

# PREDIKSI ANEMIA DARI PIXEL GAMBAR DAN LEVEL HEMOGLOBIN MENGGUNAKAN RANDOM FOREST CLASSIFIER

## PREDICTING ANAEMIA FROM IMAGE PIXELS AND HEMOGLOBIN LEVELS USING RANDOM FOREST CLASSIFIER

Huzain Azis<sup>1)</sup>, Nurul Rismayanti<sup>2)</sup>

<sup>1)</sup> Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Muslim Indonesia  
Jln Urip sumoharjo, Makassar, 90231, Indonesia

<sup>2)</sup> Program Studi Teknik Elektro, Fakultas Teknik, Universitas Negeri Malang  
Jl. Cakrawala No.5, Sumbersari, Kec. Lowokwaru, Kota Malang, Jawa Timur 65145  
Email : [huzain.azis@umi.ac.id](mailto:huzain.azis@umi.ac.id)<sup>1)</sup>, [nurul.rismayanti.2305348@students.um.ac.id](mailto:nurul.rismayanti.2305348@students.um.ac.id)<sup>2)</sup>

### Abstrak

Anemia adalah gangguan darah yang umum terjadi dan ditandai oleh kekurangan sel darah merah atau hemoglobin, yang dapat menyebabkan komplikasi kesehatan serius jika tidak segera didiagnosis dan diobati. Penelitian ini bertujuan untuk mengembangkan model machine learning untuk memprediksi anemia berdasarkan level hemoglobin dan distribusi pixel gambar, dengan memanfaatkan dataset dari Kaggle. Dataset mencakup fitur seperti persentase pixel merah, hijau, dan biru dalam gambar serta level hemoglobin. Kami menggunakan Random Forest Classifier, algoritma machine learning yang kuat, dan mengevaluasi kinerjanya menggunakan validasi silang 5-fold. Proses pre-processing data meliputi penghapusan kolom yang tidak relevan, encoding variabel kategori, dan penskalaan fitur numerik. Model ini mencapai akurasi rata-rata sebesar 97,05%, presisi 97,02%, recall 97,05%, dan F1-score 96,88%, menunjukkan keandalannya dalam memprediksi anemia. Visualisasi seperti Correlation Heatmap, 3D PCA, Parallel Coordinates Plot, 3D t-SNE, dan Violin Plot digunakan untuk memahami hubungan dan distribusi fitur. Hasil ini menyoroti potensi machine learning dalam menyediakan alat diagnostik yang non-invasif dan hemat biaya untuk anemia, terutama di daerah dengan sumber daya terbatas. Penelitian selanjutnya sebaiknya menangani ketidakseimbangan dataset dan potensi bias, mengeksplorasi fitur tambahan, serta menguji model machine learning lainnya untuk lebih meningkatkan akurasi prediksi. Studi ini berkontribusi pada bidang diagnostik medis dengan menunjukkan efektivitas integrasi level hemoglobin dan data gambar untuk prediksi anemia, membuka jalan untuk deteksi dini dan strategi pengobatan yang lebih baik.

**Kata kunci:** Anemia, Hemoglobin, Machine Learning, Random Forest.

### Abstract

Anaemia is a widespread blood disorder characterized by a deficiency of red blood cells or hemoglobin, which can lead to severe health complications if not diagnosed and treated promptly. This research aims to develop a machine learning model to predict anaemia based on hemoglobin levels and image pixel distributions, leveraging a dataset from Kaggle. The dataset includes features such as percentages of red, green, and blue pixels in images and hemoglobin levels. We applied a Random Forest Classifier, a robust machine learning algorithm, and evaluated its performance using 5-fold cross-validation. The data pre-processing involved removing irrelevant columns, encoding categorical variables, and scaling numerical features. The model achieved a mean accuracy of 97.05%, precision of 97.02%, recall of 97.05%, and F1-score of 96.88%, indicating its high reliability in predicting anaemia. Visualizations such as Correlation Heatmaps, 3D PCA, Parallel Coordinates Plots, 3D t-SNE, and Violin Plots were used to understand feature relationships and distributions. These results underscore the potential of machine learning in providing a non-invasive, cost-effective diagnostic tool for anaemia, especially in resource-limited settings. Future research should address dataset imbalance and potential biases, explore additional features, and test other machine learning models to further enhance the predictive accuracy. This study contributes to the field of medical diagnostics by demonstrating the efficacy of integrating hemoglobin levels and image data for anaemia prediction, paving the way for improved early detection and treatment strategies.

**Keywords:** Anaemia, Hemoglobin, Machine Learning, Random Forest.

## 1. PENDAHULUAN

Anemia, sebuah kondisi yang ditandai dengan kekurangan jumlah atau kualitas sel darah merah atau hemoglobin, merupakan masalah kesehatan masyarakat yang signifikan secara global. Kondisi ini mempengaruhi jutaan orang, terutama di negara-negara berkembang, dan dapat menyebabkan masalah kesehatan serius seperti kelelahan, kelemahan, dan gangguan kognitif. Metode diagnostik tradisional untuk anemia, seperti tes lengkap darah (CBC), bersifat invasif dan sering memerlukan sumber daya serta infrastruktur yang mungkin tidak tersedia di lingkungan dengan keterbatasan sumber daya [1]. Hal ini menekankan perlunya pendekatan diagnostik yang inovatif, non-invasif, dan hemat biaya yang dapat memfasilitasi deteksi dan pengobatan dini. Masalah utama dalam penelitian ini adalah pengembangan metode non-invasif untuk memprediksi anemia menggunakan data yang mudah diakses. Dengan memanfaatkan kemajuan dalam pembelajaran mesin dan pemrosesan citra, pengembangan model prediktif yang memanfaatkan level hemoglobin dan distribusi warna piksel dalam gambar terkait diupayakan. Meskipun metode yang ada untuk mendiagnosis anemia telah terbukti akurat, metode tersebut tidak selalu dapat diakses atau terjangkau oleh sebagian besar populasi global. Oleh karena itu, model yang dirancang untuk memprediksi anemia dengan akurasi tinggi menggunakan data sederhana dan non-invasif diharapkan berpotensi merevolusi cara diagnosis anemia, terutama di daerah yang kurang terlayani.

