KLASIFIKASI PENYAKIT ASMA DENGAN MENGGUNAKAN *TEACHING LEARNING BASED OPTIMIZATION* (TLBO) DAN MACHINE LEARNING

Ratu Tarisya Putri
Informatics engineering
Faculty of Computer
Sriwijaya University
South Sumatera, Indonesia
ratutarisya43@gmail.com

Annisa Darmawahyuni
Inteligent System Research Group
Faculty of Computer
Sriwijaya University
South Sumatera, Indonesia
riset.annisadarmawahyuni@gmail.com

Abstract*— Penelitian ini bertujuan mengembangkan model klasifikasi penyakit asma dengan memanfaatkan Teaching Learning Based Optimization (TLBO) sebagai metode seleksi fitur sebelum proses klasifikasi. Dataset diperoleh dari platform Kaggle dan melalui tahapan preprocessing berupa pembersihan data, normalisasi, serta penanganan ketidakseimbangan data dengan tiga pendekatan, yaitu tanpa balancing (original), random undersampling, dan Synthetic Minority Oversampling Technique (SMOTE). Tiga algoritma klasifikasi digunakan, yaitu K-Nearest Neighbor (KNN), Support Vector Machine (SVM), dan Random Forest (RF), dengan tuning parameter untuk memperoleh hasil optimal. Evaluasi dilakukan menggunakan metrik akurasi, sensitivitas, spesifisitas, presisi, dan f1-score. Hasil pengujian menunjukkan bahwa model SVM dengan kernel RBF, C=10, dan gamma=auto pada skenario SMOTE mencapai performa terbaik dengan akurasi di atas 99%, sensitivitas dan spesifisitas di atas 90%. Temuan ini menunjukkan bahwa kombinasi TLBO dengan algoritma klasifikasi serta strategi balancing data yang tepat mampu meningkatkan performa prediksi penyakit asma secara signifikan.

Keywords*—Klasifikasi Asma, Seleksi Fitur, Teaching Learning Based Optimization (TLBO),Machine Learning, SMOTE, Support Vector Machine (SVM)

I. PENDAHULUAN

Penyakit asma merupakan salah satu penyakit pernapasan kronis yang menyerang saluran pernapasan dan dapat menurunkan kualitas hidup penderitanya. Menurut World Health Organization (WHO), lebih dari 260 juta orang di seluruh dunia menderita asma, dan jumlah ini terus meningkat setiap tahun. Gejala asma sangat bervariasi, mulai dari sesak napas ringan hingga serangan akut yang berpotensi membahayakan nyawa. Oleh karena itu, deteksi dini dan diagnosis yang akurat menjadi kunci dalam penanganan asma secara efektif

Namun demikian, proses diagnosis penyakit asma tidak selalu mudah dilakukan. Gejalanya sering kali mirip dengan penyakit pernapasan lainnya, dan dalam beberapa wilayah, keterbatasan alat serta minimnya pengalaman tenaga medis turut menjadi hambatan. Hal ini mendorong perlunya solusi alternatif berbasis teknologi untuk mendukung proses diagnosis, khususnya melalui pendekatan klasifikasi data medis secara otomatis.

Perkembangan teknologi informasi, khususnya dalam bidang kecerdasan buatan, telah membawa kemajuan signifikan dalam dunia medis. Teknologi ini memudahkan proses diagnosis penyakit, meningkatkan efisiensi kerja tenaga medis, dan mempercepat pengambilan keputusan berbasis data. Berdasarkan penelitian[1], penerapan teknologi digital terbukti mampu mendukung proses klasifikasi dan prediksi penyakit dalam skala besar secara akurat dan efisien.

Salah satu teknologi yang berkembang pesat dalam bidang kesehatan adalah Machine Learning (ML). ML memungkinkan sistem untuk mempelajari pola dari data historis dan melakukan prediksi terhadap data baru secara akurat. Dalam konteks klasifikasi penyakit, ML terbukti efektif dalam mengidentifikasi pola tersembunyi dari data klinis dan memberikan hasil prediksi yang cepat dan efisien[2]. Berbagai algoritma seperti K-Nearest Neighbor (KNN), Support Vector Machine (SVM), dan Random Forest (RF) telah banyak digunakan dalam studi klasifikasi penyakit dan menunjukkan performa yang cukup baik [3][4].

Salah satu tantangan dalam penerapan algoritma klasifikasi pada data medis adalah ketidakseimbangan kelas (imbalanced data). Kondisi ini terjadi ketika jumlah sampel pada kelas mayoritas jauh lebih besar dibandingkan kelas minoritas. Situasi ini dapat menyebabkan model cenderung bias terhadap kelas mayoritas dan mengabaikan kelas minoritas, yang justru sering merepresentasikan kasus penting dalam medis. Penelitian [5]menegaskan bahwa permasalahan ketidakseimbangan data perlu ditangani melalui pendekatan balancing, baik dengan oversampling maupun undersampling, agar performa model dapat lebih stabil.

Selain itu, salah satu aspek penting dalam meningkatkan akurasi klasifikasi adalah proses seleksi fitur (*feature selection*). Fitur yang tidak relevan dapat menurunkan performa model, sehingga diperlukan teknik untuk memilih fitur-fitur yang paling berpengaruh. Salah satu metode yang banyak digunakan untuk seleksi fitur adalah Teaching Learning Based Optimization (TLBO). TLBO merupakan

`

algoritma metaheuristik yang meniru proses belajar mengajar dalam kelas dan terdiri dari dua fase, yaitu *teacher phase* dan *learner phase*, untuk menghasilkan solusi optimal tanpa parameter kontrol kompleks.[6][7]

Penelitian ini bertujuan untuk menerapkan metode TLBO dalam proses seleksi fitur pada klasifikasi penyakit asma, dan selanjutnya menguji performa lima algoritma *machine learning*, yaitu KNN, SVM, dan RF terhadap fitur terpilih. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam peningkatan akurasi klasifikasi data medis serta mendukung pengembangan sistem prediksi penyakit berbasis data.

II. TINJAUAN PUSTAKA

A. Machine learning dalam Klasifikasi Medis

Machine learning (ML) adalah cabang kecerdasan buatan yang telah banyak dimanfaatkan dalam bidang kesehatan, terutama dalam tugas klasifikasi dan prediksi. ML mampu mengenali pola dalam data medis dan menghasilkan model prediktif yang efektif. Dalam studi[8], ML telah banyak digunakan untuk mengidentifikasi pola dari data klinis, membantu diagnosis, serta memprediksi perkembangan penyakit berdasarkan data historis pasien. Dalam penelitian [9] juga algoritma *machine learning* dapat mengidentifikasi pola tersembunyi dari data klinis, mempercepat proses klasifikasi penyakit, serta membantu tenaga medis dalam pengambilan keputusan yang lebih objektif dan akurat. Hal ini menjadi dasar dalam mengembangkan sistem klasifikasi berbasis data yang adaptif dan efisien untuk berbagai penyakit, termasuk asma.

Penelitian[10] juga melakukan tinjauan komprehensif terhadap berbagai algoritma *supervised learning* yang digunakan dalam klasifikasi penyakit jantung. Studi ini menyoroti algoritma seperti Support Vector Machine (SVM), Random Forest (RF), dan K-Nearest Neighbor (KNN) serta mengevaluasi performanya berdasarkan sejumlah metrik seperti akurasi, sensitivitas, dan spesifisitas. Hasilnya menunjukkan bahwa algoritma RF dan SVM cenderung memberikan akurasi yang lebih, sementara KNN tetap relevan tergantung karakteristik dataset yang digunakan. Temuan ini memperkuat pemanfaatan model klasifikasi dalam bidang medis karena mampu menghasilkan prediksi yang cukup andal. Kemampuan ini menjadikan ML sangat relevan dalam membangun sistem klasifikasi penyakit, termasuk asma..

