

ANALISIS KOMPARATIF EKSTRAKTOR FITUR VGG-16, VGG-19, INCEPTION V3 UNTUK KLASIFIKASI KONDISI BAN MENGGUNAKAN METODE EVALUASI VISUAL DAN BERBASIS KLASTER

Ridwan Ramadhan¹, Imam Yuadi²

¹Magister of Human Resource Development, Graduate School, Airlangga University, Surabaya

²Department of Information and Library Science, Faculty of Social and Political Sciences, Airlangga University, Surabaya

Keywords:

VGG-16, VGG-19, Inception V3

Correspondent Email:

ridwan.ramadhan-2025@pasca.unair.ac.id

Abstrak. Mendeteksi kondisi ban kendaraan secara otomatis menjadi aspek penting dalam meningkatkan keselamatan berkendara. Penelitian ini bertujuan mengembangkan model klasifikasi untuk membedakan kondisi ban dalam dua kategori, yaitu good dan defective, dengan tingkat akurasi yang optimal. Dataset yang digunakan terdiri dari 110 citra ban yang seimbang antar kelas. Tahapan penelitian meliputi prapemrosesan citra (resizing, normalisasi, dan augmentasi), ekstraksi fitur menggunakan model deep learning (VGG-16, VGG-19, dan Inception V3), serta klasifikasi menggunakan algoritma machine learning seperti Logistic Regression, Support Vector Machine (SVM), dan Neural Network. Hasil penelitian menunjukkan bahwa VGG-16 merupakan model yang paling stabil dan andal, didukung oleh performa klasifikasi yang lebih konsisten serta visualisasi MDS yang relatif terstruktur dengan nilai silhouette positif meskipun masih terjadi tumpang tindih antar kelas. Sebaliknya, VGG-19 menunjukkan performa terendah dengan kemampuan ekstraksi fitur yang kurang diskriminatif. Inception V3 berada pada tingkat kinerja menengah, cukup baik dalam mendeteksi kerusakan besar namun kurang sensitif terhadap kerusakan kecil. Temuan ini mengindikasikan bahwa penggunaan citra visual saja belum cukup untuk menghasilkan pemisahan kelas yang kuat. Oleh karena itu, integrasi data multimodal seperti tekanan dan kedalaman tapak ban direkomendasikan untuk meningkatkan akurasi sistem.



Copyright © [JITET](http://www.jitet.org) (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

Abstract. Automatically detecting the condition of vehicle tires is a critical aspect of improving driving safety. This study aims to develop a classification model to distinguish tire conditions into two categories—good and defective—with optimal accuracy. The dataset used consists of 110 tire images that are class-balanced. The research stages include image preprocessing (resizing, normalization, and augmentation), feature extraction using deep learning models (VGG-16, VGG-19, and Inception V3), and classification using machine learning algorithms such as Logistic Regression, Support Vector Machine (SVM), and Neural Network. The results show that VGG-16 is the most stable and reliable model, supported by more consistent classification performance and relatively structured MDS visualizations with positive silhouette values, although there is still some overlap between classes. Conversely, VGG-19 exhibits the lowest performance with less discriminative feature extraction capabilities. Inception V3 falls into the mid-range performance category, performing well in detecting major damage but lacking

sensitivity to minor damage. These findings indicate that visual images alone are insufficient to achieve strong class separation. Therefore, the integration of multimodal data such as tire pressure and tread depth is recommended to improve system accuracy.

1. PENDAHULUAN

Sebuah Proses klasifikasi kondisi ban kendaraan sangatlah penting untuk para pengguna kendaraan. Sebuah ban dengan klasifikasi kondisi good and defective adalah indikator dari kemampuan ban untuk digunakan dalam berkendara, hal tersebut sangat berpengaruh besar terhadap keselamatan dan kerugian apabila terjadi kecelakaan. Sebagai contoh, hubungan antara kondisi ban dengan karakteristik ban telah diidentifikasi, di mana kualitas ban yang baik berkontribusi pada keselamatan serta memenuhi aspek safety driving [1].

Analisis berbasis kecerdasan buatan seperti Convolutional Neural Networks (CNN) dan algoritma pengolahan gambar menjadi semakin populer di era digital karena mereka dapat menggabungkan penilaian ban kendaraan untuk meningkatkan ketepatan klasifikasi kondisi ban tersebut hingga bisa dinyatakan laik jalan [2]. Metode ini tidak hanya mempercepat proses klasifikasi tetapi juga meningkatkan akurasi penilaian kondisi ban kendaraan. Model VGG-16 dan VGG-19 telah terbukti efektif dalam ekstraksi fitur dari citra ban kendaraan. Dari sebuah pendekatan tersebut dapat memungkinkan sebuah proses klasifikasi yang lebih cepat dan lebih akurat dibandingkan dengan metode tradisional yang mengandalkan penilaian secara [3].

Lebih lanjut, penelitian sebelumnya menunjukkan bahwa pemanfaatan teknik segmentasi yang dikombinasikan dengan pembelajaran dalam seperti deep learning meningkatkan kemampuan dalam mengevaluasi kondisi ban kendaraan [2]. Dengan menggunakan teknologi ini, peneliti mampu mengidentifikasi kondisi laik dan tidaknya pada ban kendaraan lebih mendalam [4]. Pendekatan evaluasi yang memanfaatkan teknologi tersebut menawarkan pandangan baru terkait potensi dalam analisis ban kendaraan secara real-time, yang sangat dibutuhkan dalam kehidupan sehari-hari maupun pelaku usaha.

Untuk mengategorikan kondisi ban kendaraan secara akurat, pendekatan berbasis

regresi logistik, Support Vector Machine (SVM), dan Neural Networks menawarkan kerangka analitis yang kuat. Metode ini mampu merepresentasikan hubungan probabilistik antarvariabel [5], sementara SVM efektif menangani data non-linear yang umum pada kondisi ban dengan banyak variabel [6]. Kedua pendekatan menunjukkan tingkat akurasi yang tinggi dibanding metode lain.

Tantangan utama dalam identifikasi ban kendaraan adalah tingginya variasi bentuk dari perubahan kondisi pada ban kendaraan. Penelitian menunjukkan bahwa evaluasi kondisi keausan pada posisi tertentu dapat memprediksi kualitas ban secara keseluruhan, namun tetap perlu mempertimbangkan variasi latar belakang [7]. Klasifikasi kondisi ban kendaraan menjadi tantangan dan membutuhkan pendekatan yang lebih inovatif serta menyeluruh, terutama dalam memahami karakteristik serta keunikan ban dari berbagai tipe. Penggunaan model-model canggih dalam mengidentifikasi keausan ban sangat penting, mengingat adanya praktik pemalsuan kondisi ban seperti membuat alur ban baru yang berkaitan langsung dengan isu keamanan safety driving [8]. Oleh karena itu, metode analisis yang mampu memastikan keaslian kondisi ban sangat penting untuk melindungi konsumen dan menjaga kepercayaan serta integritas pengguna maupun pelaku usaha [9].

Dari uraian yang sudah disampaikan, maka penelitian ini bertujuan untuk membuat sebuah model klasifikasi terkait kondisi ban kendaraan berbasis kecerdasan buatan yang lebih efisien dan akurat daripada metode klasifikasi manual. Tujuan dari penelitian ini didasarkan pada uraian yang sudah diberikan. Struktur penulisan jurnal ini dimulai dengan tinjauan pustaka yang membahas teori, penelitian terdahulu, dan teknologi pendukung dalam klasifikasi kondisi ban. Pada bagian metodologi telah dijelaskan sebuah tahapan pengolahan data citra, kemudian ekstraksi fitur menggunakan arsitektur CNN seperti VGG-16, VGG-19 dan Inception V3, serta penerapan algoritma klasifikasi seperti regresi logistik, SVM, dan

Neural Network. Bagian hasil dan pembahasan menguraikan temuan empiris, seperti tingkat akurasi model, seberapa baik suatu algoritma. Akhir sekali, bagian kesimpulan membahas kontribusi utama penelitian untuk meningkatkan keselamatan dengan mengklasifikasikan kondisi ban berbasis teknologi cerdas

2. TINJAUAN PUSTAKA

2.1 Klasifikasi Good dan Defective

Klasifikasi citra ke dalam kategori good dan defective merupakan salah satu bentuk penerapan computer vision yang sangat relevan dalam proses inspeksi kualitas. Pada dasarnya, proses ini bertujuan untuk membedakan objek yang masih berada dalam kondisi baik dengan objek yang telah mengalami kerusakan atau cacat melalui karakteristik visual yang tampak pada citra. Perbedaan tersebut dapat muncul dalam bentuk perubahan tekstur, pola permukaan, bentuk, maupun struktur objek. Dalam konteks ban kendaraan, keberadaan cacat bukan hanya berkaitan dengan kualitas produk, tetapi juga berkaitan langsung dengan aspek keselamatan berkendara. Kerusakan ban dapat memengaruhi stabilitas kendaraan, kenyamanan pengendalian, bahkan meningkatkan risiko kecelakaan apabila tidak terdeteksi sejak dini [2], [10]. Oleh sebab itu, sistem klasifikasi otomatis menjadi semakin penting untuk mendukung proses inspeksi yang lebih cepat, konsisten, dan objektif.

