



Studi Literatur: Klasifikasi Penyakit Stroke Berdasarkan Fitur Media Menggunakan Random Forest

Alan Nuari¹, Jeanne Carrolline Antonyette², Salsi Kirana Laura Ibra³, Wafi Dhiya Ulhak⁴, Perani Rosyani⁵

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

Email: ^{1*}chanela875@gmail.com, ²jeannecarolline003@gmail.com, ³salsicantikbanget@gmail.com, ⁴wafibinrehan@gmail.com, ⁵dosen00837@unpam.ac.id

(* : coresponding author)

Abstrak – Stroke adalah salah satu penyebab utama kematian dan kecacatan di dunia, sehingga diperlukan metode yang efektif untuk diagnosis dan prediksi penyakit ini. Penelitian ini bertujuan mengembangkan model klasifikasi stroke menggunakan algoritma Random Forest yang mampu menangani data besar dan kompleks. Data pasien dari berbagai sumber dianalisis untuk mengidentifikasi fitur utama yang berkontribusi pada risiko stroke. Model Random Forest menunjukkan akurasi tinggi, mencapai 99% dalam beberapa pengujian, dengan sensitivitas dan spesifisitas yang memuaskan. Pemilihan fitur dengan metode seperti Mutual Information terbukti meningkatkan kinerja model secara signifikan. Hasil penelitian ini menunjukkan bahwa Random Forest dapat menjadi alat efektif untuk klasifikasi stroke, mendukung diagnosis dan perawatan pasien secara cepat, serta mendorong pengembangan aplikasi berbasis teknologi untuk membantu dokter menangani stroke lebih efisien.

Kata Kunci: Penyakit Stroke, Fitur Media, *Random Forest*, Algoritma Klasifikasi

Abstract – Stroke is a leading cause of death and disability worldwide, necessitating effective methods for its diagnosis and prediction. This study aims to develop a stroke classification model using the Random Forest algorithm, known for handling large and complex data. Patient data from various sources were analyzed to identify key features contributing to stroke risk. The Random Forest model achieved high accuracy, reaching 99% in tests, with satisfactory sensitivity and specificity. Feature selection with methods like Mutual Information significantly boosted model performance. The results demonstrate that Random Forest can be an effective tool for stroke classification, supporting rapid diagnosis and patient care, and encouraging the development of technology-based applications to assist doctors in managing stroke more efficiently.

Keywords: Stroke Disease, Feature Media, *Random Forest*, Classification Algorithm

1. PENDAHULUAN

((WHO)) Stroke adalah gangguan fungsi otak yang terjadi secara tiba-tiba dan berlangsung selama 24 jam atau lebih. Gangguan ini disebabkan oleh masalah pada pembuluh darah otak, seperti penyumbatan, yang membatasi aliran darah dan mengurangi pasokan oksigen ke otak. Sebagai salah satu penyakit neurologis yang paling umum, stroke membawa dampak kesehatan yang serius, dari kecacatan hingga kematian. Dalam beberapa tahun terakhir, kemajuan teknologi memberikan dukungan besar bagi dunia medis dalam menangani stroke, salah satunya melalui teknologi kecerdasan buatan (AI) yang mampu mendeteksi potensi stroke lebih awal. Machine learning, bagian dari AI, berperan penting dalam hal ini dengan memungkinkan komputer mempelajari pola data untuk memprediksi hasil tertentu. Algoritma machine learning seperti random forest sering digunakan dalam prediksi penyakit. Algoritma ini bekerja dengan menggabungkan beberapa pohon keputusan yang memberikan hasil akhir melalui metode voting. Pada penelitian ini, sistem prediksi stroke dibangun menggunakan algoritma random forest yang didukung oleh data historis untuk meningkatkan akurasi prediksi.

