

Systematic Literature Review

Analisis SWOT dan Gap Analysis: Model Prediksi Diabetes dengan Fokus pada Kalibrasi Probabilitas dan Interpretabilitas (XAI)

Latar Belakang

Penelitian ini menganalisis artikel ilmiah terkait penerapan machine learning (ML) untuk prediksi diabetes. Analisis ini mengevaluasi lanskap penelitian terkini dalam prediksi diabetes menggunakan model *machine learning*, dengan fokus khusus pada aspek kalibrasi probabilitas dan interpretabilitas model (menggunakan teknik seperti SHAP dan LIME) untuk aplikasi dalam prediksi risiko klinis. Tujuannya adalah untuk mengidentifikasi kekuatan, kelemahan, peluang, ancaman, serta kesenjangan penelitian yang ada.

1. SWOT Analysis

Faktor Internal	Strengths (Kekuatan)	Weaknesses (Kelemahan)
Aspek Metodologis & Kinerja	<div><div>– Beberapa model, seperti XGBoost dan Random Forest, menunjukkan kinerja diskriminatif yang sangat tinggi (akurasi hingga 97.88%) dalam memprediksi diabetes.</div><div>– Adanya penelitian yang secara khusus mengembangkan metode untuk meningkatkan kalibrasi probabilitas, seperti penggunaan <i>Focal Loss</i> dan <i>Gradient-Boosted Trees</i> (DOI:10.3390/electronics14091838) serta metode koreksi berbasis probabilitas (DOI:10.1109/ICCSC66714.2025.11134984).</div><div>– Teknik <i>Explainable AI</i> (XAI) seperti SHAP dan LIME telah</div></div>	<div><div>– Fokus pada Akurasi vs. Kalibrasi: Mayoritas penelitian (seperti DOI:10.32409/jikstik.23.1.3507, DOI:10.32520/STMSI.V10I1.1129) lebih berfokus pada metrik akurasi dan diskriminasi, sementara kalibrasi probabilitas sering diabaikan. Probabilitas yang tidak terkalibrasi dapat menyesatkan dalam pengambilan keputusan klinis.</div><div>– Validasi dan Generalisasi Terbatas: Banyak model dikembangkan dan</div></div>

Faktor Internal	Strengths (Kekuatan)	Weaknesses (Kelemahan)
	<p>diadopsi untuk membuka "kotak hitam" model kompleks, meningkatkan kemampuan interpretasi (DOI:10.1109/IEEECONF6499 2.2025.10962890, DOI:10.1016/j.artmed.2025.103132).</p>	<p>dievaluasi pada dataset tunggal tanpa validasi eksternal yang memadai. Beberapa studi sistematis (DOI:10.1016/j.ijmedinf.2020.104268) mencatat perlunya peningkatan dalam metodologi dan pelaporan.</p>
Aspek Aplikasi Klinis	<ul style="list-style-type: none"> <li>– Rekalibrasi untuk Populasi Berbeda: Penelitian menunjukkan pentingnya dan keefektifan merekalibrasi model yang ada (seperti <i>Diabetes Population Risk Tool</i>) untuk populasi baru (DOI:10.1109/EMBC.2018.8513554, DOI:10.1016/j.dsx.2024.103007).</li> <li>– Pengembangan Kalkulator Klinis: Ada upaya untuk mengembangkan alat prediksi yang siap pakai di klinik, seperti kalkulator probabilitas MODY yang menunjukkan kalibrasi dan diskriminasi yang sangat baik (DOI:10.1038/s41598-024-60160-0).</li> </ul>	<ul style="list-style-type: none"> <li>– Integrasi XAI dan Kalibrasi: Terdapat kesenjangan yang jelas antara penelitian yang fokus pada interpretabilitas (XAI) dan penelitian yang fokus pada keandalan probabilitas (kalibrasi). Kedua aspek ini jarang dibahas secara integratif dalam satu studi.</li> </ul>



Faktor Eksternal	Opportunities (Peluang)	Threats (Ancaman)
Lingkungan Penelitian & Teknologi	<ul style="list-style-type: none"> <li>– Konvergensi Teknik: Terdapat peluang besar untuk mengintegrasikan teknik XAI (SHAP/LIME) dengan metode kalibrasi probabilitas (seperti <i>Conformal Prediction</i>, <i>Platt Scaling</i>) untuk menciptakan model yang sekaligus andal, dapat diinterpretasi, dan terpercaya (terlihat dari DOI:10.3390/app15147925 yang melakukan ini di domain kesehatan lain).</li> <li>– Peningkatan Standar: Tinjauan sistematis (seperti DOI:10.1016/j.ijmedinf.2020.104268) menyoroti kebutuhan akan peningkatan standar, yang membuka peluang untuk penelitian dengan metodologi yang lebih ketat, pelaporan yang transparan, dan validasi yang komprehensif.</li> </ul>	<ul style="list-style-type: none"> <li>– Kompleksitas vs. Kegunaan Klinis: Model yang semakin kompleks dan canggih (seperti XGBoost dengan XAI) dapat menjadi terlalu rumit untuk diadopsi secara langsung oleh tenaga medis jika tidak disajikan dalam antarmuka yang user-friendly.</li> <li>– Over-reliance pada Metrik Diskriminatif: Budaya penelitian yang masih sangat mengutamakan AUC dan akurasi dapat mengancam upaya penelitian yang berfokus pada kalibrasi, karena dianggap kurang meskipun sangat penting secara klinis.</li> </ul>