2.2 Machine Learning dan Deep Learning

Perkembangan teknologi pengolahan data telah mendorong penggunaan machine learning dalam berbagai bidang, termasuk klasifikasi citra. Machine learning pada dasarnya merupakan pendekatan yang memungkinkan sistem mempelajari pola dari data, lalu menggunakan pola tersebut untuk melakukan prediksi atau pengelompokan. Dalam klasifikasi citra, metode machine learning tradisional biasanya memerlukan tahapan ekstraksi fitur terlebih dahulu sebelum proses klasifikasi dilakukan. Artinya, kualitas hasil klasifikasi sangat dipengaruhi oleh seberapa baik fitur yang digunakan dapat mewakili karakteristik objek [11], [12].

Di sisi lain, deep learning hadir sebagai pendekatan yang lebih maju karena mampu mempelajari fitur secara otomatis langsung dari

data citra. Melalui jaringan saraf yang terdiri dari banyak lapisan, deep learning dapat menangkap pola visual mulai dari elemen sederhana seperti garis dan tepi hingga pola yang lebih kompleks dan abstrak. Kemampuan inilah yang membuat deep learning sangat efektif untuk menyelesaikan permasalahan klasifikasi citra, terutama ketika objek yang dibedakan memiliki kemiripan visual yang cukup tinggi [13], [14]. Dalam penelitian-penelitian terbaru, deep learning tidak hanya digunakan sebagai model klasifikasi penuh, tetapi juga sering dimanfaatkan sebagai pengekstrak fitur yang kuat.

2.3. Arsitektur VGG-16, VGG-19, dan Inception V3

Dalam studi klasifikasi citra, beberapa arsitektur CNN yang sering digunakan sebagai feature extractor adalah VGG-16, VGG-19, dan Inception V3. VGG-16 dan VGG-19 dikenal sebagai model yang memiliki struktur dalam namun tetap sederhana dari sisi desain. Keduanya menggunakan susunan filter konvolusi berukuran kecil secara berulang, sehingga mampu menangkap detail visual secara bertahap. VGG-16 terdiri atas 16 lapisan berbobot, sedangkan VGG-19 memiliki 19 lapisan berbobot. Dalam banyak penelitian, kedua model ini dikenal mampu menghasilkan fitur yang kuat dan stabil untuk berbagai tugas klasifikasi citra [15], [16].

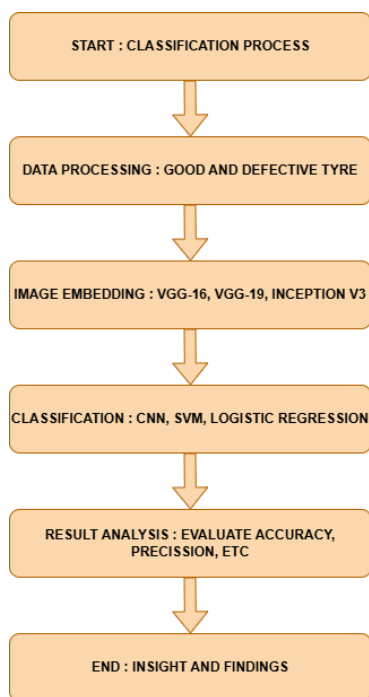
Sementara itu, Inception V3 dirancang dengan pendekatan yang berbeda. Model ini berusaha meningkatkan efisiensi komputasi tanpa mengorbankan kemampuan representasi fitur. Melalui modul Inception, model ini mampu mengekstraksi pola visual dalam berbagai skala secara bersamaan. Hal tersebut membuat Inception V3 menjadi salah satu arsitektur yang cukup andal untuk menangani citra dengan variasi pola yang kompleks [16]. Perbandingan antara VGG-16, VGG-19, dan Inception V3 menjadi penting karena masing-masing arsitektur memiliki karakteristik yang berbeda dalam menghasilkan fitur, sehingga performanya dapat berbeda tergantung pada jenis data yang digunakan [15], [9].

3. METODE PENELITIAN

Dari dataset didapatkan 110 gambar ban dengan dua kondisi berbeda: good and defective. Gambar-gambar ini kami peroleh

dari Kaggle di tautan <https://www.kaggle.com/code/mateuszk013/tyre-quality-90-acc-with-custom-se-resnet/input>. Metodologi ini dirancang untuk menciptakan sistem klasifikasi gambar yang tangguh dan terukur yang dapat mengenali pola keausan pada ban kendaraan. Adapun Langkah-langkahnya adalah pemrosesan data, pembuatan embedding, klasifikasi dengan algoritma, dan penggunaan berbagai metrik untuk melihat seberapa baik model tersebut bekerja. Seluruh proses ditunjukkan dalam diagram alir dari pra-pemrosesan hingga evaluasi kinerja.

Pengaturan percobaan dalam penelitian ini dapat dijelaskan melalui sebuah kombinasi dari berbagai ekstraktor fitur dan pengklasifikasi yang secara konsisten diuji untuk memastikan perbandingan hasil yang seimbang. Kinerja model dievaluasi menggunakan metrik klasifikasi standar seperti akurasi, presisi, recall, F1-score, AUC, dan Koefisien Korelasi Matthews (MCC), yang berfungsi sebagai dasar untuk menilai dan membandingkan efektivitas model. Diagram metodologi penelitian berikut memberikan penjelasan lebih lanjut.



Gambar 1. Research Methodology

3.1.Start

Penelitian dimulai dengan menetapkan tujuan utama yaitu melakukan klasifikasi ban kendaraan. Pada titik ini, sasaran penelitian, ruang lingkup, dan metode yang akan digunakan untuk menjalankan proses secara sistematis telah ditetapkan. Selain itu, dataset yang terdiri dari 110 gambar ban kendaraan dengan pola kondisi ban yang berbeda sedang disiapkan. Gambar-gambar ini dikategorikan ke dalam dua kategori kualitas, yaitu good dan defective.

3.2.Data Processing

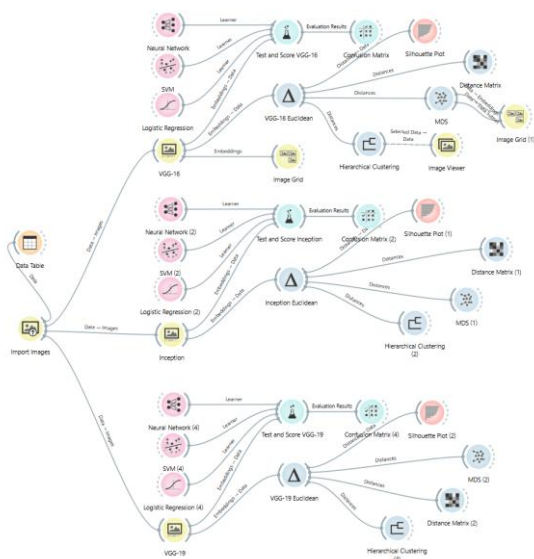
Setelah tahap awal, data dikumpulkan dan dikategorikan menurut kualitasnya (good and defective). Tahap ini sangat penting untuk memastikan bahwa sistem dapat membedakan variasi pola keausan ban berdasarkan kualitas data sebelum memasuki tahap analisis yang lebih mendalam. Secara umum, ada tiga kategori kualitas, yaitu:

- Kualitas good : 55 gambar
- Kualitas defective : 55 gambar

Untuk mempersiapkan dataset yang akan dianalisis, data harus melalui beberapa tahap prapemrosesan agar seragam dan siap digunakan dalam proses ekstraksi fitur. Tahapan tersebut meliputi:

- Resizing : semua gambar dilakukan penyesuaian pixel menjadi tidak lebih dari 2048x2048
- Normalisasi : membagi nilai pixel dengan 255 sehingga berada pada rentang [0,1]
- Augmentasi data : digunakan agar dalam pemrosesan model lebih adaptif terhadap variasi sudut pandang dan kondisi pencahayaan.

3.3. Image Embedding



Gambar 2. Image Embedding

Data gambar kemudian diekstraksi menggunakan metode image embedding. Model deep learning yang digunakan meliputi VGG-16, VGG-19, dan Inception V3. Tujuannya adalah untuk menangkap pola dan fitur yang relevan dari gambar sehingga dapat dipelajari oleh algoritma klasifikasi dengan lebih efektif. VGG-16 dan VGG-19 adalah model yang dikembangkan oleh Visual Geometry Group di University of Oxford. Keduanya memiliki kedalaman yang berbeda VGG-16 memiliki 16 lapisan dan VGG-19 memiliki 19 lapisan, dengan arsitektur yang sederhana dan konsisten yang memanfaatkan lapisan konvolusi dan pooling berulang.

Kedua model ini dikenal karena kemampuannya untuk menghasilkan representasi fitur yang kaya untuk berbagai tugas pengenalan gambar. Namun, mereka biasanya tidak dapat menangani variasi data yang ekstrim, seperti latar belakang yang kompleks atau kondisi pencahayaan yang berbeda [14]. Menurut penelitian, model VGG-16 dan VGG-19 keduanya dapat mencapai akurasi pengenalan gambar yang cukup baik; namun, VGG-19 tampak sedikit lebih baik dalam beberapa uji banding [17].

Inception V3 adalah model yang lebih kompleks yang mengimplementasikan arsitektur Inception untuk mengatasi berbagai masalah dalam proses pelatihan dengan mengizinkan lebih banyak variasi dalam ukuran

kernel dalam lapisan konvolusi. Berbagai studi menunjukkan bahwa Inception V3 secara konsisten memberikan hasil yang superior dibandingkan dengan VGG dalam banyak aplikasi, termasuk deteksi kanker dan pengenalan jenis jaringan pada gambar medis.

