

IMPLEMENTASI SUPPORT VECTOR MACHINE (SVM) DALAM KLASIFIKASI PENYAKIT ASMA BERDASARKAN DATA REKAM MEDIS

M. Ray Togubu¹, Fahrir Irhamna Rachman², Darniati³

^{1,2,3}Universitas Muhammadiyah Makassar

105841111322@student.unismuh.ac.id, Fachrim141020@unismuh.ac.id

darniati@unismuh.ac.id

Received: 03-07- 2023

Revised: 12-07-2026

Approved: 23-07-2026

ABSTRACT

Penyakit asma memiliki pola gejala, faktor risiko, dan fungsi paru yang heterogen sehingga identifikasi berbasis rekam medis memerlukan pendekatan yang objektif dan terukur. Penelitian ini mengimplementasikan Support Vector Machine (SVM) kernel Radial Basis Function (RBF) untuk mengklasifikasikan pasien tidak asma/normal dan asma berdasarkan 1.087 rekam medis RSUD Tidore periode 1 Januari 2021–31 Desember 2025 dengan 13 fitur klinis. Data dibagi secara stratified menjadi 80% data latih dan 20% data uji; MinMaxScaler ditempatkan dalam pipeline untuk mencegah information leakage. Hyperparameter C dan γ dioptimasi menggunakan Grid Search dengan stratified 5-fold cross-validation pada data latih, dan kestabilan model dievaluasi menggunakan nested stratified 5-fold cross-validation. Parameter terbaik pada data latih adalah $C=30$ dan $\gamma=0,01$. Pada holdout test set, model menghasilkan $TN=89$, $FP=1$, $FN=6$, dan $TP=122$ dengan akurasi 96,79%, presisi 99,19%, sensitivitas 95,31%, spesifisitas 98,89%, F1-score 97,21%, Matthews Correlation Coefficient (MCC) 0,935, dan ROC-AUC 0,9949. Nested cross-validation menghasilkan ROC-AUC $0,9898 \pm 0,0049$ dan akurasi $95,21 \pm 1,56\%$, yang menunjukkan performa relatif stabil pada pembagian data yang berbeda. Permutation importance menempatkan FEV1, dada terasa berat, riwayat keluarga asma, dan sesak napas sebagai fitur dengan kontribusi prediktif terbesar. Enam false negative terutama menunjukkan ketiadaan sesak napas dan gejala malam hari, sehingga penekanan false negative tetap penting dari perspektif patient safety. Keterbatasan penelitian meliputi data satu pusat, distribusi kelas yang tidak sepenuhnya seimbang, ketidaktersediaan provenance rinci label diagnosis, dan metadata unit spirometri. Model diposisikan sebagai Clinical Decision Support System dan tidak menggantikan diagnosis profesional dokter.

Kata Kunci : Asma, Rekam Medis Elektronik, Support Vector Machine, Kernel RBF, Fungsi Paru.

PENDAHULUAN

Asma merupakan penyakit inflamasi kronis saluran pernapasan yang ditandai gejala berulang, seperti sesak napas, batuk, mengi, dan rasa berat di dada. Gejala tersebut dapat berbeda dalam frekuensi, intensitas, dan pemicunya sehingga penilaian status asma tidak selalu sederhana. Tantangan pengendalian dan penilaian klinis asma juga dipengaruhi oleh variasi karakteristik pasien dan tingkat keparahan penyakit [1], [2]. Faktor risiko dan karakteristik klinis, termasuk usia, riwayat alergi, kebiasaan merokok, riwayat keluarga, serta fungsi paru, berkontribusi terhadap heterogenitas kondisi asma [3].

Dalam praktik pelayanan kesehatan, rekam medis elektronik menyimpan data demografis, gejala, hasil pemeriksaan, diagnosis, dan riwayat pengobatan. Pemanfaatan sekunder rekam medis membuka peluang analisis komputasional, tetapi juga menghadapi masalah kualitas data, kelengkapan variabel, dan representativitas populasi [4]. Selain itu, dokumentasi gejala seperti mengi, sesak napas, batuk, dan dada terasa berat dapat dipengaruhi oleh proses anamnesis, interpretasi pemeriksa, serta format pencatatan. Studi dokumentasi asma menunjukkan bahwa variasi pencatatan tetap terjadi pada rekam medis elektronik dan dapat berbeda antarpraktisi, sehingga berpotensi menjadi sumber bias bagi analisis hilir dan model informatika [18], [19].