B. Algoritma Klasifikasi

Berbagai algoritma ML telah digunakan dalam klasifikasi data medis, seperti K-Nearest Neighbor (KNN), Support Vector Machine (SVM), dan Random Forest (RF).

Algoritma KNN bekerja dengan menghitung jarak antara data uji dan data latih, kemudian mengklasifikasikan berdasarkan mayoritas kelas tetangga terdekat. Penelitian [8]menunjukkan bahwa KNN memberikan akurasi tertinggi dibanding Naive Bayes dan Decision Tree dalam klasifikasi pasien rawat jalan dan rawat inap. Hal ini diperkuat oleh penelitian [4] yang mengaplikasikan KNN untuk klasifikasi penyakit paru-paru pada anak-anak dan mencapai akurasi sebesar 83,33%.

Support Vector Machine (SVM) adalah algoritma klasifikasi berbasis teori statistik yang efektif untuk memisahkan data ke dalam dua kelas menggunakan hyperplane optimal. Studi [11] menunjukkan bahwa kombinasi SVM dengan seleksi fitur dapat meningkatkan akurasi dan efisiensi klasifikasi, khususnya pada data berskala.

RF adalah algoritma berbasis *ensemble* yang terdiri dari banyak pohon keputusan. Penelitian oleh [12] menunjukkan bahwa penggunaan algoritma Random Forest dalam klasifikasi penyakit jantung mampu menghasilkan akurasi tinggi, yaitu sebesar 94%. Keberhasilan ini dicapai dengan menerapkan tahapan *pre-processing*, normalisasi data, dan evaluasi performa yang optimal terhadap dataset *Heart Disease* dari Kaggle

C. Teaching Learning Based Optimization (TLBO)

Teaching Learning Based Optimization (TLBO) merupakan algoritma optimasi berbasis populasi yang meniru proses pembelajaran di dalam kelas, terdiri dari dua tahap yaitu *teacher phase* dan *learner phase*. Algoritma ini tidak memerlukan parameter kontrol yang kompleks sehingga mudah diterapkan. Studi [7] menunjukkan bahwa TLBO efektif dalam seleksi fitur untuk klasifikasi, terutama saat dikombinasikan dengan model machine learning seperti SVM.saat dikombinasikan dengan model ML seperti SVM.

Selain itu, studi komprehensif yang dilakukan oleh Boveiri dan Khayami [13] juga menegaskan performa unggul TLBO dibandingkan metaheuristik lainnya seperti GA, PSO, dan ABC dalam menyelesaikan permasalahan optimasi kombinatorial. Penelitian tersebut menunjukkan bahwa TLBO mampu menghasilkan solusi yang lebih akurat dan stabil, terutama pada fungsi-fungsi benchmark unimodal dan multimodal. Dengan tidak adanya parameter yang perlu disesuaikan, TLBO dianggap lebih mudah diimplementasikan dan menunjukkan kecepatan konvergensi yang kompetitif. Hasil penelitian ini memperkuat potensi TLBO dalam digunakan secara luas, termasuk dalam konteks seleksi fitur dan klasifikasi data dalam domain medis.

D. Penanganan Ketidakseimbangan Data dengan Undersampling

Ketidakseimbangan kelas pada data medis dapat menyebabkan model klasifikasi cenderung bias terhadap kelas mayoritas. Salah satu pendekatan untuk mengatasi hal ini adalah undersampling, yaitu teknik yang mengurangi jumlah sampel dari kelas mayoritas agar seimbang dengan kelas minoritas. Metode ini relatif sederhana dan mampu memperbaiki distribusi data sehingga model lebih fokus dalam mengenali kelas minoritas.

Menurut Healthcare Journal (2022)[14], *undersampling* dapat meningkatkan kinerja model dalam kondisi data yang sangat timpang. Namun, kelemahan utama dari metode ini adalah potensi hilangnya informasi penting karena penghapusan sebagian data dari kelas mayoritas. Penelitian [15] juga menunjukkan bahwa metode undersampling memang dapat menyeimbangkan distribusi kelas, tetapi performanya cenderung lebih rendah dibandingkan metode oversampling seperti SMOTE. Hal ini disebabkan karena undersampling mengurangi keragaman informasi pada data, sementara SMOTE mampu menghasilkan sampel sintetis baru pada kelas minoritas sehingga distribusi data menjadi lebih seimbang tanpa mengorbankan data mayoritas.

Dengan demikian, *undersampling* dapat digunakan sebagai pendekatan awal untuk menangani ketidakseimbangan kelas, tetapi penerapannya perlu

dipertimbangkan secara hati-hati, terutama pada dataset medis yang berukuran terbatas.

E. Penanganan Ketidakseimbangan Data dengan SMOTE

Ketidakseimbangan distribusi kelas sering menjadi tantangan dalam klasifikasi medis, khususnya ketika jumlah sampel dari kelas minoritas sangat rendah. Untuk mengatasi hal ini, digunakan Synthetic Minority Over-sampling Technique (SMOTE). metode penanganan data tidak seimbang yang bekerja dengan membangkitkan sampel sintetis dari kelas minoritas. Teknik ini membantu meningkatkan representasi kelas yang kurang tanpa menghapus data dari kelas mayoritas. Berdasarkan studi [16] SMOTE terbukti mampu meningkatkan kinerja model klasifikasi, khususnya ketika dikombinasikan dengan teknik *ensemble learning*. Pendekatan ini tidak hanya memperbaiki distribusi kelas, tetapi juga meningkatkan akurasi, *recall*, dan stabilitas prediksi, menjadikannya strategi yang efektif dalam menangani ketidakseimbangan data medis.

Selain digunakan secara luas di berbagai domain, penerapan SMOTE dalam prediksi penyakit juga terbukti memberikan peningkatan performa model secara signifikan. Penelitian[17] menerapkan SMOTE pada dataset penyakit jantung dari UCI dan menunjukkan bahwa penggunaan SMOTE mampu meningkatkan kinerja model Random Forest dengan akurasi hingga 96,6%, sensitivitas 90%, serta spesifisitas dan presisi mencapai 100%

Studi ini menunjukkan bahwa teknik oversampling seperti SMOTE efektif dalam menyeimbangkan distribusi kelas pada data medis yang cenderung tidak seimbang, sehingga model dapat mengenali kelas minoritas secara lebih baik. Hal ini memperkuat alasan digunakannya SMOTE dalam penelitian ini sebagai bagian dari tahap pra-pemrosesan data untuk meningkatkan kualitas prediksi klasifikasi asma.