Inception V3 menunjukkan hasil yang lebih baik dalam menangani data kompleks dalam satu penelitian dengan Area Under Curve (AUC) sebesar 0.846 untuk klasifikasi resistensi pada kanker ovarium [18]. Selain itu, dengan menggunakan teknik transfer learning, Inception V3 dapat memanfaatkan pelatihan sebelumnya pada dataset yang lebih besar, yang menjadikannya pilihan yang sangat baik untuk berbagai aplikasi praktis [19]. Sebuah penelitian laju akurasi menemukan bahwa Inception V3 dapat mencapai 93,2% dari total jumlah data yang dipelajari [9].

Dalam satu framework pengembangan model, menggabungkan VGG-16, VGG-19, dan Inception V3 dapat menghasilkan keunggulan yang signifikan. Misalnya, ketika digunakan dalam konteks klasifikasi gambar medis, model gabungan dapat menghasilkan klasifikasi yang lebih akurat dan efisien dibandingkan dengan menggunakan satu model secara terpisah [15]. Selain itu, metode ini juga memanfaatkan kekuatan masing-masing model: VGG untuk deteksi awal dengan fitur yang stabil dan Inception V3 untuk mengeksplorasi detail yang lebih dalam pada data yang lebih kompleks [20].

Secara keseluruhan, penggunaan VGG-16, VGG-19, dan Inception V3 dalam berbagai aplikasi deep learning menunjukkan bahwa mereka adalah alat yang sangat berharga dalam pengolahan dan analisis data visual, terutama dalam bidang kesehatan, klasifikasi objek, dan analisis perilaku gambar.

3.4. Classification

Beberapa metode yang umum digunakan dalam klasifikasi untuk analisis data dan pengolahan citra antara lain Convolutional Neural Networks (CNN), Support Vector Machines (SVM), dan regresi logistik. Ketiga pendekatan ini memiliki keunggulan dan keterbatasan masing-masing, serta dapat diterapkan dalam berbagai situasi sesuai kebutuhan.

-Convolutional Neural Networks (CNN) merupakan pendekatan yang sangat efektif

untuk tugas klasifikasi citra. CNN dapat menangkap representasi spasial dan hierarkis dari data citra dengan menggunakan beberapa lapisan, termasuk lapisan konvolusi, pooling, dan fully connected. Bagaimana CNN digunakan untuk klasifikasi penyakit mata dengan menggunakan arsitektur AlexNet dan mencapai akurasi yang tinggi [13]. Penelitian lain juga mencatat bahwa CNN dapat digunakan untuk klasifikasi tingkat kematangan buah pepaya dengan hasil akurasi mencapai 100%, yang menunjukkan potensi besar CNN dalam pengolahan citra [12].

-Support Vector Machines (SVM) adalah metode yang baik untuk klasifikasi dengan data yang berukuran kecil hingga menengah. Penerapan SVM untuk analisis sentimen terkait aplikasi PeduliLindungi dan mencatat keefektifan SVM dalam menangani data yang tidak seimbang [11]. Selain itu, dalam konteks kesehatan, penelitian lain juga mengindikasikan bahwa SVM dapat efektif dalam mendeteksi risiko tuberkulosis, meskipun memiliki keterbatasan ketika harus menangani banyak kelas [21]

-Regresi logistik adalah teknik statistik yang digunakan untuk memprediksi kemungkinan dari dua atau lebih kategori hasil berdasarkan satu atau lebih variabel independen. Metode ini telah ditunjukkan efektif dalam banyak aplikasi pengklasifikasian. Misalnya, mereka membandingkan analisis regresi logistik dengan metode lain untuk menentukan kredit dan menemukan bahwa analisis regresi logistik yang menangani ketidakseimbangan data lebih baik [22]. Selain itu, menunjukkan bahwa regresi logistik biner dapat membuat klasifikasi program pemerintah yang lebih akurat, menunjukkan bahwa teknik ini dapat digunakan di dunia nyata [22].

Karena kemampuan untuk mengekstraksi fitur secara otomatis, CNN dianggap paling cocok untuk analisis citra dari ketiga metode tersebut. Pada akhirnya, jenis dataset dan tujuan analisis menentukan metode terbaik; namun, regresi logistik dan SVM masih relevan, terutama untuk dataset dan data terstruktur berukuran kecil. Selain itu, hal-hal seperti akurasi, efisiensi, dan kemudahan untuk menjelaskan hasil dipertimbangkan.

3.5.Result Analysis

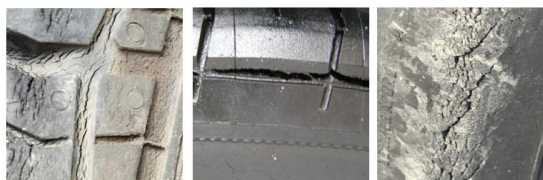
Hasil klasifikasi kemudian dianalisis dengan menggunakan berbagai metrik evaluasi, seperti accuracy, precision, recall, F1-score, maupun metrik pendukung lainnya. Evaluasi model klasifikasi sangat penting dalam berbagai bidang, khususnya dalam pembelajaran mesin dan kecerdasan buatan. Metrik kunci seperti akurasi, presisi, recall, dan skor-F1 banyak digunakan untuk menilai efektivitas model-model ini. Akurasi adalah ukuran proporsi hasil yang benar (baik true positive maupun true negative) di antara total jumlah kasus yang diperiksa [23].

Presisi, di sisi lain, mengukur keakuratan prediksi positif dan didefinisikan sebagai rasio true positive terhadap jumlah true positive dan false positive [24]. Recall, yang juga dikenal sebagai sensitivitas, mengukur kemampuan model untuk mengidentifikasi semua instance yang relevan dan didefinisikan sebagai rasio true positive terhadap jumlah true positive dan false negative [25]. Skor-F1 adalah rata-rata harmonik dari presisi dan recall, memberikan satu ukuran yang menyeimbangkan kedua aspek tersebut, yang sangat berguna ketika distribusi kelas tidak seimbang [26]. Sumber data untuk ban dengan kondisi good mencakup 55 item data, dengan contoh data sebagai berikut:



Gambar 3. good tyre

Sumber data untuk ban dengan kondisi defective 55 item data, dengan contoh data sebagai berikut:



Gambar 4. defective tyre

3.6.End

Tahap terakhir adalah merumuskan insight dan temuan dari keseluruhan proses. Dari sini dapat ditarik kesimpulan mengenai performa

model, keunggulan pendekatan yang dipakai, serta potensi pengembangan lebih lanjut.

4. HASIL DAN PEMBAHASAN

Tabel 1 menggambarkan kinerja tiga model klasifikasi: Logistic Regression, SVM, dan Neural Network, dikombinasikan dengan ekstraksi fitur dari tiga arsitektur pembelajaran mendalam (VGG-16, VGG-19, dan Inception V3). Kinerja ini dievaluasi dengan enam metrik: AUC, CA, F1-score, Precision, Recall, dan MCC.

Tabel 1. Research Model

Model	Feature Extraction	AUC	CA	F1	Precision	Recall	MCC
Logistic Regression	VGG-16	0.912	0.873	0.873	0.873	0.873	0.746
	VGG-19	0.902	0.818	0.818	0.820	0.818	0.638
	Inception V3	0.924	0.836	0.836	0.836	0.836	0.673
SVM	VGG-16	0.904	0.782	0.781	0.785	0.782	0.567
	VGG-19	0.872	0.782	0.782	0.783	0.782	0.565
	Inception V3	0.902	0.845	0.844	0.855	0.845	0.700
Neural Network	VGG-16	0.912	0.873	0.873	0.873	0.873	0.746
	VGG-19	0.872	0.782	0.782	0.783	0.782	0.565
	Inception V3	0.906	0.809	0.809	0.809	0.809	0.618

Inception V3 menghasilkan performa paling unggul pada hampir semua model. Karakteristik arsitektur Inception yang menggunakan *multi-scale convolutional processing* mampu mengekstraksi fitur visual dengan lebih kaya dan representatif.

Contoh terbaik muncul pada kombinasi : Logistic Regression + Inception V3 (AUC: 0.924 tertinggi di semua kombinasi) CA: 0.836 Precision/Recall/F1: 0.836 MCC: 0.673 Model ini menunjukkan keseimbangan sangat baik antara sensitivitas dan spesifisitas, serta kestabilan prediksi antar kelas.

VGG-16 juga menunjukkan performa solid, terutama dalam MCC—indikator terbaik untuk melihat keseimbangan prediksi pada data tidak seimbang.

Dua kombinasi dengan MCC tertinggi adalah : Neural Network + VGG-16 → MCC = 0.746 dan Logistic Regression + VGG-16 → MCC = 0.746. Hal ini menunjukkan bahwa fitur yang dihasilkan VGG-16 bersifat stabil dan informatif, terutama untuk klasifikasi biner yang membutuhkan keseimbangan antar kelas.

VGG-19 menghasilkan hasil yang lebih rendah dibandingkan VGG-16. Hal ini lazim terjadi karena : Kedalaman jaringan yang lebih besar tidak selalu meningkatkan kualitas fitur, Risiko overfitting lebih tinggi dan Fitur cenderung lebih redundan. Namun demikian, hasil tetap stabil, meskipun tidak sebaik VGG-16 atau Inception V3.

Logistic Regression menunjukkan performa paling konsisten tinggi di semua feature extractor, hal ini terjadi karena : Fitur CNN sudah sangat informatif, sehingga model sederhana bekerja sangat baik, Logistic Regression tidak rentan terhadap overfitting pada fitur high-dimension, Kestabilan pada AUC, F1, dan MCC menunjukkan generalisasi yang baik.

