

Tinjauan Komparatif Klasifikasi Penyakit Stroke Berdasarkan Fitur Medis Menggunakan *Random Forest*

Anggarini Puspita Rarasati^{1*}, Deytiza Putri Nur Azizah¹, Puput Arie Sugiyanto¹,
Arif Suryadin¹

¹Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspipetek No. 46,
Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: ^{1*}puspitararasati25@gmail.com, ²putrideytiza@gmail.com, ³puputarie88@gmail.com,
⁴arif Suryadin05@gmail.com

(* : coresponding author)

Abstrak– Stroke adalah salah satu penyakit yang mengakibatkan tingkat kecacatan dan kematian tinggi, yang memerlukan prediksi dan diagnosis dini untuk meningkatkan kualitas penanganan pasien. Salah satu algoritma yang banyak digunakan dalam klasifikasi penyakit ini adalah Random Forest, yang memiliki keunggulan dalam mengolah data kompleks dan menghasilkan akurasi yang tinggi. Penelitian ini bertujuan untuk melakukan tinjauan komparatif mengenai penerapan algoritma Random Forest dalam klasifikasi stroke berdasarkan fitur medis. Analisis komparatif dilakukan terhadap beberapa jurnal ilmiah untuk menilai efektivitas, akurasi, dan kinerja algoritma ini dalam berbagai pengaturan parameter dan teknik pemrosesan data. Hasil dari studi ini diharapkan memberikan gambaran mengenai variasi implementasi Random Forest dalam klasifikasi penyakit stroke serta mengidentifikasi potensi penelitian lanjutan untuk optimasi metode ini.

Kata Kunci: Stroke, *Random Forest*, Klasifikasi, Fitur Medis, *Machine Learning*

Abstract– Stroke is a leading cause of disability and mortality, requiring early prediction and diagnosis to improve patient care quality. One widely used algorithm in this classification is Random Forest, known for its advantages in processing complex data and yielding high accuracy. This study aims to conduct a comparative review of Random Forest applications for stroke classification based on medical features. A comparative analysis is performed across several scholarly journals to assess the algorithm's effectiveness, accuracy, and performance under various parameter settings and data processing techniques. The results of this study are expected to provide insights into the different implementations of Random Forest in stroke classification and identify potential areas for further research to optimize this method.

Keywords: Stroke, *Random Forest*, Classification, Medical Features, *Machine Learning*

1. PENDAHULUAN

Machine learning adalah cabang ilmu komputer yang mengembangkan algoritma agar komputer dapat belajar dan meningkatkan kinerja berdasarkan data. Konsep utamanya adalah membangun model yang mampu mengenali pola dalam data untuk menghasilkan prediksi atau keputusan (Sidik & Ansawarman, 2022). Kecerdasan Buatan (AI), teknologi yang memungkinkan mesin meniru kemampuan kognitif manusia, turut berperan dalam memproses informasi dan membuat prediksi dalam berbagai bidang (Fahrur Rozi, & Mochammad Bagoes Satria Junianto, 2023). Kecerdasan buatan telah meningkatkan sejumlah industri secara signifikan, termasuk industri di bidang kesehatan (Anggraini, Y., Mohamad Indra, M. Khoirusofi, Ibnu Nur Azis, & Perani Rosyani, 2023). Salah satu aplikasi penting AI dalam kesehatan adalah prediksi dan diagnosis penyakit, khususnya penyakit dengan tingkat morbiditas dan mortalitas yang tinggi, seperti stroke.

Stroke adalah penyakit otak yang menyebabkan gangguan fungsi saraf akibat gangguan peredaran darah otak non-traumatik dan menjadi salah satu penyebab utama kecacatan serta kematian. Tingginya angka kejadian stroke menunjukkan pentingnya upaya untuk mengidentifikasi faktor risiko dan memprediksi potensi stroke sejak dini. Namun, variasi karakteristik klinis pasien menyulitkan identifikasi faktor utama risiko stroke, sehingga dibutuhkan metode klasifikasi yang andal untuk membantu pengambilan keputusan medis.