F. Kesimpulan

Berdasarkan tinjauan pustaka di atas, dapat disimpulkan bahwa kombinasi metode seleksi fitur TLBO dengan algoritma klasifikasi seperti KNN, SVM, dan RF memiliki potensi besar dalam meningkatkan akurasi klasifikasi penyakit. Penggunaan Undersampling dan SMOTE juga mendukung peningkatan performa model dalam situasi data yang tidak seimbang, menjadikan pendekatan ini sangat relevan untuk aplikasi klasifikasi penyakit asma

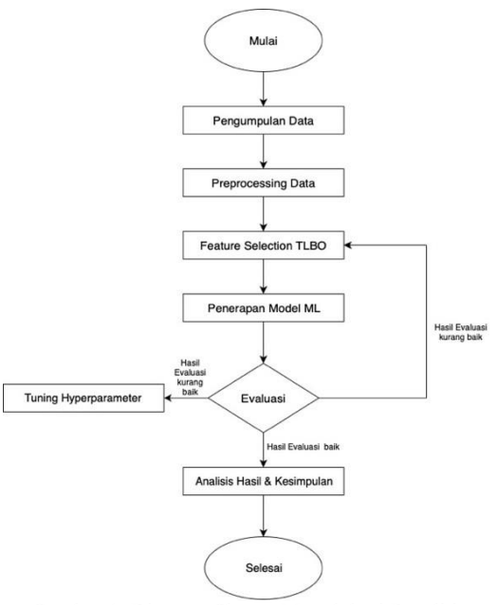
III. METODOLOGI PENELITIAN

A. DATASET

Penelitian ini menggunakan dataset penyakit asma yang diperoleh dari situs Kaggle. Dataset ini berisi 3.628 data latih, 454 data validasi, dan 454 data pengujian setelah dilakukan proses pembagian dan penyeimbangan data.

Dataset ini terdiri dari 26 fitur (variabel input) yang mencakup informasi demografis dan kondisi klinis seperti usia, jenis kelamin, indeks massa tubuh (BMI), aktivitas fisik, kualitas tidur, paparan polusi, serta riwayat keluarga. Label target adalah kolom Diagnosis yang terdiri dari dua kelas, yaitu:

- 0 untuk pasien tidak menderita asma
- 1 untuk pasien menderita asma.



Gambar 1. Diagram Skema Metodologi Penelitian

B. PRA-PEMROSESAN DATA

Pra-pemrosesan data merupakan tahapan awal yang esensial dalam proses klasifikasi untuk memastikan bahwa data berada dalam kondisi optimal sebelum digunakan dalam pelatihan model machine learning. Pada penelitian ini, eksperimen dilakukan dalam tiga skenario balancing data, yaitu:

- Model 1 (Original): dataset digunakan tanpa penanganan ketidakseimbangan kelas (imbalanced)
- Model 2 Undersampling dilakukan random undersampling pada kelas mayoritas (kelas 0) agar distribusi kelas lebih seimbang
- Model 3 (SMOTE): dilakukan Synthetic Minority Over-sampling Technique (SMOTE) untuk menambah sampel sintetis pada kelas minoritas sehingga distribusi kelas menjadi seimbang

Ketiga skenario ini digunakan untuk membandingkan pengaruh metode balancing terhadap kinerja algoritma klasifikasi yang digunakan.

Tahapan pra-pemrosesan mencakup pembersihan data, penanganan ketidakseimbangan dengan 3 skenario model diatas , pembagian dataset, serta normalisasi fitur numerik. Tahap pertama adalah pembersihan data untuk memeriksa nilai kosong (missing values) atau duplikat. Dataset yang digunakan tidak memiliki nilai kosong sehingga dapat langsung diproses ke tahap berikutnya.

Selanjutnya, dilakukan penanganan ketidakseimbangan kelas. Pada skenario Model 2 (Undersampling), dilakukan penghapusan sebagian data kelas mayoritas sehingga distribusi kelas menjadi seimbang, Sedangkan Pada skenario Model 3 (SMOTE), dilakukan pembangkitan data sintetis pada kelas minoritas langsung pada dataset asli sebelum proses pembagian data. Dengan cara ini, jumlah data antara kelas mayoritas dan minoritas menjadi lebih seimbang sejak awal. Proses sintesis sampel baru pada SMOTE dijelaskan melalui rumus berikut:

$$x_{new} = x_i + \delta \times (x_{nn} - x_i) \qquad (1)$$

Keterangan :

- x_i = sampel minoritas,
- x_{nn} =tetangga terdekat dari x_i,
- δ = bilangan acak antara 0 dan 1

Menurut penelitian[15] penggunaan SMOTE mampu meningkatkan performa klasifikasi secara signifikan pada data yang tidak seimbang, khususnya pada data medis penelitian[18] juga menunjukkan bahwa SMOTE efektif dalam meningkatkan performa model, terutama dalam mendeteksi kelas minoritas yang sering diabaikan oleh model konvensional.

Setelah dilakukan balancing sesuai dengan masing-masing skenario (original, undersampling, dan SMOTE), dataset kemudian dibagi menjadi tiga bagian: 80% untuk data latih, 10% untuk data validasi, dan 10% untuk data uji. Proses pemisahan dilakukan secara stratified agar distribusi kelas tetap proporsional di setiap subset data. Strategi ini memungkinkan model untuk dilatih secara optimal dan dievaluasi pada data yang representatif terhadap distribusi populasi sebenarnya.

Langkah terakhir adalah melakukan normalisasi terhadap fitur numerik menggunakan metode *Min-Max Scaling* untuk memastikan bahwa semua fitur berada dalam rentang nilai yang seragam antara 0 dan 1. Normalisasi penting diterapkan karena perbedaan skala antar fitur dapat menyebabkan algoritma seperti K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) menjadi bias terhadap fitur yang memiliki nilai lebih besar. Transformasi ini dilakukan menggunakan Rumus dari normalisasi *Min-Max* adalah sebagai berikut:

$$x^{\%} = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad (2)$$

Keterangan :

- X = nilai asli,
- $X^{\%}$ = nilai hasil normalisasi,
- X_{min} dan X_{max} = nilai minimum dan maksimum

Metode ini membantu mencegah dominasi fitur berskala besar dan mempercepat proses pelatihan model, sebagaimana dijelaskan di dalam peneltiian [11]. Normalisasi diterapkan secara konsisten pada seluruh data latih, validasi, dan uji.

C. SELEKSI FITUR MENGGUNAKAN TLBO

Teaching Learning Based Optimization (TLBO) merupakan salah satu algoritma optimasi berbasis populasi yang meniru proses pembelajaran di dalam kelas, di mana setiap individu dalam populasi berperan sebagai siswa, dan solusi terbaik berperan sebagai guru [2][6]. TLBO bekerja dalam dua fase utama, yaitu *teacher phase* dan *learner phase*, yang merepresentasikan dua mekanisme pembelajaran. Dalam *fase teacher individu terbaik dalam populasi berperan* sebagai teacher yang bertugas meningkatkan rata-rata kualitas seluruh populasi. Setiap individu diperbarui dengan mengacu pada selisih antara nilai rata-rata populasi dan nilai individu terbaik tersebut. Tujuannya adalah agar kemampuan keseluruhan populasi meningkat secara bertahap mendekati kualitas solusi terbaik yang telah ditemukan.

Sementara itu, learner phase menggambarkan proses pembelajaran antarindividu di mana setiap individu secara acak berinteraksi dengan individu lain yang memiliki performa lebih baik. Jika individu yang dipilih memiliki hasil lebih unggul, maka individu yang lebih lemah akan menyesuaikan solusinya mengikuti individu tersebut. Mekanisme ini memungkinkan proses peningkatan pengetahuan terjadi secara dua arah — baik dari guru maupun antar-siswa — sehingga populasi tetap beragam dan mampu mencapai konvergensi menuju solusi optimal tanpa memerlukan parameter kontrol tambahan seperti pada algoritma evolusioner lainnya [2].

