

Pendekatan *Machine Learning* untuk Menganalisis Faktor-Faktor yang Berkontribusi Terhadap Kesalahan Pengkodean ICD-10 di Semen Padang Hospital

Rury Moryanda^{*1}, Nurul Abdillah², Denos Imam Fratama¹, Dicky Fatrias³

¹Program Studi Manajemen Informasi Kesehatan, Universitas Syedza Saintika

rurymorzz@gmail.com, denos.fratama09@gmail.com

²Program Studi Teknologi Informasi, Universitas Jambi

abdillahadik15@gmail.com

³Program Studi Departemen Teknik Industri, Universitas Andalas

dickyf@eng.unand.ac.id

Keywords:

*ICD-10 Coding,
Machine learning,
Electronic Medical Records,
Neural Network,
Coding Accuracy*

ABSTRACT

ICD-10 coding errors remain a major challenge in healthcare facilities, affecting the validity of morbidity data, the accuracy of BPJS Kesehatan claims processing, and the quality of health information management. This study aimed to identify the main factors contributing to ICD-10 coding errors and evaluate the effectiveness of machine learning algorithms in detecting such errors. Electronic medical record data were analyzed through data cleaning, preprocessing, descriptive analysis, and machine learning modeling. Three algorithms were applied: Random Forest, Support Vector Machine (SVM), and Neural Network. Model performance was evaluated using accuracy, precision, and sensitivity metrics. The findings revealed an ICD-10 coding error rate of 33.8%, primarily caused by nonspecific diagnoses and insufficient clinical information. Among the tested models, the Neural Network achieved the highest accuracy (72%), followed by SVM (68%) and Random Forest (60%). These results suggest that machine learning techniques can effectively support the early detection of ICD-10 coding errors and enhance the quality of health data management. The adoption of machine learning-based predictive models may improve coding accuracy and facilitate evidence-based decision-making in health information management.

Kata Kunci

*Pengkodean ICD-10,
Machine learning,
Rekam Medis Elektronik,
Neural Network,
Akurasi Pengkodean*

ABSTRAK

Kesalahan pengkodean ICD-10 masih menjadi permasalahan penting di fasilitas pelayanan kesehatan karena dapat memengaruhi validitas data morbiditas, proses klaim BPJS Kesehatan, dan kualitas manajemen informasi kesehatan. Penelitian ini bertujuan mengidentifikasi faktor dominan penyebab kesalahan pengkodean ICD-10 serta mengevaluasi efektivitas algoritma *machine learning* dalam mendeteksi kesalahan tersebut. Metode penelitian menggunakan data rekam medis elektronik (RME) yang dianalisis melalui tahapan pembersihan data, pra-proses, analisis deskriptif, dan pemodelan *machine learning*. Algoritma yang digunakan meliputi *Random Forest*, *Support Vector Machine* (SVM), dan *Neural Network*. Kinerja model dievaluasi menggunakan metrik akurasi, presisi, dan sensitivitas. Hasil penelitian menunjukkan tingkat kesalahan pengkodean ICD-10 sebesar 33,8%, yang terutama disebabkan oleh diagnosis yang tidak spesifik dan keterbatasan informasi klinis. *Neural Network* menunjukkan kinerja terbaik dengan akurasi 72%, diikuti SVM

sebesar 68% dan *Random Forest* sebesar 60%. Temuan ini menunjukkan bahwa pendekatan *machine learning* berpotensi mendukung deteksi dini kesalahan pengkodean ICD-10 dan meningkatkan kualitas pengelolaan data kesehatan. Penerapan model prediktif berbasis *machine learning* dapat menjadi strategi untuk meningkatkan akurasi pengkodean serta mendukung pengambilan keputusan dalam manajemen informasi kesehatan.

Korespondensi Penulis:

Rury Moryanda,
Universitas Syedza Saintika,
Jl. Prof. Dr. Hamka No. 228
Telepon: +6281268514631
Email: rurymorzz@mail.com

Submitted : 12-12-2025; Accepted : 02-04-2026;
Published : 28-06-2026

Copyright (c) 2026 The Author (s)

This article is distributed under a Creative Commons
Attribution-ShareAlike 4.0 International License (CC BY-SA
4.0)

1. PENDAHULUAN

Kesalahan dalam pengkodean diagnosis medis menggunakan sistem *International Classification of Diseases, Tenth Revision* (ICD-10) masih menjadi permasalahan yang signifikan di berbagai fasilitas pelayanan kesehatan. Sistem ICD-10 memiliki peran penting dalam mendokumentasikan dan mengklasifikasikan penyakit serta kondisi medis pasien ke dalam rekam medis elektronik (*Electronic Medical Record* atau EMR). Penggunaan kode ICD-10 tidak hanya mendukung pelayanan klinis, tetapi juga menjadi sumber data penting untuk keperluan penelitian, manajemen kesehatan, pelaporan statistik morbiditas, serta proses klaim asuransi kesehatan.

Kesalahan dalam penetapan kode ICD-10 dapat menimbulkan berbagai dampak negatif, seperti ketidaktepatan diagnosis dan pengobatan, penurunan kualitas data kesehatan, kesalahan dalam analisis statistik kesehatan, serta kerugian finansial bagi fasilitas pelayanan kesehatan. Salah satu faktor utama yang menyebabkan terjadinya kesalahan pengkodean adalah kompleksitas dan ambiguitas dalam pemilihan kode diagnosis yang sesuai, terutama pada kasus dengan gejala atau karakteristik klinis yang serupa [1].

Beberapa penelitian di Indonesia menunjukkan bahwa tingkat kesalahan pengkodean diagnosis masih relatif tinggi. Berbagai studi melaporkan bahwa tingkat ketidakakuratan kode diagnosis ICD-10 berkisar antara 25% hingga 80% pada berbagai fasilitas pelayanan kesehatan. Temuan tersebut menunjukkan bahwa pengkodean diagnosis masih menjadi tantangan dalam pengelolaan rekam medis dan peningkatan kualitas data kesehatan. Ringkasan hasil penelitian terkait tingkat kesalahan pengkodean ICD-10 di Indonesia disajikan pada Tabel 1.

Tabel 1. Ringkasan Hasil Penelitian Terdahulu Mengenai Tingkat Kesalahan Pengkodean ICD-10 di Indonesia

No	Referensi	Lokasi Penelitian	Sampel	Temuan Utama	Tingkat Kesalahan (%)
1	Kartika dan Sukawan (2025) [2]	RS Jasa Kartini	77 rekam medis	Hanya 33,8% kode diagnosis injury akurat	66,2% tidak akurat
2	Andriani (2021) [3]	Klinik Obstetri dan Ginekologi	138 rekam medis pemeriksaan kehamilan	Ketidaktepatan dalam pemilihan kode ICD-10	60% tidak akurat
3	Hastuti dan Ali (2019) [4]	5 Puskesmas Kota Malang	Rekam medis rawat jalan	Ditemukan masih rendahnya akurasi kode diagnosis sesuai ICD-10	±35% tidak akurat
4	Ekawati <i>et al.</i> (2024) [5]	RS Pelamonia Makassar	Dokumen klaim BPJS	Masih ditemukan kesalahan kode akibat diagnosis tidak jelas dan singkatan	±25% tidak akurat
5	Ophivania dan Putri (2026) [6]	Berbagai rumah sakit	5 artikel <i>Literature Review</i>	Rata-rata akurasi kode diagnosis sekitar 67–70%	±30–33% kesalahan

Meskipun berbagai penelitian telah membahas kesalahan pengkodean ICD-10 dan pemanfaatan teknologi dalam sistem informasi kesehatan, sebagian besar studi masih berfokus pada audit manual, analisis deskriptif, atau pengembangan sistem *auto-coding* berbasis *Natural Language Processing* (NLP).