Random Forest adalah algoritma *machine learning* yang efektif dalam klasifikasi data kompleks, termasuk data medis, karena menggabungkan banyak pohon keputusan untuk menghasilkan prediksi akurat (Fadli, M., & Saputra, R. A, 2023). Dalam hal ini, teknologi informasi dapat memberikan solusi untuk mengatasi tantangan medis tersebut. Salah satu solusinya adalah dengan menggunakan algoritma random forest untuk mengklasifikasi penyakit stroke berdasarkan

fitur medis (Maulana Iqbal Baihaqi, Darmawati, Ari Syaripudin, Fajar Agung Nugroho, 2023). Metode ini juga memungkinkan penyesuaian parameter seperti jumlah pohon untuk mencapai kinerja optimal. Banyak penelitian menunjukkan bahwa *Random Forest* mampu mendukung diagnosis dini stroke dengan akurasi tinggi (S. Rahman, D. Siregar, R. B. Y. Syah, H. Setiawan, A. E. Maulana and Hamsiah, 2023).

Tugas ini menganalisis penerapan algoritma *Random Forest* dalam klasifikasi risiko stroke berdasarkan data fitur medis, melalui studi komparatif terhadap lima jurnal ilmiah terpilih. Analisis difokuskan pada perbandingan performa *Random Forest* yang dilaporkan dalam masing-masing jurnal, serta metode yang digunakan, termasuk pemilihan fitur, teknik *preprocessing data*, dan pengaturan parameter. Dengan membandingkan temuan dari kelima jurnal tersebut, tugas ini bertujuan untuk memberikan gambaran tentang variasi implementasi dan hasil prediksi stroke menggunakan *Random Forest*, serta mengidentifikasi potensi area penelitian lanjutan berdasarkan kesenjangan yang ditemukan.

2. METODOLOGI PENELITIAN

2.1 Pengumpulan Data

Pencarian literatur dilakukan melalui studi literatur *review* dengan menggunakan *database* ilmiah. Kami mencari data tersebut pada *google scholar*, e-jurnal universitas, dan lainnya. Setelah pencarian awal, judul dan abstrak disaring untuk mengidentifikasi artikel yang relevan. Artikel terpilih kemudian dibaca penuh untuk memastikan kesesuaian dengan kriteria inklusi. Lima jurnal yang paling relevan dan berkualitas dipilih untuk *review* mendalam.

2.2 Pemrosesan Data

Informasi yang relevan diekstrak dari jurnal yang terpilih. Data yang diekstrak kemudian ditabulasikan untuk memudahkan perbandingan dan analisis komparatif antar studi.

2.3 Hasil dan Analisis

Data yang telah diproses dianalisis secara deskriptif dan komparatif. Hasil analisis akan disajikan dalam bentuk naratif yang terstruktur dan didukung oleh tabel yang relevan untuk perbandingan antar studi.

3. ANALISA DAN PEMBAHASAN

Berikut hasil dari kegiatan penelitian yang sudah dilakukan

3.1 Analisa

Di dalam *tabel* ini tercantum beberapa artikel yang telah dianalisis untuk dilakukan *studi literature review* (SLR). Analisa ini bertujuan untuk mengevaluasi efektivitas, akurasi, dan performa algoritma *Random Forest* dalam mengklasifikasikan penyakit stroke. Dengan menggunakan pendekatan SLR, penelitian ini mengidentifikasi dan menyaring berbagai studi yang telah dilakukan terkait penggunaan *Random Forest* dalam klasifikasi stroke. Algoritma ini dipilih karena kemampuannya dalam menangani dataset besar dan kompleks, serta menghasilkan akurasi yang tinggi dalam klasifikasi data medis. Melalui analisis yang komprehensif, diharapkan dapat diperoleh pemahaman yang mendalam mengenai keandalan *Random Forest* dalam mendukung deteksi dini penyakit stroke serta memberikan kontribusi pada peningkatan kualitas prediksi dalam bidang kesehatan.

Tabel 1. Hasil Pengumpulan dan Pemrosesan

No	Nama Peneliti dan Tahun	Metode yang dibahas	Tujuan Penelitian	Hasil yang di dapat
1.	Iskandar, N. A., Ernawati, I., Widiastiwi, Y. (2022)	Metode <i>Random Forest</i> untuk klasifikasi diagnosis penyakit	Bertujuan untuk mengetahui seberapa baik model <i>Random</i>	Metode <i>Random Forest</i> dapat digunakan untuk mendiagnosis stroke, namun memiliki