Tujuan dari penelitian ini yaitu untuk mengembangkan dan memvalidasi model pembelajaran mesin yang mampu memprediksi anemia berdasarkan tingkat hemoglobin dan data piksel gambar. Secara khusus, Random Forest Classifier, sebuah algoritma pembelajaran mesin yang kuat dan banyak digunakan, dimaksudkan untuk digunakan dalam menganalisis dataset yang disediakan. Dataset ini mencakup fitur-fitur seperti persentase piksel merah, hijau, dan biru dalam gambar, beserta tingkat hemoglobin. Dengan menerapkan pendekatan sistematis pada *pre-processing data* [2], pemilihan fitur, dan validasi model, kami berupaya mencapai model prediktif berkinerja tinggi [3], [4]. Beberapa pertanyaan penelitian yang membimbing studi ini adalah sebagai berikut: Apakah anemia dapat diprediksi secara akurat melalui kombinasi tingkat hemoglobin dan distribusi piksel gambar? Bagaimana kinerja Random Forest Classifier dalam konteks ini? Bagaimana perbandingan kinerja model ini dengan metode diagnostik tradisional? Pertanyaan-pertanyaan ini dianggap penting untuk dipahami guna mengeksplorasi potensi dan keterbatasan pendekatan yang diusulkan. Hipotesis yang mendorong penelitian ini mengemukakan bahwa *Random Forest Classifier* [5] akan mencapai akurasi, presisi, recall, dan F1-score yang tinggi dalam memprediksi anemia, sehingga menawarkan alternatif yang layak bagi metode diagnostik tradisional [6].

Lingkup penelitian ini dibatasi pada dataset yang disediakan oleh Kaggle, yang mencakup berbagai fitur yang relevan untuk prediksi anemia. Namun, penelitian ini dibatasi oleh ketidakseimbangan yang melekat dalam dataset dan potensi bias yang muncul dari proses pengumpulan data gambar. Dataset dibagi menjadi subset pelatihan dan pengujian untuk memastikan kekokohan model, dan berbagai langkah pra-pemrosesan diterapkan untuk mempersiapkan data untuk analisis [7], [8]. Terlepas dari keterbatasan ini, penelitian ini bertujuan untuk memberikan wawasan yang signifikan mengenai potensi model pembelajaran mesin dalam diagnosis medis. Sebagai kesimpulan, penelitian ini berkontribusi pada bidang diagnostik medis dengan mengeksplorasi metode baru dan non-invasif untuk memprediksi anemia. Dengan integrasi tingkat hemoglobin dan data piksel gambar, pendekatan berbasis pembelajaran mesin diusulkan untuk meningkatkan deteksi dini dan pengobatan anemia, terutama di lingkungan dengan keterbatasan sumber daya. Temuan studi ini diharapkan dapat membuka jalan untuk penelitian lebih lanjut di bidang ini, mendorong pengembangan alat diagnostik yang lebih mudah diakses dan efisien. Karya ini tidak hanya meningkatkan pemahaman tentang prediksi anemia tetapi juga menyoroti potensi transformatif dari pembelajaran mesin dalam layanan kesehatan.

## 2. DASAR TEORI

Algoritma Random Forest adalah salah satu teknik pembelajaran mesin berbasis ensemble yang sangat populer untuk tugas-tugas klasifikasi dan regresi, termasuk dalam aplikasi

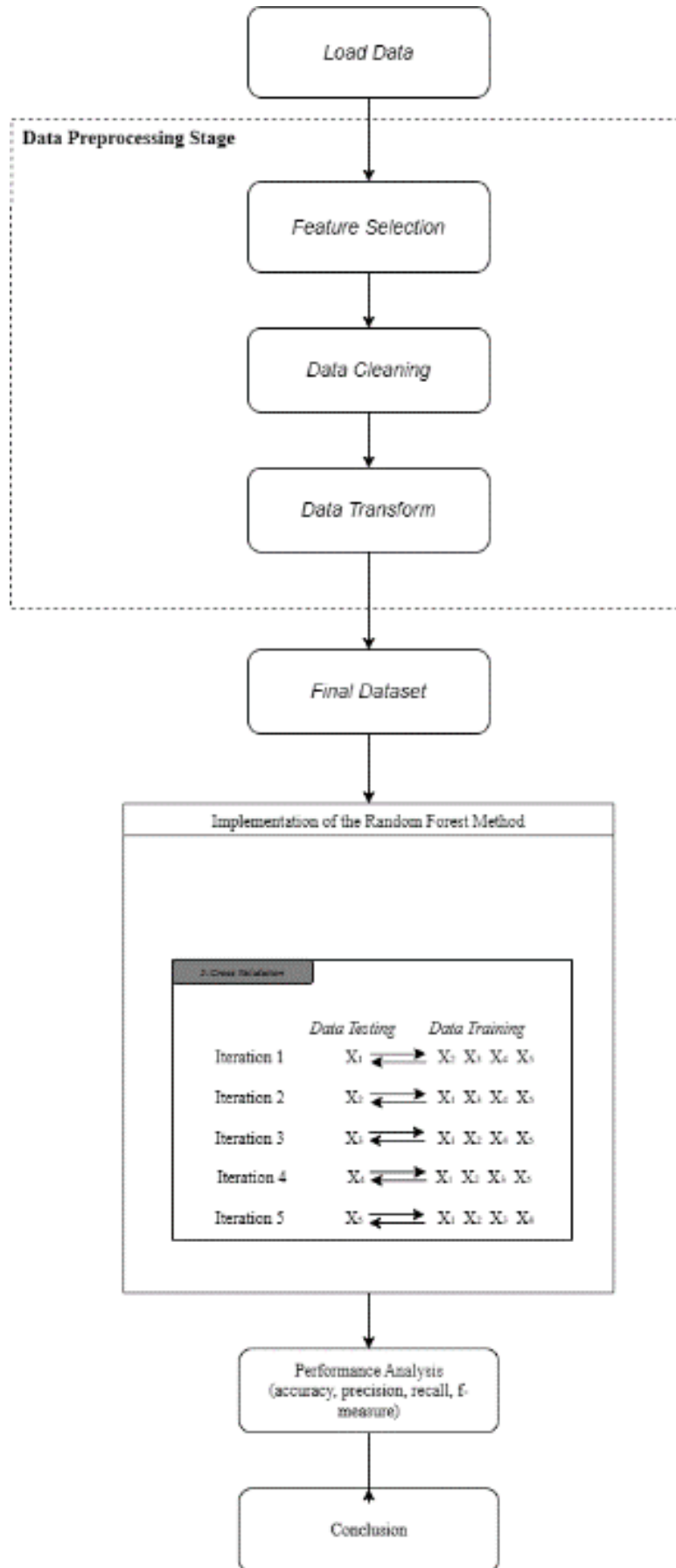
medis seperti prediksi anemia. Dalam penelitian ini, algoritma Random Forest digunakan untuk memprediksi anemia dengan memanfaatkan data distribusi warna piksel gambar dan level hemoglobin. Random Forest bekerja dengan cara membentuk banyak pohon keputusan (*decision trees*) yang masing-masing dilatih pada subset acak dari data asli (dengan teknik yang dikenal sebagai “*bootstrap aggregating*” atau “*bagging*”) [9], [10]. Setiap pohon kemudian memberikan prediksinya secara independen, dan hasil akhir ditentukan melalui *voting* mayoritas untuk tugas klasifikasi, seperti dalam prediksi status anemia ini. Pendekatan ini membantu mengurangi kemungkinan *overfitting*, yang sering terjadi pada model pohon tunggal, dan meningkatkan akurasi prediksi secara keseluruhan.

