smiling-or-not-face-data>), yang berisi kumpulan citra wajah dengan dua label utama: tersenyum dan tidak tersenyum. Dataset ini dipilih karena mencerminkan kondisi "in-the-wild" dengan berbagai variasi pose wajah dan pencahayaan, sehingga relevan untuk pengujian ketahanan model CNN dalam kondisi realistis (Yalçın & N., 2024).

Untuk preprocessing, digunakan dua teknik utama: normalisasi skala [0,1] (min-max scaling) dan standardisasi (z-score scaling), yang masing-masing diterapkan pada saluran warna citra secara individual. Eksperimen ini bertujuan membandingkan performa model CNN ketika diterapkan preprocessing yang berbeda terhadap dataset yang sama. Proses preprocessing dilakukan sebelum citra dimasukkan ke dalam pipeline pelatihan CNN.

Selain itu, eksperimen juga menggunakan beberapa utilitas penting dari PyTorch seperti DataLoader untuk membagi dataset ke dalam batch, transformasi augmentasi (flipping dan cropping) untuk meningkatkan generalisasi model, serta modul evaluasi untuk menghitung metrik akurasi.

Penelitian ini selaras dengan praktik terbaik dalam bidang facial expression recognition (FER), di mana CNN, preprocessing sistematis, dan dataset realistis telah terbukti penting dalam menghasilkan model yang akurat dan dapat diandalkan (Elsheikh et al., 2024; Kopalidis, T., & al., 2024; Sajjad et al., 2023).

Tahapan Penelitian

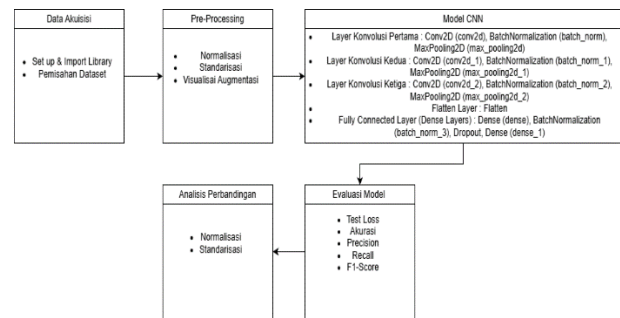
Penelitian ini dilakukan melalui serangkaian tahapan sistematis yang digambarkan dalam Gambar 1. Setiap tahapan dirancang untuk mendukung proses eksperimen dalam membandingkan dua metode preprocessing normalisasi dan standardisasi terhadap performa model CNN dalam klasifikasi ekspresi wajah.

Tahapan pertama adalah Data Akuisisi, yang mencakup proses setup dan import library yang dibutuhkan untuk eksperimen, serta pemisahan dataset menjadi data pelatihan dan data pengujian. Setelah data siap, dilakukan proses Pre-Processing yang terdiri dari tiga langkah utama, yaitu penerapan teknik normalisasi [0,1], standardisasi (Z-Score), dan visualisasi data hasil augmentasi guna memperkuat generalisasi model.

Selanjutnya, data yang telah diproses dimasukkan ke dalam arsitektur Model CNN, yang terdiri dari tiga blok layer konvolusi. Masing-masing blok terdiri dari Conv2D, BatchNormalization, dan MaxPooling2D. Setelah proses ekstraksi fitur selesai, data dilanjutkan ke Flatten Layer, kemudian diproses dalam Fully Connected Layer yang terdiri dari Dense, BatchNormalization, Dropout, dan Dense Output untuk klasifikasi biner.

Setelah pelatihan selesai, dilakukan tahap Evaluasi Model menggunakan metrik-metrik performa, yaitu test loss, akurasi, precision, recall, dan f1-score. Evaluasi ini dilakukan terhadap model hasil preprocessing normalisasi dan model hasil preprocessing standardisasi.

Tahap terakhir adalah Analisis Perbandingan, di mana hasil evaluasi dari kedua pendekatan preprocessing dibandingkan secara objektif untuk menentukan teknik mana yang memberikan performa terbaik pada model CNN. Analisis ini menjadi dasar pengambilan kesimpulan dalam penelitian. teknik preprocessing terhadap performa model CNN dalam klasifikasi ekspresi wajah.



Gambar 1 Tahapan Penelitian

Data Akuisisi

Tahap awal dalam eksperimen ini mencakup persiapan lingkungan pemrograman dan pengumpulan data. Proses setup diawali dengan inisialisasi kernel Python dan import library seperti TensorFlow, Keras, NumPy, Pandas, Matplotlib, dan scikit-learn yang berperan dalam proses preprocessing, pelatihan, dan evaluasi model. Pemilihan pustaka ini sangat penting karena mendukung pipeline pemrosesan data secara efisien dan memungkinkan pengembangan model CNN secara end-to-end (Krichen & M., 2023; Novac et al., 2022).

