KOMPUTA : Jurnal Ilmiah Komputer dan Informatika
Vol. 14, No. 2, Oktober 2025, ISSN : 2089-9033, e- ISSN : 2715-7849
DOI: 10.34010/komputa.v14i2.16642

Systematic Literature Review: Penerapan *Machine Learning* dalam Diagnosis dan Prediksi Penyakit Diabetes

Oktavia Putri Handayani[1], Purwono[2], Imam Ahmad Ashari[3], Rian Ardianto[4]
[1,2,4] Program Studi Informatika, Universitas Harapan Bangsa
[3] Program Teknologi Informasi, Universitas Harapan Bangsa
E-mail : purwono@uhb.ac.id

Abstrak

Diabetes mellitus merupakan penyakit kronis dengan prevalensi yang terus meningkat secara global, sehingga menjadi tantangan besar bagi sistem pelayanan kesehatan di seluruh dunia. Teknologi Machine Learning (ML) menawarkan solusi yang menjanjikan dalam diagnosis dan prediksi dini dengan menganalisis data medis yang kompleks secara efisien. Penelitian ini menggunakan metode Systematic Literature Review (SLR) berdasarkan protokol PRISMA untuk menganalisis 15 artikel open access yang dipublikasikan antara tahun 2022 hingga 2025 dari database ScienceDirect. Studi-studi tersebut membahas penerapan berbagai algoritma ML—termasuk Support Vector Machine (SVM), Random Forest (RF), dan Convolutional Neural Network (CNN)—dalam diagnosis diabetes. Tujuan utama penelitian ini adalah mengevaluasi efektivitas, keunggulan, dan keterbatasan masing-masing algoritma dalam konteks klinis. Kajian ini menyoroti tren terbaru, perbandingan kinerja, serta tantangan dalam implementasi model ML untuk diagnosis diabetes. Temuan ini diharapkan dapat memberikan wawasan berharga bagi pengembangan sistem diagnostik berbasis ML yang lebih akurat, efisien, dan aplikatif dalam penanganan diabetes.
Kata kunci : Diabetes; Mechine Learning; Artificial Intelligence; Diagnosis.

Literature Review Study of the Application of Machine Learning in Diagnosis and Prediction of Diabetes Disease

Abstract

Diabetes mellitus is a chronic disease with a growing global prevalence, posing significant challenges for healthcare systems worldwide. Machine Learning (ML) offers promising solutions for early diagnosis and prediction by analyzing complex medical data efficiently. This study adopts a Systematic Literature Review (SLR) method guided by the PRISMA protocol to analyze 15 open-access articles published between 2022 and 2025 from the ScienceDirect database. These studies explore the use of various ML algorithms—including Support Vector Machine (SVM), Random Forest (RF), and Convolutional Neural Network (CNN)—in diagnosing diabetes. The main objective is to evaluate the effectiveness, strengths, and limitations of each algorithm in clinical applications. The review highlights current trends, performance comparisons, and challenges in implementing ML models for diabetes diagnosis. The findings are expected to provide valuable insights for researchers and practitioners aiming to develop more accurate, efficient, and applicable ML-based diagnostic systems for improved diabetes management and early intervention.
Keywords : *Diabetes; Mechine Learning; Artificial Intelligence; Diagnosis.*

1. Pendahuluan

Diabetes mellitus adalah salah satu penyakit kronis yang paling umum di seluruh dunia, yang diperkirakan mencapai 537 juta orang pada tahun 2021 dan diproyeksikan meningkat menjadi 783 juta pada tahun 2045, terutama karena meningkatnya obesitas dan gaya hidup sedentari [1]. Deteksi awal dan pengelolaan yang tepat sangat krusial untuk menghindari komplikasi berat seperti penyakit jantung, gagal ginjal, dan neuropati. Sistem monitoring berbasis big data dan kecerdasan buatan memungkinkan diagnosis dan pengelolaan lebih cepat untuk menghindari komplikasi ini [2]. Seiring berkembangnya teknologi, metode berbasis machine learning (ML) mulai digunakan secara luas untuk meningkatkan akurasi diagnosis

Naskah dikirim: 03-07-2025; direvisi 28-10-2025; diterima: 28-10-2025

serta memprediksi risiko diabetes lebih awal menggunakan algoritma seperti Logistic Regression, SVM, dan lain-lain [3]

Kemajuan teknologi kecerdasan buatan (AI) dan ML telah merevolusi penanganan penyakit kronis, seperti diabetes dengan menyediakan model prediktif dan alat pengambilan keputusan yang efisien [4]. Machine learning mampu menganalisis data dalam jumlah besar untuk mengidentifikasi pola yang kompleks dan tidak mudah dikenali oleh manusia, termasuk dari rekam medis elektronik, data laboratorium, dan data dari perangkat wearable [5]. Model ML ini memprediksi kemungkinan penyakit berdasarkan parameter klinis seperti glukosa darah, tekanan darah, BMI, usia, dan riwayat keluarga dapat memprediksi risiko diabetes dengan akurasi tinggi [6].