Random Forest juga memungkinkan identifikasi pentingnya setiap fitur melalui metrik “*feature importance*,” yang mengukur kontribusi relatif setiap variabel dalam model prediksi [11]. Pada kasus prediksi anemia, metrik ini dapat digunakan untuk mengetahui seberapa besar pengaruh setiap fitur, seperti level hemoglobin atau distribusi piksel warna tertentu, dalam memprediksi status anemia. Hal ini membantu dalam memahami peran setiap faktor dan dapat menjadi dasar bagi pengembangan model yang lebih akurat dan efisien, dengan fokus pada fitur yang memiliki pengaruh paling besar.

Dengan kemampuan untuk menangani dataset yang tidak seimbang, seperti yang sering terjadi dalam klasifikasi medis, Random Forest menjadi algoritma yang ideal untuk kasus ini. Dalam penelitian ini, ketidakseimbangan kelas antara kasus anemia dan *non-anemia* dapat diatasi dengan pembobotan internal pada algoritma, yang memungkinkan model untuk belajar dari kedua kelas dengan efektif. Keunggulan lain dari Random Forest adalah kemampuannya menghasilkan prediksi yang akurat tanpa memerlukan pemrosesan data yang kompleks. Hal ini membuat algoritma ini cocok untuk diterapkan pada deteksi anemia secara *non-invasif*, yang dapat dilakukan dengan efisiensi tinggi dan biaya rendah. Dengan demikian, penerapan Random Forest pada kasus ini diharapkan mampu memberikan solusi diagnostik yang praktis, andal, dan dapat diakses di wilayah dengan sumber daya medis yang terbatas.

### 3. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan pembelajaran mesin yang diawasi untuk memprediksi anemia menggunakan dataset yang mencakup tingkat hemoglobin dan distribusi piksel gambar. Model yang dipilih untuk tugas ini adalah *Random Forest Classifier*, yang dikenal karena ketangguhannya dan kemampuannya untuk menangani berbagai jenis data. Desain penelitian melibatkan beberapa langkah utama: *pre-processing data*, *feature selection*, pelatihan model, dan evaluasi [12], [13]. Kinerja model dinilai menggunakan validasi silang, khususnya pendekatan *5-fold cross-validation* [14], [15], yang membantu memperkirakan kemampuan generalisasi model dan mengurangi *overfitting* [16]. Penelitian kami dirancang dalam lima tahap utama yang terstruktur dengan baik, dan aspeknya diilustrasikan dalam Gambar 1.



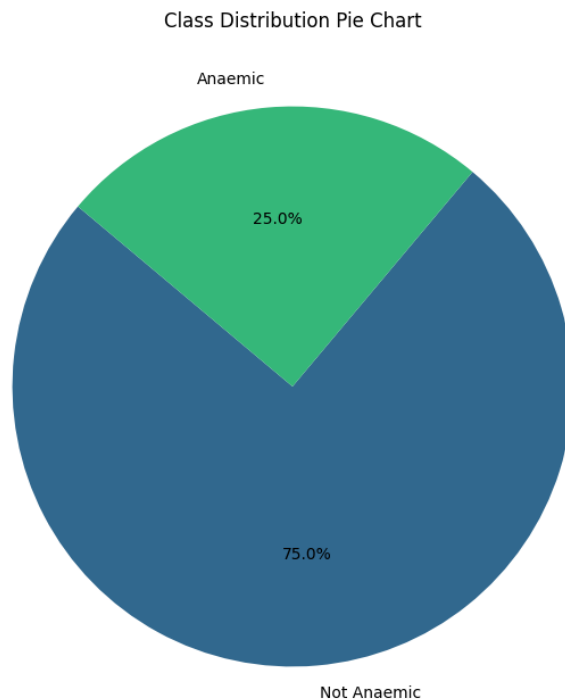
**Gambar 1.** Tahapan Penelitian

**a. Proses Pengumpulan Data**

Dataset yang digunakan dalam penelitian ini dikumpulkan dari sumber-sumber di internet dan dikurasi di Kaggle. Dataset ini mencakup berbagai fitur seperti persentase piksel merah, hijau, dan biru dalam gambar, tingkat hemoglobin, dan status anemia individu. Dataset ini tidak seimbang, dengan lebih banyak kasus non-anemia dibandingkan anemia. Ketidakseimbangan ini membutuhkan pra-pemrosesan yang hati-hati untuk memastikan model dapat mempelajari prediksi untuk kedua kelas secara efektif. Dataset ini terdiri dari fitur-fitur berikut:

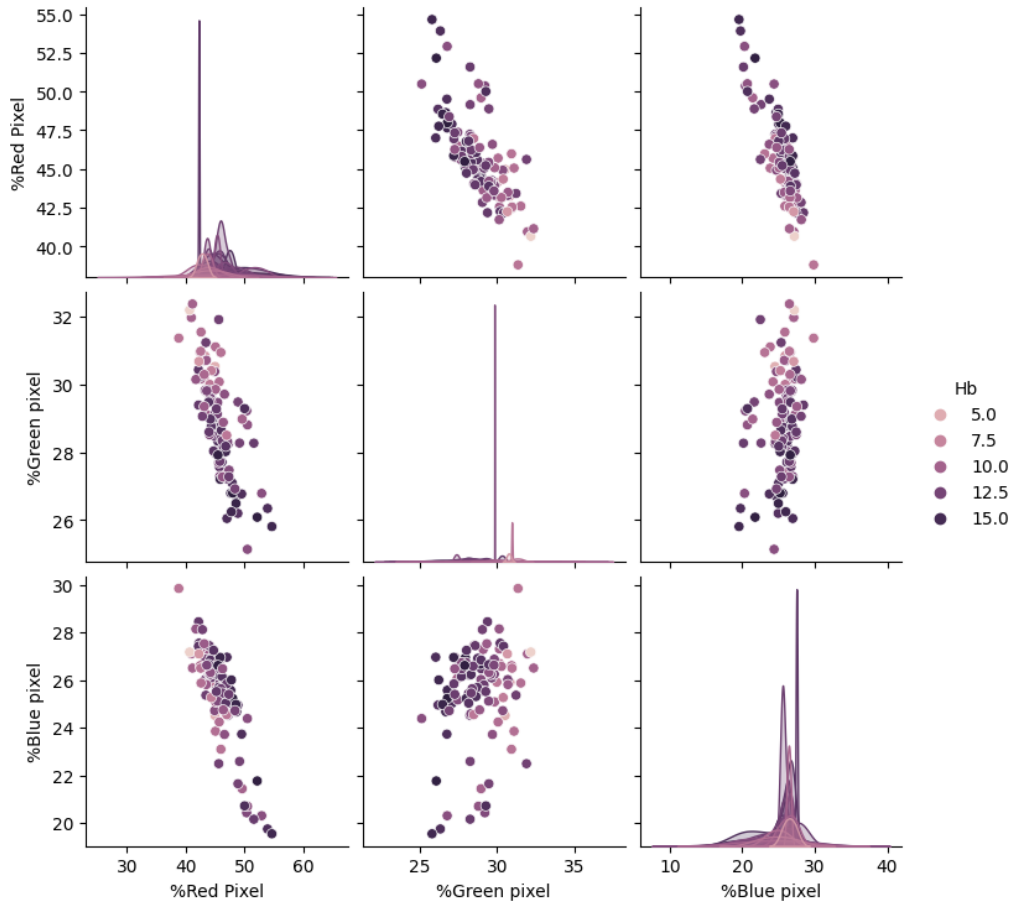
**Tabel 1. Feature Description**

Nama Kolom	Type	Deskripsi
<i>Number</i>	<i>Integer</i>	Pengidentifikasi unik untuk setiap kasus individu
<i>Sex</i>	<i>Categorical</i>	Jenis kelamin
<i>%Red Pixel</i>	<i>Float</i>	Persentase piksel merah dalam gambar.
<i>%Green Pixel</i>	<i>Float</i>	Persentase piksel hijau dalam gambar
<i>%Blue Pixel</i>	<i>Float</i>	Persentase piksel biru dalam gambar.
<i>Hb</i>	<i>Float</i>	Tingkat hemoglobin dalam gram per desiliter (g/dL).
<i>Anaemic</i>	<i>Categorical</i>	Status Anemia (Yes=1, No=0).

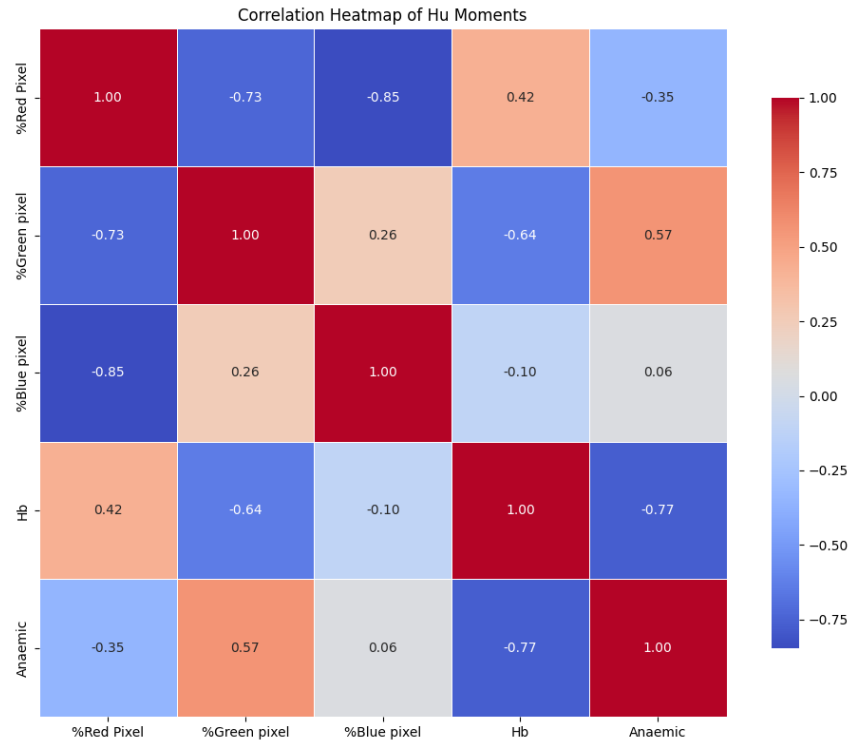


**Gambar 2. Class Distribution Pie Chart by Engagement Level**

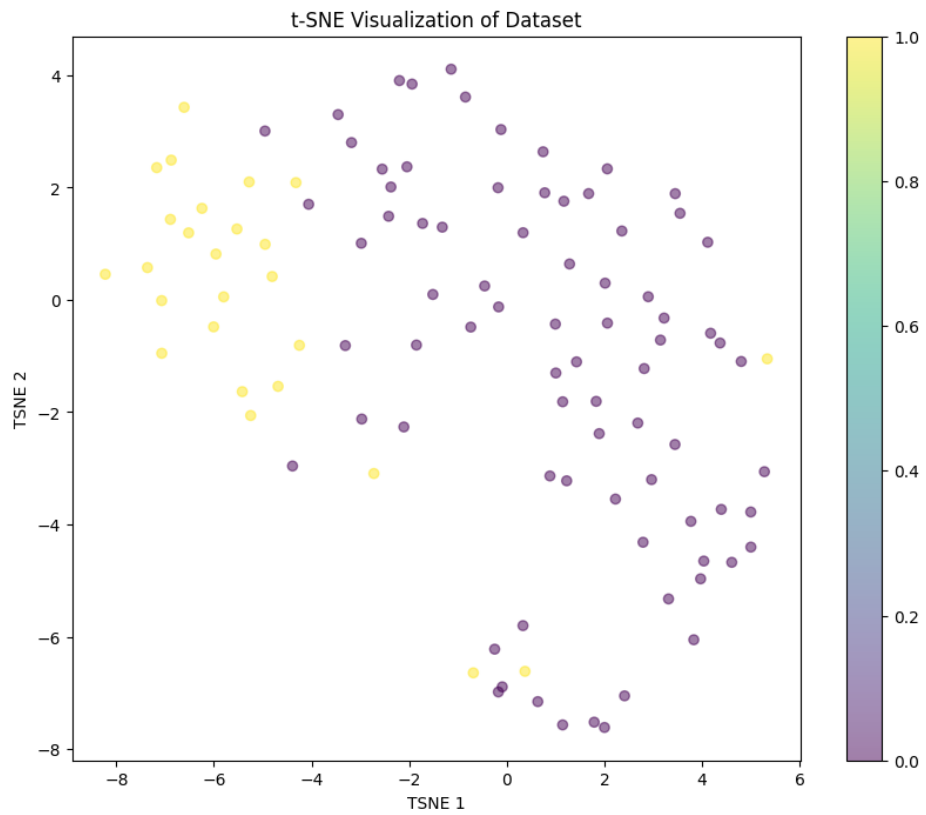
Untuk memfasilitasi pemahaman data, berbagai visualisasi dibuat. Ini termasuk *Scatter Plots*, *Correlation Heatmap* dari *Hu Moments*, *Parallel Coordinates Plot* dari fitur dataset, dan visualisasi 3D *t-SNE*. Visualisasi ini membantu dalam memahami hubungan antar fitur dan distribusinya, yang penting untuk pemilihan fitur yang efektif dan pelatihan model.