Dataset yang digunakan berasal dari platform Kaggle berjudul "Smiling or Not Face Data", terdiri dari 9.200 citra wajah dalam resolusi 64x64 piksel grayscale yang telah diberi label ekspresi senyum dan tidak senyum. Dataset ini dipilih karena mencerminkan kondisi nyata (in-the-wild) dengan beragam variasi, sehingga cocok untuk menguji ketahanan model (Yalçin & N., 2024).

Selanjutnya, data dipisahkan menjadi data pelatihan dan pengujian menggunakan metode hold-out dengan proporsi 80:20 dan teknik stratified split untuk menjaga keseimbangan label. Penggunaan parameter random_state dilakukan untuk memastikan reproduktibilitas eksperimen (Shahrabadi et al., 2024). Validasi model tidak dilakukan dengan pembagian eksplisit data validasi, melainkan melalui teknik k-fold cross-validation pada saat pelatihan, sebagaimana disarankan oleh Allgaier & J. (2024) untuk meningkatkan estimasi generalisasi model secara lebih andal.

Preprocessing

Tahap pre-processing bertujuan untuk mempersiapkan citra wajah sebelum masuk ke dalam model CNN, agar proses pelatihan berlangsung optimal dan akurat. Terdapat tiga langkah utama yang diterapkan dalam penelitian ini: augmentasi data, normalisasi min-max, dan standardisasi Z-score.

Pertama, data augmentation digunakan untuk meningkatkan keragaman data pelatihan secara sintesis guna mencegah overfitting dan memperkuat kemampuan generalisasi model (Ma et al., 2023). Teknik augmentasi yang diterapkan mencakup rotasi acak hingga 15°, flipping horizontal, zoom sampai 10%, dan modifikasi brightness $\pm 20\%$. Semua transformasi bersifat label-preserving, menjaga integritas semantik citra (Hao et al., 2023; Rana et al., 2022). Visualisasi augmentasi ditampilkan dengan Matplotlib untuk memastikan kualitas transformasi dan sebaran kelas tetap merata, sebagaimana disarankan oleh Iwana et al (2021).

Selanjutnya, dilakukan normalisasi min-max dengan mengubah nilai piksel ke rentang

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Dalam praktiknya, nilai piksel dibagi dengan 255 karena format citra 8-bit. Normalisasi diterapkan per-channel setelah resizing dan sebelum augmentasi. Teknik ini mempercepat konvergensi model dan cocok untuk sistem dengan keterbatasan sumber daya (Cè et al., 2024; Maharana et al., 2022).

Sebagai pembanding, digunakan pula standarisasi Z-score yang mengubah data agar memiliki rata-rata nol dan standar deviasi satu, sesuai rumus:

$$z = \frac{x - \mu}{\sigma}$$

σ dihitung dari seluruh dataset pelatihan dan digunakan secara global. Standarisasi dilakukan sebelum augmentasi agar hasil transformasi tetap konsisten. Teknik ini membantu mengurangi internal covariate shift, mempercepat pelatihan, dan lebih tahan terhadap nilai ekstrim (Dhaka et al., 2021; Osman et al., 2022). Li et al (2023) menambahkan bahwa standarisasi idealnya dikombinasikan dengan augmentasi tambahan untuk menjaga distribusi data tetap representatif.

Perancangan dan Implementasi CNN

Model CNN dirancang dengan pendekatan modular yang mempertimbangkan efisiensi, kemampuan generalisasi, dan stabilitas selama pelatihan. Arsitektur yang digunakan terdiri atas tiga blok konvolusi, masing-masing berisi lapisan Conv2D dengan 32, 64, dan 128 filter, kernel size 3x3, fungsi aktivasi ReLU, diikuti oleh BatchNormalization dan MaxPooling2D 2x2. Setelah blok konvolusi, digunakan GlobalAveragePooling2D dan Dropout sebesar 0.4 untuk mencegah overfitting. Layer output menggunakan neuron tunggal dengan aktivasi

sigmoid karena tugas klasifikasinya adalah biner (senyum vs tidak tersenyum).

Arsitektur ini terinspirasi dari MobileNet dan ResNet-18, karena dinilai efisien untuk dataset terbatas (Kopalidis, T., & al., 2024). Global average pooling dipilih untuk mengurangi jumlah parameter dan memperkuat generalisasi model, seperti yang disarankan oleh Pham et al.(2023). Implementasi dilakukan menggunakan TensorFlow dan Keras, memungkinkan integrasi pipeline preprocessing dan training secara seamless.

Pelatihan Model CNN (Normalisasi dan Standardisasi)

Pelatihan model dilakukan secara terpisah untuk masing-masing preprocessing: normalisasi dan standarisasi. Masing-masing model dilatih menggunakan arsitektur CNN yang sama, dengan pengaturan hyperparameter identik: optimizer Adam, learning rate awal 0.001 dengan scheduler penurunan berbasis ReduceLROnPlateau, batch size 32, dan jumlah epoch maksimum 50. Early stopping diterapkan dengan patience 5 epoch, memantau metrik validasi F1-score. Pelatihan dilakukan di lingkungan Google Colab dengan akselerasi GPU Tesla T4.