		<p>stroke. Evaluasi model dilakukan menggunakan Confusion Matrix, ROC Curve, serta metrik seperti accuracy, sensitivity, specificity, precision, dan F-measure.</p>	<p><i>Forest</i> dapat melakukan klasifikasi penyakit stroke berdasarkan <i>dataset</i> yang tersedia, dengan mengevaluasi hasil klasifikasi menggunakan berbagai metrik. Selain itu, penelitian ini juga bertujuan untuk mencari jumlah pohon optimal yang memberikan hasil klasifikasi terbaik, serta menganalisis dampak dari jumlah pohon yang berbeda pada kinerja model.</p>	<p>keterbatasan. Model dengan 90 pohon menghasilkan akurasi 95,2%, <i>specificity</i> 99,8%, dan <i>precision</i> 66,7%, namun <i>sensitivity</i> sangat rendah (4,1%), sehingga kesulitan mendeteksi pasien stroke. Meskipun ROC AUC mencapai 0,8048, rendahnya <i>sensitivity</i> menjadikan model kurang optimal untuk prediksi stroke. Penelitian menunjukkan bahwa ketidakseimbangan <i>dataset</i> memengaruhi performa, sehingga teknik <i>oversampling</i> atau model lain lebih disarankan.</p>
2.	<p>Sari, W. J., Melyani, N. A., Arrazak, T., Anahar, M. A, B., Adini, E., Al-Sawaff, Z. H., Manickam, S. (2024)</p>	<p>- Random Forest - Support Vector Machine (SVM) - Neural Network (NN).</p>	<p>Membandingkan efektivitas tiga algoritma <i>machine learning</i>—<i>Neural Network (NN)</i>, <i>Support Vector Machine (SVM)</i>, dan <i>Random Forest</i>—dalam mengklasifikasikan penyakit stroke menggunakan <i>dataset</i> Kaggle untuk menentukan algoritma terbaik berdasarkan akurasi, presisi, dan <i>recall</i>, serta memberikan wawasan praktis dalam pemilihan model untuk perawatan kesehatan.</p>	<p>Hasilnya, <i>Random Forest</i> unggul dengan akurasi tertinggi (98,58%) dan presisi terbaik (98,65%), diikuti oleh ANN (akurasi 95,72%, presisi 95,15%). SVM, meski memiliki akurasi terendah (94,11%), unggul dalam <i>recall</i> (99,41%). Secara keseluruhan, <i>Random Forest</i> dianggap model terbaik karena keseimbangan akurasi dan presisi, meskipun SVM lebih baik dalam <i>recall</i>. Pemilihan model harus mempertimbangkan <i>trade-off</i> antara metrik dan faktor lain seperti kompleksitas komputasi.</p>
3.	<p>Fernández-Lozano, C., Hervella, P., MatoñAbad, V., Rodríguez-Yáñez, M.,</p>	<p>Metode Random Forest (RF) untuk mempertimbangkan variabel klinis, biokimia, dan neuroimaging.</p>	<p>Penelitian ini bertujuan mengembangkan model prediktif menggunakan <i>Random Forest</i> untuk memprediksi</p>	<p>Hasilnya menunjukkan bahwa model <i>Random Forest (RF)</i> efektif memprediksi mortalitas dan morbiditas pasien stroke. Dari 6022 pasien, 4922 mengalami stroke</p>

	Suárez Garaboa, S., dkk. (2021)		mortalitas dan morbiditas pasien stroke 3 bulan setelah dirawat. Fokusnya adalah mengidentifikasi faktor klinis, biokimia, dan <i>neuroimaging</i> yang terkait dengan hasil jangka panjang pada pasien stroke iskemik (IS) dan perdarahan intraserebral non-traumatik (ICH).	iskemik (IS) dan 1100 mengalami perdarahan intraserebral (ICH). Pada pasien IS, 47,5% mengalami hasil buruk, sementara pada pasien ICH, 58,6%. Model RF memberikan AUC 0,904 ± 0,025 dan akurasi 0,825 untuk IS + ICH, namun prediksi untuk pasien ICH lebih bervariasi, dengan AUC antara 0,9837 dan 0,7104. Secara keseluruhan, RF menunjukkan potensi besar meski lebih menantang pada pasien ICH.
4.	Putri, M., Erlin (2024)	Algoritma <i>Random Forest</i>	Mengembangkan model prediksi risiko stroke menggunakan algoritma Random Forest. Dengan tingginya angka kasus stroke, penelitian ini diharapkan dapat mendukung deteksi dini dan pencegahan stroke. Evaluasi model dilakukan dengan matriks konfusi untuk menilai akurasi dan efektivitasnya dalam klasifikasi medis, mendukung pencegahan stroke melalui <i>machine learning</i> .	Algoritma <i>Random Forest</i> memiliki kinerja sangat baik. Model berhasil mengidentifikasi 933 kasus stroke dengan benar (<i>True Positive</i>) dan hanya 11 kesalahan prediksi (<i>False Positive</i>), tanpa kesalahan <i>False Negative</i> . Hasil AUC 1.00 dari kurva ROC menunjukkan kemampuan model yang hampir sempurna dalam membedakan kelas positif dan negatif, menjadikannya andal untuk pencegahan dan diagnosis dini stroke.
5.	Banjar, M. F., Irawati, Umar, F., Hayati L. N. (2022)	<i>Random Forest</i> dan <i>Synthetic Minority Oversampling Technique</i> (SMOTE).	Mengembangkan model prediksi akurat untuk mengklasifikasikan risiko stroke pada pasien menggunakan algoritma <i>Random Forest</i> . Dengan meningkatnya	Hasil penelitian menunjukkan bahwa model <i>Random Forest</i> dengan 100 pohon memberikan kinerja terbaik dengan akurasi 86,82%. Namun, meskipun akurasinya tinggi, metrik lain seperti presisi (15,76%), <i>recall</i>