Perkembangan machine learning memungkinkan data pasien dipelajari untuk menghasilkan klasifikasi atau prediksi penyakit. Pendekatan ini telah digunakan pada berbagai aplikasi kesehatan karena mampu mengolah banyak variabel secara simultan dan menghasilkan pola keputusan yang tidak mudah diperoleh melalui pemeriksaan manual [5]. Salah satu algoritma supervised learning yang banyak diterapkan untuk klasifikasi medis adalah Support Vector Machine (SVM). SVM membentuk hyperplane dengan margin maksimum dan dapat

menangani hubungan nonlinier melalui fungsi kernel [6]. Penggunaannya pada rekam medis juga menunjukkan potensi untuk menghasilkan model klasifikasi yang stabil pada data dengan banyak fitur [7].

Pada klasifikasi asma, machine learning dapat membantu mengintegrasikan pola gejala, faktor risiko, dan hasil pemeriksaan fungsi paru dalam satu proses analisis. Penelitian sebelumnya menunjukkan bahwa definisi operasional dan pemodelan berbasis machine learning dapat meningkatkan ketepatan identifikasi asma [8]. Validasi definisi kasus asma pada rekam medis juga penting agar model belajar dari label target yang konsisten [9]. Namun, kinerja model sangat dipengaruhi oleh tahapan praproses, skala fitur, distribusi kelas, konfigurasi parameter, dan metode evaluasi.

Penelitian ini menggunakan data rekam medis RSUD Tidore dan memfokuskan klasifikasi pada dua kategori, yaitu tidak asma/normal dan asma. Kontribusi penelitian diarahkan pada analisis kemampuan SVM RBF dalam memodelkan kombinasi 13 fitur klinis lokal yang mencakup karakteristik pasien, faktor risiko, gejala, dan fungsi paru, serta pada pengujian kestabilan, diskriminasi, interpretabilitas, dan pola misklasifikasinya. Verifikasi fungsi keputusan menggunakan Microsoft Excel diposisikan hanya sebagai quality control implementasi, bukan sebagai scientific novelty. Penelitian bertujuan: (1) mengimplementasikan SVM RBF pada kombinasi 13 fitur klinis rekam medis; (2) mengoptimasi C dan gamma serta mengevaluasi kestabilan model menggunakan cross-validation dan metrik yang sesuai untuk klasifikasi tidak seimbang; dan (3) menganalisis kontribusi fitur serta karakteristik kasus false negative dan false positive.

METODE PENELITIAN

A. Desain, Sumber Data, dan Variabel Penelitian

Penelitian menggunakan pendekatan kuantitatif komputasional dengan skema klasifikasi biner. Data berupa data sekunder rekam medis RSUD Tidore, Kota Tidore Kepulauan, Maluku Utara, periode 1 Januari 2021 sampai 31 Desember 2025, telah dianonimkan dan berjumlah 1.087 pasien berusia 5–79 tahun. Variabel target Diagnosis_Asma terdiri atas label 0 untuk tidak asma/normal dan label 1 untuk asma. Dataset teranonimkan yang tersedia untuk analisis tidak memuat metadata tentang kualifikasi klinis yang menetapkan label, mekanisme verifikasi diagnosis, atau provenance kode ICD secara terpisah. Oleh karena itu, Diagnosis_Asma diperlakukan sebagai label diagnosis yang tercatat pada dataset, bukan diklaim sebagai gold standard independen. Keterbatasan provenance label ini dipertimbangkan dalam interpretasi hasil.

Tiga belas fitur masukan terdiri atas Usia, Status_Merokok, Riwayat_Alergi, Alergi_Hewan, Riwayat_Keluarga_Asma, Mengi, Sesak_Napas, Dada_Terasa_Berat, Batuk, Gejala_Malam_Hari, Dipicu_Olahraga, FEV1, dan FVC. Kolom ID_Pasien hanya digunakan sebagai identitas dan dikeluarkan dari pemodelan. Mengacu pada pentingnya rasio FEV1/FVC dalam interpretasi spirometri klinis [17], kelayakan pembentukan rasio tersebut diperiksa terlebih dahulu. Pada dataset, FEV1 berada pada rentang 35–112 dan FVC 45–116 tanpa metadata unit; sebanyak 326 dari 1.087 data (29,99%) memiliki FEV1>FVC. Karena kondisi ini tidak konsisten dengan asumsi kedua kolom sebagai volume spirometri mentah yang langsung dapat dirasioikan, FEV1/FVC tidak dibentuk untuk mencegah interpretasi klinis yang tidak valid. FEV1 dan FVC dipertahankan sebagai fitur numerik sebagaimana tercatat pada sumber data.