Pendekatan tersebut umumnya ditujukan untuk meningkatkan efisiensi proses pengkodean, namun belum banyak dimanfaatkan untuk menganalisis faktor-faktor penyebab kesalahan pengkodean secara prediktif dan berbasis pola data. Selain itu, penelitian yang mengkaji kesalahan pengkodean ICD-10 dalam konteks fasilitas pelayanan kesehatan di negara berkembang masih relatif terbatas, terutama yang berkaitan dengan kualitas dokumentasi klinis, kompetensi sumber daya manusia, dan sistem pembiayaan berbasis klaim. Kondisi tersebut menunjukkan masih adanya kesenjangan penelitian terkait pemanfaatan *machine learning* untuk mengidentifikasi faktor dominan dan memprediksi risiko kesalahan pengkodean ICD-10 secara sistematis [7].

Pengkodean diagnosis yang dilakukan secara manual masih rentan terhadap *human error*. Keterbatasan pemahaman terhadap aturan pengkodean, tingginya beban kerja, serta ketidakkonsistenan dokumentasi medis merupakan beberapa faktor yang sering dikaitkan dengan terjadinya kesalahan pengkodean [7]. Berbagai penelitian menunjukkan bahwa teknologi seperti NLP dan *deep learning* memiliki potensi untuk meningkatkan akurasi pengkodean melalui proses otomatisasi dan analisis data klinis [8], [9]. Namun demikian, penerapan teknologi tersebut masih menghadapi berbagai tantangan, seperti keberagaman data medis, kualitas data yang tidak terstruktur, serta keterbatasan dalam interpretasi hasil model [10]. Di sisi lain, kesalahan pengkodean diagnosis ICD-10 berdampak langsung terhadap validitas data morbiditas, ketepatan klaim pembiayaan, serta kualitas manajemen informasi kesehatan [11], [12]. Oleh karena itu, diperlukan pendekatan yang mampu mengidentifikasi faktor-faktor penyebab kesalahan pengkodean sekaligus mendukung peningkatan akurasi pengkodean secara lebih efektif.

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan penerapan algoritma *machine learning* untuk menganalisis faktor-faktor penyebab kesalahan pengkodean ICD-10 dan memprediksi risiko terjadinya kesalahan pengkodean secara otomatis. Berbeda dengan penelitian sebelumnya yang lebih berfokus pada pengembangan sistem *auto-coding*, penelitian ini menitikberatkan pada identifikasi faktor dominan penyebab kesalahan pengkodean dan evaluasi kinerja berbagai model *machine learning* dalam mendeteksi kesalahan pengkodean diagnosis. Pendekatan ini didukung oleh penelitian Kim *et al.* (2021) yang menunjukkan bahwa kesalahan pengkodean ICD-10 sering dipengaruhi oleh kemiripan kode diagnosis dan ketidaksesuaian antara interpretasi manusia dengan sistem otomatis [13].

Tujuan penelitian ini adalah mengidentifikasi faktor-faktor dominan yang menyebabkan kesalahan pengkodean ICD-10 pada RME, mengembangkan model *machine learning* untuk memprediksi risiko kesalahan pengkodean, serta mengevaluasi kinerja beberapa algoritma *machine learning* dalam mendeteksi kesalahan pengkodean diagnosis. Melalui pendekatan tersebut, penelitian ini diharapkan dapat menghasilkan rekomendasi berbasis data untuk meningkatkan akurasi pengkodean diagnosis dan mendukung pengelolaan informasi kesehatan yang lebih efektif. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan metode berbasis data untuk meningkatkan akurasi pengkodean diagnosis, mengurangi beban kerja petugas pengkodean, serta mendukung pengambilan keputusan dalam manajemen informasi kesehatan.

2. METODE PENELITIAN

2.1 Desain Penelitian

Penelitian ini menggunakan desain kuantitatif dengan pendekatan retrospektif berbasis analisis data sekunder. Data penelitian diperoleh dari rekam medis elektronik (RME) yang diekstraksi dari sistem informasi rumah sakit. Pendekatan retrospektif dipilih untuk menganalisis data historis yang telah tersedia sehingga memungkinkan identifikasi faktor-faktor yang berkontribusi terhadap kesalahan pengkodean ICD-10 tanpa memengaruhi proses pelayanan kesehatan yang sedang berlangsung.

2.2 Sumber Data dan Variabel Penelitian

Sumber data penelitian berasal dari RME yang tersimpan dalam sistem informasi rumah sakit. Variabel yang dikumpulkan meliputi diagnosis klinis dalam bentuk teks naratif, kode ICD-10 yang diberikan oleh koder, karakteristik demografis pasien, serta indikator kelengkapan dokumentasi klinis. Seluruh data telah dianonimkan sebelum proses analisis untuk menjaga kerahasiaan dan privasi pasien sesuai dengan prinsip etika penelitian kesehatan.

2.3 Populasi dan Sampel Penelitian

Populasi penelitian mencakup seluruh RME pasien yang tersedia selama periode penelitian. Teknik *sampling* yang digunakan adalah *total sampling*, yaitu seluruh rekam medis yang memenuhi kriteria

penelitian diikutsertakan dalam analisis. Sebanyak 600 rekam medis digunakan sebagai sampel penelitian. Kriteria inklusi meliputi rekam medis yang memiliki diagnosis utama yang telah diberikan kode ICD-10 dan didukung oleh dokumentasi klinis yang dapat diverifikasi. Rekam medis dengan data yang tidak lengkap secara signifikan atau teridentifikasi sebagai duplikasi dikeluarkan dari penelitian. Data dikumpulkan secara retrospektif pada periode Oktober hingga Desember 2024 untuk memperoleh jumlah observasi yang memadai serta meminimalkan potensi bias temporal.

2.4 Validasi Data Pengkodean ICD-10

Validasi data dilakukan melalui proses *re-coding* oleh *expert coder* yang memiliki kompetensi dalam pengkodean diagnosis berdasarkan ICD-10. Hasil pengkodean ulang tersebut dibandingkan dengan kode yang terdapat pada rekam medis untuk menilai tingkat konsistensi dan kesepakatan antar penilai. Kode yang telah tervalidasi kemudian digunakan sebagai label referensi (*ground truth*) dalam pengembangan dan evaluasi model *machine learning*. Proses validasi ini bertujuan untuk meningkatkan validitas internal penelitian serta memastikan keandalan data yang digunakan pada tahap pelatihan dan pengujian model.