			<p>kasus stroke global, penelitian ini menganalisis variabel kunci yang mempengaruhi klasifikasi untuk mendukung diagnosis dini dan pencegahan. Model diuji dengan berbagai jumlah pohon keputusan (50, 100, 200, dan 500 pohon) serta menggunakan teknik SMOTE untuk mengatasi ketidakseimbangan data.</p>	<p>(38,15%), dan F1-score (22,30%) menunjukkan hasil yang lebih rendah, mengindikasikan ketidakseimbangan antara presisi dan recall dalam mengidentifikasi kasus stroke.</p>
--	--	--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

3.2 Pembahasan

Berdasarkan hasil tinjauan literatur ini memberikan pemahaman yang komprehensif tentang metode *Random Forest* (Dzaky Nafis Alfarizi, Agung Pangestu, R., Dimas Aditya, Muhammad Adi Setiawan, & Rosyani, P., 2023), penulis mendapatkan algoritma *Random Forest* menunjukkan potensi yang besar dalam klasifikasi penyakit stroke, namun efektivitasnya bervariasi tergantung pada kondisi dataset dan pendekatan yang digunakan. Secara umum, *Random Forest* berhasil mencapai tingkat akurasi yang tinggi dalam beberapa penelitian.

4. KESIMPULAN

4.1 Kesimpulan

Berdasarkan analisis yang dilakukan terhadap lima jurnal ilmiah terpilih, dapat disimpulkan bahwa algoritma *Random Forest* merupakan metode yang sangat efektif dan andal dalam klasifikasi risiko stroke berdasarkan data fitur medis. Hasil komparatif menunjukkan bahwa *Random Forest* memiliki kinerja yang cukup konsisten dengan akurasi yang tinggi dalam berbagai studi yang dianalisis. Keberhasilan algoritma ini terletak pada kemampuannya untuk menangani data yang kompleks dan bervariasi, serta kemampuannya dalam mengidentifikasi pola yang mendasari faktor risiko stroke.

Beberapa faktor yang mempengaruhi kinerja algoritma *Random Forest*, seperti pemilihan fitur, teknik preprocessing data, serta pengaturan parameter, ditemukan bervariasi di tiap studi. Meskipun demikian, pemilihan fitur yang tepat dan penggunaan teknik preprocessing yang sesuai seperti normalisasi dan penanganan missing values terbukti meningkatkan akurasi model secara signifikan.

Secara keseluruhan, *Random Forest* terbukti dapat mendukung diagnosis dini stroke dengan akurasi yang tinggi, namun terdapat beberapa area yang masih bisa ditingkatkan, seperti pemilihan fitur yang lebih optimal, pengujian model dengan dataset yang lebih besar dan beragam, serta eksplorasi parameter yang lebih mendalam. Penelitian lanjutan dapat difokuskan pada pengembangan teknik feature selection dan optimasi hyperparameter untuk meningkatkan performa model, serta penggunaan data yang lebih bervariasi untuk meningkatkan generalisasi hasil prediksi.

4.1 Saran

Untuk meningkatkan penerapan *Random Forest* dalam prediksi risiko stroke meliputi: optimasi pemilihan fitur dengan teknik seleksi fitur yang lebih canggih, peningkatan preprocessing data untuk menangani data hilang dan ketidakseimbangan kelas, serta pengujian dengan dataset yang lebih besar dan beragam. Selain itu, pengaturan *hyperparameter* yang lebih baik, integrasi dengan metode lain, dan pengembangan sistem prediksi yang lebih interaktif untuk tenaga medis juga sangat dianjurkan. Kolaborasi antara ilmuwan data, sistem pakar dan tenaga medis akan mendukung pengembangan model yang lebih relevan dan efektif.