Beragam algoritma ML telah digunakan dalam kajian-kajian terbaru untuk memprediksi diabetes, seperti Support Vector Machine (SVM), Random Forest (RF), dan K-Nearest Neighbors (KNN) semua algoritma ini menunjukkan performa klasifikasi yang layak dalam konteks prediksi diabetes [7] [8]. Penelitian oleh mengungkapkan bahwa algoritma Random Forest memperoleh akurasi maksimal sebesar 75,25% dalam memprediksi diabetes dengan menggunakan dataset Diabetes Pima Indians [9]. Studi lain juga melaporkan angka yang bervariasi dari 76% hingga lebih dari 91% tergantung metode praproses dan evaluasi [10] [11]. Selain itu, metode deep learning seperti Convolutional Neural Network (CNN) dan Long Short-Term Memory (LSTM) juga mulai diterapkan, bahkan dalam pengembangan model hibrida untuk populasi Pima Indian, menunjukkan potensi dalam mengolah data citra maupun time-series [12]

Meskipun hasilnya menjanjikan, Beberapa tantangan utama yang diidentifikasi dalam penggunaan ML untuk diabetes meliputi kualitas data yang rendah, ketidakseimbangan kelas (class imbalance), serta keterbatasan dalam interpretabilitas model, yang membatasi penerapan klinis secara luas [13] [14]. Penelitian oleh menunjukkan bahwa meskipun metode ensemble dalam ML dapat meraih akurasi sampai 95,6%, tetapi mereka juga menyoroti tantangan besar dalam memahami model secara klinis, khususnya untuk digunakan oleh tenaga medis [15]. Oleh karena itu, penelitian ini bertujuan untuk mengklasifikasikan berbagai metode ML yang telah digunakan dalam diagnosis dan prediksi diabetes, serta mengevaluasi kelebihan seperti akurasi tinggi dan kelemahan seperti generalisasi buruk antar populasi dan kurangnya transparansi model [16] [17].

2. Metodologi

2.1 Desain Penelitian

Penelitian ini menerapkan pendekatan Systematic Literature Review (SLR) guna mengidentifikasi, menilai, dan mensintesis penelitian-penelitian sebelumnya yang berhubungan dengan penggunaan machine learning dalam diagnosa penyakit diabetes. Proses ini mengikuti panduan PRISMA (Preferensi Laporan untuk Tinjauan Sistematik dan Meta-Analisis) sebagai acuan untuk meningkatkan transparansi dan reproduksi proses pemilihan literatur.

Tabel berikut menjelaskan tahapan-tahapan dalam melakukan penelitian Systematic Literature Review (SLR) yang diterapkan untuk meninjau penerapan machine learning dalam diagnosis dan prediksi penyakit diabetes. Langkah-langkah dalam gambar mencakup identifikasi masalah, penentuan kriteria inklusi dan eksklusi, pencarian dan seleksi literatur, hingga analisis sistematis dan penarikan kesimpulan. Proses ini penting untuk menjamin bahwa semua artikel yang dianalisis relevan, berkualitas, dan dapat memberikan wawasan yang berarti terhadap topik yang ditinjau.

Gambar 1 Metode Penelitian

Tabel 1 Kriteria Inklusi dan Eksklusi

Kriteria	Inklusi	Eksklusi
Topik Penelitian	Studi yang membahas penerapan machine learning dalam diagnosis atau prediksi diabetes	Studi yang tidak berfokus pada diabetes atau tidak menggunakan machine learning
Sumber Database	Artikel yang tersedia di ScienceDirect	Artikel dari luar ScienceDirect (misalnya, PubMed, Springer, dll.) jika tidak relevan
Kualitas Relevansi	Artikel yang secara langsung menjawab pertanyaan penelitian (RQ)	Artikel yang tidak menjawab RQ atau hanya menyebut ML tanpa penerapan nyata
Metode	Menggunakan atau membandingkan algoritma machine learning	Tidak menggunakan metode ML sama sekali atau hanya teori umum
Tahun Publikasi	2022 hingga 2025	Sebelum tahun 2022

2.2 Pengumpulan Data

Tahapan ini dilakukan dengan mengumpulkan artikel dari database terpercaya seperti ScienceDirect, menggunakan kata kunci pencarian: "implementation" AND "machine learning" AND "diagnosis" AND "diabetes". Kriteria inklusi meliputi: tahun publikasi 2022–2025, artikel berbahasa Inggris, full text tersedia (open access), dan relevan dengan topik. Artikel yang tidak memenuhi kriteria tersebut akan dikecualikan

Daftar pertanyaan penelitian disusun untuk memahami metode, keuntungan, dan tantangan dalam penerapan machine learning untuk diagnosis dan prediksi penyakit diabetes. Berikut ini adalah pertanyaan penelitian dalam Tinjauan Pustaka ini:

a. R Q1. Bagaimana algoritma machine learning dapat meningkatkan akurasi diagnosis dan prediksi dini diabetes berdasarkan data klinis?
b. R Q2. Algoritma machine learning apa yang paling umum digunakan dalam studi diagnosis dan prediksi diabetes, dan bagaimana performanya dibandingkan satu sama lain?
c. RQ3. Apa saja tantangan dan keterbatasan umum yang dihadapi dalam penerapan machine learning untuk diagnosis diabetes, serta solusi yang diusulkan oleh peneliti?