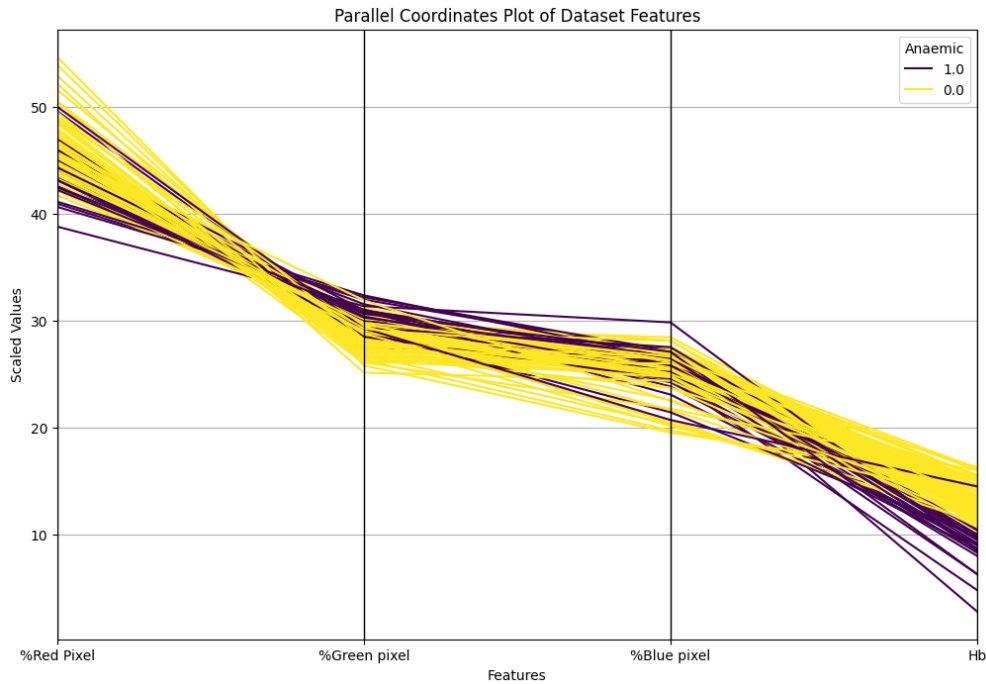
**Gambar 3.** *Scatter Plots of All Features*



**Gambar 4.** Correlation Heatmap of Hu Moments



**Gambar 5.** t-SNE Visualization of Dataset



**Gambar 6.** *Parallel Coordinates Plot of Dataset Features*

Gambar 3 menampilkan distribusi setiap fitur dalam dataset, menyoroti perbedaan antara kasus anemia dan non-anemia. Gambar 4 memvisualisasikan korelasi antara berbagai fitur dalam dataset, membantu mengidentifikasi variabel yang sangat berkorelasi yang dapat memengaruhi kinerja model karena multikolinearitas. Gambar 6 memvisualisasikan fitur di berbagai kasus, memungkinkan identifikasi pola dan *outliers* dalam data. Gambar 5 adalah teknik yang digunakan untuk memvisualisasikan data berdimensi tinggi dalam ruang berdimensi lebih rendah, menunjukkan kluster dan pemisahan antara kasus anemia dan non-anemia.

**b. Metode Analisis Data**

Proses analisis data melibatkan beberapa langkah untuk memastikan efektivitas dan keandalan model. Langkah pertama adalah *pre-processing data*, yang mencakup penanganan nilai hilang, pengkodean variabel kategorikal, dan skala fitur numerik. Secara khusus, kolom 'Nuimber' dan 'Sex' dihapus, dan kolom 'Anaemic' dikodekan ke nilai biner (*No*=0, *Yes*=1). *Feature scaling* diterapkan untuk menstandarkan data, memastikan mean 0 dan varian 1.

$$X_{scaled} = \frac{X - \mu}{\sigma} \tag{1}$$

Dimana  $\mu$  adalah rata-rata dan  $\sigma$  adalah simpangan baku dari fitur tersebut.

**Tabel 2.** *Features Descriptions* setelah *Pre-processing*

Nama Kolom	Type	Deskripsi
<i>%Red Pixel</i>	<i>Float</i>	Persentase piksel merah dalam gambar.
<i>%Green Pixel</i>	<i>Float</i>	Persentase piksel hijau dalam gambar
<i>%Blue Pixel</i>	<i>Float</i>	Persentase piksel biru dalam gambar.
<i>Hb</i>	<i>Float</i>	Tingkat hemoglobin dalam gram per desiliter (g/dL).

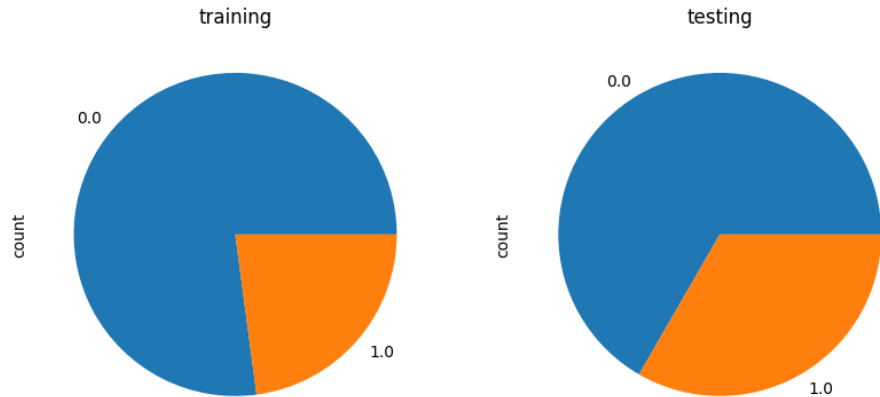


Anaemic	Float	Status Anemia (Yes=1, No=0).
---------	-------	------------------------------

c. **Model Training and Evaluation**

*Random Forest Classifier* dilatih menggunakan dataset yang telah diproses [17], [18]. Kinerja model dievaluasi menggunakan *5-fold cross-validation* [19], dimana dataset dibagi menjadi lima bagian, dan model dilatih dan divalidasi lima kali, setiap kali menggunakan bagian yang berbeda sebagai himpunan validasi dan bagian lainnya sebagai himpunan pelatihan.