SVM menunjukkan hasil baik namun fluktuatif. Kombinasi terbaiknya adalah : SVM + Inception V3 maka MCC = 0.700 (sangat tinggi), Namun SVM cenderung tidak stabil pada VGG-16 dan VGG-19, kemungkinan karena margin yang sensitif terhadap distribusi fitur yang sangat kompleks.

Dengan VGG-16, Neural Network mencapai MCC tertinggi (0.746), tetapi ketika menggunakan Inception V3 dan VGG-19, kinerjanya menurun. Hal ini disebabkan oleh kemungkinan bahwa sampel data tidak cukup besar untuk melatih lapisan akhir NN dengan cara yang ideal.

Secara keseluruhan, untuk dataset ban kendaraan ini, Inception V3 adalah Extractor Fitur terbaik karena menawarkan fitur yang paling kaya dan paling mudah dipisahkan oleh classifier. NN lebih mudah mengalami overfitting dibandingkan dengan classifier linear. Logistic Regression adalah model klasifikasi terbaik karena paling stabil, akurat, dan tidak overfitting. Sementara itu, VGG-16 adalah MCC terbaik karena menunjukkan keseimbangan prediksi yang tinggi baik pada kelas minoritas maupun mayoritas.

		Predicted		
		defective	good	Σ
Actual	defective	92.0 %	15.0 %	55
	good	8.0 %	85.0 %	55
Σ		50	60	110

Gambar 6. confusion matrix result (VGG-16)

Confusion matrix pada Gambar 6 merupakan hasil dari VGG-16 model Neural Network menunjukkan hasil klasifikasi kondisi ban kendaraan ke dalam dua kategori: good and defective, dengan total 110 sampel (55 per kelas).

True Defective terhadap Predicted Defective (92%) Model berhasil mengidentifikasi 92% ban yang benar-benar rusak. Ini menunjukkan tingkat sensitivitas yang sangat baik. Dalam konteks keselamatan kendaraan, keberhasilan mendeteksi ban rusak sangat penting untuk mencegah kecelakaan akibat pecah ban atau hilangnya traksi.

True Defective terhadap Predicted Good (15%) False Negative, Sebanyak 15% ban yang sebenarnya rusak diklasifikasikan sebagai ban baik. Kesalahan tipe ini sangat kritis karena : Ban rusak yang tidak terdeteksi dapat tetap digunakan, kemudian risiko kecelakaan meningkat, terutama pada kecepatan tinggi serta potensi terjadinya blow-out, selip, atau kehilangan kendali kendaraan. Dengan demikian, False Negative merupakan prioritas utama untuk diminimalkan pada sistem deteksi kondisi ban.

True Good terhadap Predicted Defective (8%) False Positive, Sebanyak 8% ban yang baik salah diklasifikasikan sebagai rusak. Kesalahan ini tidak membahayakan keselamatan, tetapi : Dapat meningkatkan biaya operasional karena penggantian ban lebih cepat dari seharusnya. Menurunkan efisiensi logistik jika praktik pemeriksaan dilakukan pada skala besar. Namun demikian, tingkat False Positive ini masih dapat diterima apabila orientasi utama sistem adalah keselamatan.

True Good terhadap Predicted Good (85%) dan Model mampu mengidentifikasi 85% ban layak jalan secara benar. Hal ini menunjukkan stabilitas model dalam mengenali pola keausan atau kondisi ban yang normal.

Pada akhirnya, model ini menunjukkan hasil yang baik dalam mendeteksi kondisi ban kendaraan, terutama dalam menemukan ban yang kondisinya sudah tidak laik jalan. Model ini bisa menjadi bagian penting dari sistem pemeriksaan ban berbasis kecerdasan buatan, meskipun masih ada tingkat False Negative sebesar 15%. Namun, kesalahan kritis tersebut harus diperbaiki lebih lanjut. Model ini cocok digunakan sebagai sistem pendukung keputusan dalam perawatan dan keselamatan kendaraan.

Untuk meningkatkan kinerja model dan mengurangi tingkat False Negative, langkah-langkah berikut dapat dilakukan: Mengubah ambang klasifikasi agar model lebih peka terhadap indikasi kerusakan ban. tingkatkan data, terutama dengan menambah variasi

gambar ban yang rusak dari ringan hingga sedang.

		Predicted		Σ
		defective	good	
Actual	defective	84.3 %	20.3 %	55
	good	15.7 %	79.7 %	55
Σ		51	59	110

Gambar 7. confusion matrix result (VGG-19)

Confusion matrix pada Gambar 7 merupakan hasil dari VGG-19 model Logistic Regression menunjukkan hasil klasifikasi kondisi ban kendaraan ke dalam dua kategori: good and defective, dengan total 110 sampel (55 per kelas).

True Defective terhadap Predicted Defective (84.3%) Model mampu mendeteksi 84.3% ban yang benar-benar rusak. Hal ini merupakan kemampuan penting karena ban rusak yang tidak terdeteksi dapat menyebabkan kecelakaan, kehilangan stabilitas, atau pecah ban (blowout). Model cukup baik dalam mengidentifikasi ban bermasalah, meskipun masih bisa ditingkatkan.

True Defective terhadap Predicted Good (20.3%) False Negative. Ini adalah kategori paling berbahaya untuk keselamatan. Artinya sebanyak 20.3% ban yang sebenarnya rusak justru diklasifikasikan sebagai ban baik. Dampak fatal di dunia nyata : ban rusak tetap digunakan dan potensi kecelakaan meningkat. Kemudian risiko pecah ban terutama pada kecepatan tinggi, kendaraan logistik berat, cuaca panas, dan kondisi jalan buruk. Kemudian terdapat risiko tergelincir pada jalan basah atau tidak rata.

True Good terhadap Predicted Defective (15.7%) False Positive. Sebanyak 15.7% ban yang baik dinilai sebagai rusak. Hal ini berdampak tidak mengancam keselamatan namun meningkatkan biaya perawatan karena ban yang masih layak diganti lebih cepat. Sehingga siklus maintenance menjadi tidak efisien. Pada sebuah industri/logistik, false positive masih dapat ditoleransi dibanding false negative.

True Good terhadap Predicted Good (79.7%) adalah sebuah model untuk mengenali

ban yang layak jalan dengan akurasi 79.7%. Meskipun cukup tinggi, angka ini lebih rendah dibanding performa model VGG-16 sebelumnya yang sudah dijelaskan.

Dengan basis VGG-19, model telah menunjukkan kinerja yang cukup baik, tetapi tidak ideal untuk aplikasi keselamatan karena memiliki False Negative yang tinggi (20.3%) dan True Positive untuk ban rusak yang lebih rendah daripada model lain.

Untuk meningkatkan kinerja model dan mengurangi tingkat False Negative, beberapa langkah berikut disarankan: mengubah ambang klasifikasi agar model lebih peka terhadap tanda-tanda kerusakan ban. tingkatkan data, terutama dengan menambah variasi gambar ban yang rusak dari ringan hingga sedang.

		Predicted		Σ
		defective	good	
Actual	defective	91.3 %	20.3 %	55
	good	8.7 %	79.7 %	55
Σ		46	64	110

Gambar 8. confusion matrix result (Inception)

Confusion matrix pada Gambar 8 merupakan hasil dari Inception V3 model SVM menunjukkan hasil klasifikasi kondisi ban kendaraan ke dalam dua kategori: good and defective, dengan total 110 sampel (55 per kelas).

True Defective terhadap Predicted Defective (91.3%) Model berhasil mendeteksi 91.3% ban yang benar-benar rusak. Terdapat makna untuk keselamatan kendaraan adalah model memiliki sensitivitas tinggi, sangat baik dalam mengidentifikasi ban yang tidak layak jalan dan risiko ban rusak yang lolos deteksi lebih rendah dibanding VGG-19. Ini adalah performansi yang sangat penting untuk mencegah kecelakaan akibat pecah ban (blowout), kehilangan grip, tergelincir pada kecepatan tinggi.

True Defective terhadap Predicted Good (20.3%) False Negative. Model gagal mendeteksi 20.3% ban yang sebenarnya rusak, dan menganggapnya sebagai ban baik. Ini adalah kategori kesalahan paling kritis, karena ban rusak yang tidak terdeteksi dapat tetap

digunakan. Kemudian risiko kecelakaan meningkat signifikan lalu bahaya lebih besar pada kendaraan berat, bus, atau kendaraan bermuatan tinggi dan potensi kerusakan fatal terutama pada jalan panas atau tidak rata. Meskipun Inception V3 lebih baik daripada VGG-19 dalam hal True Positive, False Negative masih perlu ditekan.

True Good terhadap Predicted Defective (8.7%) False Positive sebanyak 8.7% ban baik salah dinyatakan rusak. Dampaknya ban yang masih layak diganti lebih cepat sehingga biaya meningkat akan tetapi tidak mengancam keselamatan, namun menurunkan efisiensi pemeliharaan. Tingkat False Positive ini relatif rendah dan masih dapat diterima dalam sistem yang mengutamakan keselamatan

True Good terhadap Predicted Good (79.7%) Model mengenali 79.7% ban layak jalan dengan benar. Artinya model cukup stabil dalam mengidentifikasi ban dalam kondisi normal. Angka ini mirip dengan VGG-19, tetapi Inception V3 memiliki performa keseluruhan lebih baik terutama untuk pendeteksian ban rusak.