Tabel 2 Kriteria Inklusi dan Eksklusi

NO	Penulis dan tahun	Judul	Algoritma ML	Dataset	Akurasi	Kelebihan	Kekurangan
1.	Gabriel Aguilera-Venegas, Amador López-Molina, Gemma Rojo-Martínez, José Luis Galán-García (2023) [18]	Comparing and tuning machine learning algorithms to predict type 2 diabetes mellitus	Membandingkan Decision Tree, RF, kNN & Neural Networks. RF & Neural Networks menunjukkan kinerja terbaik.	Data dari studi kohort nasional (Spanyol).	Nilai spesifik tidak disebutkan, namun RF & Neural Networks paling unggul.	Pendekatan komparatif pada data kohort nasional yang relevan.	Akurasi tidak dicantumkan; generalisasi hasil terbatas karena fokus pada populasi Spanyol.
2.	Gowthami S, R. Venkata Siva Reddy, Mohammed Riyaz Ahmed (2023) [19]	Exploring the effectiveness of machine learning algorithms for early detection	Membandingkan LR, KNN, DT, RF, dan SVM. Random Forest (RF) menunjukk	520 data pasien dari sebuah rumah sakit di Bangladesh.	RF mencapai akurasi 98%, mengungguli model lainnya.	Menyoroti potensi besar RF untuk deteksi dini T2DM dengan akurasi tinggi.	Ukuran dataset yang kecil serta tidak adanya validasi eksternal atau uji klinis,

NO	Penulis dan tahun	Judul	Algoritma ML	Dataset	Akurasi	Kelebihan	Kekurangan
		of Type-2 Diabetes Mellitus	an kinerja terbaik.				yang membatasi generalisasi hasil.
3.	Iqra Nissar, Waseem Ahmad Mir, Tawseef Ayoub Shaikh, Tuba Areen, Mohammad Kashif, Simran Khiani, dan Asif Hussain (2024) [20]	An Intelligent Healthcare System for Automated Diabetes Diagnosis and Prediction using Machine Learning	Membandingkan Random Forest (RF), SVM, Naïve Bayes, dan KNN. Model RF menghasilkan akurasi tertinggi	Pima Indian Diabetes Dataset (PIDD)	RF mencapai akurasi 98.07% (dengan rasio data uji 0.2).	Akurasi tinggi didukung oleh penggunaan seleksi fitur mRMR untuk meningkatkan performa model.	Generalisasi terbatas karena hanya menggunakan satu dataset (PIDD) dan kurangnya validasi klinis.
4.	Gangani Dharmarathne, Thilini N. Jayasinghe, Madhusha Bogahawaththa, D.P.P. Meddage, Upaka Rathnayake (2024) [21]	A novel machine learning approach for diagnosing diabetes with a self-explainable interface	Membandingkan DT, KNN, SVC, dan XGB. XGBoost (XGB) menunjukkan keunggulan.	Sumber dataset tidak dijelaskan secara rinci.	Semua model memiliki akurasi yang baik, dengan XGB sebagai yang terbaik.	Menggabungkan model XGB dengan metode interpretasi SHAP, menghasilkan antarmuka yang dapat menjelaskan prediksinya (self-explainable).	Generalisasi terbatas karena menggunakan satu dataset publik tanpa verifikasi eksternal dan tidak adanya evaluasi klinis.
5.	D. Kaviyaadharshani, M. Nivedhidha, R. Jeyarohini, J. Lece Elizabeth Rani, M. P. Ramkumar, G. S. R. Emil Selvan (2024) [22]	Diagnosing Diabetes using Machine Learning-based Predictive Models	Random Forest, Decision Tree, dan Support Vector Machine	Sumber dataset tidak dijelaskan secara rinci.	Mencapai akurasi sebesar 91.41%.	Pemilihan fitur berhasil meningkatkan relevansi dan kinerja sistem secara keseluruhan.	Tidak secara eksplisit menyebutkan dataset yang digunakan, sehingga menghambat replikasi dan verifikasi hasil.
6.	Mohammad Mihrab Chowdhury, Ragib	An investigation of machine	Random Forest, Support Vector	Behavioral Risk Factor Surveillanc	Nilai akurasi spesifik tidak	Menyediakan kerangka kerja	Tidak mencantumkan nilai akurasi