2.5 Pengolahan dan Analisis Data

Data yang telah tervalidasi selanjutnya diproses untuk pengembangan model *machine learning*. Variabel diagnosis klinis, karakteristik pasien, dan indikator dokumentasi klinis digunakan sebagai fitur (*features*), sedangkan hasil validasi kode ICD-10 digunakan sebagai label referensi. Model *machine learning* dikembangkan untuk mengidentifikasi faktor-faktor yang berkontribusi terhadap kesalahan pengkodean ICD-10 serta mengevaluasi pola yang berhubungan dengan ketidakakuratan kode diagnosis.

2.6 Tahapan Penelitian

Tahapan penelitian ini disusun secara sistematis untuk mengidentifikasi faktor-faktor yang berkontribusi terhadap kesalahan pengkodean ICD-10 melalui pendekatan *machine learning*. Tahapan penelitian dijelaskan sebagai berikut.

1. Pengumpulan Data

Penelitian diawali dengan proses pengumpulan data sebagai tahap utama dalam pengembangan model *machine learning*. Dataset penelitian diperoleh dari RME yang berasal dari rumah sakit. Data yang dikumpulkan mencakup informasi diagnosis pasien, kode ICD-10, serta variabel pendukung yang berhubungan dengan proses pengkodean diagnosis. Untuk memastikan kualitas data yang digunakan, dilakukan proses pembersihan data (*data cleaning*) dengan menghapus data yang tidak lengkap, tidak valid, atau terduplikasi. Tahap ini bertujuan untuk meningkatkan kualitas dataset sehingga proses analisis dan pengembangan model *machine learning* dapat dilakukan secara lebih akurat dan reliabel.

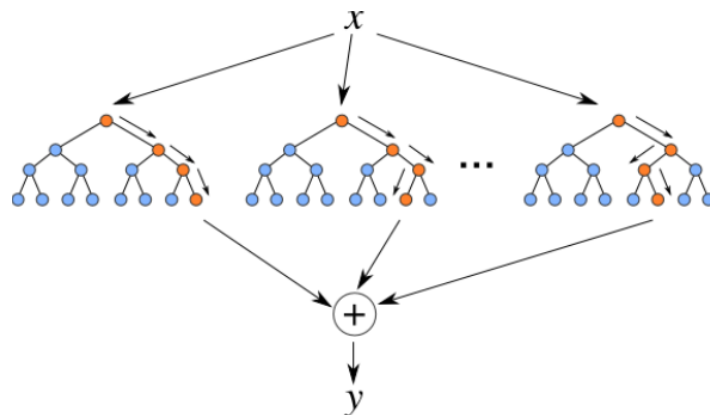
2. Processing Data

Setelah proses pembersihan data selesai dilakukan, tahap berikutnya adalah pemrosesan data untuk menyiapkan fitur yang akan digunakan dalam pemodelan *machine learning*. Data diagnosis klinis yang berbentuk teks naratif diolah menggunakan teknik *Natural Language Processing* (NLP) agar dapat dikonversi menjadi representasi numerik yang dapat diproses oleh algoritma *machine learning* [8]. Proses ekstraksi fitur dilakukan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Metode ini mengubah teks diagnosis menjadi vektor numerik berdasarkan frekuensi kemunculan kata dan tingkat kepentingannya dalam keseluruhan dokumen. Representasi numerik tersebut memungkinkan algoritma *machine learning* mengidentifikasi pola-pola yang berkaitan dengan kesalahan pengkodean ICD-10 secara lebih efektif.

Setelah proses ekstraksi fitur selesai, dataset dibagi menjadi data pelatihan (*training set*) dan data pengujian (*testing set*) dengan proporsi 80:20. Data pelatihan digunakan untuk membangun model *machine learning*, sedangkan data pengujian digunakan untuk mengevaluasi kemampuan model dalam mendeteksi kesalahan pengkodean pada data yang belum pernah diproses sebelumnya. Untuk meningkatkan reliabilitas model dan mengurangi risiko *overfitting*, penelitian ini menerapkan teknik *k-fold cross-validation*. Pada metode ini, dataset dibagi menjadi beberapa bagian (*fold*), kemudian proses pelatihan dan pengujian dilakukan secara bergantian hingga seluruh bagian data digunakan sebagai data uji. Pendekatan ini memungkinkan evaluasi model yang lebih stabil dan objektif dalam mengukur performa prediksi kesalahan pengkodean ICD-10.

3. Pemilihan dan Pengembangan Model *Machine learning*

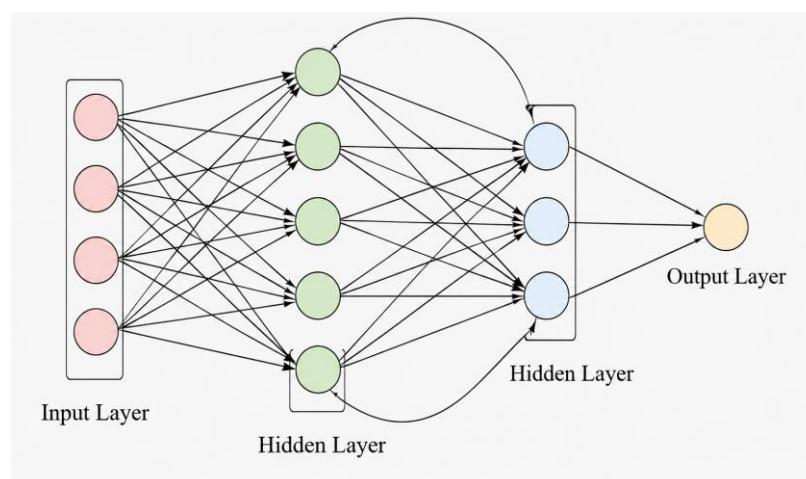
Setelah data diproses dan fitur yang diperlukan telah disiapkan, tahap selanjutnya adalah pemilihan dan pengembangan model *machine learning*. Beberapa model *machine learning* yang umum digunakan dalam pengkodean ICD-10 akan dipilih, termasuk model berbasis *Neural Network* seperti *Recurrent Neural Network* (RNN) serta model *Support Vector Machine* (SVM) [8]. Model-model tersebut akan digunakan untuk memprediksi kode ICD-10 berdasarkan data yang telah diproses. Selanjutnya, teknik seperti *cross-validation* dan *hyperparameter tuning* akan diterapkan untuk memperoleh model dengan kinerja terbaik.



Gambar 1. Desain Konseptual Model Algoritma *Random Forest*

Model *Random Forest* dikembangkan sebagai pendekatan ensemble learning yang menggabungkan sejumlah pohon keputusan (*decision tree*) untuk menghasilkan prediksi yang lebih stabil dan akurat. Secara konseptual, proses diawali dengan tahap praproses data yang meliputi pembersihan data, normalisasi, seleksi fitur, dan transformasi variabel guna memastikan kualitas data masukan yang optimal. Selanjutnya, dataset dibagi menjadi beberapa subset menggunakan teknik *bootstrap sampling* yang digunakan untuk melatih setiap pohon keputusan secara independen. Pada setiap pohon, pemilihan fitur dilakukan secara acak pada setiap simpul (*node*) melalui mekanisme *random feature selection*, sehingga meningkatkan keberagaman struktur model dan mengurangi korelasi antarpohon.