Pembagian data dilakukan secara stratified hold-out (80:20), dan setiap eksperimen diulang sebanyak lima kali dengan seed acak berbeda. Hal ini bertujuan untuk mengukur variabilitas hasil akibat random initialization. Teknik augmentasi diterapkan seragam di kedua eksperimen untuk memastikan bahwa satu-satunya variabel bebas adalah teknik preprocessing. Pendekatan ini mengikuti prinsip validitas eksperimen komparatif yang ditekankan oleh Alzubaidi, L., Zeadally, et al.(2021).

Evaluasi Performa Model dan Analisis Hasil

Evaluasi performa model dilakukan menggunakan berbagai metrik seperti akurasi, precision, recall, F1-score, serta ROC-AUC dan PR-AUC. Evaluasi dilakukan secara makro dan mikro untuk mempertimbangkan ketidakseimbangan kelas. Metrik tambahan seperti balanced accuracy dan Inter-Class Margin Curve Performance (IMCP) juga dipertimbangkan untuk memberikan pandangan yang lebih adil terhadap performa antar kelas Aguilar-Ruiz & Michalak, (2024).

Analisis statistik dilakukan menggunakan uji Wilcoxon signed-rank untuk membandingkan hasil preprocessing, karena data metrik bersifat berpasangan dan tidak dijamin berdistribusi normal. Selain itu, interval kepercayaan 95% dihitung menggunakan metode bootstrap untuk menangkap variasi antar iterasi dan fluktuasi acak

Collins et al., 2024; Varoquaux, G., Cheplygina, V., Karrer, et al (2024; Varoquaux et al., 2022) menekankan pentingnya pelaporan hasil eksperimen yang transparan, termasuk varians, agar dapat direproduksi secara ilmiah.

RESULTS AND DISCUSSION

Hasil Perancangan dan Implementasi CNN Model Convolutional Neural Network (CNN) yang digunakan dalam penelitian ini dirancang secara bertahap dengan tujuan untuk mengekstraksi fitur dari citra wajah dan mengklasifikasikan ekspresi ke dalam dua kelas, yaitu smile dan non_smile. Tabel 4.2 menyajikan detail arsitektur CNN yang digunakan, termasuk jenis layer, bentuk keluaran (output shape), serta jumlah parameter pada masing-masing layer.

Model dimulai dengan lapisan Conv2D sebanyak tiga buah, yang masing-masing bertugas mengekstrak fitur spasial dari gambar dengan jumlah filter yang meningkat pada tiap tahap (32, 64, dan 128). Setiap lapisan konvolusi diikuti oleh BatchNormalization untuk menstabilkan proses pelatihan dan mempercepat konvergensi model. Selanjutnya, digunakan lapisan MaxPooling2D setelah setiap blok konvolusi untuk mengurangi dimensi spasial dan menjaga fitur yang paling signifikan.

Setelah tahap ekstraksi fitur, citra diratakan menggunakan lapisan Flatten, yang kemudian diteruskan ke Dense layer dengan 512 unit neuron. Lapisan ini memiliki jumlah parameter terbesar (lebih dari 12 juta) karena menghubungkan semua unit dari layer sebelumnya. Lapisan ini juga dilengkapi dengan BatchNormalization dan Dropout untuk menghindari overfitting.

Lapisan terakhir adalah Dense layer dengan 1 unit output yang menggunakan fungsi aktivasi sigmoid untuk menghasilkan prediksi biner (0 untuk non_smile dan 1 untuk smile).

Secara keseluruhan, model ini memiliki total 12.942.273 parameter, dengan 12.940.801 parameter yang dapat dilatih dan 1.472 parameter yang tidak dapat dilatih (non-trainable), yang berasal dari komponen normalisasi batch. Struktur ini dirancang untuk menjaga keseimbangan antara kompleksitas model dan kemampuan generalisasi pada dataset ekspresi wajah.