NO	Penulis dan tahun	Judul	Algoritma ML	Dataset	Akurasi	Kelebihan	Kekurangan
	Shahariar Ayon, Md. Sakhawat Hossain (2024) [23]	learning algorithms and data augmentation techniques for diabetes diagnosis using class imbalanced BRFSS dataset	Machine (SVM), dan K-Nearest Neighbors (KNN)	e System (BRFSS)	disebutkan dalam publikasi.	machine learning yang menyeluruh bagi para profesional kesehatan.	model dan tidak adanya evaluasi klinis, sehingga efektivitas praktisnya belum terbukti.
7.	N. P. Shetty, J. Shetty, V. Hegde, S. D. Dharne, dan M. Kv (2024) [24]	A Machine Learning-Based Clinical Decision Support System For Effective Stratification Of Gestational Diabetes Mellitus And Management Through Ayurveda	Menerapkan berbagai algoritma klasifikasi machine learning untuk prediksi faktor risiko.	Data pasien untuk mengidentifikasi faktor risiko Gestational Diabetes Mellitus (GDM).	Sebagian besar model mencapai akurasi antara 75% hingga 82%.	Menggabungkan deteksi dini menggunakan ML dengan manajemen penyakit melalui pendekatan Ayurveda (gaya hidup dan herbal).	Ringkasan yang tersedia tidak secara spesifik menyebutkan keterbatasan dari metodologi penelitian yang digunakan.
8.	Imran Rehan, Kamran Rehan, Sabiha Sultana, Mujeeb Ur Rehman (2024) [25]	Fingernail Diagnostics: Advancing type II diabetes detection using machine learning algorithms and laser spectroscopy	Menggabungkan teknik spektroskopi LIBS (laser-induced breakdown spectroscopy) dengan machine learning.	Sampel kuku jari dari individu dengan dan tanpa diabetes.	Model hibrida yang dikembangkan mencapai akurasi 95%.	Menawarkan metode diagnostik non-invasif dan cepat yang sangat cocok untuk skrining massal.	Ukuran sampel yang kecil dapat membatasi generalisasi, serta tidak adanya validasi eksternal atau uji klinis.
9.	Bhuvaneswari Amma N.G. (2024) [26]	En-RfRsK: An ensemble machine learning	Menggunakan teknik ensemble En-RfRsK yang menggabun	Pima Indians Diabetes Dataset.	Model ensemble En-RfRsK mencapai	Pendekatan ensemble menunjukkan peningkatan presisi	Meskipun akurasi tinggi, studi ini tidak mencantu

NO	Penulis dan tahun	Judul	Algoritma ML	Dataset	Akurasi	Kelebihan	Kekurangan
		technique for prognostication of diabetes mellitus	gkan Random Forest (RF), Radial Support Vector Machine (R-SVM), dan K-NN.		akurasi 91%.	dengan memanfaatkan keunggulan masing-masing algoritma.	mkan adanya validasi eksternal atau uji klinis nyata.
10.	Gangani Dharmarathne, Madhusha Bogahawaththa, Marion McAfee, Upaka Rathnayake, D.P.P. Meddage (2024) [27]	On the diagnosis of chronic kidney disease using a machine learning-based interface with explainable artificial intelligence	Menggunakan Extreme Gradient Boosting (XGBoost) yang didukung oleh Shapley Additive Explanations (SHAP) dan Partial Dependency Plots (PDP).	Sumber dataset tidak dijelaskan secara rinci.	XGBoost menunjukkan akurasi tertinggi, meskipun nilai akurasi spesifiknya tidak disebutkan.	Menggabungkan machine learning dengan AI yang dapat dijelaskan (explainable AI) untuk meningkatkan kepercayaan pengguna dan mendukung pengambilan keputusan klinis.	Nilai akurasi model tidak dicantumkan dan tidak adanya uji klinis membuat penerapan praktisnya belum terverifikasi.
11.	Pawan Kumar Patidar, Savita Shiwani, Shruti Garg (2025) [28]	Prediction of Diabetes using Machine Learning Classifiers with Polar Bear Optimization	Menggunakan Logistic Regression (LR), Decision Tree (DT), dan Random Forest (RF) dengan teknik optimasi Polar Bear Optimization (PBO) untuk pemilihan fitur.	Sumber dataset tidak dijelaskan secara rinci.	Nilai akurasi spesifik tidak disebutkan publikasi.	Menggabungkan optimasi alam (PBO) untuk pemilihan fitur, sehingga meningkatkan efisiensi dan akurasi model prediksi.	Tidak disebutkannya akurasi spesifik model dan absennya evaluasi klinis, sehingga efektivitas praktisnya belum terbukti.
12.	Zhizhong Sun, Zidong Cao, Limin Ge, Yifan Li, Haoming Huang, Mingrui Li,	Applications of Resting-State fMRI and Machine Learning	Scoping review yang menganalisis berbagai algoritma machine	Menganalisis berbagai studi yang menggunakan dataset rs-fMRI dari pasien	Nilai akurasi spesifik tidak disebutkan publikasi.	Membahas potensi kombinasi rs-fMRI dan ML untuk deteksi	Kurangnya informasi spesifik tentang algoritma dan metrik

