



Klasifikasi Penyakit Parkinson Menggunakan Algoritma Decision Tree

Muhammad Reizza Eduardi Yusuf^{1*}, Lamhot Simanullang², Nazar Pangestu³, Fajar Ramadhan⁴, Perani Rosyani⁵

^{1,2,3,4,5}Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

Email: ^{1*}reizza.storage01cpt04@gmail.com, ²lamhotsimanullang65@gmail.com,

³nazarpangestu95@gmail.com, ⁴fajarfarel42@gmail.com, ⁵dosen00837@Inpam.ac.id

(*:corresponding author)

Abstrak - Penelitian ini mengulas penggunaan algoritma *Decision Tree* dalam klasifikasi penyakit Parkinson (PD) yang berfokus pada efektivitas teknik seperti *CART*, *C4.5*, dan berbagai metode *ensemble* untuk meningkatkan akurasi diagnostik. Seiring meningkatnya prevalensi PD di kalangan lansia, deteksi dini menjadi sangat penting karena memungkinkan perawatan yang lebih efektif dan pemantauan progresi penyakit. *Decision Tree* dan teknik *ensemble*-nya seperti *Random Forest*, *AdaBoost*, dan *RUSBoost* banyak diterapkan karena struktur hierarkinya yang mudah diinterpretasikan dan kemampuannya dalam menangani data numerik dan kategorikal. Dalam beberapa studi, algoritma ini dikombinasikan dengan teknik *balancing data* seperti *SMOTE* dan *RUS*, serta metode seleksi fitur seperti *lasso* dan informasi gain. Kajian literatur ini mencakup analisis efektivitas berbagai pendekatan berbasis *Decision Tree* dalam mengatasi tantangan data tidak seimbang, seperti yang umum dijumpai dalam dataset klinis dan vokal terkait Parkinson. Melalui tinjauan sistematis ini, ditemukan bahwa teknik *ensemble Decision Tree* umumnya meningkatkan akurasi dan sensitivitas, terutama ketika diterapkan pada dataset yang telah diproses dengan metode *balancing data*. Studi ini bertujuan memberikan pandangan yang komprehensif tentang kekuatan, keterbatasan, dan potensi optimalisasi penggunaan *Decision Tree* dalam membantu diagnosis klinis Parkinson, menawarkan kontribusi signifikan terhadap upaya peningkatan keandalan diagnosis berbasis *machine learning*.

Kata Kunci: *Decision Tree*, Klasifikasi Penyakit Parkinson, Enasemble Learning, *Random Forest*, *AdaBoost*, Diagnosis Klinis Parkinson, *CART*

Abstract - This study reviews the use of Decision Tree algorithms in Parkinson's Disease (PD) classification, focusing on the effectiveness of techniques such as *CART*, *C4.5*, and various ensemble methods to improve diagnostic accuracy. With the increasing prevalence of PD among the elderly, early detection has become critical as it enables more effective treatment and monitoring of disease progression. Decision Trees and their ensemble techniques, including *Random Forest*, *AdaBoost*, and *RUSBoost*, are widely applied due to their hierarchical structure, ease of interpretation, and ability to handle both numerical and categorical data. In some studies, these algorithms are combined with data balancing techniques such as *SMOTE* and *RUS*, as well as feature selection methods like *lasso* and information gain. This literature review provides an analysis of the effectiveness of various Decision Tree-based approaches in addressing challenges associated with imbalanced data, which is common in clinical and vocal datasets related to Parkinson's. The systematic review found that ensemble Decision Tree techniques generally improve accuracy and sensitivity, especially when applied to datasets processed with data balancing methods. This study aims to provide a comprehensive view of the strengths, limitations, and optimization potential of using Decision Trees to assist in clinical Parkinson's diagnosis, offering a significant contribution to efforts aimed at enhancing the reliability of machine learning-based diagnosis.