Tabel 1. Fitur Masukan dan Target Klasifikasi

Kelompok	Variabel	Tipe
Karakteristik	Usia	Numerik
Faktor/riwayat	Status merokok; riwayat alergi; alergi hewan; riwayat keluarga asma	Biner
Gejala klinis	Mengi; sesak napas; dada terasa berat; batuk; gejala malam hari; dipicu olahraga	Biner
Fungsi paru	FEV1; FVC	Numerik
Target	Diagnosis_Asma: 0=tidak asma/normal; 1=asma	Biner

B. Praproses dan Pembagian Data

Dataset diperiksa berdasarkan jumlah baris dan kolom, nama variabel, tipe data, nilai kosong, duplikasi, validitas fitur biner, serta konsistensi rentang FEV1 dan FVC. Hasil pemeriksaan menunjukkan tidak terdapat nilai kosong dan tidak ditemukan data duplikat. Audit nilai fungsi paru digunakan untuk menentukan bahwa rasio FEV1/FVC tidak dapat diturunkan secara klinis tanpa klarifikasi unit sumber.

Normalisasi dilakukan menggunakan MinMaxScaler karena SVM sensitif terhadap perbedaan skala fitur. Untuk mencegah information leakage, MinMaxScaler dan SVM ditempatkan dalam satu pipeline sehingga parameter penskalaan pada setiap proses pelatihan hanya dipelajari dari data training/fold terkait [10], [11].

Data dibagi secara stratified dengan rasio 80:20 dan random_state=52 menjadi 869 data latih dan 218 data uji. Data uji dipertahankan sebagai final holdout test set dan tidak digunakan dalam optimasi hyperparameter. Pada data latih, Grid Search mengevaluasi $C=\{0,01; 0,1; 1; 10; 30; 100; 300; 1.000; 3.000; 10.000\}$ dan $\gamma=\{0,0001; 0,0003; 0,001; 0,003; 0,01; 0,03; 0,1; 1; 10; \text{scale}; \text{auto}\}$ menggunakan stratified 5-fold cross-validation dengan ROC-AUC sebagai fungsi seleksi. Selain itu, nested stratified 5-fold cross-validation diterapkan pada keseluruhan dataset; setiap outer fold memiliki Grid Search 5-fold pada bagian training untuk mengurangi bias optimasi dan menilai kestabilan model pada pembagian data yang berbeda.

Tabel 2. Pembagian dan Distribusi Kelas Dataset

Jenis Data	Tidak Asma/Normal	Asma	Total
Data latih	359	510	869
Data uji	90	128	218
Total	449	638	1.087