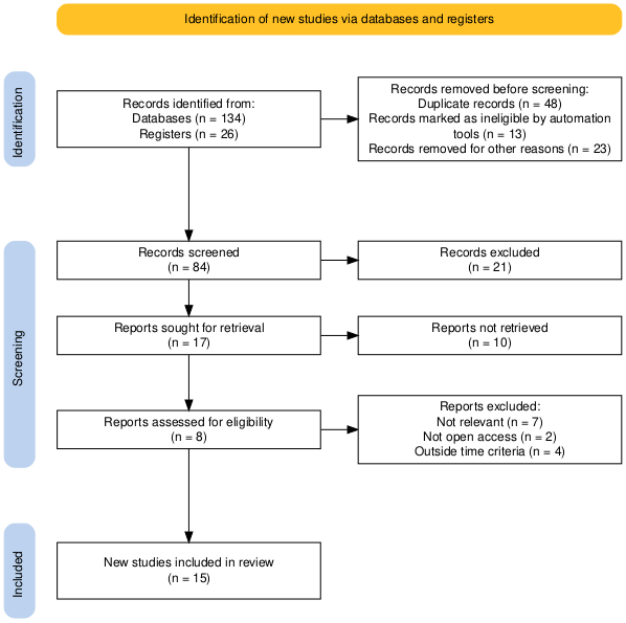
NO	Penulis dan tahun	Judul	Algoritma ML	Dataset	Akurasi	Kelebihan	Kekurangan
	Shijun Qiu (2025) [29]	in Cognitive Impairment in Type 2 Diabetes Mellitus: A Scoping Review	learning untuk data rs-fMRI.	T2DM dengan gangguan kognitif.		gangguan kognitif pada pasien T2DM, membuka peluang untuk alat diagnostik non-invasif.	kinerja (seperti akurasi) menghambat pemahaman efektivitas metode.
13.	Sulaiman Afolabi, Nurudeen Ajadi, Afeez Jimoh, Ibrahim Adenekan (2025) [30]	Predicting diabetes using supervised machine learning algorithms on E-health records	Menerapkan berbagai algoritma machine learning terawasi (supervised).	Data E-health records (EHR) yang mencakup informasi klinis pasien.	Nilai akurasi spesifik tidak disebutkan publikasi.	Meningkatkan deteksi dini diabetes dengan memanfaatkan data EHR untuk mengenali pola yang sulit dideteksi secara konvensional.	Akurasi model tidak dicantumkan dan tidak adanya uji klinis, sehingga efektivitas praktisnya belum teruji.
14.	Siddhant Kumar Jena, Dayal Kumar Behera, Ajay Kumar Jena, Jitendra Kumar Rout (2025) [31]	A Novel Ensemble Machine Learning Framework for Improved Diabetes Prediction and Complication Prevention	Menerapkan kerangka kerja ensemble machine learning yang menggabungkan berbagai model.	Sumber dataset tidak dijelaskan secara rinci.	Nilai akurasi spesifik tidak disebutkan publikasi.	Pendekatan ensemble meningkatkan akurasi prediksi dengan memungkinkan deteksi pola yang lebih luas dan kinerja yang lebih baik.	Nilai akurasi tidak dicantumkan dan tidak adanya uji klinis, sehingga efektivitas praktisnya belum teruji.
15.	Blessing Oluwatobi Olorunfemi, Abidemi Emmanuel Adeniyi, Adewale Opeoluwa Ogunde, Israel Korede Adeyanju, Federick Oscar, Nabeela T.	A Comparative Analysis of the Performance of Parallel Ensemble and Sequential Ensemble Machine Learning Methods in the	Membandingkan metode ensemble sekuensial (XGBoost, AdaBoost M1) dan paralel (Random Forest).	Pima Indian Diabetes dari UCI Machine Learning Repository.	Metode ensemble sekuensial mencapai 100%, sementara metode paralel (RF) mencapai 96%.	Menggabungkan metode ensemble dengan seleksi fitur maju-mundur untuk meningkatkan akurasi, mengurangi overfitting, dan meningkatkan	Terbatas pada dataset Pima yang kurang mewakili populasi luas, dan akurasi 100% perlu divalidasi lebih lanjut untuk menghind

NO	Penulis dan tahun	Judul	Algoritma ML	Dataset	Akurasi	Kelebihan	Kekurangan
	Adebola (2025) [32]	Detection of Diabetes Miletus				generalisasi.	ari overfitting.

2.3 Teknik Analisis Sistematik

Analisis sistematik dilakukan menggunakan diagram PRISMA untuk menyeleksi artikel berdasarkan tahapan identifikasi, penyaringan, penilaian kelayakan, dan inklusi akhir ke dalam tinjauan. Proses ini menghasilkan 15 artikel yang sesuai dan digunakan dalam studi ini.