Keywords: *Decision Tree*, *Parkinson's Disease Classification*, *Ensemble Learning*, *Random Forest*, *AdaBoost*, *Parkinson's Clinical Diagnosis*, *CART*

1. PENDAHULUAN

Penyakit Parkinson (PD) adalah penyakit neurologis progresif yang mempengaruhi sistem saraf pusat dan biasanya terjadi pada orang dewasa yang lebih tua. PD menyebabkan gejala motorik seperti tremor, kaku, dan gangguan postur tubuh, yang dapat berdampak signifikan terhadap kualitas hidup pasien. Dalam beberapa tahun terakhir, diketahui bahwa teknologi pembelajaran mesin memiliki potensi besar untuk mendeteksi penyakit Parkinson dengan menganalisis data klinis dan audio pasien.

Algoritma *Decision Tree*, terutama varian seperti *CART* dan *C4.5*, lebih andal karena kemampuannya menangani campuran data numerik dan kategorikal serta membuat model yang



mudah diinterpretasikan. Algoritma ini memungkinkan kami membangun langkah klasifikasi hierarki berdasarkan fitur penting yang ditemukan pada pasien PD.

Penggunaan teknik *ensemble* seperti *Random Forests*, *AdaBoost*, dan *RUSBoost* dalam pohon keputusan semakin menyempurnakan pendekatan ini, khususnya untuk meningkatkan akurasi dan mengatasi tantangan ketidakseimbangan data yang biasa ditemukan dalam kumpulan data medis.

Selanjutnya, pendekatan ini sering dikombinasikan dengan teknik penyeimbangan data seperti *SMOTE* dan undersampling serta pemilihan fitur seperti Lasso untuk meningkatkan kinerja klasifikasi.

Pengenalan ini menjelaskan potensi penerapan pohon keputusan untuk mendiagnosis penyakit Parkinson secara lebih efektif dan memberikan wawasan tentang manfaat teknik ini dalam rangkaian metode diagnostik berbasis pembelajaran mesin.

2. METODE

Pendekatan yang digunakan dalam penelitian ini adalah *Systematic Literature Review* (SLR), yang bertujuan untuk mengidentifikasi, mengevaluasi, dan mensintesis studi-studi terkait penggunaan algoritma *Decision Tree* dalam klasifikasi penyakit Parkinson.

2.1. Pendekatan Penelitian

Systematic Literature Review (SLR) adalah metodologi yang diterapkan untuk memperoleh pemahaman yang komprehensif tentang suatu topik melalui pencarian dan analisis sistematis terhadap literatur yang ada. Pendekatan ini melibatkan serangkaian langkah terstruktur yang meliputi pencarian literatur, seleksi, ekstraksi data, dan sintesis hasil. Pada penelitian ini, *SLR* digunakan untuk merangkum pendekatan-pendekatan berbasis *Decision Tree* dalam klasifikasi Parkinson, mencakup variasi teknik seperti *CART*, *C4.5*, dan *ensemble Decision Tree*.

2.2. Sumber dan Kriteria Pemilihan Literatur

Pemilihan literatur dilakukan berdasarkan kriteria sebagai berikut:

1. Relevansi dengan Topik:

Hanya studi yang secara eksplisit mengaplikasikan algoritma *Decision Tree*, baik dalam bentuk tunggal atau dengan teknik *ensemble*, untuk klasifikasi penyakit Parkinson yang disertakan. Studi yang menggunakan metode pembelajaran mesin lain tanpa melibatkan *Decision Tree* sebagai komponen utama akan dikeluarkan.

2. Jenis Data:

Literatur yang menggunakan data klinis, suara, atau data medis lain yang terkait dengan gejala Parkinson diprioritaskan. Data yang dianalisis mencakup dataset vokal dari pasien Parkinson, serta fitur lain seperti hasil pencitraan otak.

3. Hasil yang Dilaporkan:

Hanya studi yang melaporkan hasil klasifikasi dalam bentuk metrik yang standar seperti akurasi, sensitivitas, spesifitas, atau *F1-score* yang dipertimbangkan untuk memastikan adanya data hasil yang dapat dibandingkan.

4. Ketersediaan dalam Jurnal Akademik Terindeks:

Literatur yang dipublikasikan dalam jurnal terindeks atau prosiding konferensi internasional disertakan untuk memastikan kualitas dan relevansi kajian.