Keunggulan utama dari TLBO adalah tidak memerlukan parameter kontrol tambahan seperti laju mutasi atau *crossover*, berbeda dengan algoritma *evolusioner* lainnya. Hal ini menjadikannya lebih sederhana dan efisien dalam penerapannya [6]. Dalam konteks klasifikasi medis, TLBO telah terbukti efektif dalam meningkatkan akurasi model dengan melakukan seleksi fitur secara optimal, seperti yang ditunjukkan dalam studi Dubey et al. (2022) [2]

Dalam penelitian ini, setiap solusi direpresentasikan sebagai vektor biner, di mana nilai 1 menunjukkan bahwa fitur dipilih, dan 0 berarti tidak dipilih. Proses optimasi dilakukan melalui dua tahap, yaitu *teacher phase* dan *learner phase*. Pada *teacher phase*, solusi diperbarui dengan merujuk pada guru atau individu terbaik dalam populasi. Formula yang digunakan adalah:

$$X_{new} = X_{old} + r.(X_{teacher} - T_F.X_{mean}) \qquad (3)$$

Selanjutnya, *learner phase* memungkinkan setiap individu untuk belajar dari individu lain yang dipilih secara acak. Jika individu lain memiliki solusi lebih baik, maka individu tersebut akan belajar darinya, sebagaimana dijelaskan dalam formula:

$$X_{new} = \begin{cases} X_i + r.(X_j - X_i), & jika\ f(X_j) > f(X_i) \\ X_i + r.(X_i - X_j), & jika\ f(X_j) \le f(X_i) \end{cases} \qquad (4)$$

D. PEMODELAN KLASIFIKASI

Setelah dilakukan tahapan pra-pemrosesan data dan seleksi fitur menggunakan algoritma Teaching Learning Based Optimization (TLBO), langkah berikutnya adalah membangun model klasifikasi. Dalam penelitian ini, lima algoritma pembelajaran mesin digunakan untuk mengklasifikasikan data penyakit asma berdasarkan fitur hasil seleksi TLBO, yaitu K-Nearest Neighbor (KNN), Support Vector Machine (SVM), dan Random Forest (RF).

- K-Nearest Neighbor (KNN)

KNN adalah metode klasifikasi berbasis kedekatan, yang memprediksi kelas sebuah data uji berdasarkan mayoritas

label dari k tetangga terdekatnya. Karima dan Fatah [3] menyatakan bahwa KNN efektif dalam klasifikasi penyakit paru-paru dengan akurasi cukup tinggi. Dalam penelitian ini, model KNN diuji dengan nilai k = 3, 5, 7, 9, dan 11 Pemilihan parameter terbaik dilakukan berdasarkan performa tertinggi pada data validasi dan uji.

- Support Vector Machine (SVM)

SVM adalah algoritma klasifikasi yang memisahkan dua kelas data menggunakan *hyperplane* optimal. Untuk meningkatkan performa, dilakukan tuning parameter *C* dan *gamma*. Penelitian oleh Fadli et al. [9] menunjukkan bahwa optimasi *hyperparameter* dan seleksi fitur dapat meningkatkan akurasi SVM secara signifikan. Dalam eksperimen ini, kernel RBF digunakan dengan nilai *C* = 0.1, 1, dan 10, serta parameter *gamma* = "scale" dan "auto".

- Random Forest (RF)

Random Forest adalah algoritma *ensemble* yang terdiri dari sejumlah pohon keputusan dan menggabungkan hasilnya untuk mendapatkan prediksi akhir. Hidayat et al. [10] mengungkapkan bahwa RF memiliki akurasi tinggi dalam klasifikasi penyakit jantung dan robust terhadap data yang bervariasi. Dalam penelitian ini, RF diuji dengan jumlah pohon (n_estimators) sebanyak 100 dan 200, serta kombinasi konfigurasi populasi dan iterasi pada TLBO untuk seleksi fitur.

E. EVALUASI MODEL

Untuk menilai performa model klasifikasi, digunakan lima metrik evaluasi utama, yaitu akurasi, presisi, recall (sensitivitas), spesifisitas, dan f1-score. Evaluasi dilakukan terhadap data validasi dan data uji untuk mengukur kemampuan generalisasi model. Khususnya, perhatian lebih diberikan pada kelas minoritas (penderita asma) karena ketidakseimbangan data yang dapat memengaruhi akurasi secara keseluruhan.Berikut adalah rumus-rumus dari masing-masing metrik evaluasi:

- Akurasi

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (5)$$

- Presisi

$$presisi = \frac{TP}{TP + FP} \times 100 \quad (6)$$

- Recall

$$Recall = \frac{TP}{TP + FN} \times 100 \quad (7)$$

- Spesifisitas

$$spesifisitas = \frac{TN}{TN + FP} \times 100 \quad (8)$$

- F1-Score

$$F1 = 2 \times \frac{Presisi \times Recall}{Presisi + Recall} \times 100 \quad (9)$$

Evaluasi performa model merupakan tahap penting dalam proses klasifikasi, khususnya pada data medis yang sering kali memiliki distribusi kelas tidak seimbang. Oleh karena itu, tidak hanya akurasi yang diperhitungkan, namun juga metrik lain seperti precision, recall, f1-score, sensitivitas, dan spesifisitas.

Menurut penelitian [19] penggunaan metrik evaluasi yang beragam menjadi krusial untuk menilai efektivitas algoritma machine learning dalam bidang kesehatan. Dalam kasus prediksi penyakit, misalnya, nilai akurasi tinggi belum tentu mencerminkan performa yang optimal apabila model tidak mampu mengenali kelas minoritas secara tepat. Oleh karena itu, kombinasi antara recall dan precision, yang direpresentasikan dalam f1-score, menjadi indikator yang lebih representatif untuk kasus klasifikasi medis.

Metrik-metrik tersebut tidak hanya memberikan gambaran seberapa banyak prediksi yang benar, tetapi juga mengukur kemampuan model dalam mendeteksi kasus positif secara akurat. Hal ini menjadi sangat penting dalam diagnosis penyakit seperti asma, di mana kesalahan klasifikasi dapat berdampak pada keputusan klinis dan penanganan pasien.

IV. HASIL PENELITIAN

A. Hasil Penelitian

Pada tahap ini dipaparkan hasil implementasi dari alur penelitian yang telah dirancang, meliputi pra-pemrosesan data, penanganan ketidakseimbangan kelas, pembagian dataset, serta seleksi fitur menggunakan Teaching Learning Based Optimization (TLBO). Seluruh tahapan dilakukan untuk memastikan data berada dalam kondisi optimal sebelum dimanfaatkan pada proses pelatihan model klasifikasi.