Setelah tahap penyaringan, diperoleh 84 artikel untuk diseleksi lebih lanjut. Dari jumlah tersebut, 21 artikel dieksklusi karena tidak sesuai topik atau abstrak tidak relevan. Sebanyak 17 laporan dicari untuk diakses, namun hanya 8 laporan yang berhasil diakses penuh. Dari laporan yang berhasil dianalisis, 15 artikel akhir dimasukkan ke dalam tinjauan sistematik ini, karena relevan dan sesuai kriteria inklusi



Gambar 2. Diagram Prisma

3. Hasil dan Pembahasan

Berdasarkan kajian sistematis terhadap 15 artikel, ditemukan bahwa Random Forest (RF) merupakan algoritma yang paling dominan dan konsisten dalam diagnosis serta prediksi diabetes, dengan beberapa studi seperti oleh Gowthami et al. (2023) dan Iqra Nissar et al. (2024) melaporkan pencapaian akurasi hingga 98%. Dominasi ini didukung oleh berbagai penelitian yang juga menerapkan algoritma lain seperti SVM, KNN, dan Decision Tree dengan hasil kompetitif. Lebih dari sekadar perbandingan algoritma, analisis literatur juga menyoroti adanya tren inovasi yang signifikan. Sebagai contoh, penelitian oleh Imran Rehan et al. (2024) memperkenalkan metode diagnosis non-invasif menggunakan data spektroskopi kuku dengan akurasi 95%, yang berpotensi merevolusi skrining massal.

Sejalan dengan itu, muncul pula penerapan Explainable AI (XAI), seperti pada studi oleh Gangani Dharmarathne et al. (2024), yang menggabungkan model XGBoost dengan teknik interpretasi (SHAP dan PDP) untuk meningkatkan transparansi dan kepercayaan tenaga medis. Meskipun berbagai temuan ini menunjukkan hasil yang sangat menjanjikan, kajian ini secara konsisten mengidentifikasi adanya tantangan metodologis fundamental yang menghambat implementasi praktis. Mayoritas studi masih bergantung pada dataset publik yang terbatas seperti Pima Indian Diabetes Dataset (PIDD) dan BRFSS, atau menggunakan

dataset privat berukuran kecil yang tidak disebutkan sumbernya secara eksplisit. Keterbatasan ini, diperparah oleh kurangnya validasi eksternal dan minimnya uji klinis di dunia nyata pada hampir semua penelitian, menimbulkan keraguan besar terhadap kemampuan generalisasi model pada populasi yang lebih luas.

Dengan demikian, teridentifikasi adanya kesenjangan yang jelas antara performa teoretis model yang tinggi dengan kesiapannya untuk adopsi klinis, sehingga penelitian di masa depan harus memprioritaskan validasi yang lebih ketat dan transparansi pelaporan yang lebih baik.

4. Kesimpulan

Kajian ini menunjukkan bahwa penerapan algoritma machine learning memiliki potensi besar dalam meningkatkan akurasi diagnosis dan prediksi penyakit diabetes, baik tipe 2 maupun gestasional. Random Forest muncul sebagai algoritma paling andal dan umum digunakan, diikuti oleh SVM, KNN, dan Decision Tree. Meskipun demikian, akurasi model sangat dipengaruhi oleh kualitas dan jumlah data, teknik praproses, serta metode seleksi fitur yang digunakan. Penggunaan metode seleksi fitur seperti mRMR dan PCA terbukti mampu meningkatkan performa model dengan mengurangi noise data yang tidak relevan.

Studi juga mulai mengadopsi pendekatan yang lebih inovatif seperti penggunaan data non-konvensional (misalnya kuku jari) dan integrasi dengan explainable AI untuk memberikan transparansi pada hasil prediksi. Hal ini penting agar sistem berbasis ML tidak hanya akurat tetapi juga dapat dipercaya dan dimengerti oleh tenaga medis sebagai pengguna akhir. Tantangan utama dalam penerapan klinis meliputi keterbatasan validasi eksternal, generalisasi model antar populasi yang berbeda, serta kurangnya uji coba dalam setting dunia nyata.

Oleh karena itu, meskipun algoritma ML telah menunjukkan performa tinggi secara teoritis, adopsi luasnya dalam sistem kesehatan masih membutuhkan pendekatan yang lebih sistematis, validasi yang lebih ketat, serta kolaborasi antara peneliti teknologi dan tenaga medis. Ke depan, penerapan ML dalam diagnosis diabetes berpotensi menjadi alat bantu yang sangat berguna, tetapi hal ini harus dibarengi dengan upaya memastikan keandalan, transparansi, dan keterpakaian model dalam konteks klinis yang sesungguhnya.

Daftar Pustaka

[1] E. K. Hoogeveen, “The Epidemiology Of Diabetic Kidney Disease,” *Kidney And Dialysis*, Vol. 2, No. 3, Pp. 433–442, Aug. 2022, Doi: 10.3390/Kidneydial2030038.

[2] S. Alzu’bi *Et Al.*, “Diabetes Monitoring System In Smart Health Cities Based On Big Data Intelligence,” *Future Internet*, Vol. 15, No. 2, P. 85, Feb. 2023, Doi: 10.3390/Fi15020085.

[3] Md. S. Reza, A. R. S. Fakir, Md. R. Islam, A. M. T. Alom, T. Sen, And M. S. Islam, “Early Stage Diabetes Prediction Using Machine Learning Techniques,” In *2023 26th International Conference On Computer And Information Technology (Iccit)*, Ieee, Dec. 2023, Pp. 1–6. Doi: 10.1109/Iccit60459.2023.10441427.