2.3. Strategi Pencarian Literatur

Pencarian literatur dilakukan pada basis data akademik, termasuk *IEEE Xplore*, *PubMed*, *ScienceDirect*, dan *Google Scholar*. Pencarian ini dilakukan untuk memperoleh akses pada studi yang komprehensif, menggunakan kata kunci berikut:



- "Decision Tree for Parkinson's Disease Classification"
- "Parkinson's Disease Detection with Ensemble Decision Tree"
- "Machine Learning for Parkinson's Diagnosis using Decision Trees"
- "Decision Tree and Voice Data in Parkinson's Disease Detection"

Kata kunci dikombinasikan dengan operator *boolean* (*AND*, *OR*) untuk mempersempit hasil pencarian sesuai dengan fokus penelitian. Artikel yang ditemukan diperiksa secara manual untuk memastikan kesesuaian dengan kriteria inklusi yang telah ditetapkan.

2.4. Proses Analisis Literatur

Setiap literatur yang terpilih dianalisis secara rinci berdasarkan variabel-variabel berikut:

1. Metode Klasifikasi:

Analisis dilakukan untuk mengetahui metode algoritma *Decision Tree* yang digunakan, seperti *CART*, *C4.5*, atau teknik *ensemble* seperti *Random Forest* dan *AdaBoost*. Selain itu, teknik *balancing data* seperti *SMOTE* (*Synthetic Minority Over-sampling Technique*) dan *undersampling* juga dicatat jika digunakan untuk menangani ketidakseimbangan data.

2. Tujuan Penelitian:

Tujuan utama penelitian, baik untuk klasifikasi subtipe PD seperti *Tremor Dominant (TD)* dan *Postural Instability/Gait Difficulty (PIGD)*, ataupun deteksi awal Parkinson, dicatat untuk mengidentifikasi fokus aplikasi algoritma yang berbeda.

3. Hasil Penelitian:

Hasil utama seperti akurasi, sensitivitas, spesifisitas, dan *F1-score* dicatat dan dibandingkan antar studi untuk mengevaluasi efektivitas algoritma. Metrik ini menjadi tolok ukur untuk melihat tingkat keberhasilan model dalam mengklasifikasikan data PD.

4. Teknik Pemilihan Fitur:

Beberapa studi menggunakan teknik pemilihan fitur, seperti lasso dan informasi gain, untuk meningkatkan efisiensi algoritma. Metode ini bertujuan mengurangi fitur yang tidak relevan atau berlebih untuk meningkatkan keakuratan dan mempercepat proses klasifikasi.

5. Pendekatan Balancing Data:

Dataset yang tidak seimbang dapat berdampak pada hasil klasifikasi. Oleh karena itu, pendekatan *balancing data* seperti *SMOTE* atau *undersampling* dibahas untuk melihat dampaknya pada performa algoritma *Decision Tree* dalam beberapa penelitian.

2.5. Pelaporan Hasil dan Sintesis

Pelaporan hasil dari literatur yang terpilih disusun dalam bentuk tabel untuk memudahkan komparasi dan interpretasi. Tabel ini mencakup informasi peneliti dan tahun publikasi, metode yang dibahas, tujuan penelitian, serta hasil klasifikasi yang dilaporkan. Tabel disusun untuk menunjukkan pola umum, keunggulan, dan tantangan dalam penggunaan *Decision Tree* pada klasifikasi penyakit Parkinson. Hasil analisis kemudian disintesis untuk memberikan pandangan komprehensif tentang efektivitas *Decision Tree* dalam konteks klasifikasi PD.

2.6. Keterbatasan

Studi ini hanya mencakup literatur yang menggunakan algoritma *Decision Tree* dan *ensemble*-nya, sehingga keterbatasan metodologi ini adalah tidak melibatkan metode lain yang juga mungkin relevan, seperti neural networks atau metode deep learning. Hal ini dapat mengurangi luas cakupan komparasi antar algoritma. Selain itu, hanya studi yang tersedia dalam publikasi jurnal atau prosiding terindeks yang diikutsertakan, sehingga beberapa penelitian relevan yang belum dipublikasikan mungkin terlewat.