Tahap pra-pemrosesan dimulai dengan pemeriksaan kelengkapan data. Dataset yang digunakan tidak memiliki nilai kosong maupun duplikasi, sehingga dapat langsung diproses. Selanjutnya dilakukan penanganan ketidakseimbangan kelas melalui tiga skenario berbeda, yaitu:

- Model 1 (Original): dataset digunakan tanpa perubahan distribusi kelas
- Model 2 (Undersampling): sebagian data pada kelas mayoritas dihapus secara acak agar distribusi kelas lebih seimbang
- Model 3 (SMOTE): dilakukan penambahan sampel sintetis pada kelas minoritas menggunakan Synthetic Minority Over-sampling Technique (SMOTE) langsung pada dataset asli sebelum pemisahan data

Setelah proses balancing, dataset kemudian dibagi menjadi tiga subset dengan perbandingan 80% data latih, 10% data validasi, dan 10% data uji. Proses pemisahan dilakukan secara stratified sampling agar proporsi kelas tetap konsisten di setiap subset. Tahap berikutnya adalah

normalisasi fitur numerik menggunakan metode Min–Max Scaling agar semua atribut berada dalam rentang [0,1]. Normalisasi diperlukan untuk mencegah bias pada algoritma berbasis jarak seperti KNN maupun algoritma dengan margin seperti SVM

Proses seleksi fitur dilakukan dengan algoritma TLBO untuk memperoleh subset fitur yang paling relevan. TLBO bekerja dengan meniru mekanisme interaksi pengajar dan pelajar dalam suatu kelas, sehingga fitur terpilih merupakan variabel yang memberikan kontribusi terbesar terhadap pemisahan kelas. Hasil seleksi menunjukkan bahwa jumlah fitur yang dipilih berbeda pada tiap model klasifikasi, namun terdapat beberapa fitur yang konsisten muncul, antara lain Age, Gender, Shortness of Breath, Wheezing, dan Coughing. Fitur-fitur tersebut memiliki relevansi klinis karena berhubungan langsung dengan faktor risiko maupun gejala utama asma

Dengan demikian, tahap hasil penelitian menunjukkan bahwa dataset yang telah diproses, diseimbangkan, dan direduksi fiturnya siap untuk digunakan pada tahap evaluasi model klasifikasi.

B. Evaluasi Model

Evaluasi model dilakukan untuk menilai kinerja algoritma klasifikasi pada berbagai skenario balancing data. Penilaian menggunakan metrik akurasi, precision, recall (sensitivitas), specificity, dan F1-score, dengan fokus utama pada kemampuan model dalam mengenali kelas positif (penderita asma). Selain itu, confusion matrix digunakan untuk menggambarkan distribusi prediksi benar maupun salah pada masing-masing kelas. Hasil pengujian terhadap seluruh konfigurasi algoritma disajikan dalam tabel-tabel berikut, yang memuat performa setiap model beserta analisis perbandingannya.

Hasil evaluasi pada data uji terhadap seluruh model dan konfigurasi yang diuji disajikan dalam Tabel-tabel berikut:

- Model 1 (Original)

Pada skenario Model Original, dataset digunakan tanpa dilakukan penyeimbangan kelas sehingga distribusi antara kelas mayoritas dan minoritas tetap apa adanya. Kondisi ini merepresentasikan keadaan awal data medis di dunia nyata yang sering kali tidak seimbang. Evaluasi pada skenario ini bertujuan untuk melihat bagaimana performa masing-masing algoritma klasifikasi—KNN, SVM, dan Random Forest—setelah melalui tahap seleksi fitur dengan TLBO, namun tanpa intervensi balancing data. Hasil dari model original ini menjadi tolok ukur dasar (baseline) untuk kemudian dibandingkan dengan skenario undersampling dan SMOTE pada tahap selanjutnya.

Model	Varian / Parameter	Precision (0/1)	Recall (0/1)	F1-Score (0/1)	Accuracy
TLBO-KNN	k = 3	0.966/ 0.000	0.995/ 0.000	0.980/0.000	0.96
	k = 5	0.966/ 0.000	0.995/ 0.000	0.980/0.000	0.96
	k = 7	0.966 / 0.000	1.000 / 0.000	0.983 / 0.000	0.96
	k = 9	0.966 / 0.000	1.000 / 0.000	0.983/ 0.000	0.96
	k = 11	0.966 / 0.000	1.000 / 0.000	0.983/ 0.000	0.96
TLBO-RF	n= 100	0.970/ 0.166	0.978/ 0.125	0.974/ 0.142	0.95
	n= 200	0.973 / 0.166	0.956 / 0.250	0.965/ 0.200	0.93
TLBO-SVM	Kernel=rbf	0.97/ 0.00	1.00 / 0.00	0.98 / 0.00	0.97
	C = 10 Kernel= rbf	0.97/ 0.00	1.00 / 0.00	0.98 / 0.00	0.97
	C = 10, Kernel=rbf gamma= auto	0.97 / 0.00	0.98 / 0.00	0.97/ 0.00	0.95

Tabel 1. Hasil Evaluasi Model 1

Berdasarkan tabel di atas, seluruh algoritma pada skenario Model Original masih menunjukkan nilai akurasi yang tinggi, berkisar antara 93% hingga 97%. Namun, jika dianalisis lebih dalam, terlihat bahwa performa pada kelas minoritas (penderita asma) sangat rendah. Hampir semua konfigurasi algoritma hanya mampu mengenali kelas mayoritas, ditunjukkan dengan nilai precision, recall, dan F1-score pada kelas 1 yang mendekati nol. Kondisi ini merupakan bentuk accuracy paradox, yaitu situasi ketika akurasi tampak tinggi tetapi sebenarnya tidak mencerminkan kemampuan model dalam mendeteksi kelas yang penting.

Dengan demikian, meskipun Model Original terlihat cukup baik dari sisi akurasi keseluruhan, kualitas prediksinya terhadap penderita asma tidak dapat diandalkan. Hal ini menegaskan bahwa strategi penyeimbangan kelas perlu diterapkan agar model dapat memberikan hasil yang lebih adil dan representatif dalam mengklasifikasikan kedua kelas.

- Model 2 (Random Undersampling)

Pada skenario Model 2 (Random Undersampling), penanganan ketidakseimbangan kelas dilakukan dengan cara mengurangi jumlah sampel pada kelas mayoritas secara acak hingga jumlahnya seimbang dengan kelas minoritas. Proses ini menghasilkan dataset berukuran lebih kecil, yakni 248 sampel dengan distribusi masing-masing kelas sebanyak 124. Selanjutnya, data dibagi menjadi set latih, validasi, dan uji dengan proporsi 80:10:10 menggunakan stratified sampling untuk menjaga keseimbangan antar kelas. Strategi undersampling ini diharapkan mampu mengurangi bias model terhadap kelas mayoritas, meskipun dengan konsekuensi hilangnya sebagian informasi dari data asli. Evaluasi performa algoritma KNN, SVM, dan RF pada skenario ini disajikan dalam tabel berikut.

Model	Varian / Parameter	Precision (0/1)	Recall (0/1)	F1-Score (0/1)	Accuracy
TLBO-KNN	k = 3	0.50/ 0.72	0.70/ 0.53	0.58/0.61	0.60
	k = 5	0.42/ 0.63	0.60/ 0.46	0.50/0.53	0.52
	k = 7	0.20 / 0.46	0.20 / 0.46	0.20/ 0.46	0.36

	k = 9	0.55 / 0.68	0.50 / 0.73	0.52/ 0.70	0.64
	k = 11	0.36 / 0.57	0.40 / 0.53	0.38/ 0.55	0.48
TLBO-RF	n= 100	0.33/ 0.53	0.40/ 0.46	0.36/ 0.50	0.44
	n= 200	0.00 / 0.60	0.00 / 1.00	0.00/ 0.75	0.60
TLBO-SVM	Kernel=rbf	0.50/ 0.61	0.10 / 0.93	0.17 / 0.74	0.60
	C = 10 Kernel= rbf	0.33/ 0.59	0.10 / 0.87	0.15 / 0.70	0.56
	C = 10, Kernel=rbf gamma= auto	0.33 / 0.56	0.30 / 0.60	0.32/ 0.58	0.48

Tabel 2. Hasil Evaluasi Model 2

Berdasarkan Tabel di atas, performa model pada skenario Random Undersampling mengalami penurunan signifikan dibandingkan Model Original. Nilai akurasi tertinggi hanya mencapai sekitar 64% pada KNN dengan k=9, sementara konfigurasi lain umumnya berada pada kisaran 48–60%. Random Forest dan SVM juga menunjukkan pola serupa, di mana akurasi tidak stabil dan metrik precision serta recall pada kelas 1 (penderita asma) masih rendah. Kondisi ini terjadi karena undersampling mengurangi jumlah data mayoritas secara drastis, sehingga informasi yang terkandung dalam dataset menjadi terbatas. Meskipun strategi ini membantu menyeimbangkan distribusi kelas, pengurangan data justru berdampak pada menurunnya kemampuan generalisasi model.