[4] M. A. Alam, A. Sohel, K. M. Hasan, And M. A. Islam, “Machine Learning And Artificial Intelligence In Diabetes Prediction And Management: A Comprehensive Review Of Models,” *Innovatech Engineering Journal*, Vol. 1, No. 01, Pp. 107–124, Dec. 2024, Doi: 10.70937/Jnes.V1i01.41.

[5] S. Ziajor *Et Al.*, “The Use Of Artificial Intelligence In The Diagnosis And Detection Of Complications Of Diabetes,” *Journal Of Education, Health And Sport*, Vol. 65, Pp. 11–27, Apr. 2024, Doi: 10.12775/Jehs.2024.65.001.

[6] L. P. Zhuhadar And M. D. Lytras, “The Application Of Automl Techniques In Diabetes Diagnosis: Current Approaches, Performance, And Future Directions,” *Sustainability*, Vol. 15, No. 18, P. 13484, Sep. 2023, Doi: 10.3390/Su151813484.

[7] M. Salem Alzboon, M. Subhi Al-Batah, M. Alqaraleh, A. Abuashour, And A. F. Hamadah Bader, “Early Diagnosis Of Diabetes: A Comparison Of Machine Learning Methods,” *International Journal Of Online And Biomedical Engineering (Ijoe)*, Vol. 19, No. 15, Pp. 144–165, Oct. 2023, Doi: 10.3991/Ijoe.V19i15.42417.

[8] R. Islam, A. Sultana, Md. N. Tuhin, Md. S. H. Saikat, And M. R. Islam, "Clinical Decision Support System For Diabetic Patients By Predicting Type 2 Diabetes Using Machine Learning Algorithms," *J Healthc Eng*, Vol. 2023, No. 1, Jan. 2023, Doi: 10.1155/2023/6992441.

[9] T. Tominanto, A. Syukur, Affandy, And A. Marjuni, "Performance Analysis Of Machine Learning Algorithms For Diabetes Diagnosis Prediction," In *2023 International Seminar On Application For Technology Of Information And Communication (Isemantic)*, Ieee, Sep. 2023, Pp. 227–232. Doi: 10.1109/Isemantic59612.2023.10295368.

[10] M. S. Alzboon, M. Al-Batah, M. Alqaraleh, A. Abuashour, And A. F. Bader, "A Comparative Study Of Machine Learning Techniques For Early Prediction Of Diabetes," In *2023 Ieee Tenth International Conference On Communications And Networking (Comnet)*, Ieee, Nov. 2023, Pp. 1–12. Doi: 10.1109/Comnet60156.2023.10366688.

[11] S. Chakraborty, A. R. Panda, P. Vadiya, S. Samal, I. Gupta, And N. K. Ray, "Predicting Diabetes: A Comparative Study Of Machine Learning Models," In *2023 Oits International Conference On Information Technology (Ocit)*, Ieee, Dec. 2023, Pp. 743–748. Doi: 10.1109/Ocit59427.2023.10431372.

[12] F. K. S And A. Poulose, "Hybrid Convolutional Neural Networks For Pima Indians Diabetes Prediction," In *2024 Fifteenth International Conference On Ubiquitous And Future Networks (Icufn)*, Ieee, Jul. 2024, Pp. 268–273. Doi: 10.1109/Icufn61752.2024.10624950.

[13] J. Musa And A. M. Abdulazeez, "A Review On Diabetes Classification Based On Machine Learning Algorithms," *The Indonesian Journal Of Computer Science*, Vol. 13, No. 2, Apr. 2024, Doi: 10.33022/Ijcs.V13i2.3886.

[14] C. Wei, "Application Of Machine Learning In The Field Of Diabetes," *Science And Technology Of Engineering, Chemistry And Environmental Protection*, Vol. 1, No. 8, Aug. 2024, Doi: 10.61173/Kc05df27.

[15] Q. Saihood And E. Sonuç, "A Practical Framework For Early Detection Of Diabetes Using Ensemble Machine Learning Models," *Turkish Journal Of Electrical Engineering And Computer Sciences*, Vol. 31, No. 4, Pp. 722–738, Jul. 2023, Doi: 10.55730/1300-0632.4013.

[16] J. Musa And A. M. Abdulazeez, "A Review On Diabetes Classification Based On Machine Learning Algorithms," *The Indonesian Journal Of Computer Science*, Vol. 13, No. 2, Apr. 2024, Doi: 10.33022/Ijcs.V13i2.3886.

[17] M. A. Alam, A. Sohel, K. M. Hasan, And M. A. Islam, "Machine Learning And Artificial Intelligence In Diabetes Prediction And Management: A Comprehensive Review Of Models," *Innovatech Engineering Journal*, Vol. 1, No. 01, Pp. 107–124, Dec. 2024, Doi: 10.70937/Jnes.V1i01.41.

[18] G. Aguilera-Venegas, A. López-Molina, G. Rojo-Martínez, And J. L. Galán-García, "Comparing And Tuning Machine Learning Algorithms To Predict Type 2 Diabetes Mellitus," *J Comput Appl Math*, Vol. 427, P. 115115, Aug. 2023, Doi: 10.1016/J.Cam.2023.115115.