REFERENCES

- Sidik, A. D., & Ansawarman, A. (2022). Prediksi Jumlah Kendaraan Bermotor Menggunakan Machine Learning. *Formosa Journal of Multidisciplinary Research*, 1(3), 559–568. <https://doi.org/10.55927/fjmr.v1i3.745>
- Fahrur Rozi, & Mochammad Bagoes Satria Junianto. (2023). Penerapan Machine Learning Untuk Prediksi Harga Saham PT.Telekomunikasi Indonesia Tbk Menggunakan Algoritma K-Nearest Neighbors. *Jurnal Informatika Multi*, 1(1), 18–24. Retrieved from <https://jurnal.publikasitecno.id/index.php/multi/article/view/4>
- Anggraini, Y., Mohamad Indra, M. Khoirusofi, Ibnu Nur Azis, & Perani Rosyani. (2023). Systematic Literature Review: Sistem Pakar Diagnosa Penyakit Gigi Menggunakan Metode Forward Chaining. *BINER : Jurnal Ilmu Komputer, Teknik Dan Multimedia*, 1(1), 1–7. Retrieved from <https://journal.mediapublikasi.id/index.php/Biner/article/view/2520>
- Fadli, M., & Saputra, R. A. (2023). Klasifikasi Dan Evaluasi Performa Model Random Forest Untuk Prediksi Stroke Classification And Evaluation Of Performance Models Random Forest For Stroke Prediction. 12(02), 72–80.)
- Maulana Iqbal Baihaqi, Darmawati, Ari Syaripudin, Fajar Agung Nugroho. (2023). Implementasi Algoritma Random Forest Pada Prediksi Harga Saham Berdasarkan Data Historis. *Jubitek: Jurnal Big Data Dan Teknologi Informasi*, 1(3), 42–51. Retrieved from <http://jurnal.jcosbida.com/index.php/jubitek/article/view/244>
- S. Rahman, D. Siregar, R. B. Y. Syah, H. Setiawan, A. E. Maulana and Hamsiah, "The Effective Breast Cancer Classification with the Random Forest Algorithm," 2023 *International Conference of Computer Science and Information Technology (ICOSNIKOM)*, Binjia, Indonesia, 2023, pp. 1-5, doi: 10.1109/ICOSNIKOM60230.2023.10364529.\
- Iskandar, N. A., Ernawati, I., & Widiastiwi, Y. (2022). Klasifikasi diagnosis penyakit stroke dengan menggunakan metode Random Forest. *SENAMIKA: Seminar Nasional Mahasiswa Informatika UPN "Veteran" Jakarta*, 3(2), 1–10. <https://conference.upnvj.ac.id/index.php/senamika/article/view/2190>
- Fernández-Lozano, C., Hervella, P., MatoñAbad, V., Rodríguez-Yáñez, M., SuárezGaraboa, S., & otros. (2021). Random forest-based prediction of stroke outcome. *Random Forest-Based Prediction of Stroke Outcome*, 11, 1–12.
- Sari, W. J., Melyani, N. A., Arrazak, F., Anahar, M. A. B., Addini, E., Al-Sawaff, Z. H., & Manickam, S. (2024). Performance Comparison of Random Forest, Support Vector Machine and Neural Network in Health Classification of Stroke Patients, 2(1), 34–43 : <https://journal.irpi.or.id/index.php/predatecs>
- Putri, M., & Erlin. (2024). Prediksi penyakit stroke menggunakan machine learning dengan algoritma Random Forest. *Jurnal Infomedia: Teknik Informatika, Multimedia & Jaringan*, 9(2), 1–10. <https://e-jurnal.pnl.ac.id/infomedia/article/view/5199>
- Banjar, M., Irawati, I., Umar, F., & Hayati, L. (2022). Analysis of Stroke Classification Using Random Forest Method. *ILKOM Jurnal Ilmiah*, 14(3), 186-193. doi:<https://doi.org/10.33096/ilkom.v14i3.1252.186-193>
- Dzaky Nafis Alfarizi, Agung Pangestu, R., Dimas Aditya, Muhammad Adi Setiawan, & Rosyani, P. (2023). Penggunaan Metode YOLO Pada Deteksi Objek: Sebuah Tinjauan Literatur Sistematis. *AI Dan SPK : Jurnal Artificial Intelligent Dan Sistem Penunjang Keputusan*, 1(1), 54–63. Retrieved from <https://jurnalmahasiswa.com/index.php/aidanspk/article/view/144>