- 1). *Data Splitting*: Dataset dibagi menjadi himpunan pelatihan (80%) dan pengujian (20%) untuk memfasilitasi validasi dan pengujian model.



**Gambar 7.** *Splitting Dataset 20 % testing, 80% training*

- 2). *Random Forest Classification*: Algoritma *Random Forest* beroperasi sebagai metode ansambel, di mana setiap pohon dibuat menggunakan sampel bootstrap yang berbeda dari data pelatihan. Prediksi akhir diperoleh dari agregasi prediksi dari semua pohon individu [20]–[24]. Untuk tugas klasifikasi, suara mayoritas digunakan:

$$\hat{y} = \text{mode}\{T_1(x), T_2(x), \dots, T_n(x)\} \tag{2}$$

dimana:

$\hat{y}$  adalah kelas yang diprediksi.

$T_i(x)$  adalah prediksi dari pohon ke- $i$  untuk input  $x$ .

$n$  adalah jumlah total pohon.

- 3). *Cross-Validation*: Teknik validasi silang *5-fold* digunakan untuk memastikan kekokohan model [25]–[27]. Dataset dibagi menjadi 5 subset, dan model dilatih dan divalidasi sebanyak 5 kali, setiap kali menggunakan subset yang berbeda sebagai himpunan validasi dan sisanya sebagai himpunan pelatihan.

$$CV_{(K)} = \frac{1}{K} \sum_{i=1}^K \text{Error}_i \tag{3}$$

- 4). **Accuracy**: Rasio prediksi yang benar terhadap total instansi [28]:

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{4}$$

- 5). **Precision**: Rasio prediksi positif benar terhadap total prediksi positif [29]:

$$Precision = \frac{TP}{(TP + FP)} \tag{5}$$

6). **Recall:** Rasio prediksi positif benar terhadap total positif aktual [2]:

$$Recall = \frac{TP}{(TP + FN)} \tag{6}$$

7). **F1-Score:** Rata-rata dari presisi dan *recall* [22]:

$$F - measure = \frac{2(presisi \times recall)}{(presisi + recall)} \tag{7}$$

Di mana *TP* adalah *True Positive*, *TN* adalah *True Negative*, *FP* adalah *False Positive*, dan *FN* adalah *False Negative*. Metrik-metrik ini memberikan pemahaman komprehensif tentang kinerja model, menyoroti kekuatan dan area yang perlu ditingkatkan.

#### 4. PENGUJIAN DAN PEMBAHASAN

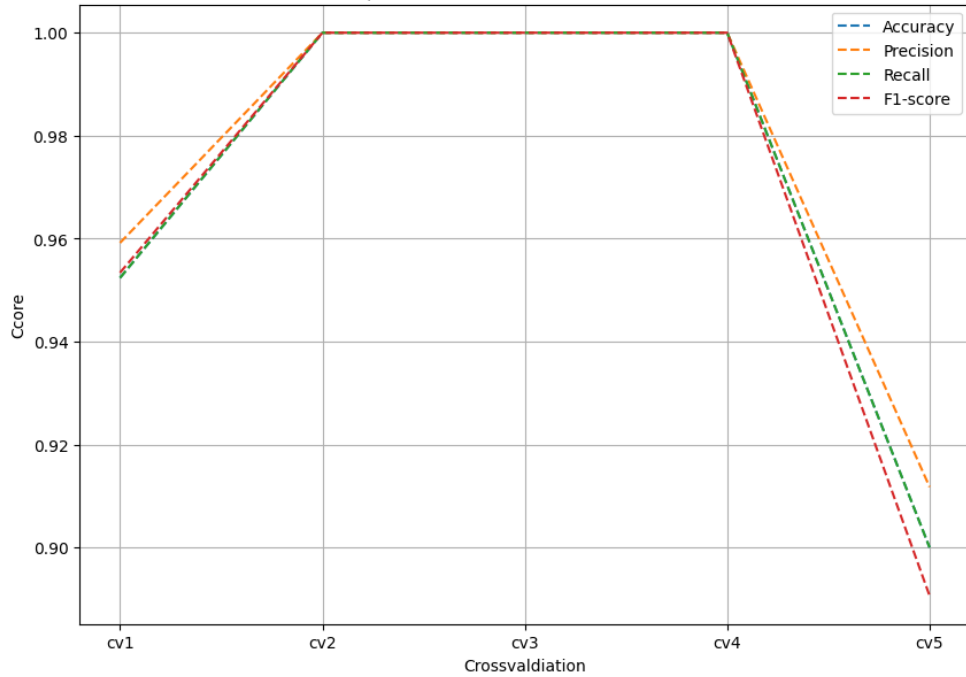
##### a. Hasil

Dataset dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Kolom ‘*Number*’ dan ‘*Sex*’ dihapus, dan kolom ‘*Anaemic*’ dikodekan menjadi nilai biner (*No*=0, *Yes*=1). Fitur numerik diskalakan agar memiliki mean 0 dan variansi 1, memastikan konsistensi dalam data yang dimasukkan ke dalam model. *Random Forest Classifier* dilatih menggunakan data yang telah diproses ini, dengan *hiperparameter* yang dioptimalkan melalui *grid search* dan dievaluasi menggunakan validasi silang 5-fold. Kinerja *Random Forest Classifier* di seluruh 5 lipatan dicatat sebagai berikut pada Tabel 3.

**Tabel 3.** Performa Metrik menggunakan Algoritma RFC

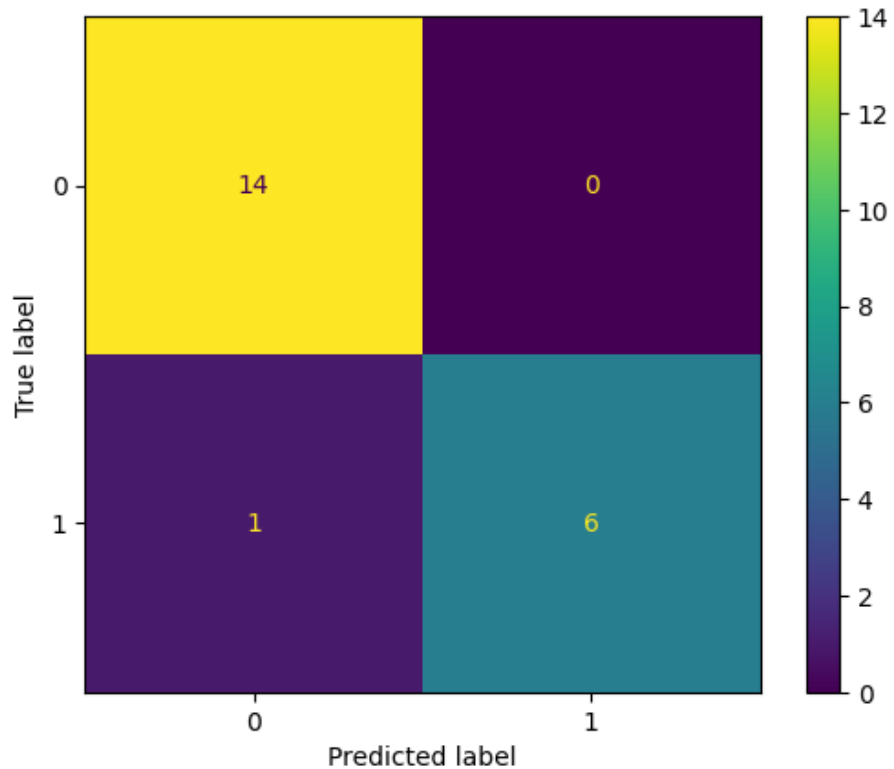
K-n	Metrik			
	Akurasi	Preisis	Recall	F-Measure
K-1	95.24%	95.92%	95.24%	95.34%
K-2	100%	100%	100%	100%
K-3	100%	100%	100%	100%
K-4	100%	100%	100%	100%
K-5	90%	91.18%	90%	89.06%
$\sum$ Avg	97.05%	97.02%	97.05%	96.88%