[19] G. S, R. Venkata Siva Reddy, And M. R. Ahmed, "Exploring The Effectiveness Of Machine Learning Algorithms For Early Detection Of Type-2 Diabetes Mellitus," *Measurement: Sensors*, Vol. 31, P. 100983, Feb. 2024, Doi: 10.1016/J.Measen.2023.100983.

[20] I. Nissar *Et Al.*, "An Intelligent Healthcare System For Automated Diabetes Diagnosis And Prediction Using Machine Learning," *Procedia Comput Sci*, Vol. 235, Pp. 2476–2485, 2024, Doi: 10.1016/J.Procs.2024.04.233.

[21] G. Dharmarathne, T. N. Jayasinghe, M. Bogahawaththa, D. P. P. Meddage, And U. Rathnayake, "A Novel Machine Learning Approach For Diagnosing Diabetes With A Self-Explainable Interface," *Healthcare Analytics*, Vol. 5, P. 100301, Jun. 2024, Doi: 10.1016/J.Health.2024.100301.

[22] D. Kaviyaadharshani, M. Nivedhidha, R. Jeyarohini, J. L. E. Rani, M. P. Ramkumar, And G. S. R. E. Selvan, "Diagnosing Diabetes Using Machine Learning-Based Predictive Models," *Procedia Comput Sci*, Vol. 233, Pp. 288–294, 2024, Doi: 10.1016/J.Procs.2024.03.218.

[23] M. M. Chowdhury, R. S. Ayon, And M. S. Hossain, "An Investigation Of Machine Learning Algorithms And Data Augmentation Techniques For Diabetes Diagnosis Using Class Imbalanced Brfss Dataset," *Healthcare Analytics*, Vol. 5, P. 100297, Jun. 2024, Doi: 10.1016/J.Health.2023.100297.

[24] N. P. Shetty, J. Shetty, V. Hegde, S. D. Dharne, And M. Kv, "A Machine Learning-Based Clinical Decision Support System For Effective Stratification Of Gestational Diabetes Mellitus And

Management Through Ayurveda," *J Ayurveda Integr Med*, Vol. 15, No. 6, P. 101051, Nov. 2024, Doi: 10.1016/J.Jaim.2024.101051.

[25] I. Rehan, K. Rehan, S. Sultana, And M. Ur Rehman, "Fingernail Diagnostics: Advancing Type Ii Diabetes Detection Using Machine Learning Algorithms And Laser Spectroscopy," *Microchemical Journal*, Vol. 201, P. 110762, Jun. 2024, Doi: 10.1016/J.Microc.2024.110762.

[26] B. Amma N.G., "En-Rfrsk: An Ensemble Machine Learning Technique For Prognostication Of Diabetes Mellitus," *Egyptian Informatics Journal*, Vol. 25, P. 100441, Mar. 2024, Doi: 10.1016/J.Eij.2024.100441.

[27] G. Dharmarathne, M. Bogahawaththa, M. Mcafee, U. Rathnayake, And D. P. P. Meddage, "On The Diagnosis Of Chronic Kidney Disease Using A Machine Learning-Based Interface With Explainable Artificial Intelligence," *Intelligent Systems With Applications*, Vol. 22, P. 200397, Jun. 2024, Doi: 10.1016/J.Iswa.2024.200397.

[28] P. K. Patidar, S. Shiwani, And S. Garg, "Prediction Of Diabetes Using Machine Learning Classifiers With Polar Bear Optimization," *Procedia Comput Sci*, Vol. 258, Pp. 1338–1347, 2025, Doi: 10.1016/J.Procs.2025.04.367.

[29] Z. Sun *Et Al.*, "Applications Of Resting-State Fmri And Machine Learning In Cognitive Impairment In Type 2 Diabetes Mellitus: A Scoping Review," *Meta-Radiology*, P. 100136, Mar. 2025, Doi: 10.1016/J.Metrad.2025.100136.

[30] S. Afolabi, N. Ajadi, A. Jimoh, And I. Adenekan, "Predicting Diabetes Using Supervised Machine Learning Algorithms On E-Health Records," *Informatics And Health*, Vol. 2, No. 1, Pp. 9–16, Mar. 2025, Doi: 10.1016/J.Infoh.2024.12.002.

[31] S. K. Jena, D. K. Behera, A. K. Jena, And J. K. Rout, "A Novel Ensemble Machine Learning Framework For Improved Diabetes Prediction And Complication Prevention," *Procedia Comput Sci*, Vol. 258, Pp. 4008–4017, 2025, Doi: 10.1016/J.Procs.2025.04.652.

[32] B. O. Olorunfemi, A. E. Adeniyi, A. O. Ogunde, I. K. Adeyanju, F. Oscar, And N. T. Adebola, "A Comparative Analysis Of The Performance Of Parallel Ensemble And Sequential Ensemble Machine Learning Methods In The Detection Of Diabetes Miletus," *Procedia Comput Sci*, Vol. 258, Pp. 1038–1049, 2025, Doi: 10.1016/J.Procs.2025.04.340.