Dalam mendeteksi kondisi ban kendaraan, model berbasis Inception V3 menunjukkan kinerja yang lebih baik dan akurat dibandingkan dengan fitur extractor lainnya. Keunggulan utama termasuk sensitivitas yang sangat tinggi (91.3%) untuk ban yang rusak, false positif yang rendah, stabilitas, dan cocok sebagai sistem screening awal. Kelemahan utama adalah false negatif yang masih cukup tinggi (20.3%), yang berarti bahwa inspeksi manual masih diperlukan. Kesimpulannya, Inception V3 adalah salah satu model deteksi ban terbaik, tetapi memerlukan prosedur keselamatan tambahan untuk beroperasi secara mandiri.

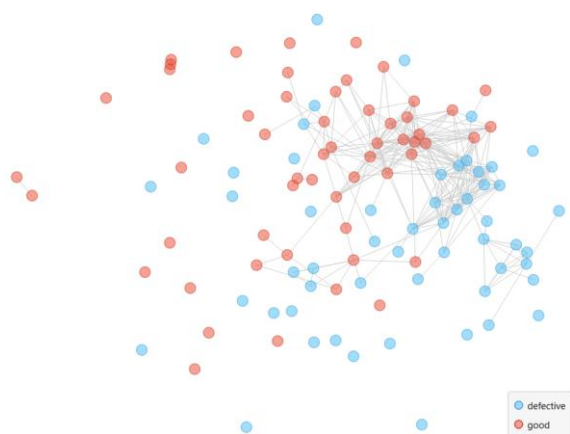
Secara keseluruhan, model menunjukkan kemampuan klasifikasi yang kuat pada semua kategori. VGG-16 adalah model terbaik dan paling stabil untuk mendeteksi kondisi ban kendaraan secara keseluruhan.

False Negative rendah sehingga aman untuk keselamatan, Performa konsisten sehingga dapat digunakan sebagai sistem screening awal yang andal. VGG-19 memiliki performa terlemah dan tidak direkomendasikan untuk aplikasi yang mengutamakan keselamatan karena False Negative sangat tinggi. Sedangkan Inception V3 memiliki performa tinggi pada deteksi kerusakan jelas,

tetapi tetap menghasilkan FN yang cukup besar sehingga harus disesuaikan threshold-nya sebelum diterapkan.

Hasilnya menunjukkan bahwa metode ekstraksi fitur dan klasifikasi yang telah digunakan sudah efektif. Namun, pengembangan lebih lanjut diperlukan, seperti penyempurnaan model deep learning, penggunaan dataset yang lebih besar, dan rekayasa fitur yang lebih canggih.

Terbukti bahwa model klasifikasi yang menggabungkan ekstraksi fitur berbasis CNN (VGG-16, VGG-19, dan Inception V3) dengan algoritma regresi logistik, SVM, dan neural networking dapat secara konsisten mengidentifikasi kondisi ban kendaraan.



Gambar 9. MDS Result (VGG-16)

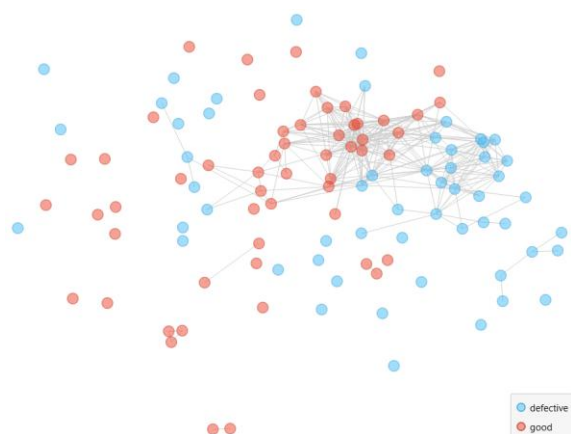
Gambar 9 menampilkan hasil analisis Multidimensional Scaling (MDS) yang memproyeksikan data ban kendaraan ke dalam ruang dua dimensi berdasarkan tingkat kemiripan (similarity) antar sampel. Setiap titik mewakili satu kondisi ban, dengan warna yang menunjukkan kategori yaitu kondisi good (merah), defective (biru) garis abu menunjukkan kedekatan atau hubungan antar sampel (neighborhood graph).

Visualisasi MDS menunjukkan bahwa data tidak terpisah secara sempurna, baik untuk kelas defective maupun good. Terdapat tumpang tindih (overlap) yang cukup besar antara kedua kelas di wilayah pusat plot. Akan tetapi terdapat juga kelompok data yang menyebar lebih jauh (outliers), kemudian membentuk sub-clusters yang relatif konsisten. Hal ini umum terjadi pada data citra ban karena kerusakan ban sering memiliki variasi tekstur, bentuk, dan pencahayaan berbeda.

Struktur kluster berdasarkan MDS terdiri dari kluster utama (wilayah tengah) dan kluster pinggir/kelompok terpisah. Di bagian tengah kluster utama, terdapat kumpulan titik merah dan biru yang berdekatan, menunjukkan bahwa fitur VGG-16 menangkap beberapa karakteristik visual yang serupa antara ban yang rusak dan ban yang baik, sehingga model masih melihat banyak kesamaan. Penggabungan ini menjelaskan mengapa VGG-16, meskipun akurat, masih menghasilkan hasil false positive dan false negatif. Overlap ini terjadi karena beberapa ban rusak memiliki pola visual yang mirip ban baik (kerusakan halus), beberapa ban baik memiliki tekstur yang menyerupai ban aus atau bercak kotor sehingga dianggap defective.

Pada Kluster Pinggir Kelompok Terpisah, terdapat beberapa titik defective dan good yang berada jauh dari pusat. Outlier defective menunjukkan kemungkinan kerusakan unik, misalnya sobekan sangat jelas, benjolan ekstrem, permukaan rusak berat. Outlier good menunjukkan kemungkinan ban sangat mulus dan baru, pencahayaan sangat berbeda sehingga fitur menariknya unik. Interpretasi VGG-16 sangat baik dalam mengenali pola kerusakan besar (major defects), tampak dari kelompok defective yang terpisah, namun sulit membedakan kerusakan ringan dari kondisi normal (minor defects).

Secara keseluruhan, hasil MDS menunjukkan bahwa VGG-16 telah menghasilkan sebuah pemetaan fitur yang cukup baik, terlihat dari cluster defective dan good yang mulai terpisah, terutama untuk kasus yang jelas. Akan tetapi terdapat overlap yang signifikan, kemudian menunjukkan beberapa ban rusak dan ban baik memiliki kemiripan visual sehingga sulit dibedakan hanya dengan fitur CNN. Deteksi kerusakan besar lebih akurat dibanding kerusakan ringan termasuk kluster outlier mendukung hal ini. Hasil MDS mendukung performa confusion matrix yaitu akurasi tinggi, namun masih ada false positive dan false negative. Model VGG-16 cocok sebagai sistem screening, namun tidak cukup untuk menjadi satu-satunya alat penilaian keselamatan ban.



Gambar 10. MDS Result (VGG-19)

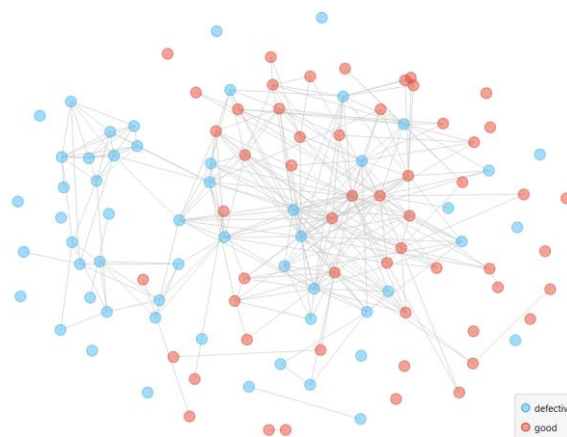
Gambar 10 menunjukkan hasil analisis Multidimensional Scaling (MDS) yang memproyeksikan data ban kendaraan ke dalam ruang dua dimensi berdasarkan tingkat kemiripan (similarity) antar sampel. Setiap titik mewakili satu kondisi ban, dengan warna yang menunjukkan kategori yaitu kondisi good (merah), defective (biru) dan garis abu menunjukkan kedekatan atau hubungan antar sampel (neighborhood graph).

Visualisasi ini memperlihatkan bahwa kelas defective dan good banyak tumpang-tindih (overlap). Tidak ada pemisahan cluster yang jelas antara dua kelas. titik menyebar cukup acak dan tidak membentuk kluster terdefinisi kuat. Hal tersebut telah mengindikasikan bahwa VGG-19 mengalami sebuah kesulitan dalam membedakan karakteristik visual antara ban rusak dan ban baik, hanya sedikit sub-cluster kompak yang terbentuk. Sebagian besar titik tersebar luas, baik untuk kelas defective maupun good. Struktur kluster berdasarkan MDS terbagi menjadi Kluster utama (wilayah Tengah) dan Kluster pinggir/ kelompok terpisah (Outlier). Kluster Utama (Wilayah Tengah) Overlap Berat di bagian tengah terlihat banyak titik merah dan biru bercampur, terhubung oleh garis abu, ini menunjukkan bahwa Fitur yang diekstrak VGG-19 sering kali tidak cukup diskriminatif untuk membedakan ban rusak vs ban baik, banyak citra ban memiliki tekstur atau pola visual yang dianggap mirip oleh jaringan, sehingga menghasilkan overlap tinggi. Model kemungkinan mengalami redundansi fitur karena kedalaman VGG-19 yang lebih tinggi tidak selalu berarti semantik yang lebih baik. Ini

adalah alasan utama tingginya False Negative dan False Positive pada VGG-19.