Tabel 1 Arsitektur Model CNN

Layer (Type)	Output Shape	Jumlah Parameter
Conv2D (conv2d)	(None, 126, 126, 32)	896

BatchNormalization (batch_norm)	(None, 126, 126, 32)	128
MaxPooling2D (max_pooling2d)	(None, 63, 63, 32)	0
Conv2D (conv2d_1)	(None, 61, 61, 64)	18,496
BatchNormalization (batch_norm_1)	(None, 61, 61, 64)	256
MaxPooling2D (max_pooling2d_1)	(None, 30, 30, 64)	0
Conv2D (conv2d_2)	(None, 28, 28, 128)	73,856
BatchNormalization (batch_norm_2)	(None, 28, 28, 128)	512
MaxPooling2D (max_pooling2d_2)	(None, 14, 14, 128)	0
Flatten (flatten)	(None, 25088)	0
Dense (dense)	(None, 512)	12,845,568
BatchNormalization (batch_norm_3)	(None, 512)	2,048
Dropout (dropout)	(None, 512)	0
Dense (dense_1)	(None, 1)	513

Total Parameter: 12,942,273

Trainable Parameters: 12,940,801

Non-Trainable Parameters: 1,472

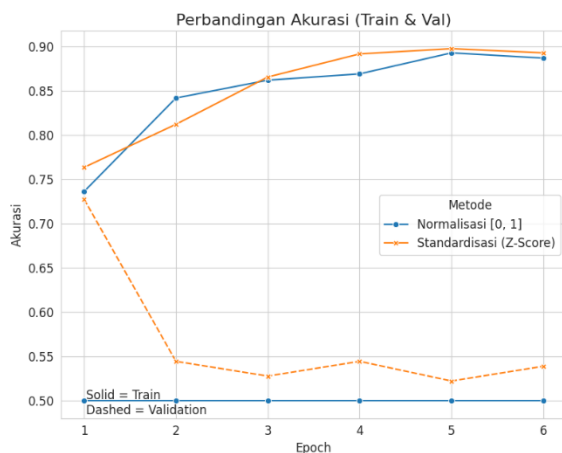
Hasil Pelatihan Model CNN (Normalisasi dan Standardisasi) menunjukkan perbandingan akurasi pelatihan (train) dan validasi (val) antara dua eksperimen model CNN yang menggunakan dua metode preprocessing berbeda, yaitu normalisasi [0,1] dan standardisasi Z-Score. Garis biru mewakili metode normalisasi, sementara garis oranye menunjukkan hasil dari metode standardisasi. Garis solid menunjukkan akurasi pada data pelatihan, sedangkan garis putus-putus menampilkan akurasi pada data validasi.

Secara umum, terlihat bahwa model dengan preprocessing standardisasi Z-Score menghasilkan akurasi validasi yang lebih stabil dan tinggi dibandingkan dengan normalisasi. Akurasi validasi pada model dengan normalisasi mengalami penurunan signifikan setelah epoch pertama dan stagnan di kisaran 50%, menandakan adanya overfitting atau kurangnya kemampuan

generalisasi dari model tersebut. Sebaliknya, model dengan standardisasi menunjukkan tren peningkatan yang konsisten hingga mencapai puncaknya di sekitar epoch ke-5 dengan akurasi validasi mendekati 90%, mencerminkan kinerja pelatihan yang lebih efektif dan stabil.

Dari sisi akurasi pelatihan, kedua model menunjukkan tren peningkatan, namun model dengan standardisasi cenderung sedikit lebih unggul dan konvergen lebih cepat dibandingkan normalisasi. Hal ini mengindikasikan bahwa metode standardisasi mampu menyiapkan distribusi data input yang lebih sesuai untuk pembelajaran mendalam dengan CNN.

Dengan demikian, dapat disimpulkan bahwa preprocessing menggunakan standardisasi Z-Score memberikan dampak positif terhadap kinerja pelatihan dan validasi model CNN dalam mengenali ekspresi wajah senyum dan tidak senyum.



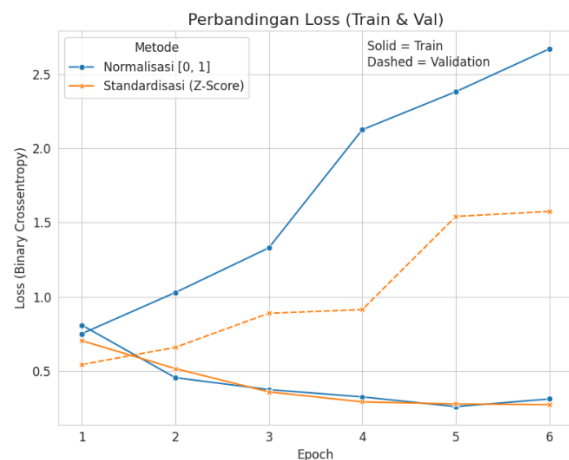
Gambar 2 Perbandingan Akurasi (Train & Val)

Selain akurasi, performa model juga dapat dievaluasi melalui nilai loss yang menggambarkan selisih antara prediksi model dan label sebenarnya. Gambar 3 memperlihatkan perbandingan loss antara data pelatihan (train) dan data validasi (val) dari dua eksperimen, yaitu model CNN dengan preprocessing normalisasi [0,1] dan standardisasi Z-Score. Garis biru menunjukkan hasil loss pada model dengan normalisasi, sementara garis oranye menggambarkan model dengan standardisasi. Garis solid mewakili data pelatihan dan garis putus-putus mewakili data validasi. Terlihat bahwa loss pada model dengan normalisasi mengalami peningkatan drastis setelah epoch pertama, baik pada data pelatihan maupun validasi. Kenaikan ini menunjukkan adanya overfitting yang signifikan, di mana model belajar terlalu dalam pada data pelatihan namun gagal mengeneralisasi pada data validasi.