Dengan demikian, Model 2 belum mampu memberikan performa yang optimal, sehingga diperlukan pendekatan alternatif seperti SMOTE yang dapat menjaga keseimbangan kelas tanpa mengorbankan jumlah data.

- Model 3 (SMOTE)

Pada skenario ketiga, penanganan ketidakseimbangan kelas dilakukan menggunakan Synthetic Minority Over-sampling Technique (SMOTE) yang diterapkan langsung pada dataset asli sebelum dilakukan proses pembagian data. Penerapan SMOTE menghasilkan dataset yang seimbang antara penderita asma (kelas 1) dan non-asma (kelas 0), sekaligus meningkatkan jumlah total sampel dari 2.392 menjadi 4.536. Setelah proses balancing, data dibagi menjadi 80% data latih, 10% data validasi, dan 10% data uji dengan distribusi kelas yang tetap seimbang pada setiap subset. Strategi ini diharapkan mampu memperbaiki sensitivitas model terhadap penderita asma, karena representasi data minoritas menjadi lebih proporsional tanpa mengorbankan ukuran dataset secara signifikan. Evaluasi performa algoritma KNN, SVM, dan RF pada skenario ini disajikan dalam tabel berikut.

Model	Varian / Parameter	Precision (0/1)	Recall (0/1)	F1-Score (0/1)	Accuracy
TLBO-KNN	k = 3	1.00/ 0.79	0.75/ 1.00	0.85/0.88	0.87
	k = 5	1.00/ 0.70	0.61/ 1.00	0.76/0.82	0.80
	k = 7	0.98 / 0.74	0.67 / 0.99	0.80/ 0.84	0.82
	k = 9	0.98 / 0.72	0.64 / 0.99	0.77/ 0.83	0.81
	k = 11	0.97 / 0.70	0.61 / 0.98	0.75/ 0.82	0.79
TLBO-RF	n= 100	0.94/ 0.98	0.98/ 0.94	0.96/ 0.96	0.96
	n= 200	0.95 / 0.96	0.97 / 0.93	0.95/ 0.95	0.95
TLBO-SVM	Kernel=rbf	0.86/ 0.83	0.84 / 0.85	0.85 / 0.84	0.85
	C = 10 Kernel= rbf	0.90/ 0.86	0.86 / 0.90	0.88/ 0.88	0.88
	C = 10, Kernel=rbf gamma= auto	1.00 / 0.97	0.97 / 1.00	0.99/ 0.99	0.99

Tabel 3. Hasil Evaluasi Model 3

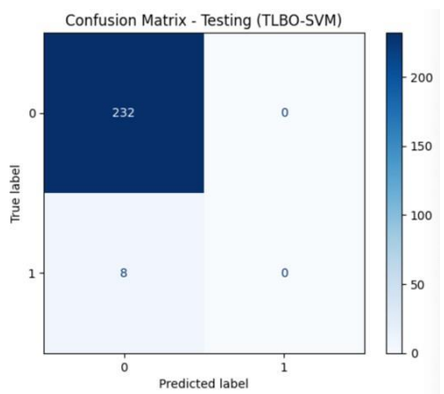
Berdasarkan Tabel di atas, terlihat bahwa penerapan SMOTE pada dataset asli memberikan dampak positif yang signifikan terhadap performa model. Pada algoritma KNN, meskipun akurasi berkisar antara 79–87%, nilai precision dan recall untuk kelas minoritas menunjukkan perbaikan yang jelas dibandingkan Model Original maupun Undersampling, sehingga prediksi terhadap penderita asma menjadi lebih seimbang. Random Forest tampil stabil dengan akurasi tinggi di kisaran 95–96% serta precision dan recall yang relatif seimbang, menunjukkan kemampuannya dalam memanfaatkan fitur hasil seleksi TLBO secara konsisten.

Peningkatan paling mencolok terjadi pada SVM. Konfigurasi kernel RBF dengan C=10 dan gamma=auto menghasilkan akurasi 99% serta nilai precision, recall, dan F1-score yang hampir sempurna. Hal ini menegaskan bahwa kombinasi SMOTE, TLBO, dan tuning parameter pada SVM mampu mengoptimalkan pemisahan kelas secara signifikan.

Dengan demikian, Model 3 (SMOTE) dapat disimpulkan sebagai skenario terbaik dalam penelitian ini karena berhasil mengatasi bias terhadap kelas mayoritas sekaligus meningkatkan sensitivitas model dalam mengenali penderita asma.

Untuk melengkapi hasil evaluasi yang telah disajikan dalam bentuk tabel, pada penelitian ini juga ditampilkan confusion matrix dari algoritma terbaik pada masing-masing skenario (Model 1, Model 2, dan Model 3). Visualisasi confusion matrix ini digunakan untuk menunjukkan secara lebih rinci distribusi prediksi benar dan salah pada kelas positif (penderita asma) maupun kelas negatif (non-asma),

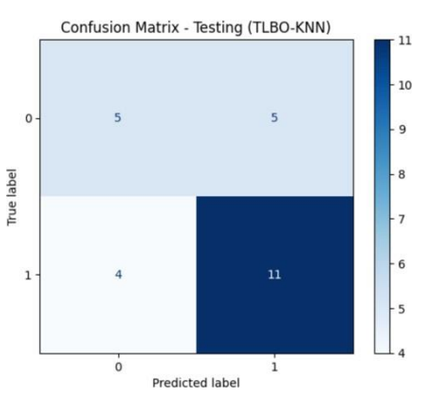
sehingga dapat memperkuat analisis performa model yang diperoleh.



Gambar 2. *Confusion Matrix TLBO-SVM pada Model 1*

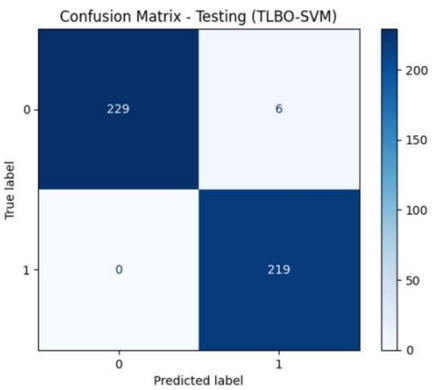
Confusion matrix pada Model 1 (original tanpa penerapan undersampling maupun SMOTE) dengan algoritma SVM menunjukkan akurasi yang tinggi, yaitu sebesar 0,97. Namun, tingginya akurasi tersebut tidak diikuti dengan kemampuan yang baik dalam mengenali seluruh kelas. Model cenderung hanya mampu mengklasifikasikan kelas mayoritas dengan tepat, sementara kelas minoritas sama sekali tidak teridentifikasi.