C. Pelatihan dan Pengujian Model

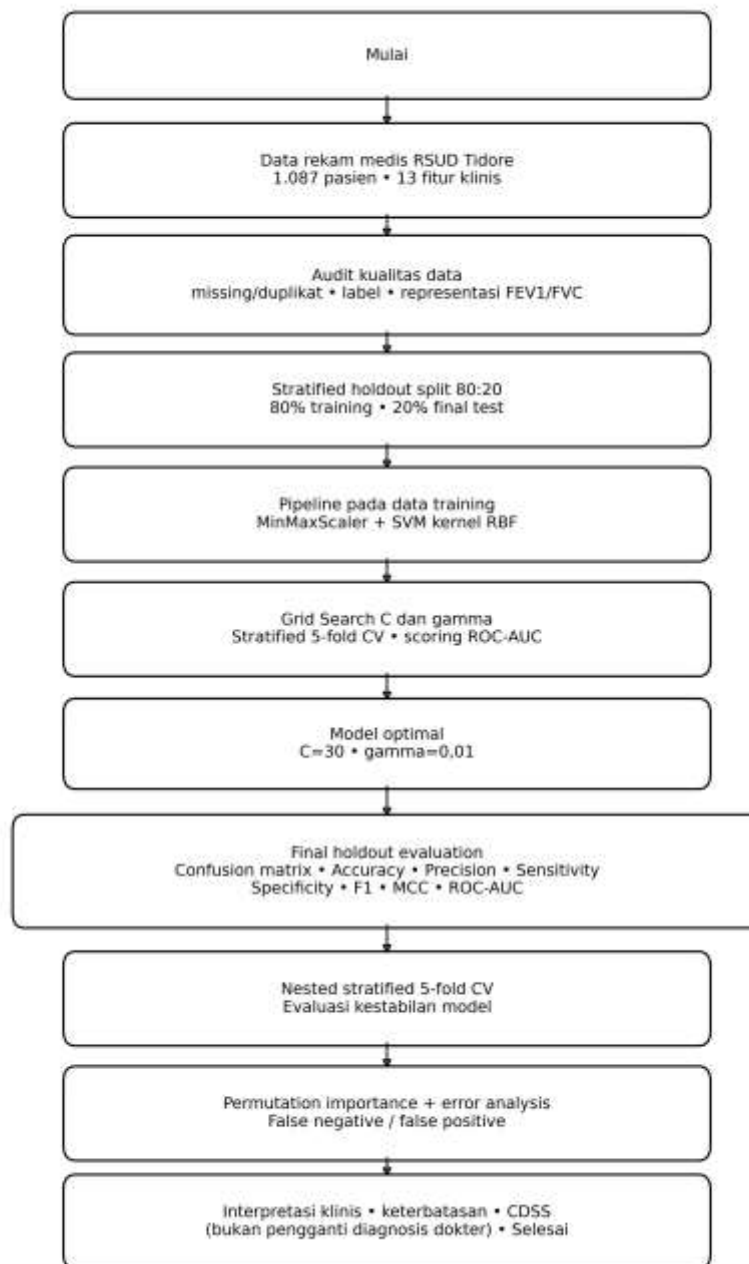
Model menggunakan SVM kernel RBF untuk menangkap hubungan nonlinier antarfitur klinis [12]. Grid Search pada data latih memilih $C=30$ dan $\gamma=0,01$ dengan rata-rata ROC-AUC internal 0,9889. Fungsi keputusan model dinyatakan sebagai $f(x)=\sum \alpha_i y_i \exp(-\gamma||x_i-x||^2)+b$; $f(x)<0$ diklasifikasikan sebagai kelas 0 dan $f(x)\geq 0$ sebagai kelas 1.

Evaluasi final dilakukan pada 218 data uji menggunakan confusion matrix dan metrik akurasi, presisi, sensitivitas/recall, spesifisitas, F1-score, Matthews Correlation Coefficient (MCC), dan Receiver Operating Characteristic–Area Under the Curve (ROC-AUC). MCC digunakan karena merangkum keempat komponen confusion matrix dan relevan untuk evaluasi klasifikasi biner dengan distribusi kelas yang tidak sepenuhnya seimbang [13], sedangkan ROC-AUC menilai kemampuan diskriminasi model pada berbagai threshold keputusan.

Sebagai quality control implementasi, fungsi keputusan satu sampel diuji secara independen dengan parameter model yang sama. Pemeriksaan ini hanya digunakan untuk memastikan konsistensi numerik implementasi dan tidak diposisikan sebagai kontribusi ilmiah utama atau pengganti evaluasi performa model.

Tabel 3. Konfigurasi Model SVM

Parameter	Nilai
Algoritma	Support Vector Machine
Kernel	Radial Basis Function
C	30
Gamma	0,01
Normalisasi	MinMaxScaler
Pembagian data	80:20 stratified; final holdout
Random state	52
Optimasi hyperparameter	Grid Search pada data latih
Validasi internal	Stratified 5-fold CV
Validasi kestabilan	Nested stratified 5-fold CV
Kriteria optimasi	ROC-AUC



Gambar 1. Alur Penelitian dan Evaluasi Model SVM

HASIL DAN PEMBAHASAN

A. Kondisi Awal Dataset dan Praproses

Dataset terdiri atas 1.087 baris dan 15 kolom, yaitu satu kolom identitas, 13 fitur masukan, dan satu target klasifikasi. Pemeriksaan menunjukkan nilai kosong sebanyak 0 dan data duplikat sebanyak 0 sehingga seluruh data dapat dilanjutkan ke tahap pemodelan. Kolom ID_Pasien tidak digunakan sebagai prediktor karena tidak memiliki makna klinis terhadap status asma.