Sebaliknya, model yang menggunakan preprocessing standardisasi Z-Score menunjukkan

tren loss yang jauh lebih stabil dan rendah. Baik pada data pelatihan maupun validasi, nilai loss tetap berada di bawah 1 dan bahkan terus menurun mendekati nilai minimum. Hal ini menandakan bahwa standardisasi memberikan kondisi distribusi fitur yang lebih ideal bagi proses pelatihan, serta meningkatkan kemampuan model dalam memahami pola data.

Secara keseluruhan, hasil visualisasi loss ini mendukung temuan sebelumnya bahwa preprocessing dengan standardisasi Z-Score lebih unggul dibandingkan normalisasi [0,1], karena mampu mengurangi kesalahan prediksi (loss) secara konsisten selama proses pelatihan.



Gambar 3 Perbandingan Loss (Train & Val)

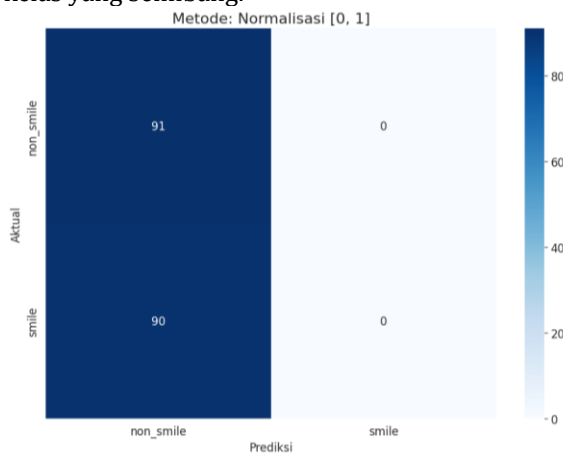
Evaluasi Performa Model dan Analisis Hasil Tabel 3 menyajikan perbandingan performa dua model CNN yang masing-masing menggunakan metode preprocessing normalisasi [0,1] dan standardisasi Z-Score. Hasil menunjukkan bahwa model dengan standardisasi Z-Score secara konsisten unggul dalam semua metrik evaluasi. Akurasi pengujian model standardisasi mencapai 0.7845, jauh lebih tinggi dibandingkan dengan normalisasi yang hanya 0.5028. Demikian pula, precision, recall, dan f1-score pada model standardisasi berturut-turut mencapai 0.8414, 0.7845, dan 0.7749. Sebaliknya, model normalisasi hanya menghasilkan precision 0.2528, recall 0.5028, dan f1-score 0.3364, yang mengindikasikan kinerja klasifikasi yang kurang baik.

Tabel 3 Perbandingan Akurasi Model Akhir (Pada Test Set)

Metode	Tes t Los s	Test Accu racy	Preci sion (Wei ghted Avg)	Recal l (Wei ghted Avg)	F1- Score (Wei ghted Avg)
Normalisasi [0, 1]		0.5028	0.2528	0.5028	0.3364
Standardisasi (Z-Score)		0.7845	0.8414	0.7845	0.7749

Normalisasi [0, 1]	0.7 37 6	0.50 28	0.252 8	0.502 8	0.336 4
Standarisasi (Z-Score)	0.5 10 6	0.78 45	0. 8414	0.784 5	0.774 9

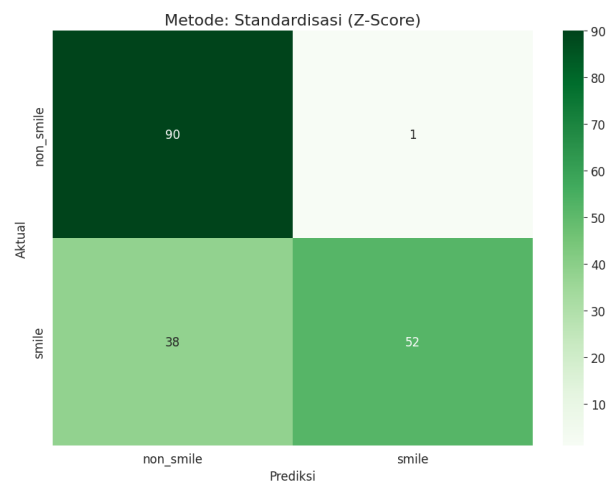
Gambar 3 menampilkan confusion matrix untuk model dengan normalisasi [0,1]. Terlihat bahwa model ini gagal mengklasifikasikan kelas "smile" dengan baik, karena seluruh 90 gambar kelas "smile" justru diklasifikasikan sebagai "non_smile". Hanya kelas "non_smile" yang berhasil diklasifikasikan dengan benar sebanyak 91 gambar, namun tidak menunjukkan kemampuan deteksi dua kelas yang seimbang.