Setiap pohon menghasilkan prediksi secara mandiri, kemudian seluruh prediksi tersebut digabungkan melalui mekanisme agregasi berbasis *majority voting* untuk tugas klasifikasi. Proses agregasi ini memungkinkan *Random Forest* menghasilkan keputusan akhir yang lebih *robust* terhadap *noise* dan mengurangi risiko *overfitting*. Dalam penelitian ini, *Random Forest* digunakan sebagai model analitik untuk mengidentifikasi pola hubungan antarvariabel, menentukan faktor-faktor dominan penyebab kesalahan pengkodean ICD-10, serta memprediksi risiko kesalahan pengkodean secara sistematis dan berbasis data. Dengan demikian, model ini diharapkan dapat mendukung peningkatan akurasi pengkodean dan kualitas manajemen informasi kesehatan.

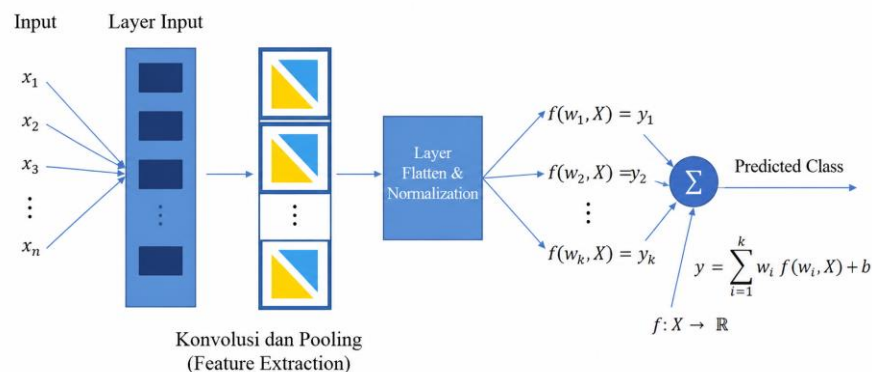


Gambar 2. Desain Konseptual Model Algoritma *Neural Network*

Model *Neural Network* dalam penelitian ini dirancang sebagai pendekatan *supervised learning* untuk mengidentifikasi pola kompleks dan hubungan *nonlinier* antarvariabel yang berkontribusi terhadap kesalahan pengkodean ICD-10. Secara konseptual, arsitektur model terdiri atas tiga komponen utama, yaitu *input layer*, satu atau lebih *hidden layer*, dan *output layer*. Tahap awal meliputi praproses data yang mencakup pembersihan data, penanganan *missing value*, normalisasi atau standardisasi variabel, serta pembagian dataset menjadi data latih dan data uji.

Variabel yang telah diproses kemudian dimasukkan ke dalam *input layer* untuk diteruskan ke *hidden layer*. Pada setiap neuron dalam *hidden layer*, dilakukan perhitungan menggunakan bobot (*weight*), bias, dan fungsi aktivasi, seperti ReLU atau *sigmoid*, untuk menghasilkan representasi pola yang lebih abstrak. Proses pelatihan model dilakukan menggunakan algoritma *backpropagation* dengan optimisasi berbasis *gradient descent* guna meminimalkan nilai fungsi kerugian (*loss function*). Hasil prediksi kemudian dihasilkan pada *output layer* dalam bentuk probabilitas klasifikasi risiko kesalahan pengkodean.

Secara konseptual, model ini memungkinkan identifikasi hubungan kompleks antar faktor penyebab kesalahan pengkodean ICD-10 serta meningkatkan akurasi prediksi melalui proses pembelajaran iteratif berbasis data. Dengan demikian, model *Neural Network* dapat mendukung peningkatan mutu pengkodean dan kualitas manajemen informasi kesehatan.



Gambar 3. Desain Konseptual Model *Support Vector Machine*

Model *Support Vector Machine* (SVM) dalam penelitian ini dirancang sebagai metode *supervised learning* yang bertujuan untuk mengklasifikasikan data secara optimal dengan membentuk batas pemisah (*hyperplane*) terbaik antarkelas. Secara konseptual, SVM bekerja dengan mencari garis atau bidang pemisah yang memiliki margin maksimum, yaitu jarak terjauh antara *hyperplane* dan titik data terdekat dari masing-masing kelas yang dikenal sebagai *support vector*. Tahap awal model meliputi praproses data, seperti pembersihan data, penanganan nilai hilang, serta normalisasi variabel untuk memastikan keseragaman skala data. Selanjutnya, dataset dibagi menjadi data latih dan data uji.

Pada tahap pelatihan, algoritma SVM mengoptimalkan fungsi objektif untuk menemukan *hyperplane* yang mampu memisahkan kelas kesalahan dan nonkesalahan pengkodean ICD-10 secara optimal. Dengan pendekatan tersebut, model SVM dapat mengidentifikasi pola klasifikasi pada data dan digunakan untuk memprediksi risiko kesalahan pengkodean berdasarkan karakteristik data yang telah diproses.

4. Evaluasi Model

Setelah model dipilih, tahap selanjutnya adalah evaluasi model. Tahap ini sangat penting untuk memastikan bahwa model yang dikembangkan mampu memprediksi kode ICD-10 secara akurat. Evaluasi dilakukan menggunakan metrik yang umum digunakan dalam *machine learning*, yaitu akurasi (*accuracy*), presisi (*precision*), *recall*, dan skor F1 (*F1-Score*) [7]. Metrik-metrik tersebut memberikan gambaran yang lebih komprehensif mengenai kinerja model.

Akurasi mengukur proporsi prediksi yang benar terhadap seluruh data yang dievaluasi. Presisi mengukur proporsi prediksi positif yang benar dibandingkan dengan seluruh prediksi positif yang dihasilkan oleh model. Sementara itu, *recall* mengukur kemampuan model dalam mengidentifikasi seluruh kasus positif yang sebenarnya, sedangkan skor F1 merupakan rata-rata harmonis antara presisi dan *recall* yang digunakan untuk menilai keseimbangan kinerja model pada kedua metrik tersebut.

5. Analisis Faktor Penyebab Kesalahan Diagnosa

Setelah model dievaluasi menggunakan data latih dan data uji, tahap selanjutnya adalah melakukan analisis terhadap faktor-faktor yang menyebabkan kesalahan pengkodean diagnosis. Salah satu penyebab kesalahan adalah adanya kode diagnosis yang memiliki kemiripan atau tumpang tindih karakteristik, sehingga berpotensi menimbulkan kesalahan dalam pemilihan kode. Selain itu, faktor manusia juga dapat berkontribusi terhadap terjadinya kesalahan, seperti kurangnya ketelitian atau keterbatasan dalam menginterpretasikan data yang kompleks [14]. Oleh karena itu, analisis kesalahan menjadi penting untuk mengidentifikasi jenis kesalahan yang paling sering terjadi serta menentukan upaya yang dapat dilakukan untuk mengurangi atau mencegah kesalahan tersebut pada masa mendatang.