Kondisi ini mengindikasikan adanya ketidakseimbangan performa, di mana metrik akurasi tampak tinggi tetapi sebenarnya tidak representatif untuk menggambarkan kinerja model secara menyeluruh. Oleh karena itu, diperlukan pendekatan penyeimbangan data agar model mampu bekerja lebih optimal pada kedua kelas.



Gambar 3. *Confusion Matrix TLBO-KNN (k =9) pada model 2*

Pada Model 2 yang hanya menggunakan teknik undersampling, performa terbaik diperoleh pada algoritma KNN dengan nilai k=9, yang menghasilkan akurasi sebesar 0,64. Meskipun nilai akurasinya tidak terlalu tinggi, KNN relatif lebih stabil dibandingkan algoritma lain karena mampu menjaga keseimbangan antara nilai precision dan recall pada kedua kelas. Hal ini menunjukkan bahwa meskipun undersampling dapat mengurangi ketidakseimbangan data, jumlah data yang berkurang cukup signifikan sehingga berdampak pada menurunnya performa keseluruhan model. Dengan demikian, Model 2 masih memiliki keterbatasan, terutama dalam generalisasi hasil prediksi, namun KNN k=9 tetap menjadi pilihan terbaik di antara algoritma yang diuji.



Gambar 4. *Confusion Matrix TLBO-SVM (C = 10, kernel = rbf, gamma = auto) pada model 3*

Berdasarkan confusion matrix pada Gambar, model TLBO-SVM dengan parameter C=10 dan gamma=auto merupakan model terbaik dibandingkan dengan variasi parameter maupun algoritma lainnya. Hal ini terlihat dari hasil klasifikasi yang hampir sempurna, dengan 229 data negatif dan 219 data positif teridentifikasi dengan benar. Kesalahan hanya terjadi pada 6 data negatif yang terklasifikasi sebagai positif, sementara tidak terdapat kesalahan pada kelas positif (False Negative = 0).

Dibandingkan dengan model TLBO-KNN maupun TLBO-RF yang masih menghasilkan kesalahan klasifikasi lebih besar, model ini mampu mencapai akurasi 99%, sensitivitas yang sangat tinggi, serta keseimbangan precision dan recall yang unggul. Dengan demikian, TLBO-SVM dapat dianggap sebagai model paling optimal untuk klasifikasi penyakit asma dalam penelitian ini.

C. Perbandingan dengan Penelitian Sebelumnya

Untuk menilai posisi kontribusi penelitian ini terhadap penelitian terdahulu, dilakukan perbandingan hasil klasifikasi dengan beberapa metode metaheuristik lain yang telah diterapkan pada domain medis. Perbandingan ini difokuskan pada algoritma yang berada pada tingkat yang sama, yaitu metode optimasi berbasis metaheuristik yang digabungkan dengan algoritma klasifikasi Support Vector Machine (SVM). Hasil perbandingan ditunjukkan pada Tabel 5 berikut.

Peneliti	metode	kasus	akurasi	keterangan
Yang et al. (2024)	PSO-SVM	Kanker	97.50	Kombinasi Particle Swarm Optimization dan SVM untuk diagnosis kanker.
Lee et al. (2024)	GA-SVM	Sarcopenia	78.50	Seleksi fitur menggunakan Genetic Algorithm untuk klasifikasi penyakit sarcopenia.
Model TLBO-SVM (usulan 2025)	TLBO-SVM	Asma	99.00	Kombinasi optimasi TLBO dan penyeimbangan data SMOTE menghasilkan performa terbaik.

Tabel 4. Perbandingan Hasil Penelitian dengan metode lain

Berdasarkan hasil perbandingan yang ditunjukkan pada Tabel 4, dapat diamati bahwa model TLBO–SVM yang dikembangkan dalam penelitian ini menghasilkan akurasi tertinggi, yaitu sebesar 99%, pada kasus klasifikasi penyakit asma. Nilai ini melampaui hasil yang diperoleh oleh Yang et al. (2024) yang mengusulkan kombinasi Particle Swarm Optimization (PSO) dengan SVM untuk diagnosis kanker dan hanya mencapai akurasi sebesar 97.5%[20]

Selain itu, penelitian yang dilakukan oleh Lee et al. (2024) menggunakan pendekatan Genetic Algorithm (GA) untuk seleksi fitur pada klasifikasi penyakit sarcopenia memperoleh akurasi sebesar 78.5%[21], yang masih jauh di bawah performa model pada penelitian ini.

Perbedaan kinerja tersebut menunjukkan bahwa mekanisme dua fase pada TLBO, yaitu teacher phase dan learner phase, mampu menyeleksi fitur yang lebih relevan tanpa memerlukan parameter kontrol seperti pada PSO atau GA. Selain itu, penerapan teknik SMOTE dalam tahap pra-pemrosesan turut meningkatkan sensitivitas model terhadap kelas minoritas, sehingga prediksi terhadap penderita asma menjadi lebih seimbang.

Dengan demikian, kombinasi TLBO dan SVM pada penelitian ini terbukti memberikan peningkatan signifikan dibandingkan metode optimasi berbasis PSO dan GA dalam konteks klasifikasi data medis yang bersifat tidak seimbang.

Selain itu, penelitian ini juga sejalan dengan perkembangan terkini dalam algoritma Teaching Learning Based Optimization (TLBO). Penelitian [22] mengusulkan Enhanced TLBO (ETLBO) dengan penerapan adaptive weight strategy dan Kent chaotic map untuk meningkatkan kemampuan eksplorasi serta mencegah terjebak pada local minima.

Hasil penelitian tersebut menunjukkan bahwa ETLBO memiliki kinerja yang lebih baik dibandingkan beberapa algoritma metaheuristik lain seperti PSO, WOA, dan HHO pada kasus seleksi fitur dan klasifikasi berbasis SVM. Hal ini menegaskan bahwa TLBO masih menjadi metode optimasi yang aktif dikembangkan hingga saat ini, sehingga penerapannya dalam penelitian ini berada dalam jalur yang konsisten dengan tren penelitian terkini di bidang optimasi berbasis pembelajaran.

V. KESIMPULAN

Sistem klasifikasi penyakit asma dibangun dengan pendekatan Teaching Learning Based Optimization (TLBO) sebagai metode seleksi fitur dan tiga algoritma klasifikasi, yaitu K-Nearest Neighbor (KNN), Support Vector Machine (SVM), dan Random Forest (RF). Tahapan pra-pemrosesan meliputi penyeimbangan data menggunakan undersampling dan SMOTE, serta normalisasi dengan Min-Max Scaling..

Sama halnya dengan penelitian [23] yang menunjukkan bahwa algoritma machine learning memiliki akurasi tinggi dalam klasifikasi penyakit berbasis data klinis, penelitian ini membuktikan bahwa kombinasi TLBO dengan algoritma klasifikasi mampu mengenali pola data pasien secara efektif. Hal ini dapat membantu dalam proses diagnosis dan pengambilan keputusan medis yang lebih cepat dan tepat.