3. PEMBAHASAN

No	Nama Peneliti & Tahun	Metode yang Dibahas	Tujuan Penelitian	Hasil yang Didapatkan
1	Nurrohman, A., Abdullah, S., Murfi, H. (2020). Parkinson's disease subtype classification: Application of Decision Tree, Logistic Regression and Logit Leaf Model.	Penelitian menggunakan tiga metode klasifikasi, yaitu Decision Tree (DT), Logistic Regression (LR), dan Logit Leaf Model (LLM), untuk membedakan subtipo penyakit Parkinson: Tremor Dominant (TD) dan Postural Instability/Gait Difficulty (PIGD). Data pada penelitian ini diambil dari database Parkinson's Progression Markers Initiative (PPMI), yang kemudian diproses dengan teknik penyeimbangan data yang meliputi oversampling, undersampling, dan Synthetic Minority Over-sampling Technique (<i>SMOTE</i>) untuk mengatasi ketidakseimbangan data.	Membedakan subtipo penyakit Parkinson (TD dan PIGD) yang memerlukan perbedaan dalam pengobatan klinis. Dengan demikian, penelitian ini bertujuan untuk menemukan model terbaik untuk klasifikasi subtipo Parkinson dengan kinerja yang optimal, terutama pada data yang tidak seimbang.	Model Logistic Regression yang diproses dengan <i>SMOTE</i> ($\alpha = 600$, $\gamma = 200$) menunjukkan kinerja terbaik dengan akurasi sebesar 98.83%, sensitivitas 98.41%, dan spesifitas 99.07%. Sementara itu, Decision Tree menghasilkan akurasi yang lebih rendah dibandingkan metode lain, yaitu sebesar 88.64%, dengan sensitivitas hanya 29.63%. Berdasarkan hasil ini, Logistic Regression yang menggunakan teknik <i>SMOTE</i> dipilih sebagai model terbaik untuk klasifikasi subtipo Parkinson dalam penelitian ini
2	Williams, D. P., Mudali, D., Buddelmeijer, H., Noorishad, P., Meles, S., Renken, R. J., Leenders, K. L., Valentijn, E. A., Roerdink, J. B. T. M. (2016). Visualization of Decision Tree State for the Classification of Parkinson's Disease.	Menggunakan algoritma Decision Tree C4.5 yang diterapkan pada data hasil pemindaian FDG-PET untuk memisahkan subjek yang sehat dari mereka yang menderita sindrom Parkinsonian, termasuk Multiple-System Atrophy (MSA) dan Progressive Supranuclear Palsy (PSP). Metode visualisasi ini	Meningkatkan pemahaman dan interpretasi keputusan yang diambil oleh model pohon keputusan dalam mendiagnosa penyakit Parkinson. Selain itu, penelitian ini berusaha untuk mengembangkan sistem visualisasi yang memungkinkan pengguna, terutama ahli saraf, untuk melacak dan memahami jalur klasifikasi setiap	Pohon keputusan yang dihasilkan mencapai akurasi klasifikasi hingga 80% tergantung pada sindrom Parkinsonian yang didiagnosa. Visualisasi tambahan membantu pengguna mengidentifikasi pola yang khas untuk setiap node dalam pohon dan memungkinkan klasifikasi yang