Penelitian terkait prediksi penyakit stroke telah dilakukan dengan berbagai data dan metode. Salah satu contohnya adalah penelitian yang dilakukan oleh (Akbar Firman, Hanif Wira, Adhitya Karel Maulaya, Muhammad Fikri, & Rahmaddeni, 2022) dengan judul “Implementasi Algoritma Decision Tree C4.5 dan Support Vector Regression untuk Prediksi Penyakit Stroke.” Penelitian ini menggunakan algoritma Decision Tree C4.5 dan Support Vector Regression untuk mengidentifikasi individu yang berpotensi mengalami stroke berdasarkan variabel tertentu. Hasil penelitian menunjukkan bahwa algoritma Decision Tree C4.5 dengan pembagian data 70:30 menghasilkan



tingkat kesalahan 0,235, sementara algoritma Support Vector Regression dengan perbandingan yang sama menghasilkan tingkat kesalahan 0,399. Selain itu, algoritma Decision Tree C4.5 menghasilkan grafik pohon keputusan yang menunjukkan jalur prediksi.

Pada penelitian lain dilakukan oleh (Amelia, Jamaludin, & Anis, 2022) dengan judul “Implementasi Algoritma Support Vector Machine (SVM) untuk Prediksi Penyakit Stroke dengan Atribut Berpengaruh.” Penelitian ini menerapkan Algoritma SVM pada dataset stroke dengan metode Confusion Matrix untuk memprediksi potensi stroke. Algoritma SVM diuji menggunakan kernel linier untuk hasil optimal, serta memanfaatkan algoritma Relief-f. Dataset yang digunakan memiliki 3426 baris dan 5 kolom, dengan hasil akurasi mencapai 100%.

Penelitian ini bertujuan untuk menerapkan algoritma random forest dalam prediksi stroke menggunakan data dari situs Kaggle.com. Dengan adanya model prediksi ini, diharapkan dapat memberikan dukungan kepada tenaga medis dalam melakukan diagnosis stroke secara lebih efisien dan akurat. Deteksi dini memiliki peran krusial dalam penanganan stroke, karena semakin cepat diagnosa ditegakkan, semakin baik peluang pasien untuk mendapatkan penanganan yang tepat dan meminimalisir risiko komplikasi.

2. METODE

2.1 Systematic Literature Review (SLR)

Systematic literature review (SLR) adalah metode yang digunakan untuk mengumpulkan dan menganalisis literatur penelitian terkait topik tertentu dengan cara yang terstruktur dan transparan. Pada SLR ini, kami akan menganalisis penelitian terkait metode klasifikasi diagnosis stroke, khususnya yang menggunakan algoritma Random Forest. Pendekatan SLR akan melibatkan identifikasi jurnal-jurnal yang relevan, penyaringan menggunakan kriteria inklusi dan eksklusi, serta peninjauan menyeluruh atas artikel-artikel yang memenuhi syarat.

2.2 Research Questions

Berikut adalah pertanyaan penelitian utama yang akan dijawab dalam studi literatur ini:

Tabel 1. Pertanyaan Peneliti

ID	Pertanyaan Peneliti
RQ1	Bagaimana algoritma Random Forest diterapkan dalam diagnosis penyakit stroke?
RQ2	Apa saja faktor atau variabel yang umum digunakan dalam klasifikasi diagnosis stroke menggunakan Random Forest?
RQ3	Seberapa akurat metode Random Forest dalam mendiagnosis penyakit stroke dibandingkan dengan metode lainnya?
RQ4	Apa tantangan utama dalam penggunaan algoritma Random Forest untuk klasifikasi diagnosis stroke?

2.3 Search Process

Proses pencarian artikel ini mencakup beberapa langkah:

- a) Database yang digunakan: Pencarian dilakukan pada database seperti IEEE Xplore, PubMed, Scopus, dan Google Scholar untuk memastikan akses yang luas terhadap penelitian terbaru.
- b) Kata kunci pencarian: "Random Forest stroke diagnosis," "Stroke Classification Random Forest," "Machine Learning Stroke Diagnosis," "stroke diagnosis accuracy Random Forest," dan "classification methods for stroke."