Pada Kluster Titik Terpisah (Outliers) terdapat beberapa titik defective dan good yang berada jauh dari kluster tengah. Kemudian pada Outlier defective menunjukkan adanya kemungkinan ban rusak parah (kerusakan sangat jelas). Outlier good menunjukkan mungkin ban sangat bersih atau memiliki pola unik. Namun jumlah titik terpisah ini lebih sedikit dibanding VGG-16, menunjukkan bahwa VGG-19 kurang mampu menangkap pola khas kerusakan yang ekstrem.

VGG-19 menghasilkan pemetaan fitur yang kurang baik, ditunjukkan oleh cluster yang bercampur dan minim pemisahan kelas. Sebuah Overlap besar antara defective dan good dapat menjelaskan bahwa performa VGG-19 yang rendah dibanding VGG-16. Model ini tidak mampu menangkap sebuah pola visual yang membedakan kerusakan ban secara kuat, terutama pada ban dengan kerusakan yang ringan. VGG-19 berisiko tinggi menghasilkan False Negative yang menunjukkan sangat tidak ideal untuk aplikasi keselamatan ban kendaraan.



Gambar 11. MDS Result (Inception V3)

Gambar 11 menampilkan hasil analisis Multidimensional Scaling (MDS) yang memproyeksikan data ban kendaraan ke dalam ruang dua dimensi berdasarkan tingkat kemiripan (similarity) antar sampel. Setiap titik mewakili satu kondisi ban, dengan warna yang menunjukkan kategori yaitu kondisi good (merah), defective (biru) garis abu menunjukkan kedekatan atau hubungan antar sampel (neighborhood graph).

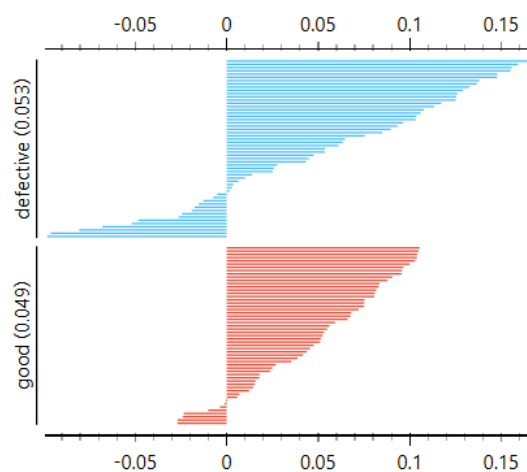
Pada sebuah tampilan data yang dihasilkan telah menunjukkan sebuah Sebaran yang lebih terorganisir dibandingkan VGG-19, namun tidak sebersih dari VGG-16. Titik-titiknya masih bercampur (tumpang tindih), tetapi pola distribusinya lebih terorganisir. Ada tanda-tanda subcluster kecil yang mulai muncul, terutama untuk barang-barang yang cacat. Kedekatan titik-titik di tengah menunjukkan bahwa area fitur sangat mirip antar kelas. Jaringan yang digambarkan sangat padat di tengah, yang berarti banyak sampel memiliki kedekatan fitur yang tinggi. Hal ini sesuai dengan kemampuan Inception V3 untuk mengekstrak tekstur ban, tetapi belum menghasilkan pemisahan kelas yang jelas.

Struktur kluster berdasarkan MDS terbagi menjadi Kluster Tengah – Overlap tinggi dan Kluster tepi/ data yang mulai terpisah. Kluster Tengah Di area inti, titik defective dan good bercampur sangat rapat, dihubungkan oleh banyak garis abu, hal ini menunjukkan bahwa Inception V3 menangkap fitur visual secara kuat, namun karakteristik ban rusak vs baik masih sangat mirip. Perbedaan antar kelas tidak cukup tegas ketika diproyeksikan ke ruang 2D. Banyak titik defective berada tepat di tengah cluster good, dan sebaliknya. Hal ini konsisten dengan False Negative $\approx 20\%$ pada confusion matrix.

Pada Kluster tepi Beberapa titik defective dan good tersebar lebih jauh hal ini menjelaskan bahwa Outlier defective menunjukkan kemungkinan kerusakan besar dan jelas. Outlier good menunjukkan ban sangat mulus, minim noise visual. Hal ini menunjukkan Inception V3 lebih efektif daripada VGG-19 dalam memisahkan sampel ekstrem, namun pemisahan tidak sekuat VGG-16 yang lebih stabil.

Pola MDS menunjukkan kluster tak terpisah dengan jelas, meskipun terdapat struktur lebih baik dibanding VGG-19. Titik defective lebih terkumpul di beberapa area, tetapi tumpang tindih besar dengan good tetap terjadi. Visualisasi ini mendukung hasil confusion matrix bahwa recall defective tinggi, FN tetap tinggi ($\approx 20\%$) dan FP rendah. Model baik dalam mendeteksi kerusakan besar, namun kesulitan pada kerusakan halus. Inception V3 adalah feature extractor kuat namun belum ideal sebagai pemisah sempurna kondisi ban. Sehingga Inception V3 lebih baik dari VGG-19

namun tidak se-stabil VGG-16 dalam pemisahan cluster.



Gambar 12. Silhouette Plot Result (VGG-16)

Gambar 12 Silhouette plot digunakan untuk mengevaluasi sejauh mana data dalam satu kluster serupa satu sama lain, serta seberapa baik model memisahkan kluster yang berbeda. Pada plot ini memperlihatkan warna Biru = cluster defective, Merah = cluster good, Nilai silhouette berada pada rentang -1 hingga $+1$, Caption cluster Defective 0.053 dan Good 0.049.

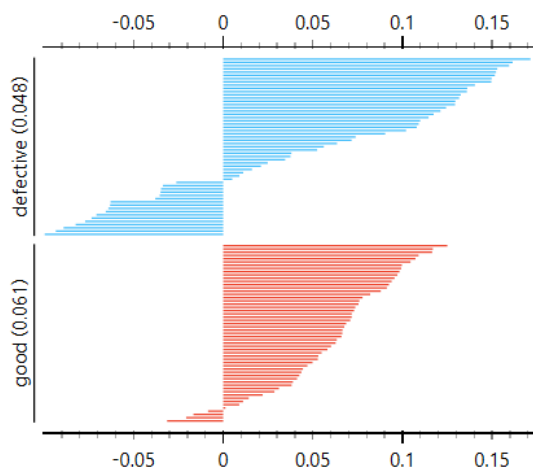
Nilai rata-rata silhouette dari kedua kelas sangat rendah, defective = 0.053 dan good = 0.049. Angka ini berada dekat 0, yang berarti pemisahan antar kelas lemah cluster tumpang-tindih, fitur yang dihasilkan VGG-16 tidak mampu memisahkan dua kelas secara tegas. Namun nilainya masih positif, artinya pemisahan ada, meskipun tipis serta masih terdapat struktur dasar yang bisa dipelajari classifier.

Pada Cluster Defective (0.053) Banyak sampel defective berada di nilai silhouette ≥ 0 , namun dekat titik nol, ini menandakan data defective cenderung berdekatan satu sama lain, tetapi juga sangat dekat dengan cluster good dan hanya sedikit sampel defective yang memiliki nilai silhouette tinggi (>0.1), artinya kerusakan yang sangat jelas baru mudah dipisahkan. Sehingga Interpretasi adalah VGG-16 mampu mengenali pola kerusakan besar (major defect), tetapi kesulitan mengenali kerusakan ringan yang visualnya mirip ban normal.

Pola pada cluster Good (0.049) sangat mirip dengan cluster yang defective. Hampir

seluruh nilai silhouette berada di kisaran rendah (0.00–0.15). Tidak banyak sampel good yang jauh dari cluster defective. Sehingga interpretasi adalah Ban dalam kondisi baik memiliki variasi tekstur, cahaya, dan pola permukaan yang sering menyerupai ban rusak ringan. Ini menghasilkan overlap besar dengan cluster defective.

Hasil visualisasi menunjukkan bahwa Nilai silhouette sangat rendah namun positif (0.053 dan 0.049) menunjukkan pemisahan cluster lemah tetapi tetap ada struktur dasar. VGG-16 menangkap pola kerusakan besar, tetapi tidak efektif membedakan kerusakan ringan vs ban normal. Sebuah Overlap yang tinggi antar cluster selaras dengan confusion matrix dan MDS. Model berbasis VGG-16 masih layak digunakan sebagai alat screening, akan tetapi tidak cukup kuat sebagai dasar tunggal untuk keputusan keselamatan kendaraan. Serta sebuah peluang improvement tambahan diperlukan agar model menjadi lebih sensitif terhadap variasi kerusakan minor.



Gambar 13. Silhouette Plot Result (VGG-19)

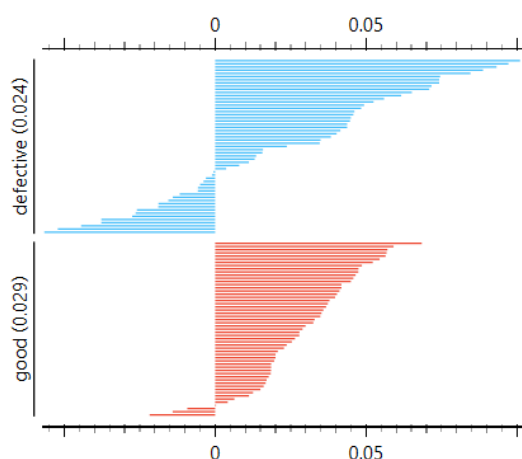
Gambar 13 Silhouette plot digunakan untuk mengevaluasi sejauh mana data dalam satu kluster serupa satu sama lain, serta seberapa baik model memisahkan kluster yang berbeda. Pada plot ini memperlihatkan warna Biru = cluster defective, Merah = cluster good, Nilai silhouette berada pada rentang -1 hingga $+1$, Caption cluster Defective 0.048 dan Good 0.061. Nilai ini berada dalam kategori sangat rendah, tetapi masih positif, mengindikasikan pemisahan yang lemah antara dua kelas.