Fitur numerik Usia, FEV1, dan FVC memiliki rentang berbeda dibandingkan fitur biner sehingga normalisasi diperlukan. Audit tambahan menunjukkan rentang FEV1 35–112 dan FVC 45–116; 326 data (29,99%) memiliki FEV1 lebih besar daripada FVC. Karena unit dan representasi spirometri tidak terdokumentasi pada dataset analisis, rasio FEV1/FVC tidak diturunkan. Keputusan ini mencegah pengenalan fitur turunan yang dapat menyesatkan secara klinis.

Tabel 4. Hasil Pemeriksaan Awal Dataset

Pemeriksaan	Hasil
Jumlah data	1.087
Jumlah kolom	15
Nilai kosong	0
Data duplikat	0
Kolom identitas	ID_Pasien
Kolom target	Diagnosis_Asma

B. Pembagian Data dan Pembentukan Model

Kelas target pada dataset awal terdiri atas 449 data tidak asma/normal dan 638 data asma. Setelah pembagian stratified, data latih memuat 359 kelas tidak asma dan 510 kelas asma, sedangkan data uji memuat 90 kelas tidak asma dan 128 kelas asma. Distribusi tersebut mempertahankan komposisi kelas awal, sehingga evaluasi pada data uji lebih representatif terhadap dataset yang tersedia.

Grid Search pada 869 data latih menghasilkan kombinasi terbaik $C=30$ dan $\gamma=0,01$ dengan rata-rata ROC-AUC 0,9889 pada stratified 5-fold cross-validation. Model teroptimasi kemudian dikunci dan dievaluasi sekali pada 218 data uji yang tidak digunakan selama pemilihan hyperparameter. Pendekatan ini memisahkan proses model selection dari final model assessment.

C. Hasil Evaluasi Model

Tabel 5. Confusion Matrix Hasil Pengujian

Kondisi Aktual/Prediksi	Prediksi Tidak Asma	Prediksi Asma
Aktual Tidak Asma	89	1
Aktual Asma	6	122

Confusion matrix model teroptimasi menunjukkan 89 pasien tidak asma diprediksi benar dan 122 pasien asma diprediksi benar. Kesalahan terdiri atas 1 false positive dan 6 false negative. Dengan demikian, 211 dari 218 data uji diklasifikasikan secara benar. Dibanding konfigurasi awal $C=1$ dan $\gamma=scale$, optimasi mempertahankan akurasi yang sama tetapi menurunkan false positive dari 3 menjadi 1 dan meningkatkan false negative dari 4 menjadi 6. Perubahan ini menunjukkan adanya trade-off antara spesifisitas dan sensitivitas pada threshold keputusan standar.

False negative memiliki konsekuensi klinis lebih penting karena pasien dengan label asma dapat diklasifikasikan sebagai tidak asma. Walaupun jumlahnya enam kasus, kesalahan tersebut berpotensi menunda evaluasi atau tindak lanjut apabila keluaran model digunakan secara tunggal. Oleh karena itu, performa agregat yang tinggi tidak cukup untuk membenarkan penggunaan model sebagai penentu diagnosis; keluaran harus diperlakukan sebagai informasi pendukung dan dikonfirmasi secara klinis.

Tabel 6. Metrik Evaluasi Model SVM

Metrik	Nilai
Akurasi	96,79%
Presisi	99,19%
Sensitivitas/Recall	95,31%
Spesifisitas	98,89%
F1-score	97,21%
MCC	0,935
ROC-AUC	0,9949

Pada final holdout test set, model mencapai akurasi 96,79%, presisi 99,19%, sensitivitas 95,31%, spesifisitas 98,89%, F1-score 97,21%, MCC 0,935, dan ROC-AUC 0,9949. MCC yang mendekati 1 menunjukkan kesesuaian yang sangat kuat antara prediksi dan label aktual ketika seluruh komponen confusion matrix dipertimbangkan. ROC-AUC 0,9949 menunjukkan

kemampuan diskriminasi yang sangat tinggi pada test set, tetapi nilai ini tetap merupakan validasi internal satu institusi.