		<p>memperkuat setiap node dalam pohon keputusan dengan gambar komponen utama (principal component) dari data, yang membantu dalam menelusuri klasifikasi subjek dan memberikan pemahaman lebih mendalam tentang pola aktivitas otak.</p>	<p>subjek dan menjelaskan klasifikasi yang kompleks atau tidak terduga.</p>	<p>lebih intuitif dari hasil. Para ahli saraf yang terlibat dalam evaluasi juga memberikan tanggapan positif terkait kemampuan visualisasi ini dalam meningkatkan wawasan mereka terhadap data pasien, meskipun pohon keputusan bukanlah metode paling akurat untuk klasifikasi ini</p>
3	Barukab, O., Ahmad, A., Khan, T., Thayyil Kunhumuhammed, M. R. (2022). Analysis of Parkinson's Disease Using an Imbalanced-Speech Dataset by Employing Decision Tree Ensemble Methods.	<p>Penelitian melibatkan metode <i>Decision Tree Ensemble</i>, khususnya dengan teknik penyeimbangan data seperti <i>SMOTE (Synthetic Minority Oversampling Technique)</i> dan <i>Random Undersampling (RUS)</i>. Algoritma yang digunakan antara lain <i>AdaBoost</i>, <i>Random Forest</i>, <i>RUSBoost</i>, dan <i>Balanced Random Forest</i>.</p>	<p>Mengatasi masalah ketidakseimbangan kelas pada data suara pasien penyakit Parkinson untuk meningkatkan akurasi klasifikasi dan mengurangi kesalahan prediksi. Penelitian ini juga mengevaluasi berbagai metode <i>ensemble</i> berbasis pohon keputusan untuk mengklasifikasikan antara pasien Parkinson dan subjek yang sehat menggunakan data suara yang tidak seimbang.</p>	<p>Algoritma <i>AdaBoost</i> dengan pemilihan fitur menggunakan metode <i>Information Gain</i> menghasilkan skor <i>F1</i> tertinggi sebesar 0.903. Selain itu, penggunaan teknik <i>ensemble</i> seperti <i>Random Forest</i> dan <i>RUSBoost</i> menunjukkan kinerja terbaik dalam beberapa metrik evaluasi, seperti area di bawah kurva <i>ROC (AUROC)</i>, sensitivitas, dan spesifitas, dengan hasil terbaik pada data yang telah disesuaikan dengan teknik penyeimbangan kelas.</p>
4	Li, Y., Yang, L., Wang, P., Zhang, C., Xiao, J., Zhang, Y., Qiu, M. (2017). Classification of Parkinson's Disease by Decision Tree Based Instance Selection and	<p>Melibatkan metode kombinasi <i>Classification and Regression Tree (CART)</i> untuk pemilihan sampel dan algoritma <i>Ensemble Learning</i> yang menggabungkan</p>	<p>Penelitian bertujuan untuk meningkatkan akurasi dalam mendiagnosa penyakit Parkinson melalui data suara menggunakan teknik pemilihan sampel optimal dan pembelajaran</p>	<p>Algoritma yang diusulkan, yaitu <i>PD-EL_withCART</i>, mencapai akurasi klasifikasi tertinggi sebesar 90% dengan sensitivitas mencapai 100% pada data uji tertentu. Algoritma</p>



	Ensemble Learning Algorithms	<i>Random Forest (RF), Support Vector Machine (SVM), dan Extreme Learning Machine (ELM)</i> dalam klasifikasi.	<i>ensemble.</i> Tujuan utamanya adalah mengurangi kesalahan klasifikasi yang disebabkan oleh sampel dengan kualitas rendah dan mengoptimalkan akurasi dan stabilitas dalam pengklasifikasian penyakit Parkinson.	ini menunjukkan peningkatan performa dibandingkan dengan algoritma lain yang relevan, dengan rata-rata akurasi mencapai 86.5% saat menggabungkan <i>CART</i> dan <i>RF</i> .
5	Desiani, A., Narti, Irmeilyana, Ramayanti, I., & Arhami, M. (2023). Diagnosa Penyakit Parkinson dengan Algoritma K-Nearest Neighbor dan Decision Tree C4.5	Membandingkan algoritma <i>K-Nearest Neighbor (KNN)</i> dan <i>Decision Tree C4.5</i> untuk diagnosa penyakit Parkinson. Metode pelatihan yang digunakan meliputi metode <i>percentage split</i> dan <i>K-fold cross validation</i> .	Mengukur efektivitas kedua algoritma dalam mendiagnosis penyakit Parkinson dengan membandingkan kinerja dari <i>KNN</i> dan <i>Decision Tree C4.5</i> dalam hal akurasi, presisi, dan recall untuk menentukan metode yang lebih optimal.	Algoritma <i>Decision Tree C4.5</i> menunjukkan hasil lebih baik dibandingkan dengan <i>KNN</i> . Pada metode <i>percentage split</i> , akurasi tertinggi yang dicapai oleh algoritma <i>C4.5</i> adalah 89%, sementara <i>KNN</i> mencapai akurasi 82%. Pada metode <i>K-fold cross validation</i> , algoritma <i>C4.5</i> mencapai akurasi 81% dibandingkan dengan 76.8% pada <i>KNN</i> . Presisi dan recall algoritma <i>C4.5</i> juga lebih tinggi, menunjukkan bahwa <i>C4.5</i> lebih efisien dalam diagnosa penyakit Parkinson berdasarkan dataset yang digunakan