Pada Cluster Defective (0.048) pemisahan sangat lemah, sebagian besar nilai silhouette

mendekati 0. Banyak sampel defective berada di perbatasan antara cluster defective dan good. Hanya sedikit titik defective yang memiliki silhouette tinggi (>0.10). Ada beberapa titik yang memasuki nilai negatif tipis artinya sampel tersebut lebih mirip cluster good dibanding defective. Artinya VGG-19 tidak mampu menangkap representasi visual khas dari ban rusak dengan baik. Kerusakan dengan pola ringan atau samar sering terlihat mirip dengan ban baik, sehingga cluster defective menjadi kabur.

Pada Cluster Good (0.061) masih lemah namun sedikit lebih baik. Rata-rata silhouette sedikit lebih tinggi dibanding defective. Namun secara keseluruhan masih rendah (sekitar 0.06). Hampir semua nilai positif, tetapi tipis terhadap cluster good tidak benar-benar terpisah kuat dari defective dan tidak ada struktur cluster yang kompak. Artinya Walaupun VGG-19 sedikit lebih baik dalam mengenali ban yang masih baik, fitur yang diekstraksinya tetap menghasilkan overlap besar dengan kelas defective.

Hasil dari visualisasi menunjukkan Nilai silhouette sangat rendah sehingga menunjukkan hampir tidak adanya pemisahan cluster defective dan good. Fitur VGG-19 gagal membedakan kerusakan ban secara visual, terutama kerusakan ringan. Cluster defective dan good saling tumpang tindih secara signifikan. Ini konsisten dengan tingginya false negative dan false positive pada confusion matrix. Artinya VGG-19 kurang cocok untuk mendeteksi kondisi ban kendaraan dan tidak dapat diandalkan untuk aplikasi keselamatan. Akan tetapi secara akademik bahwa VGG-19 memiliki performa terendah di antara semua arsitektur karena memberikan representasi fitur yang kurang diskriminatif.



Gambar 14. Silhouette Plot Result (Inception)

Gambar 14 Silhouette plot digunakan untuk mengevaluasi sejauh mana data dalam satu kluster serupa satu sama lain, serta seberapa baik model memisahkan kluster yang berbeda. Pada plot ini memperlihatkan warna Biru = cluster defective, Merah = cluster good, Nilai silhouette berada pada rentang -1 hingga $+1$, Caption cluster Defective 0.024 dan Good 0.029. Nilai Keduanya sangat rendah, menunjukkan pemisahan yang sangat lemah antara kedua kelas

Pada Cluster Defective (0.024) menunjukkan pemisahan sangat lemah. Banyak nilai silhouette berada tepat di sekitar 0, menunjukkan bahwa sampel defective berada sangat dekat dengan batas cluster. Hanya sebagian kecil yang mencapai nilai >0.05 . Ada beberapa nilai negatif tipis mengindikasikan beberapa sampel defective lebih mirip fitur good. Artinya Inception V3 menangkap fitur visual global (seperti pola, struktur, tekstur besar), namun gagal membedakan kerusakan ringan yang memiliki kemiripan kuat dengan ban normal.

Pada Cluster Good (0.029) sedikit lebih baik tetapi masih lemah. Nilai rata-rata sedikit lebih tinggi dari defective, namun tetap sangat kecil. Hampir semua nilai berada di rentang $0-0.07$. Tidak ada kluster dengan separasi kuat. Artinya variasi visual ban baik cukup besar (pola alur, pencahayaan, jenis ban), sehingga sebagian terlihat mirip ban dengan kerusakan ringan.

Hasil visualisasi menunjukkan Nilai silhouette defective (0.024) dan good (0.029) sangat rendah menunjukkan cluster hampir tidak

terpisah. Fitur Inception V3 tidak memberikan representasi yang jelas untuk memisahkan dua kelas ban. Banyak sampel berada dalam posisi borderline, menunjukkan kesulitan mendeteksi kerusakan ringan. Ini sepenuhnya konsisten dengan tingginya false negative pada confusion matrix. Inception V3 lebih baik dari VGG-19 namun tetap lebih buruk dari VGG-16 dalam hal separasi cluster. Sehingga Inception V3 menghasilkan fitur kuat untuk kerusakan besar, tetapi tidak cukup diskriminatif untuk kasus kerusakan halus sehingga cluster menjadi sangat tumpang tindih.

Hasil penelitian menunjukkan bahwa performa model klasifikasi kondisi ban kendaraan sangat dipengaruhi oleh kualitas fitur yang dihasilkan oleh arsitektur CNN yang digunakan sebagai feature extractor. Ketiga model—VGG-16, VGG-19, dan Inception V3—memiliki karakteristik berbeda yang tercermin dalam confusion matrix, MDS, dan silhouette plot.

Secara umum, model VGG-16 telah menghasilkan sebuah performa yang paling stabil dan seimbang, hal ini terlihat terutama dalam mendeteksi sebuah kondisi ban rusak (defective). Dapat dibuktikan juga dari confusion matrix yang mempunyai tingkat True Defective tinggi dan False Negative lebih rendah dibanding dua model lainnya. Visualisasi MDS untuk model VGG-16 menunjukkan sebuah struktur kluster yang relatif lebih jelas, meskipun tetap terdapat kondisi overlap. Silhouette plot VGG-16 menunjukkan nilai positif, meskipun rendah. Ini berarti pemisahan kluster lemah tetapi tetap konsisten. Dapat disimpulkan bahwa VGG-16 dapat menangkap pola kerusakan yang lebih diskriminatif, terutama kerusakan dengan kondisi visual yang jelas.

Di sisi lain, dari ketiga model, VGG-19 menunjukkan hasil yang paling buruk. Hasil MDS menunjukkan bahwa titik-titik kedua kelas tersebar secara tidak terstruktur dan sangat tumpang-tindih. Meskipun nilai silhouette VGG-19 positif, nilainya sangat rendah, menunjukkan bahwa model ini tidak dapat membentuk perbedaan yang jelas antara ban baik dan ban rusak. Sehubungan dengan confusion matrix, VGG-19 menghasilkan False Negative dan False Positive tertinggi dari dua model lainnya. Dalam konteks ini, kedalaman arsitektur yang lebih besar pada VGG-19 tidak

menguntungkan, bahkan cenderung menyebabkan overfitting noise visual.

Hasil Inception V3, di sisi lain, rata-rata: lebih baik dari VGG-19 tetapi tidak stabil seperti VGG-16. Menurut matriks kekacauan, Inception V3 memiliki kemampuan yang luar biasa untuk mendeteksi kerusakan besar (kerusakan besar), tetapi karena kesulitan membedakan kerusakan kecil (kerusakan kecil), masih menghasilkan nilai negatif negatif yang tinggi. MDS mendukung ini dengan menunjukkan bahwa kelas buruk dan rusak masih sangat bercampur meskipun ada beberapa subcluster yang menunjukkan perbedaan. Selain itu, plot silhouette Inception V3 menunjukkan nilai yang sangat rendah, menunjukkan bahwa fitur yang dihasilkan dari model tersebut tidak dapat memberikan pemisahan yang kuat.

Secara keseluruhan, hasil ini menunjukkan bahwa meskipun ketiga arsitektur CNN mampu mengekstraksi fitur visual ban, hanya VGG-16 yang memberikan representasi yang lebih efektif dan stabil untuk tugas klasifikasi kondisi ban. Tingginya tingkat overlap pada MDS dan rendahnya nilai silhouette pada ketiga model menunjukkan bahwa karakteristik visual antara ban rusak dan ban baik sering kali sangat mirip—terutama untuk kerusakan ringan—sehingga pemisahan kelas menjadi sangat menantang. Oleh karena itu, performa sistem berbasis citra murni memiliki batasan inheren, terutama jika digunakan untuk tujuan keselamatan kendaraan yang membutuhkan sensitivitas tinggi terhadap kerusakan kecil.

Dengan mempertimbangkan keseluruhan hasil evaluasi, pendekatan visual murni untuk mendeteksi kondisi ban masih perlu dilengkapi dengan data sensorik tambahan seperti tekanan ban (TPMS), kedalaman tapak, atau getaran roda untuk meningkatkan akurasi dan mengurangi False Negative, yang merupakan kesalahan paling kritis untuk aplikasi keselamatan jalan raya.

5. KESIMPULAN

Studi ini menemukan bahwa kemampuan model untuk mengekstrak fitur tampilan visual yang relevan dan sangat berpengaruh pada kemampuan untuk mengklasifikasi kondisi ban kendaraan. Di antara ketiga arsitektur CNN yang diuji, VGG-16 adalah yang terbaik untuk ekstraksi fitur karena mampu menghasilkan

pemisahan cluster yang lebih stabil, nilai True Defective yang lebih tinggi, dan tingkat False Negative yang lebih rendah daripada dua model lainnya. Akibatnya, VGG-16 lebih dapat diandalkan untuk aplikasi screening kondisi ban.