Berdasarkan hasil eksperimen dengan tiga scenario model:

- Model 1 (original tanpa balancing) menunjukkan bahwa meskipun akurasi relatif tinggi pada algoritma tertentu, model ini cenderung bias karena distribusi data yang tidak seimbang.
- Model 2 (undersampling) berhasil mengurangi ketidakseimbangan kelas, namun mengorbankan sebagian informasi data sehingga performa tidak selalu lebih baik dibanding model lain
- Model 3 (SMOTE) memberikan hasil yang lebih stabil dengan peningkatan sensitivitas terhadap kelas minoritas. Pada model ini, kombinasi TLBO-SVM dengan parameter C=10 dan gamma='auto' memberikan hasil terbaik dengan akurasi 99%, recall 100%, precision 99%, dan f1-score 99%. Sedangkan TLBO-RF dengan 200 pohon juga menunjukkan performa tinggi dengan akurasi 96,04%. TLBO-KNN masih memberikan hasil cukup baik, meskipun belum melampaui performa SVM dan RF.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa integrasi TLBO dengan algoritma klasifikasi yang tepat, khususnya SVM pada data hasil balancing SMOTE, mampu meningkatkan akurasi, sensitivitas, dan f1-score model dalam mendeteksi penderita asma secara lebih optimal.

REFERENCES

[1] Mafaza Cahya Yuniar, Melisa Intan Safila, Mohamad Putra, Muhammmad Haidar Asyraf, Nahdiyah Dwi Amelia, and Dion Kunto Adi Patria, "2022+Sep_5.+Pengembangan+Teknologi+dalam+ Bidang+Kesehatan (1)," *Jurnal Teknologi Kesehatan (Journal of Health Technology)*, vol. 18, no. 2, pp. 49–52, Sep. 2022, Accessed: Jul. 10, 2025. [Online]. Available: https://www.e-journal.poltekkesjogja.ac.id/index.php/JTK/article/view/1143

[2] A. K. Dubey, U. Gupta, and S. Jain, "Medical data clustering and classification using TLBO and machine learning algorithms," *Computers, Materials and Continua*, vol. 70, no. 3, pp. 4523–4543, 2022, doi: 10.32604/cmc.2022.021148.

[3] M. Napiah and S. Heristian, "Perbandingan Algoritma Machine Learning pada Klasifikasi Penyakit Jantung." [Online]. Available: http://ejournal.bsi.ac.id/ejurnal/index.php/infortech46

[4] H. Jurnal, R. Alfiatul Karima, and Z. Fatah, "JURNAL ILMIAH MULTIDISIPLIN ILMU IMPLEMENTASI METODE K-NEAREST NEIGHBOR UNTUK KLASIFIKASI PENYAKIT PARU-PARU PADA ANAK", doi: 10.69714/yx4smf68.

[5] P. Budi Utomo, M. Faruqziddan, E. Herdika Septa Aulia, and S. Dini Azzahra, "Perbandingan Skenario Balancing Oversampling dan Undersampling dalam Klasifikasi Resiko Kambuh Kanker Tiroid menggunakan Algoritma SVM Linear," *JAMI:*

Jurnal Ahli Muda Indonesia, vol. 5, no. 2, pp. 172–182, Dec. 2024, doi: 10.46510/jami.v5i2.320.

[6] M. Allam and M. Nandhini, "Optimal feature selection using binary teaching learning based optimization algorithm," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 2, pp. 329–341, Feb. 2022, doi: 10.1016/j.jksuci.2018.12.001.

[7] M. Arashpour *et al.*, "Predicting individual learning performance using machine-learning hybridized with the teaching-learning-based optimization," *Computer Applications in Engineering Education*, vol. 31, no. 1, pp. 83–99, Jan. 2023, doi: 10.1002/cae.22572.

[8] F. Aziz *et al.*, "JOURNAL PHARMACY AND APPLICATION OF COMPUTER SCIENCES PENERAPAN METODE MACHINE LEARNING UNTUK MENGKLASIFIKASI PENANGANAN PERAWATAN PASIEN APPLICATION OF MACHINE LEARNING METHOD TO CLASSIFICATION OF PATIENT CARE TREATMENT," *Agustus*, vol. 1, no. 2, p. 2023.

[9] I. Akbar, F. Supriadi, and D. I. Junaedi, "PEMANFAATAN MACHINE LEARNING DI BIDANG KESEHATAN," 2025.

[10] A. Mohsin Abdulazeez and S. S. Hasan, "Classification of Heart Diseases Based on Machine Learning: A Review," 2025.

[11] S. Fadli, M. Ashari, P. Studi Sistem Informasi, and S. Lombok, "JISA (Jurnal Informatika dan Sains) Optimization of Support Vector Machine Method Using Feature Selection to Improve Classification Results," 2021.

[12] H. Hidayat, A. Sunyoto, and H. Al Fatta, "Klasifikasi Penyakit Jantung Menggunakan Random Forest Clasifier," *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, vol. 7, no. 1, pp. 31–40, Oct. 2023, doi: 10.47970/siskom-kb.v7i1.464.

[13] H. R. Boveiri and R. Khayami, "On the Performance of Metaheuristics: A Different Perspective," Jan. 2020, [Online]. Available: http://arxiv.org/abs/2001.08928

[14] D. H. Jeong, S. E. Kim, W. H. Choi, and S. H. Ahn, "A Comparative Study on the Influence of Undersampling and Oversampling Techniques for the Classification of Physical Activities Using an Imbalanced Accelerometer Dataset," *Healthcare (Switzerland)*, vol. 10, no. 7, Jul. 2022, doi: 10.3390/healthcare10071255.

[15] M. Khushi *et al.*, "A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data," *IEEE Access*, vol. 9, pp. 109960–109975, 2021, doi: 10.1109/ACCESS.2021.3102399.

[16] M. Naseriparsa, A. Al-Shammari, M. Sheng, Y. Zhang, and R. Zhou, "RSMOTE: improving classification performance over imbalanced medical datasets," *Health Inf Sci Syst*, vol. 8, no. 1, Dec. 2020, doi: 10.1007/s13755-020-00112-w.

[17] Narayanan and Jayashree, "Implementation of Efficient Machine Learning Techniques for Prediction of Cardiac Disease using SMOTE," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 558–569. doi: 10.1016/j.procs.2024.03.245.

[18] A. J. Mohammed, "Improving Classification Performance for a Novel Imbalanced Medical Dataset using SMOTE Method," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 3, pp. 3161–3172, Jun. 2020, doi: 10.30534/ijatcse/2020/104932020.

[19] M. A. Al-Hashem, A. M. Alqudah, and Q. Qananwah, "Performance Evaluation of Different Machine Learning Classification Algorithms for Disease Diagnosis," *International Journal of E-Health and Medical Communications*, vol. 12, no. 6, 2021, doi: 10.4018/IJEHMC.20211101.oa5.

[20] F. Yang, Z. Xu, H. Wang, L. Sun, M. Zhai, and J. Zhang, "A hybrid feature selection algorithm combining information gain and grouping particle swarm optimization for cancer diagnosis," *PLoS One*, vol. 19, no. 3 March, Mar. 2024, doi: 10.1371/journal.pone.0290332.

[21] J. Lee, Y. Yoon, J. Kim, and Y. H. Kim, "Metaheuristic-Based Feature Selection Methods for Diagnosing Sarcopenia with Machine Learning Algorithms," *Biomimetics*, vol. 9, no. 3, Mar. 2024, doi: 10.3390/biomimetics9030179.

[22] D. Wu *et al.*, "Enhance Teaching-Learning-Based Optimization for Tsallis-Entropy-Based Feature Selection Classification Approach," *Processes*, vol. 10, no. 2, Feb. 2022, doi: 10.3390/pr10020360.

[23] N. Arminarahmah and G. Mahalisa, "Implementasi Model Machine Learning pada Klasifikasi Status Penyakit Diabetes Berbasis Streamlit," *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 13, no. 3, Jul. 2024, doi: 10.30591/smartcomp.v13i3.5866.