Grafik dan tabel berikut memberikan visualisasi rinci tentang kinerja *Random Forest Classifier* selama validasi silang 5-fold dan matriks kebingungan, yang memberikan gambaran jelas tentang akurasi klasifikasi model. Visualisasi ini sangat penting untuk memahami kekuatan model serta area yang perlu ditingkatkan.



**Gambar 8.** Grafik Performa Metrik menggunakan Algoritma *RFC*

Untuk lebih memahami kinerja model, berbagai visualisasi disediakan, termasuk grafik metrik kinerja di seluruh lipatan dan matriks kebingungan yang mengilustrasikan jumlah *true positive*, *true negative*, *false positive*, dan *false negative*.



**Gambar 9.** *Confusion Matrix*

Metrik kinerja menunjukkan bahwa *Random Forest Classifier* secara konsisten mencapai akurasi, presisi, recall, dan F1-score yang tinggi di semua lipatan, yang menunjukkan ketangguhan dan keandalannya dalam memprediksi anemia. Variasi kecil pada Lipatan 5, di mana

metrik sedikit menurun, menunjukkan beberapa variasi dalam dataset, yang umum dalam aplikasi dunia nyata.

#### b. Pembahasan

Hasil menunjukkan bahwa *Random Forest Classifier* sangat efektif dalam memprediksi anemia, dengan rata-rata metrik kinerja di atas 90% untuk semua ukuran yang dievaluasi. Tingkat kinerja yang tinggi ini menyoroti potensi model pembelajaran mesin dalam diagnosis medis, terutama dalam metode non-invasif seperti yang diusulkan dalam penelitian ini. Dibandingkan dengan penelitian sebelumnya, terlihat bahwa integrasi antara level hemoglobin dan data piksel gambar memberikan dasar yang kuat untuk prediksi anemia yang akurat. Penelitian sebelumnya juga telah menyoroti efektivitas pembelajaran mesin dalam diagnosis medis, dan temuan kami sejalan dengan observasi ini, memperkuat nilai analitik canggih dalam bidang kesehatan. Implikasi praktis dari penelitian ini sangat signifikan, terutama untuk daerah dengan sumber daya terbatas di mana metode diagnostik tradisional kurang dapat diakses. Dengan memanfaatkan data yang mudah diperoleh, seperti tingkat hemoglobin dan gambar, pendekatan ini menawarkan solusi yang hemat biaya dan skalabel untuk deteksi dini anemia.

Namun, penelitian ini tidak lepas dari keterbatasan. Ketidakseimbangan dalam dataset dapat memengaruhi generalisasi model. Selain itu, potensi bias dalam proses pengumpulan data gambar dapat mempengaruhi hasil. Penelitian selanjutnya sebaiknya mengeksplorasi fitur tambahan, dataset yang lebih besar dan beragam, serta model pembelajaran mesin alternatif untuk meningkatkan ketangguhan dan akurasi prediksi. Secara keseluruhan, penelitian ini berkontribusi pada literatur yang berkembang tentang metode diagnostik non-invasif dan menyoroti potensi transformatif pembelajaran mesin dalam layanan kesehatan. Penelitian lanjutan harus membangun temuan ini dengan fokus pada penyempurnaan model dan mengeksplorasi aplikasinya dalam konteks klinis yang lebih luas..

### 5. KESIMPULAN

Secara keseluruhan, *Random Forest Classifier* menunjukkan kinerja yang luar biasa dalam memprediksi anemia menggunakan data tingkat hemoglobin dan piksel gambar, dengan rata-rata akurasi, presisi, *recall*, dan F1-score yang semuanya melebihi 90%. Tingkat akurasi yang tinggi ini menunjukkan bahwa kombinasi fitur-fitur tersebut sangat efektif untuk prediksi anemia secara non-invasif. Penelitian ini menjawab pertanyaan penelitian dengan mengonfirmasi bahwa tingkat hemoglobin dan distribusi piksel gambar dapat secara akurat memprediksi anemia, dan *Random Forest Classifier* merupakan model yang kuat untuk tujuan ini. Temuan ini memberikan kontribusi yang signifikan bagi bidang diagnostik medis, menawarkan alternatif yang hemat biaya dan mudah diakses dibandingkan metode diagnostik tradisional, terutama di daerah dengan keterbatasan sumber daya.

Penelitian ini turut berkontribusi pada literatur yang berkembang terkait aplikasi pembelajaran mesin dalam bidang kesehatan, yang menunjukkan potensi alat diagnostik non-invasif. Penelitian di masa depan harus berfokus pada upaya mengatasi keterbatasan studi ini, seperti ketidakseimbangan dataset dan potensi bias dalam pengumpulan data gambar. Selain itu, eksplorasi model pembelajaran mesin lain dan penerapan dataset yang lebih besar dan lebih beragam dapat lebih meningkatkan akurasi dan generalisasi prediksi. Upaya ini akan membantu menyempurnakan model dan memperluas penerapannya dalam berbagai konteks klinis, yang pada akhirnya meningkatkan deteksi dini dan pengobatan anemia.

### DAFTAR PUSTAKA

- [1] I. A. Liberty, I. S. Septadina, A. M. Kurniati, and E. S. Ananingsih, *Monograf Risiko Transmisi COVID-19 pada Ibu Rumah Tangga*. 2023.
- [2] A. P. Wibowo, M. Taruk, T. E. Tarigan, and ..., "Improving Mental Health Diagnostics through Advanced Algorithmic Models: A Case Study of Bipolar and Depressive Disorders," *Indones. J. ...*, 2024.