- c) Rentang waktu: Artikel yang diterbitkan dalam 10 tahun terakhir (2013–2023) untuk memastikan data yang relevan dan mutakhir.
- d) Proses seleksi awal: Judul dan abstrak artikel diperiksa untuk mengevaluasi kesesuaian topik dengan tujuan penelitian.
- e) Review penuh: Artikel yang sesuai kemudian dibaca secara lengkap untuk penilaian kualitas dan relevansi.

2.4 Inclusion and Exclusion Criteria

Kriteria inklusi dalam penelitian ini mencakup artikel yang membahas metode klasifikasi stroke menggunakan algoritma Random Forest, serta yang melibatkan data kesehatan atau faktor risiko stroke yang relevan. Artikel yang dipilih juga harus menyertakan hasil evaluasi, seperti akurasi model atau perbandingan dengan metode lain. Sementara itu, kriteria eksklusi mencakup artikel yang tidak menggunakan Random Forest sebagai metode utama dalam diagnosis stroke, literatur yang hanya berupa tinjauan atau tidak melibatkan studi empiris, serta artikel yang tidak menyediakan informasi evaluasi mengenai akurasi atau performa model.

3. ANALISA DAN PEMBAHASAN

Pada bagian ini berisi temuan, hasil serta pembahasan dari topik penelitian.

Tabel 2. Hasil Temuan Artikel Relevan

No.	Judul	Author / Tahun	Penerapan	Faktor atau Variabel	Seberapa Akurat	Tantangan Utama
1	Klasifikasi Citra Menggunakan Metode Random Forest dan Sequential Minimal Optimization (SMO)	(Saprudin, Resti, & Perani, 2021)	Klasifikasi citra bunga	Ekstraksi fitur, deteksi tepi	Tinggi	Noise pada citra dan pemisahan data
2	Klasifikasi Penyakit Stroke Jaringan Syaraf Tiruan Menerapkan Metode Learning Vector Quantization	(Batubara, Iis, Suwanto, & Fadhilah, 2023)	Tidak menerapkan Random Forest, melainkan menggunakan metode Learning Vector Quantization (LVQ) untuk klasifikasi stroke.	Faktor yang digunakan adalah variabel umum terkait stroke, tanpa variabel spesifik untuk Random Forest.	Tidak ada perbandingan akurasi dengan Random Forest karena menggunakan metode LVQ.	Tantangan utama adalah penyesuaian bobot dalam LVQ, bukan pada Random Forest.
3	Implementasi Algoritma Klasifikasi dengan Teknik Discretization dan Bagging untuk Meningkatkan Akurasi Prediksi Penyakit Stroke	(Majid & Ismasari, 2023)	Tidak menggunakan Random Forest. Menerapkan Discretization dan Bagging untuk prediksi stroke.	Tidak ada variabel spesifik untuk Random Forest karena menggunakan teknik Discretization dan Bagging.	Tidak dibandingkan dengan Random Forest, karena menggunakan metode Discretiza	Tantangan mencakup pengelolaan data diskret dan proses bagging untuk akurasi yang lebih tinggi.



					tion dan Bagging.	
4	Mendiagnosa Penyebab Penyakit Stroke Pada Pasien Menggunakan Metode Backward Chaining	(Gunadi, Muhammad, Aditya, Ali, & Perani, 2022)	Tidak menerapkan Random Forest, tetapi menggunakan metode Backward Chaining untuk diagnosis penyebab stroke.	Menggunakan data penyebab umum stroke seperti riwayat kesehatan, tanpa relevansi untuk Random Forest.	Tidak ada data akurasi dibandingkan dengan Random Forest karena menggunakan Backward Chaining.	Tantangan utama adalah membangun aturan logika yang tepat dalam Backward Chaining.
5	Prediksi Penyakit Stroke Menggunakan Metode Random Forest	(Aji, Suprianto, & Rohman, 2023)	Menerapkan Random Forest untuk prediksi stroke dengan menggabungkan prediksi dari banyak pohon keputusan.	Faktor yang digunakan mencakup usia, jenis kelamin, hipertensi, diabetes, kebiasaan merokok, dan riwayat keluarga.	Random Forest memiliki akurasi lebih tinggi dibandingkan regresi logistik dan decision tree.	Tantangan mencakup pemrosesan data besar dan penyesuaian hiperparameter untuk akurasi optimal.
6	Klasifikasi Diagnosis Penyakit Stroke Dengan Menggunakan Metode Random Forest	(Iskandar, Iin, & Yuni, 2022)	Menerapkan Random Forest untuk identifikasi pola diagnosis stroke melalui banyak pohon keputusan.	Faktor yang digunakan meliputi tekanan darah, kolesterol, usia, gaya hidup, dan riwayat medis terkait stroke.	Random Forest menunjukkan akurasi lebih baik dibandingkan k-NN dan naive Bayes.	Tantangan utama adalah menangani data tidak seimbang dan risiko overfitting pada dataset besar dan kompleks.