6. Penyusunan Rekomendasi

Setelah analisis kesalahan selesai dilakukan, tahap selanjutnya adalah implementasi model *machine learning* yang telah teruji untuk meningkatkan akurasi pengkodean ICD-10 secara otomatis. Model yang telah dievaluasi dan disempurnakan berdasarkan hasil analisis kesalahan kemudian diterapkan dalam sistem pengkodean guna mendukung proses pengkodean diagnosis yang lebih efisien dan akurat. Penyempurnaan sistem dilakukan berdasarkan temuan selama tahap evaluasi untuk mengurangi kemungkinan terjadinya kesalahan pada masa mendatang. Dengan demikian, implementasi model diharapkan dapat mendukung peningkatan kualitas pengkodean serta pengelolaan informasi kesehatan secara lebih optimal.

3. HASIL DAN ANALISIS

3.1 Analisis Deskriptif Awal

Analisis deskriptif dilakukan untuk mengidentifikasi gambaran umum kesalahan pengkodean diagnosis ICD-10.

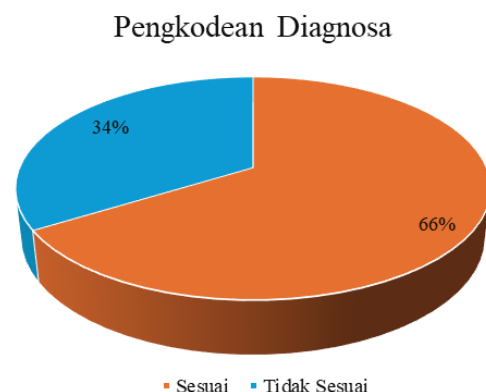
1. Persentase Kesalahan Pengkodean ICD-10

Tabel 2. Distribusi Kesesuaian dan Ketidakesuaian Akurasi Pengkodean

Kategori	Jumlah	Persentase (%)	Kasus
Kode diagnosis sesuai (akurat)	397	66,2	Diagnosis klinis jelas dan kode yang diberikan sesuai dengan ICD-10
Kode diagnosis tidak sesuai (tidak akurat)	203	33,8	Ketidakesuaian antara diagnosis dokter dan kode yang diberikan. Misalnya diagnosis “hipertensi esensial” tetapi diberi kode hipertensi sekunder

Berdasarkan Tabel 2 mengenai distribusi kesesuaian dan ketidakesuaian akurasi pengkodean, diketahui bahwa dari total 600 kasus yang dianalisis, sebanyak 397 kasus (66,2%) memiliki kode diagnosis yang sesuai dengan diagnosis klinis yang tercantum dalam rekam medis. Hasil tersebut menunjukkan bahwa sebagian besar proses pengkodean telah dilakukan sesuai dengan standar klasifikasi penyakit yang berlaku, yaitu ICD-10.

Meskipun demikian, masih terdapat 203 kasus (33,8%) dengan kode diagnosis yang tidak sesuai. Ketidakesuaian tersebut terjadi karena adanya perbedaan antara diagnosis yang ditetapkan oleh dokter dan kode yang diberikan oleh petugas pengkodean. Sebagai contoh, diagnosis hipertensi esensial pada beberapa kasus ditemukan diberi kode hipertensi sekunder. Temuan ini menunjukkan bahwa meskipun tingkat akurasi pengkodean tergolong cukup baik, peningkatan ketelitian dalam proses pengkodean serta evaluasi terhadap kualitas dokumentasi diagnosis masih diperlukan untuk meminimalkan terjadinya kesalahan pengkodean.



Gambar 4. Persentase Kesalahan Pengkodean Diagnosis ICD-10

Berdasarkan hasil analisis kesesuaian pengkodean diagnosis, diperoleh gambaran bahwa sebagian besar kode diagnosis yang diberikan telah sesuai dengan standar pengkodean yang berlaku. Hal ini ditunjukkan oleh distribusi pengkodean yang didominasi oleh kategori sesuai dibandingkan dengan kategori tidak sesuai. Pada Gambar 4 terlihat bahwa proporsi pengkodean diagnosis yang sesuai lebih besar dibandingkan dengan pengkodean diagnosis yang tidak sesuai.

2. Faktor Dominan yang Teridentifikasi Secara Deskriptif

Berdasarkan hasil observasi dan wawancara dengan verifikator kode diagnosis, beberapa faktor yang menjadi penyebab kesalahan pengkodean diagnosis meliputi diagnosis yang tidak spesifik sebanyak 92 kasus (45,3%), keterbatasan informasi klinis sebanyak 68 kasus (33,5%), dan kesalahan administratif pada saat entri data sebanyak 43 kasus (21,2%). Temuan ini menunjukkan bahwa kualitas dokumentasi klinis merupakan faktor yang paling dominan dalam memengaruhi ketepatan pengkodean diagnosis.

3. Distribusi Kesalahan Berdasarkan Jenis Kelamin dan Usia

Tabel 3. Distribusi Kesalahan Pengkodean ICD-10 Berdasarkan Jenis Kelamin

Rentang Kesalahan (%)	Laki-laki	Perempuan
< 1 atau Kosong	0	0
1–20	18	25
21–40	23	31
41–60	13	20
61–80	30	24
81–100	12	8
Total	96	108

Berdasarkan tabel di atas, terlihat bahwa jumlah kesalahan pada pasien perempuan cenderung lebih tinggi dibandingkan pada pasien laki-laki pada beberapa kelompok usia, terutama usia 21–40 tahun. Sementara itu, pada kelompok usia yang lebih tinggi, distribusi kesalahan relatif lebih bervariasi. Perbedaan tersebut tidak secara langsung menunjukkan bahwa jenis kelamin memengaruhi akurasi pengkodean.

Temuan ini lebih menggambarkan perbedaan jumlah kasus dan kompleksitas diagnosis pada masing-masing kelompok pasien. Oleh karena itu, variasi kesalahan pengkodean berdasarkan jenis kelamin dan usia kemungkinan lebih berkaitan dengan karakteristik kasus yang tercatat dalam rekam medis. Upaya peningkatan akurasi pengkodean perlu difokuskan pada peningkatan kualitas dokumentasi klinis serta kompetensi petugas dalam melakukan pengkodean diagnosis.

3.2 Analisis dengan *Machine learning*

3.2.1 Variabel Independen

Variabel independen yang digunakan meliputi usia, jenis kelamin, hasil laboratorium, riwayat penyakit, serta diagnosis klinis. Pemilihan variabel tersebut didasarkan pada pertimbangan bahwa faktor demografis maupun klinis memiliki peran penting dalam menentukan ketepatan pengkodean diagnosis. Usia dan jenis kelamin sering kali berhubungan dengan prevalensi penyakit tertentu, sedangkan riwayat

penyakit dan hasil laboratorium memberikan konteks tambahan yang mendukung penetapan kode sesuai standar ICD-10. Diagnosis klinis menjadi variabel kunci karena merupakan titik awal dalam proses pengkodean medis, sehingga kualitas data diagnosis sangat menentukan akurasi kode yang dihasilkan.