Inception V3 sangat tepat untuk digunakan sebagai pendeteksi awal, akan tetapi tidak dapat berdiri sendiri dalam sistem keselamatan yang memerlukan presisi tinggi. Namun, itu memiliki kemampuan deteksi yang sangat kuat terhadap kerusakan besar, akan tetapi tidak cukup sensitif terhadap kerusakan ringan sehingga menghasilkan false negative tinggi.

VGG-19 dapat menunjukkan performa terendah dibandingkan dengan model lainnya, baik pada confusion matrix, MDS, maupun silhouette plot. Arsitektur yang lebih dalam tidak memberikan keunggulan dalam konteks pemrosesan citra ban, bahkan menghasilkan representasi fitur yang kurang stabil dan tidak diskriminatif.

Secara keseluruhan, meskipun ketiga model memiliki kemampuan dasar dalam mengklasifikasikan ban rusak dan ban baik, tingkat overlap antar kedua kelas baik pada MDS maupun silhouette plot mengindikasikan bahwa pendekatan berbasis citra saja belum cukup memadai untuk mendeteksi kerusakan ban ringan secara konsisten. Oleh karena itu, digunakan sebagai sistem pendukung keputusan (Decision Support System), bukan sebagai alat tunggal inspeksi keselamatan.

Sebuah pengembangan yaitu Integrasi dengan data sensorik fisik (misalnya tekanan, temperatur, getaran), serta penyesuaian threshold, lalu fine-tuning lanjutan sangat direkomendasikan untuk meningkatkan sebuah akurasi dan dapat mengurangi risiko kesalahan pemilihan klasifikasi ban kendaraan yang dapat berdampak pada keselamatan kendaraan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini, terutama istri dan anak-anak yang selalu memberikan ruang dan waktu untuk menyelesaikan penelitian ini. Selain daripada itu keterlibatan dosen dalam memberikan bimbingan ilmu dan arahan untuk menyelesaikan penelitian ini.

DAFTAR PUSTAKA

- [1] S. Bazargan-Hejazi, T. Gaines, A. Ahmadi, and M. Sadjadi, "Driving behavior, risk perception, and road safety among professional drivers," *Safety Science*, vol. 133, p. 105022, 2021, doi: 10.1016/j.ssci.2020.105022.
- [2] M. A. Aguirre, F. Martinez, and J. Sánchez, "Analysis of tire wear and its impact on vehicle handling and safety," *Accident Analysis & Prevention*, vol. 161, p. 106350, 2021, doi: 10.1016/j.aap.2021.106350.
- [3] T. Hidayat and A. Kurniawan, "Identifikasi kerusakan dan kendaraan dan implikasinya terhadap keselamatan berkendara," *Jurnal Keselamatan Kerja & Lingkungan*, vol. 3, no. 2, pp. 55–63, 2020.
- [4] mateuszk013, "Tyre Quality – 90% ACC with Custom SE-ResNet," Kaggle, Mar. 21, 2025. [Online]. Available: <https://www.kaggle.com/datasets/warcoder/tyre-quality-classification>
- [5] G. Ardi Pratama, E. Yulia and H. Maulana, "convolutional neural network dan faster region convolutional neural network untuk klasifikasi kualitas biji kopiarabika," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)* Vol. 12 No. 3, 2024, <https://doi.org/10.23960/jitet.v12i3.4887>.
- [6] Y. You, "Using Elman neural network model to forecast and analyze the agricultural economy," *Journal of Mathematics*, vol. 2022, no. 1, 2022, doi: 10.1155/2022/8374696.
- [7] O. Awe and R. Dias, "Comparative analysis of ARIMA and artificial neural network techniques for forecasting non-stationary agricultural output time series," *Agris on-Line Papers in Economics and Informatics*, vol. 14, no. 4, pp. 3–9, 2022, doi: 10.7160/aol.2022.140401.
- [8] P. Papantoniou, C. Antoniou, and G. Yannis, "Review of human factors integrating behavioural theories in road safety," *Safety Science*, vol. 127, p. 104712, 2020, doi: 10.1016/j.ssci.2020.104712.
- [9] S. Sufriyanto and I. Yuadi, "Classify a path on tire by using logistic regression and support vector machine (SVM) based on VGG-16, VGG-19, and Inception V3 models," *Eduvest - Journal of Universal Studies*, vol. 5, no. 8, pp. 9711–9722, 2025, doi: 10.59188/eduvest.v5i8.50960.
- [10] A. Ghosh and A. Chawla, "A review of tyre failures and their influence on road accidents," *Safety Science*, vol. 136, p. 105148, 2021, doi: 10.1016/j.ssci.2020.105148.
- [11] R. Akmalia, I. Slamet, and H. Pratiwi, "Analisis sentimen Twitter berbahasa Indonesia terhadap aplikasi PeduliLindungi dengan algoritma SVM, KNN, dan regresi logistik," *Prosiding Seminar Nasional MIPA UNIPA*, pp. 150–156, 2022, doi: 10.30862/psnmu.v7i1.21.
- [12] A. Firlansyah, A. Kaswar, and A. Risal, "Klasifikasi tingkat kematangan buah pepaya berdasarkan fitur warna menggunakan JST," *Techno Xplore Jurnal Ilmu Komputer dan Teknologi Informasi*, vol. 6, no. 2, pp. 55–60, 2021, doi: 10.36805/technoxplore.v6i2.1438.
- [13] F. Cahya, N. Hardi, D. Riana, and S. Hadiyanti, "Klasifikasi penyakit mata menggunakan convolutional neural network (CNN)," *Sistemasi*, vol. 10, no. 3, p. 618, 2021.
- [14] M. Tsai and Y. Tao, "Deep learning technology applied to medical image tissue classification," *Diagnostics*, vol. 12, no. 10, p. 2430, 2022, doi: 10.3390/diagnostics12102430.
- [15] W. Habibi and I. Yuadi, "Evaluating logistic regression and SVM for image analysis using VGG-16, VGG-19, and Inception V3 features," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 30, no. 2, pp. 136–155, 2025, doi: 10.35760/tr.2025.v30i2.14056.
- [16] C. Szegedy *et al.*, "Rethinking the inception architecture for computer vision," in *Proc. CVPR*, 2016.
- [17] D. Özdemir and N. Arslan, "Analysis of deep transfer learning methods for early diagnosis of the COVID-19 disease with chest X-ray images," *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, vol. 10, no. 2, pp. 628–640, 2022, doi: 10.29130/dubited.976118.
- [18] H. Zhuang, B. Li, J. Ma, P. Monkam, W. Qian, and D. He, "An attention-based deep learning network for predicting platinum resistance in ovarian cancer," *IEEE Access*, vol. 12, pp. 41000–41008, 2024, doi: 10.1109/ACCESS.2024.3377560.
- [19] A. Ansari, S. Ansari, S. Prasla, and A. Naveed, "Comparative analysis of image classification methods on cat breeds and behavior using machine learning techniques," *Pakistan Journal of Engineering Technology & Science*, vol. 12, no. 1, pp. 91–103, 2024, doi: 10.22555/pjets.v12i1.1100.
- [20] Y. Qu, H. Zhu, K. Cao, X. Li, M. Ye, and Y. Sun, "Prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer using a deep learning method," *Thoracic Cancer*, vol. 11, no. 3, pp. 651–658, 2020, doi: 10.1111/1759-7714.13309.
- [21] A. Husen, A. Afiah, S. Soesanti, and F. Tempola, "Deteksi dini resiko tuberkulosis di Kota Ternate: Pelacakan dan implementasi algoritma klasifikasi," *Jurnal Coscitech (Computer Science and Information*

- Technology), vol. 3, no. 2, pp. 217–225, 2022, doi: 10.37859/coscitech.v3i2.3986.
- [22] S. Pramesti, I. Indahwati, and U. Syafitri, “Analisis regresi logistik dan CART untuk credit scoring dengan penanganan kelas tak seimbang,” *Xplore Journal of Statistics*, vol. 11, no. 3, pp. 226–237, 2022, doi: 10.29244/xplore.v11i3.1015.
- [22] M. Suhendra, D. Ispriyanti, and S. Sudarno, “Ketepatan klasifikasi pemberian kartu keluarga sejahtera di Kota Semarang menggunakan metode regresi logistik biner dan metode CHAID,” *Jurnal Gaussian*, vol. 9, no. 1, pp. 64–74, 2020, doi: 10.14710/j.gauss.v9i1.27524.
- [23] S. Mutmainah and D. H. Fudholi, “Leveraging BiLSTM and LDA for analyzing and dashboarding user feedback in applications,” *Jurnal Media Informatika Budidarma*, vol. 8, no. 1, p. 51, 2024, doi: 10.30865/mib.v8i1.7022.
- [24] M.A.A.Z. Vieira and K.L.e. Silva, “Comparison of record linkage methods,” *Revista de Gestão e Secretariado (Management and Administrative Professional Review)*, vol. 14, no. 5, pp. 7999–8004, 2023, doi: 10.7769/gesec.v14i5.2171.
- [25] A. Pambudi, “Comparative SVM and decision tree algorithm in identifying the eligibility of KIP scholarship awardee,” *Conference Series*, vol. 4, no. 1, pp. 49–57, 2023, doi: 10.34306/conferenceseries.v4i1.625.
- [26] E. Casas, L. Ramos, E. Bendek, and F. Rivas, “Assessing the effectiveness of YOLO architectures for smoke and wildfire detection,” *IEEE Access*, vol. 11, pp. 96554–96583, 2023, doi: 10.1109/ACCESS.2023.3312217.