Faktor Eksternal	Opportunities (Peluang)	Threats (Ancaman)
<p><b>Kebutuhan Klinis &amp; Regulasi</b></p>	<ul style="list-style-type: none"> <li>– Personalized Medicine: Kalibrasi untuk populasi spesifik (seperti pada populasi India, DOI:10.1016/j.dsx.2024.103007) membuka jalan bagi pendekatan kedokteran yang lebih personal dan akurat.</li> <li>– Decision Support yang Andal: Kebutuhan akan sistem pendukung keputusan klinis yang tidak hanya akurat tetapi juga dapat menjelaskan alasannya dan memberikan tingkat keyakinan (probabilitas) yang terkalibrasi semakin meningkat.</li> </ul>	<ul style="list-style-type: none"> <li>– Masalah Etika dan Kepercayaan: Jika model memberikan probabilitas risiko yang tidak terkalibrasi, dapat menyebabkan kesalahan diagnosis, penanganan yang tidak tepat, dan akhirnya merusak kepercayaan terhadap AI dalam medis.</li> <li>– Hambatan Regulasi: Model AI untuk perangkat medis (Software as a Medical Device/SaMD) kemungkinan akan membutuhkan standar ketat tidak hanya untuk akurasi tetapi juga untuk kalibrasi dan interpretabilitas, yang menjadi tantangan bagi model yang tidak memenuhinya.</li> </ul>

## 2. Gap Analysis

Berdasarkan analisis terhadap kumpulan dokumen, beberapa kesenjangan penelitian (*research gaps*) yang signifikan dapat diidentifikasi:

### 1) Gap Integratif antara Interpretabilitas dan Kalibrasi Probabilitas

#### – Kesenjangan

Sebagian besar penelitian berfokus pada salah satu aspek: *baik* interpretabilitas (menggunakan SHAP/LIME) *atau* kalibrasi probabilitas. Sangat sedikit studi yang secara eksplisit menyatukan keduanya untuk menciptakan model yang sekaligus dapat dijelaskan dan memberikan probabilitas yang andal.

#### – Bukti

DOI:10.1109/IEEECONF64992.2025.10962890 membahas SHAP dan LIME untuk interpretabilitas, sementara DOI:10.3390/electronics14091838 membahas kalibrasi. DOI:10.3390/app15147925 menunjukkan integrasi ini di domain lain (monitoring proses kesehatan), membuktikan bahwa ini adalah peluang yang belum terjamah dalam prediksi diabetes.

### 2) Gap dalam Evaluasi Model yang Komprehensif

#### – Kesenjangan

Evaluasi model seringkali terbatas pada metrik diskriminasi (akurasi, AUC, F1-Score) tanpa menyertakan metrik kalibrasi (seperti *Brier Score*, *Calibration Plot*, atau *Expected Calibration Error*).

#### – Bukti

Studi seperti DOI:10.32409/jikstik.23.1.3507 dan DOI:10.32520/STMSI.V10I1.1129 hanya melaporkan akurasi. Studi sistematis (DOI:10.1016/j.ijmedinf.2020.104268) mengonfirmasi bahwa pelaporan metrik kalibrasi masih jarang.

### 3) Gap dalam Penerapan Teknik Kalibrasi Modern pada Model Canggih

#### – Kesenjangan

Meskipun teknik seperti *Conformal Prediction* (DOI:10.3390/app15147925) dan *Focal Loss* (DOI:10.3390/electronics14091838) telah terbukti meningkatkan kalibrasi, penerapannya pada model *boosting* populer seperti XGBoost untuk prediksi diabetes masih relatif terbatas dan belum menjadi praktik standar.

#### **4) Gap dalam Penelitian Berbasis Kebutuhan Klinis yang Spesifik**

##### **– Kesenjangan**

Banyak model dikembangkan sebagai "solusi umum". Namun, bukti menunjukkan bahwa model membutuhkan rekailbrasi untuk populasi yang berbeda (DOI:10.1016/j.dsx.2024.103007). Kesenjangan terletak pada kurangnya kerangka kerja yang sistematis untuk adaptasi dan validasi model yang cepat dan andal across berbagai seting klinis dan demografi.

#### **Rekomendasi untuk Penelitian Mendatang**

- 1) Kembangkan sebuah kerangka kerja yang mengintegrasikan XGBoost (sebagai model yang kuat) dengan teknik kalibrasi probabilitas (seperti Platt Scaling atau Isotonic Regression) dan teknik XAI (SHAP) secara bersamaan.
- 2) Lakukan evaluasi yang komprehensif tidak hanya dengan metrik diskriminasi tetapi juga dengan metrik kalibrasi, dan laporkan kedua hasilnya.
- 3) Validasi model yang telah dikalibrasi dan dapat diinterpretasi ini pada dataset eksternal dari berbagai populasi untuk menguji robustitas dan kegunaan klinisnya.