Pemilihan variabel tersebut didasarkan pada konsep bahwa proses pengkodean diagnosis sangat dipengaruhi oleh kelengkapan dan kejelasan informasi klinis yang terdapat dalam rekam medis. Informasi demografi pasien, seperti usia dan jenis kelamin, dapat memengaruhi pola penyakit, sedangkan riwayat penyakit, hasil laboratorium, dan diagnosis klinis merupakan komponen penting dalam penegakan diagnosis yang menjadi dasar proses pengkodean penyakit. Hal ini sejalan dengan pedoman klasifikasi penyakit yang menyatakan bahwa pengkodean diagnosis harus didasarkan pada diagnosis yang telah ditegakkan secara klinis serta didukung oleh informasi medis yang relevan dalam rekam medis. Selain itu, penelitian sebelumnya juga menunjukkan bahwa kualitas dokumentasi klinis, termasuk kelengkapan data diagnosis dan hasil pemeriksaan penunjang, berpengaruh terhadap tingkat akurasi pengkodean penyakit dalam sistem klasifikasi penyakit internasional [15].

3.2.2 Variabel dependen

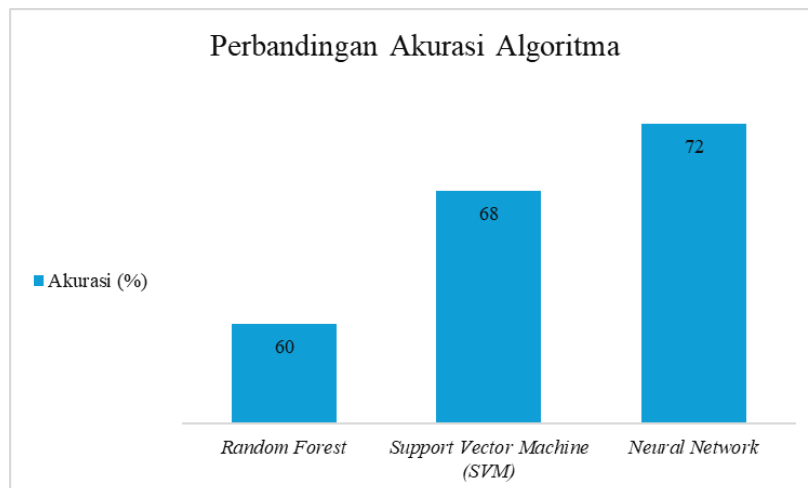
Variabel dependen ditetapkan berupa status pengkodean yang diklasifikasikan menjadi benar atau salah. Kategori ini digunakan untuk mengukur sejauh mana kode diagnosis yang dihasilkan sesuai dengan ketentuan ICD-10 dan standar pengkodean medis. Dengan adanya pengelompokan biner ini, analisis statistik maupun algoritma *machine learning* dapat diterapkan untuk mengevaluasi pengaruh variabel independen terhadap tingkat kesalahan pengkodean. Pendekatan ini sejalan dengan penelitian sebelumnya yang menekankan pentingnya validasi pengkodean diagnosis sebagai indikator mutu data kesehatan.

3.2.3 Algoritma yang digunakan

Untuk meningkatkan ketepatan analisis, beberapa algoritma *machine learning* diterapkan, yaitu *Random Forest*, *Support Vector Machine* (SVM), dan *Neural Network*. *Random Forest* dikenal memiliki kemampuan tinggi dalam menangani variabel kompleks dan interaksi antarvariabel, sedangkan SVM efektif dalam membedakan kategori data dengan margin yang jelas. *Neural Network* unggul dalam mengenali pola nonlinier yang sulit diidentifikasi oleh algoritma tradisional [1], [7]. Pemanfaatan kombinasi algoritma tersebut diharapkan dapat memberikan gambaran yang komprehensif mengenai faktor-faktor yang paling berkontribusi terhadap terjadinya kesalahan pengkodean diagnosis. Sejalan dengan penelitian sebelumnya bahwa integrasi algoritma *machine learning* telah terbukti mampu meningkatkan akurasi prediksi dan mendukung proses pengendalian mutu data kesehatan [8].

Hasil pengujian menunjukkan bahwa *Neural Network* memberikan akurasi tertinggi (72%) dalam memprediksi kesalahan pengkodean, diikuti oleh *Support Vector Machine* (68%) dan *Random Forest* (60%). Untuk mengetahui apakah perbedaan performa tersebut signifikan secara statistik, dilakukan pengujian menggunakan metode *Friedman Test*. Hasil pengujian menunjukkan nilai χ^2 sebesar 9,27 dengan nilai *p-value* sebesar 0,009. Karena nilai *p-value* lebih kecil dari tingkat signifikansi 0,05, dapat disimpulkan bahwa terdapat perbedaan performa yang signifikan antara model yang diuji.

Selanjutnya, dilakukan uji lanjutan menggunakan metode *Nemenyi Post Hoc Test* untuk mengetahui pasangan model yang memiliki perbedaan signifikan. Hasil pengujian menunjukkan bahwa terdapat perbedaan signifikan antara model *Random Forest* dan SVM ($p = 0,031$) serta antara *Random Forest* dan *Neural Network* ($p = 0,004$). Namun, perbandingan antara model SVM dan *Neural Network* tidak menunjukkan perbedaan yang signifikan secara statistik ($p = 0,087$). Temuan ini menunjukkan bahwa model *Neural Network* memiliki performa yang lebih baik dibandingkan model lainnya dalam melakukan klasifikasi data pada penelitian ini, meskipun secara statistik performanya tidak berbeda signifikan dengan model SVM.



Gambar 5. Grafik Perbandingan Algoritma *Machine learning*

Gambar 5 menunjukkan perbandingan tingkat akurasi tiga algoritma *machine learning* yang digunakan dalam proses klasifikasi atau prediksi. Berdasarkan grafik, algoritma *Random Forest* memiliki tingkat akurasi sebesar 60%, *Support Vector Machine* (SVM) sebesar 68%, dan *Neural Network* menunjukkan akurasi tertinggi, yaitu 72%. Hasil ini menunjukkan bahwa model *Neural Network* memiliki performa yang lebih baik dibandingkan dua algoritma lainnya dalam menghasilkan prediksi yang akurat. Sementara itu, SVM berada pada tingkat akurasi menengah, dan *Random Forest* memiliki tingkat akurasi paling rendah di antara ketiga algoritma yang dibandingkan. Temuan ini mengindikasikan bahwa pemilihan algoritma memengaruhi tingkat keberhasilan model dalam memprediksi atau mengklasifikasikan data.

1. *Neural Network* (Akurasi 72%, *Precision* 0,73, *Recall* 0,71, *F1-Score* 0,72)