- [3] R. F. Syam, “Performance Comparison Analysis of Classifiers on Binary Classification Dataset,” *Indones. J. Data Sci.*, 2023.
- [4] N. Rismayanti and A. P. Utami, “Improving Multi-Class Classification on 5-Celebrity-Faces Dataset using Ensemble Classification Methods,” *Indones. J. Data ...*, 2023.
- [5] X. Yu, “Random forest algorithm-based classification model of pesticide aquatic toxicity to fishes,” *Aquat. Toxicol.*, vol. 251, 2022, doi: 10.1016/j.aquatox.2022.106265.
- [6] S. Khomsah and E. Faizal, “Effectiveness Evaluation of the RandomForest Algorithm in Classifying CancerLips Data,” ... *Artif. Intell. Med. ...*, 2023.
- [7] A. R. Manga, M. A. F. Latief, A. W. M. Gaffar, and ..., “Hyperparameter Tuning of Identity Block Uses an Imbalance Dataset with Hyperband Method,” *2024 18th ...*, 2024.
- [8] H. Azis, L. Syafie, F. Fattah, and ..., “Unveiling Algorithm Classification Excellence: Exploring Calendula and Coreopsis Flower Datasets with Varied Segmentation Techniques,” *2024 18th Int. ...*, 2024.
- [9] F. D. U. Arif, *Perbandingan Kinerja Algoritma Random Forest, Xgboost Dan Lightgbm Dalam Klasifikasi Emosi Komentar Reddit*. 2023.
- [10] Nilawati, “Perbandingan Tingkat Akurasi Metode Weighted Naïve Bayes Dengan Random Forest Dalam Mengklasifikasi Penerima Program Keluarga Harapan ( PKH ) Nilawati Program Studi Statistika Universitas Sulawesi Barat Tahun 2024,” 2024.
- [11] D. Liliyawati, “Perbandingan Performa Model Prediksi Customer Churn Berbasis Machine Learning Pada Fashion E-Commerce,” 2023.
- [12] A. Faradibah, D. Widyawati, A. U. T. Syahar, and ..., “Comparison Analysis of Random Forest Classifier, Support Vector Machine, and Artificial Neural Network Performance in Multiclass Brain Tumor Classification,” *Indones. J. ...*, 2023.
- [13] L. B. C. Tanujayaa, B. Susanto, and A. Saragiha, “Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Fitur Mode Audio Spotify,” *Indones. J. data Sci.*, vol. 1, no. 3, pp. 68–78, 2020, doi: <https://doi.org/10.33096/ijodas.v1i3.16>.
- [14] M. M. Baharuddin, T. Hasanuddin, and H. Azis, “Analisis Performa Metode K-Nearest Neighbor untuk Identifikasi Jenis Kaca,” *Ilk. J. Ilm.*, vol. 11, no. 28, pp. 269–274, 2019.
- [15] H. Azis, F. Fattah, and P. Putri, “Performa Klasifikasi K-NN dan Cross-validation pada Data Pasien Pengidap Penyakit Jantung,” *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 81–86, 2020.
- [16] H. Nugroho, G. E. Yuliastuti, and Andrean Firman Pradana, “Klasifikasi Diagnosis Diabetes Melitus Menggunakan Metode Naïve Bayes Dengan Seleksi Fitur Backward Elimination,” *NERO (Networking Eng. Res. Oper.)*, vol. 8, no. 2, pp. 97–106, 2023, doi: 10.21107/nero.v8i2.21110.
- [17] I. Sulistiani, “Breast Cancer Prediction Using Random Forest and Gaussian Naïve Bayes Algorithms,” *2022 1st Int. Conf. Inf. Syst. Inf. Technol. ICISIT 2022*, pp. 170–175, 2022, doi: 10.1109/ICISIT54091.2022.9872808.
- [18] I. F. Hanbal, “Classifying Wastes Using Random Forests, Gaussian Naïve Bayes, Support Vector Machine and Multilayer Perceptron,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 803, no. 1, 2020, doi: 10.1088/1757-899X/803/1/012017.
- [19] T. E. Tarigan, E. Susanti, M. I. Siami, I. Arfiani, and ..., “Performance Metrics of AdaBoost and Random Forest in Multi-Class Eye Disease Identification: An Imbalanced Dataset Approach,” ... *Artif. Intell. ...*, 2023.
- [20] I. Alwiah, U. Zaky, and A. W. Murdiyanto, “Assessing the Predictive Power of Logistic Regression on Liver Disease Prevalence in the Indian Context,” ... *J. Data Sci.*, 2024.
- [21] R. Setiawan and H. Oumarou, “Classification of Rice Grain Varieties Using Ensemble

- Learning and Image Analysis Techniques,” *Indones. J. Data ...*, 2024.
- [22] I. P. A. Pratama, E. S. J. Atmadji, and ..., “Evaluating the Performance of Voting Classifier in Multiclass Classification of Dry Bean Varieties,” *Indones. J. ...*, 2024.
- [23] B. S. W. Poetro, E. Maria, H. Zein, and ..., “Advancements in Agricultural Automation: SVM Classifier with Hu Moments for Vegetable Identification,” *Indones. J. ...*, 2024.
- [24] F. T. Admojo and N. Rismayanti, “Estimating Obesity Levels Using Decision Trees and K-Fold Cross-Validation: A Study on Eating Habits and Physical Conditions,” *Indones. J. Data ...*, 2024.
- [25] N. Rismayanti, A. Naswin, U. Zaky, M. Zakariyah, and D. A. Purnamasari, “Evaluating Thresholding-Based Segmentation and Humoment Feature Extraction in Acute Lymphoblastic Leukemia Classification using Gaussian Naive Bayes,” *Int. J. Artif. Intell. Med. Issues*, vol. 1, no. 2, 2023.
- [26] A. Naswin and A. P. Wibowo, “Performance Analysis of the Decision Tree Classification Algorithm on the Pneumonia Dataset,” ... *Artif. Intell. Med. ...*, 2023.
- [27] R. A. Azdy, R. F. Syam, E. Faizal, and ..., “Performance Evaluation of Bagging Meta-Estimator in Lung Disease Detection: A Case Study on Imbalanced Dataset,” *Int. J. ...*, 2023.
- [28] P. S. Kumar, “Classification of skin cancer using convolutional neural network in comparison with decision tree classifier,” *AIP Conf. Proc.*, vol. 2822, no. 1, 2023, doi: 10.1063/5.0173035.
- [29] V. R. Nitha, “Lung Cancer Malignancy detection Using Voting Ensemble Classifier,” *ICCSC 2023 - Proc. 2nd Int. Conf. Comput. Syst. Commun.*, 2023, doi: 10.1109/ICCSC56913.2023.10142984.