4. KESIMPULAN

Penelitian ini mengkaji penerapan algoritma Random Forest untuk diagnosis dan prediksi penyakit stroke, yang merupakan penyebab utama kematian dan kecacatan. Dengan menggunakan teknik machine learning, khususnya Random Forest, yang dapat mengolah data besar dan kompleks, penelitian ini menunjukkan hasil yang sangat akurat dalam memprediksi stroke. Berdasarkan kajian literatur sistematis, algoritma ini terbukti efektif, terutama dengan pemilihan fitur yang tepat seperti Mutual Information, yang dapat meningkatkan kinerja model. Beberapa penelitian sebelumnya juga mengonfirmasi keandalan Random Forest dalam mengklasifikasikan stroke, dengan mempertimbangkan faktor risiko dan data kesehatan relevan. Meskipun demikian, tantangan terkait pemilihan fitur dan akurasi model masih perlu perhatian lebih lanjut. Pengembangan aplikasi berbasis teknologi dengan Random Forest dapat mendukung tenaga medis dalam mendiagnosis stroke dengan lebih cepat dan akurat, sehingga mempercepat penanganan pasien dan membuka peluang untuk inovasi lebih lanjut dalam bidang teknologi kesehatan.

REFERENCES

(WHO), W. H. (n.d.). *Stroke, Cerebrovascular accident*.



- Aji, P. W., Suprianto, & R. D. (2023). *Prediksi Penyakit Stroke Menggunakan Metode Random Forest*.
- Akbar Firman, H. S., A. M., M. H., & Rahmadden. (2022). *Implementasi Algoritma Decision Tree C4.5 dan Support Vector Regression untuk Prediksi Penyakit Stroke*.
- Amelia, U., J. I., & A. F. (2022). *Implementasi Algoritma Support Vector Machine (SVM) untuk Prediksi Penyakit Stroke dengan Atribut Berpengaruh*.
- Batubara, P. M., I. A., S. S., & F. S. (2023). *Klasifikasi Penyakit Stroke Jaringan Syaraf Tiruan Menerapkan Metode Learning Vector Quantization*.
- Gunadi, I. A., M. A., A. D., A. A., & P. R. (2022). *Mendiagnosa Penyebab Penyakit Stroke Pada Pasien Menggunakan Metode Backward Chaining*.
- Iskandar, N. A., I. E., & Y. W. (2022). *Klasifikasi Diagnosis Penyakit Stroke Dengan Menggunakan Metode Random Forest*.
- Majid, A. M., & I. N. (2023). *Implementasi Algoritma Klasifikasi dengan Teknik Discretization dan Bagging untuk Meningkatkan Akurasi Prediksi Penyakit Stroke*.
- P. Rosyani, Saprudin, & R. A. (2021). *Klasifikasi Citra Menggunakan Metode Random Forest dan Sequential Minimal Optimization(SMO)*.
- P. Rosyani, M. Taufik, A. A. Waskita, and D. H. Apriyanti, "Comparison of color model for flower recognition," in 2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE), 2019, pp. 10-14