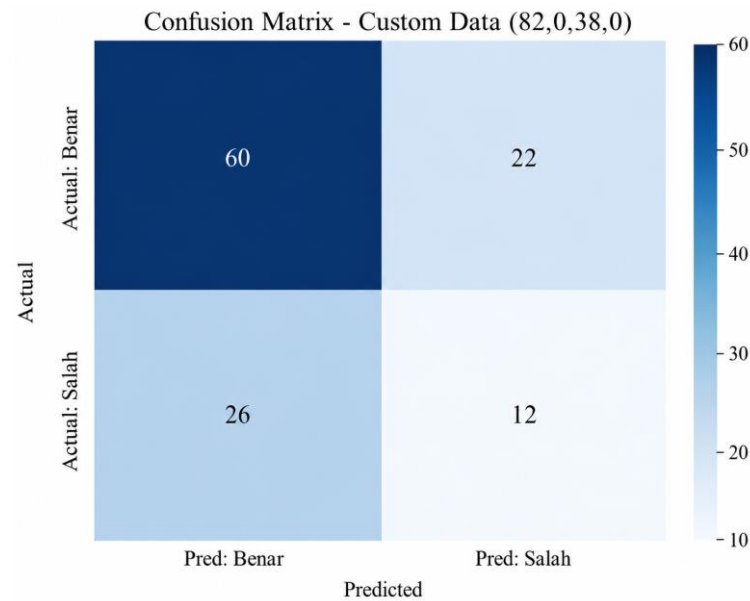
Pencapaian akurasi tertinggi menunjukkan bahwa model ini lebih mampu menangkap pola kompleks dalam data terkait faktor penyebab kesalahan pengkodean ICD-10. Nilai *precision* sebesar 0,73 menunjukkan bahwa sebagian besar prediksi yang dihasilkan model sebagai kesalahan pengkodean memang benar terjadi. Sementara itu, nilai *recall* sebesar 0,71 menunjukkan kemampuan model dalam mendeteksi sebagian besar kasus kesalahan yang sebenarnya terjadi dalam dataset. Nilai *F1-Score* sebesar 0,72 mengindikasikan keseimbangan yang cukup baik antara *precision* dan *recall*. Hal ini sesuai dengan karakteristik *Neural Network* yang unggul dalam mempelajari hubungan *nonlinier* antarvariabel, sehingga cocok diterapkan pada data kesehatan yang umumnya memiliki banyak dimensi dan variasi.

2. *Support Vector Machine* (Akurasi 68%, *Precision* 0,69, *Recall* 0,67, *F1-Score* 0,68)

SVM menunjukkan performa yang cukup baik dan relatif stabil dibandingkan *Random Forest*. Nilai *precision* sebesar 0,69 menunjukkan bahwa model cukup akurat dalam mengklasifikasikan kasus pengkodean yang salah. Nilai *recall* sebesar 0,67 menunjukkan kemampuan model dalam mengidentifikasi sebagian besar kasus kesalahan pengkodean dalam data. Nilai *F1-Score* sebesar 0,68 menunjukkan keseimbangan performa klasifikasi yang cukup baik. Kinerja SVM yang mendekati *Neural Network* menunjukkan bahwa algoritma ini efektif dalam memisahkan kelas benar dan salah dalam pengkodean, terutama pada dataset dengan margin yang jelas. Namun, keterbatasan SVM umumnya muncul pada dataset besar dengan banyak fitur karena membutuhkan waktu komputasi yang lebih tinggi.

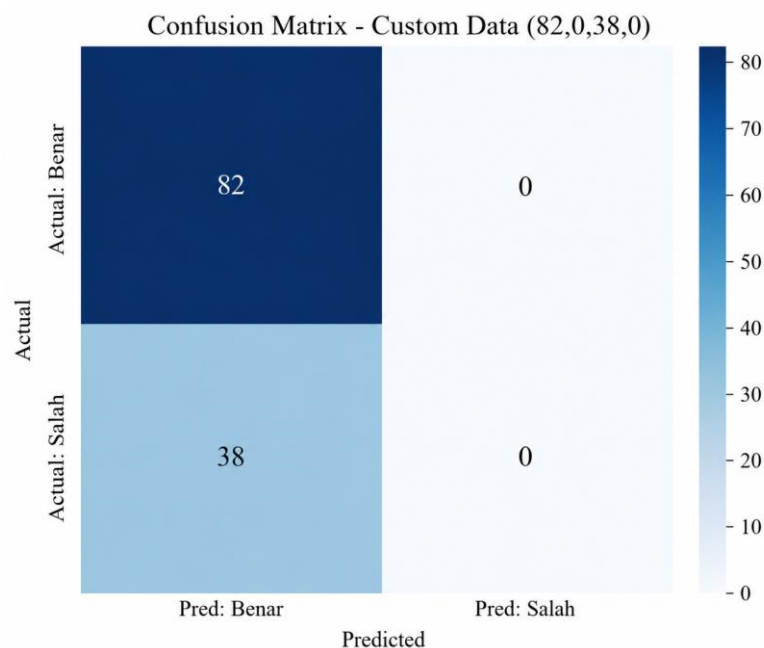
3. *Random Forest* (Akurasi 60%, *Precision* 0,61, *Recall* 0,59, *F1-Score* 0,60)

Model *Random Forest* menunjukkan performa yang lebih rendah dibandingkan dua model lainnya. Nilai *precision* sebesar 0,61 menunjukkan bahwa sebagian prediksi kesalahan pengkodean yang dihasilkan model masih kurang tepat. Nilai *recall* sebesar 0,59 mengindikasikan bahwa model belum mampu mendeteksi seluruh kasus kesalahan pengkodean secara optimal. Nilai *F1-Score* sebesar 0,60 menunjukkan keseimbangan yang masih relatif rendah antara *precision* dan *recall*. Akurasi yang lebih rendah ini menunjukkan bahwa performa *Random Forest* pada dataset penelitian ini masih berada di bawah SVM dan *Neural Network*. Meskipun *Random Forest* dikenal tangguh terhadap *overfitting* dan mampu menangani data dengan missing value, hasil penelitian ini menunjukkan bahwa model berbasis pohon keputusan kurang efektif dibandingkan model berbasis kernel (SVM) dan model berbasis jaringan (*Neural Network*) pada kasus penelitian ini.

Gambar 6. *Confusion Matrix Random Forest*

Gambar 6 menunjukkan *Confusion Matrix* yang digunakan untuk mengevaluasi kinerja model *Random Forest* dalam mengklasifikasikan kategori benar dan salah. Berdasarkan matriks tersebut, model berhasil mengklasifikasikan 60 data kategori benar secara tepat (*true positive*) dan 12 data kategori salah secara tepat (*true negative*). Namun, masih terdapat 22 data kategori benar yang diklasifikasikan sebagai salah (*false negative*) serta 26 data kategori salah yang diklasifikasikan sebagai benar (*false positive*).

Hasil tersebut menunjukkan bahwa model *Random Forest* mampu mengklasifikasikan sebagian data dengan benar, tetapi masih menghasilkan sejumlah kesalahan klasifikasi pada kedua kategori. Jumlah *false positive* dan *false negative* yang relatif tinggi menunjukkan bahwa model belum mampu membedakan secara optimal antara data dengan pengkodean benar dan salah. Kondisi ini sejalan dengan nilai akurasi, *precision*, *recall*, dan *F1-Score* yang lebih rendah dibandingkan model *SVM* dan *Neural Network*. Oleh karena itu, diperlukan penyempurnaan model atau optimalisasi parameter agar kemampuan klasifikasi dapat ditingkatkan.

Gambar 7. *Confusion Matrix Support Vector Machine (SVM)*

Gambar 7 menunjukkan *Confusion Matrix* yang digunakan untuk mengevaluasi kinerja model *Support Vector Machine* (SVM) dalam mengklasifikasikan kategori benar dan salah. Berdasarkan matriks tersebut, model berhasil mengklasifikasikan 82 data kategori benar secara tepat sebagai benar. Namun, terdapat 38 data kategori salah yang juga diklasifikasikan sebagai benar, sehingga tidak terdapat data yang diprediksi sebagai kategori salah.