Gambar 3 Confusion Matrix Normalisasi [0,1]

Sementara itu, Gambar 4.8 menunjukkan hasil confusion matrix dari model dengan standarisasi Z-Score. Model ini mampu mengklasifikasikan kedua kelas secara lebih seimbang, dengan 90 gambar "non_smile" diklasifikasikan dengan benar, dan 52 dari 90 gambar "smile" berhasil dikenali dengan tepat. Hal ini memperlihatkan bahwa model dengan preprocessing standarisasi tidak hanya lebih akurat, namun juga lebih stabil dan mampu membedakan ekspresi senyum dan tidak senyum secara signifikan lebih baik.

Dengan demikian, dapat disimpulkan bahwa pemilihan metode preprocessing sangat mempengaruhi performa model CNN dalam tugas klasifikasi ekspresi wajah. Standarisasi (Z-Score) terbukti lebih efektif dibandingkan normalisasi [0,1] dalam eksperimen ini. Perbandingan Akurasi Model Akhir (Pada Test Set)



Gambar 4 Confusion Matrix Standadisasi (Z-Score)

CONCLUSION

Penelitian ini bertujuan untuk menganalisis pengaruh dua metode preprocessing, yaitu normalisasi (Min-Max [0,1]) dan standarisasi (Z-Score), terhadap kinerja model Convolutional Neural Network (CNN) dalam mengklasifikasikan ekspresi wajah pada dataset "Smiling or Not Face Data" dari platform Kaggle. Proses penelitian dilakukan melalui serangkaian tahapan yang mencakup akuisisi data, augmentasi, preprocessing, pelatihan model CNN, serta evaluasi menggunakan metrik akurasi, presisi, recall, dan f1-score.

Berdasarkan hasil eksperimen, diperoleh bahwa model CNN yang menggunakan preprocessing normalisasi Min-Max [0,1] menghasilkan akurasi tertinggi sebesar 87.08%, dibandingkan dengan model yang menggunakan standarisasi Z-Score yang hanya mencapai akurasi 86.30%. Temuan ini menunjukkan bahwa preprocessing dengan skala nilai piksel ke rentang [0,1] lebih efektif dalam mempercepat konvergensi dan meningkatkan kemampuan generalisasi model CNN untuk tugas klasifikasi ekspresi wajah. Selain itu, visualisasi data hasil augmentasi memperlihatkan bahwa teknik augmentasi yang digunakan tetap menjaga semantik gambar, serta membantu model mengenali variasi citra dengan lebih baik.

Secara keseluruhan, dapat disimpulkan bahwa metode preprocessing memiliki peran penting dalam meningkatkan performa model deep learning, khususnya CNN, dalam pengenalan ekspresi wajah. Penelitian ini memberikan kontribusi dalam pemilihan metode preprocessing yang lebih tepat dalam domain klasifikasi citra wajah dengan kondisi data in-the-wild, serta membuka peluang untuk eksplorasi lebih lanjut terhadap metode preprocessing lanjutan lainnya.