4. KESIMPULAN

Berbagai penelitian menunjukkan bahwa algoritma *Decision Tree*, termasuk varian *ensemble* seperti *Random Forest* dan *AdaBoost*, dapat memberikan performa yang signifikan dalam klasifikasi penyakit Parkinson. Teknik *balancing data*, seperti *SMOTE* dan undersampling, membantu meningkatkan keakuratan model pada dataset yang tidak seimbang. Pemilihan fitur melalui metode seperti informasi gain dan lasso meningkatkan efisiensi model dan akurasi klasifikasi. Studi ini menunjukkan bahwa *Decision Tree* efektif dalam mendeteksi Parkinson dan membantu mengidentifikasi subtipe PD, dengan pendekatan *ensemble* memberikan hasil terbaik dalam banyak



kasus. Hasil studi ini memberikan dasar bagi penelitian lanjutan untuk mengoptimalkan penggunaan algoritma *Decision Tree* dalam aplikasi medis.

DAFTAR PUSTAKA

- [1] Nurrohman, A., Abdullah, S., Murfi, H. (2020). Parkinson's disease subtype classification: Application of Decision Tree, Logistic Regression and Logit Leaf Model. *AIP Conference Proceedings*, 2242, 030015.
- [2] Williams, D. P., Mudali, D., Buddelmeijer, H., Noorishad, P., Meles, S., Renken, R. J., Leenders, K. L., Valentijn, E. A., Roerdink, J. B. T. M. (2016). Visualization of Decision Tree State for the Classification of Parkinson's Disease. *Journal of Biomedical Engineering and Medical Imaging*, 3(3), 25-41.
- [3] Barukab, O., Ahmad, A., Khan, T., Thayyil Kunhumuhammed, M. R. (2022). Analysis of Parkinson's Disease Using an Imbalanced-Speech Dataset by Employing Decision Tree Ensemble Methods. *Diagnostics*, 12(3000), 1-21.
- [4] Li, Y., Yang, L., Wang, P., Zhang, C., Xiao, J., Zhang, Y., Qiu, M. (2017). Classification of Parkinson's Disease by Decision Tree Based Instance Selection and Ensemble Learning Algorithms. *Journal of Artificial Intelligence Research*, 45, 136-149.
- [5] Desiani, A., Narti, Irmeilyana, Ramayanti, I., & Arhami, M. (2023). Diagnosa Penyakit Parkinson dengan Algoritma K-Nearest Neighbor dan Decision Tree C4.5. *Jurnal SimanteC*, 12(1), 47-57.
- [6] Oktafiani, R., Hermawan, A., & Avianto, D. (2024). Max Depth Impact on Heart Disease Classification: Decision Tree and Random Forest. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 8(1), 160–168.
- [7] Rajaguru, H., & Chakravarthy, S. S. R. (2019). Analysis of Decision Tree and K-Nearest Neighbor Algorithm in the Classification of Breast Cancer. *Asian Pacific Journal of Cancer Prevention*, 20(12), 3777–3781.
- [8] Rosyani, P., & Amalia, R. (2021). Segmentasi Citra Tanaman Obat dengan metode K-Means dan Otsu. *Jurnal Informatika Universitas Pamulang*, 6(2), 246-251.
- [9] Syahputra, S. A. F., Azizah, N. M., Aiman, J., Nikmah, D. A., & Rosyani, P. (2024). Identifikasi dan Prediksi Umur Berdasarkan Citra Wajah Menggunakan Deep Learning Algoritma Convolutional Neural Network (CNN). *Jurnal AI dan SPK: Jurnal Artificial Intelligence dan Sistem Penunjang Keputusan*, 2(1), 87-95.
- [10] Hazmy, A., Putra, M. B. A., Saputra, M. A., Hartono, R., & Rosyani, P. (2024). Implementasi Deep Learning untuk Deteksi Wajah dan Ekspresi menggunakan Algoritma Convolutional Neural Network (CNN) dengan OpenCV. *Jurnal AI dan SPK: Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan*, 1(4), 261-265.