Hasil tersebut menunjukkan bahwa model SVM memiliki kecenderungan untuk mengklasifikasikan seluruh data ke dalam kategori benar. Meskipun kondisi ini menghasilkan jumlah prediksi benar yang tinggi pada kategori mayoritas, kemampuan model dalam mengidentifikasi kategori salah masih terbatas. Dengan demikian, model belum mampu memberikan pemisahan yang optimal antara kedua kategori pada dataset yang digunakan.

Meskipun demikian, berdasarkan hasil evaluasi secara keseluruhan, model SVM tetap menunjukkan performa yang lebih baik dibandingkan *Random Forest* dengan tingkat akurasi sebesar 68%. Hal ini menunjukkan bahwa algoritma SVM memiliki kemampuan yang cukup baik dalam mengenali pola data yang digunakan dalam penelitian ini. Namun, diperlukan penyesuaian parameter model, penyeimbangan distribusi kelas, atau optimalisasi proses pelatihan agar kemampuan deteksi terhadap kategori salah dapat ditingkatkan sehingga menghasilkan performa klasifikasi yang lebih seimbang.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa faktor demografis dan klinis, seperti usia, jenis kelamin, riwayat penyakit, hasil laboratorium, serta diagnosis klinis, berpengaruh terhadap ketepatan pengkodean diagnosis. Diagnosis klinis menjadi variabel paling penting karena merupakan dasar utama dalam penetapan kode ICD-10. Status pengkodean yang diklasifikasikan sebagai benar atau salah memungkinkan evaluasi kualitas pengkodean secara objektif. Kinerja model dievaluasi menggunakan akurasi (*accuracy*) dan presisi (*precision*). Akurasi menunjukkan proporsi prediksi yang benar terhadap seluruh data, sedangkan presisi menunjukkan ketepatan model dalam mengidentifikasi kasus pengkodean yang sesuai dengan standar ICD-10.

Penerapan algoritma *machine learning*, yaitu *Random Forest*, *Support Vector Machine* (SVM), dan *Neural Network*, menunjukkan bahwa teknologi prediktif dapat digunakan untuk mengidentifikasi faktor penyebab kesalahan pengkodean. Hasil pengujian menunjukkan bahwa *Neural Network* memiliki performa terbaik dengan akurasi 72%, diikuti SVM 68% dan *Random Forest* 60%. Temuan ini menunjukkan bahwa *Neural Network* lebih efektif dalam menangkap pola kompleks pada data klinis.

REFERENSI

- [1] E. Nurmambetova *et al.*, "Developing an Inpatient Electronic Medical Record Phenotype for Hospital-Acquired Pressure Injuries: Case Study Using Natural Language Processing Models," *JMIR AI*, vol. 2, pp. 1–15, Mar. 2023, doi: 10.2196/41264.
- [2] I. Kartika and A. Sukawan, "Akurasi Pengodean Injury dan External Causes Berdasarkan ICD-10 dan ICD-9 CM di Rumah Sakit Jasa Kartini," *J. Manaj. Inf. Kesehat. Indones.*, vol. 13, no. 2, pp. 206–211, 2025, doi: <https://doi.org/10.33560/jmiki.v13i2.676>.
- [3] R. Andriani, "Akurasi Kode ICD-10 Kasus Pemeriksaan Kehamilan Pada Rekam Medis Elektronik," *JMIK J. Manaj. Inf. dan Adm. Kesehat.*, vol. 4, no. 2, pp. 59–68, 2021, doi: <https://doi.org/10.32585/jmiak.v4i2.2015>.
- [4] E. S. D. Hastuti and M. Ali, "Faktor-faktor yang Berpengaruh pada Akurasi Kode diagnosis di Puskesmas Rawat Jalan Kota Malang," *J. Kedokt. Brawijaya*, vol. 30, no. 3, pp. 228–234, Feb. 2019, doi: 10.21776/ub.jkb.2019.030.03.12.
- [5] D. Ekawati, A. A. Nuryadin, S. Windarti, and A. Zulfikar, "Peran Coder Terkait Keakuratan Pengkodean ICD-10 Pada Pelaksanaan Sistem INA-CBG's Rumah Sakit TK. II 14.05.01 Pelamonia Makassar," *J. Manaj. Bisnis dan Kesehat.*, vol. 1, no. 2, Oct. 2024, doi: 10.70817/jmbk.v1i2.13.
- [6] M. O. D. Putri and E. Krisnawati, "Literature Review Analisis Faktor yang Mempengaruhi Ketepatan Pemberian Kodefikasi Penyakit," *J. Manaj. Inf. Kesehat.*, vol. 1, no. 1, pp. 8–16, Jan. 2026, doi: 10.36916/j-mika.v1i1.382.
- [7] X. Diao *et al.*, "Automated ICD coding for primary diagnosis via clinically interpretable machine learning," *Int. J. Med. Inform.*, vol. 153, pp. 1–9, Sep. 2021, doi: 10.1016/j.ijmedinf.2021.104543.
- [8] S.-M. Wang *et al.*, "Using Deep Learning for Automatic Icd-10 Classification from Free Text Data," *Eur. J. Biomed. Informatics*, vol. 16, pp. 1–10, 2020, doi: 10.24105/ejbi.2020.16.1.1.
- [9] M. Almagro, R. Martínez, S. Montalvo, and V. Fresno, "A cross-lingual approach to automatic ICD-10

- coding of death certificates by exploring machine translation,” *J. Biomed. Inform.*, vol. 94, pp. 1–7, Jun. 2019, doi: 10.1016/j.jbi.2019.103207.
- [10] C. Sen, B. Ye, J. Aslam, and A. Tahmasebi, “From Extreme Multi-label to Multi-class: A Hierarchical Approach for Automated ICD-10 Coding Using Phrase-level Attention,” *arXiv Prepr. arXiv2102.09136*, pp. 1–9, 2021, doi: <https://doi.org/10.48550/arXiv.2102.09136>.
- [11] L. Zhou, C. Cheng, D. Ou, and H. Huang, “Construction of a semi-automatic ICD-10 coding system,” *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, pp. 1–12, Dec. 2020, doi: 10.1186/s12911-020-1085-4.
- [12] B. Volkov and G. Kopanitsa, “Development of a Method for Automatic Matching of Unstructured Medical Data to ICD-10 Codes,” in *Proceeding of the International Conference on Wearable Micro and Nano Technologies for Personalized Health*, May 2024, pp. 93–97. doi: 10.3233/SHTI240065.
- [13] Y. Kim, C. Li, B. Ye, A. Tahmasebi, and J. Aslam, “Supervised Learning in the Presence of Noise: Application in ICD-10 Code Classification,” *arXiv Prepr. arXiv2103.07808*, pp. 1–9, 2021.
- [14] E. Cardillo and L. Frattura, “Assisted morbidity coding: the SISCO.web use case for identifying the main diagnosis in Hospital Discharge Records,” *arXiv Prepr. arXiv2412.09651*, pp. 51–78, 2024, doi: <https://doi.org/10.48550/arXiv.2412.09651>.
- [15] S. Bowman, “Impact of Electronic Health Record Systems on Information Integrity: Quality and Safety Implications,” *Perspect. Heal. Inf. Manag.*, vol. 10, no. 1c, pp. 1–19, 2013.