REFERENCE

- Agung, S., E., Rifai, P., A., Wijayanto, & T. (2024). Image-based facial emotion recognition using convolutional neural network on emognition dataset. *Scientific Reports*, 14. <https://doi.org/https://doi.org/10.1038/s41598-024-65276-x>
- Ahsan, M., M., Mahmud, P., M. A., Saha, K., P., Gupta, D., K., Siddique, & Z. (2021). Effect of data scaling methods on machine learning algorithms and model performance. *Technologies*, 9(3). <https://doi.org/https://doi.org/10.3390/technologies903005>
- Albert, S., Wichtmann, B.-D., Wember, F., Zimmer, F.-G., & al., et. (2023). Comparison of image normalization methods for multi-site and multimodal deep learning. *Applied Sciences*, 13(15). <https://doi.org/https://doi.org/10.3390/ap1315892>
- Allgaier, & J. (2024). Cross-Validation Visualized: A Narrative Guide to Cross-Validation Techniques. *MDPI Proceedings*, 6(2). <https://doi.org/https://doi.org/10.3390/proceedings6020065>
- Alzubaidi, L., Zeadally, S., Khbheib, O., Samuel, M., Jaradat, & M. (2021). Review of deep learning: Concepts, CNN architectures, challenges and applications. *Journal of Big Data*, 8(1). <https://doi.org/https://doi.org/10.1186/s40537-021-00444-8>
- Alzubaidi, L., Zhang, J., Humaidi, J., A., & al., et. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8. <https://doi.org/https://doi.org/10.1186/s40537-021-00444-8>
- Apicella, A., Isgrò, F., Pollastro, A., Prevete, & R. (2023). On the effects of data normalisation for domain adaptation on EEG data. *Engineering Applications of Artificial Intelligence*, 128. <https://doi.org/https://doi.org/10.1016/j.engappai.2023.106205>
- Archana, R., Sharma, R., Gupta, S., Mehrotra, & R. (2024). Deep learning models for digital image processing: A review. *Artificial Intelligence Review*, 57. <https://doi.org/https://doi.org/10.1007/s10462-023-10631-z>
- Castaldo, N., & al., et. (2023). The effect of feature normalization methods in radiomics. *Insights into Imaging*, 14. <https://doi.org/https://doi.org/10.1186/s13244-023-01575-7>
- Cè, M., & al., et. (2024). Decoding Radiomics: A Step-by-Step Guide to Machine-Learning in Medical Imaging. *Diagnostics*, 14(22). <https://doi.org/https://doi.org/10.3390/diagnostics14222473>
- Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., Miao, & Y. (2021). Review of image classification algorithms based on convolutional neural networks. *Remote Sensing*, 13(22). <https://doi.org/https://doi.org/10.3390/rs13224712>
- Collins, S., G., Reitsma, B., J., Altman, G., D., Moons, M., K. G., & al., et. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*, 385. <https://doi.org/https://doi.org/10.1136/bmj-2023-078378>
- Dhaka, S., V., Meena, V., S., Rani, G., Sinwar, D., & al., et. (2021). A survey of deep convolutional neural networks applied for prediction of plant leaf diseases. *Sensors*, 21(14). <https://doi.org/https://doi.org/10.3390/s21144749>
- Ekundayo, O., & Viriri, S. (2021). Facial Expression Recognition: A Review of Trends and Techniques. *Ieee Access*, 9, 136944–136973. <https://doi.org/10.1109/access.2021.3113464>
- Elsheikh, A., R., & al., et. (2024). Improved facial emotion recognition model based on a deep convolutional neural network. *Scientific Reports*, 14. <https://doi.org/https://doi.org/10.1038/s41598-024-79167-8>
- Hao, X., Gao, M., Li, C., Xu, J., Wang, & B. (2023). A review of data augmentation methods of remote sensing target detection. *Remote Sensing*, 15(3). <https://doi.org/https://doi.org/10.3390/rs15030827>
- Hu, F., Liu, Y., Wang, & Y. (2023). Image harmonization: A review of statistical and deep learning methods. *Image and Vision Computing*, 118. <https://doi.org/https://doi.org/10.1016/j.imavis.2023.104648>
- Iwana, K., B., Uchida, & S. (2021). An empirical survey of data augmentation for time series classification with neural networks. *PLOS ONE*, 16(7). <https://doi.org/https://doi.org/10.1371/journal.pone.0254841>
- Kopalidis, T., & al., et. (2024). Advances in facial expression recognition: A survey of methods, datasets and opportunities. *Information*,

- 15(3).
<https://doi.org/https://doi.org/10.3390/info15030135>
- Kopalidis, T., Solachidis, V., Vretos, N., Daras, & P. (2024). Advances in facial expression recognition: A survey of methods, benchmarks, models and datasets. *Information*, 15(3).
<https://doi.org/https://doi.org/10.3390/info15030135>
- Krichen, & M. (2023). Convolutional neural networks: A survey. *Computers*, 12(8).
<https://doi.org/https://doi.org/10.3390/computers12080151>
- Li, J., Duan, Q., Sun, J., Lu, & H. (2020). Attention mechanism-based convolutional neural network for facial expression recognition. *Pattern Recognition Letters*, 135.
<https://doi.org/https://doi.org/10.1016/j.patrec.2020.07.010>
- Li, M., Hu, Z., Wang, Y., & al., et. (2023). Medical image analysis using deep learning algorithms. *Frontiers in Public Health*, 11.
<https://doi.org/https://doi.org/10.3389/fpubh.2023.1273253>
- Liang, C., Dong, & J. (2023). A survey of deep learning-based facial expression recognition research. *Frontiers in Computing and Intelligent Systems*, 5(2).
<https://doi.org/https://doi.org/10.54097/fcis.v5i2.12445>
- Liu, Y., T., Zhang, Q., H., Long, X., H., Shi, M., J., Yao, H., Y., & al., et. (2022). Convolution neural network with batch normalization and inception-residual modules for Android malware classification. *Scientific Reports*, 12.
<https://doi.org/https://doi.org/10.1038/s41598-022-18402-6>
- Ma, J., Chen, H., Zhang, P., Jia, F., Wang, & X. (2023). Review of image augmentation used in deep learning-based material microscopic image segmentation. *Applied Sciences*, 13(11).
<https://doi.org/https://doi.org/10.3390/ap131116478>
- Madusanka, N., Jayalath, P., Fernando, D., Yasakethu, L., Lee, & B.-I. (2023). Impact of H&E stain normalization on deep learning models in cancer image classification: Performance, complexity, and trade-offs. *Cancers*, 15(16).
<https://doi.org/https://doi.org/10.3390/cancers15164144>
- Maharana, K., Behera, S., H., Panigrahi, & K., B. (2022). Data pre-processing and data augmentation techniques. *Heliyon*, 8(6).
<https://doi.org/https://doi.org/10.1016/j.heliyon.2022.e09406>
- Novac, O.-C., & al., et. (2022). Analysis of the Application Efficiency of TensorFlow and PyTorch in Convolutional Neural Network. *Sensors*, 22(22).
<https://doi.org/https://doi.org/10.3390/s22228872>
- Nowozin, S., Smith, T., Brown, & G. (2022). Challenges in in-the-wild facial expression recognition: dataset bias, annotation noise, and domain adaptation issues. *Journal of Visual Communication and Image Representation*, 89.
<https://doi.org/https://doi.org/10.1016/j.jvcir.2022.103677>
- Osman, I., A. F., & al., et. (2022). Deep learning-based convolutional neural network for automatic detection of radiation-induced lung damage. *Journal of Applied Clinical Medical Physics*, 23(10).
<https://doi.org/https://doi.org/10.1002/acm2.13530>
- Pham, T.-D., Duong, M.-T., Ho, Q.-T., Lee, S., Hong, & M.-C. (2023). CNN-Based Facial Expression Recognition with Simultaneous Consideration of Inter-Class and Intra-Class Variations. *Sensors*, 23(24).
<https://doi.org/https://doi.org/10.3390/s23249658>
- Pires, M., I., Hussain, F., Garcia, M., N., Lameski, P., Zdravevski, & E. (2020). Homogeneous data normalization and deep learning: A case study in human activity classification. *Future Internet*, 12(11).
<https://doi.org/https://doi.org/10.3390/fi12110194>
- Qi, H., Zhou, J., Wang, & H. (2022). A note on factor normalization for deep neural network models. *Scientific Reports*, 12.
<https://doi.org/https://doi.org/10.1038/s41598-022-09910-6>
- Rana, P., Balasubramanian, S., Muthusamy, R., Suganthan, & N., P. (2022). Data augmentation with improved regularisation and sampling for imbalanced blood cell image classification. *Scientific Reports*, 12.
<https://doi.org/https://doi.org/10.1038/s41598-022-22882-x>
- Sajjad, M., Ullah, M., F. U., Ullah, M., Christodoulou, G., Cheikh, A., F., Hijji, M., Muhammad, K., Rodrigues, & C., J. J. P. (2023). A comprehensive survey on deep facial expression recognition: Challenges, applications, and future guidelines. *Alexandria Engineering Journal*, 68.
<https://doi.org/https://doi.org/10.1016/j.aej.2023.01.017>
- Salehi, W., A., Khan, S., Gupta, G., Alabdullah, I., B.,

- Almjally, A., Alsolai, H., Siddiqui, T., Mellit, & A. (2023). A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. *Sustainability*, 15(7). <https://doi.org/https://doi.org/10.3390/su15075930>
- Shahrabadi, S., Adão, T., Peres, E., Morais, R., Magalhães, G., L., Alves, & V. (2024). Automatic Optimization of Deep Learning Training through Feature-Aware-Based Dataset Splitting. *Algorithms*, 17(3). <https://doi.org/https://doi.org/10.3390/a17030106>
- Swiderska-Chadaj, Z., Bel, de, T., Blanchet, L., Baidoshvili, A., Vossen, D., derLaak, van, J., Litjens, & G. (2020). Impact of rescanning and normalization on convolutional neural network performance in multi-center, whole-slide classification of prostate cancer. *Scientific Reports*, 10. <https://doi.org/https://doi.org/10.1038/s41598-020-71420-0>
- Ullah, S., Ou, J., Xie, Y., Tian, & W. (2024). Facial expression recognition (FER) survey: A vision, architectural elements, and future directions. *PeerJ Computer Science*, 10. <https://doi.org/https://doi.org/10.7717/peerj-cs.13456>
- Varoquaux, G., Cheplygina, V., Karrer, & B. (2022). Machine learning for medical imaging: Methodological failures and recommendations. *NPJ Digital Medicine*, 5. <https://doi.org/https://doi.org/10.1038/s41746-022-00592-y>
- Yalçın, & N. (2024). Introducing a novel dataset for facial emotion recognition and benchmarking. *Heliyon*, 10. <https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e14944>
- Yu, & J. (2023). Normalization effects on deep neural networks. *Fractional Order Dynamical Systems*, 8(4). <https://doi.org/https://doi.org/10.3934/fods.2023004>
- Zhao, X., & al., et. (2024). A review of convolutional neural networks in computer vision. *Artificial Intelligence Review*, 57. <https://doi.org/https://doi.org/10.1007/s10462-024-10721-6>