Nested stratified 5-fold cross-validation menghasilkan rata-rata $\pm$ SD: akurasi 95,21 $\pm$ 1,56%, presisi 97,27 $\pm$ 1,42%, sensitivitas 94,51 $\pm$ 1,94%, spesifisitas 96,21 $\pm$ 2,05%, F1-score 95,86 $\pm$ 1,36%, MCC 0,903 $\pm$ 0,032, dan ROC-AUC 0,9898 $\pm$ 0,0049. Variasi antarfold yang relatif kecil mendukung bahwa performa tidak semata-mata bergantung pada satu random split. Hasil ini sejalan dengan studi machine learning untuk identifikasi asma yang menunjukkan potensi diskriminatif tinggi, meskipun validasi eksternal tetap diperlukan [8], [15].

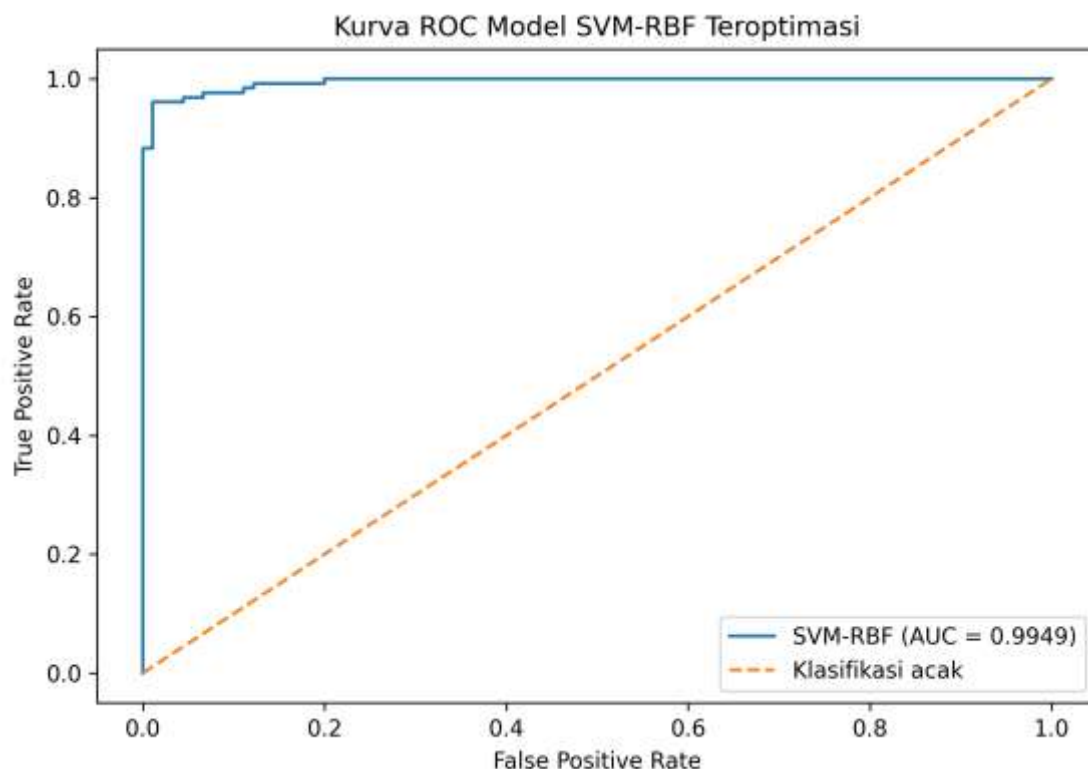
D. Validasi Silang, ROC-AUC, dan Interpretabilitas Model

Tabel 7. Hasil Nested Stratified 5-Fold Cross-Validation

Metrik	Rata-rata $\pm$ SD
Akurasi	95,21 $\pm$ 1,56%
Presisi	97,27 $\pm$ 1,42%
Sensitivitas	94,51 $\pm$ 1,94%
Spesifisitas	96,21 $\pm$ 2,05%
F1-score	95,86 $\pm$ 1,36%
MCC	0,903 $\pm$ 0,032
ROC-AUC	0,9898 $\pm$ 0,0049

Kurva ROC pada final holdout test set menghasilkan AUC 0,9949, yang mengindikasikan kemampuan model membedakan kedua kelas pada berbagai threshold keputusan. Namun, ROC-AUC tidak secara langsung menentukan threshold klinis yang paling aman; peningkatan false negative setelah tuning menunjukkan bahwa pemilihan threshold dan cost-sensitive learning perlu dipertimbangkan bila model diarahkan untuk skrining.

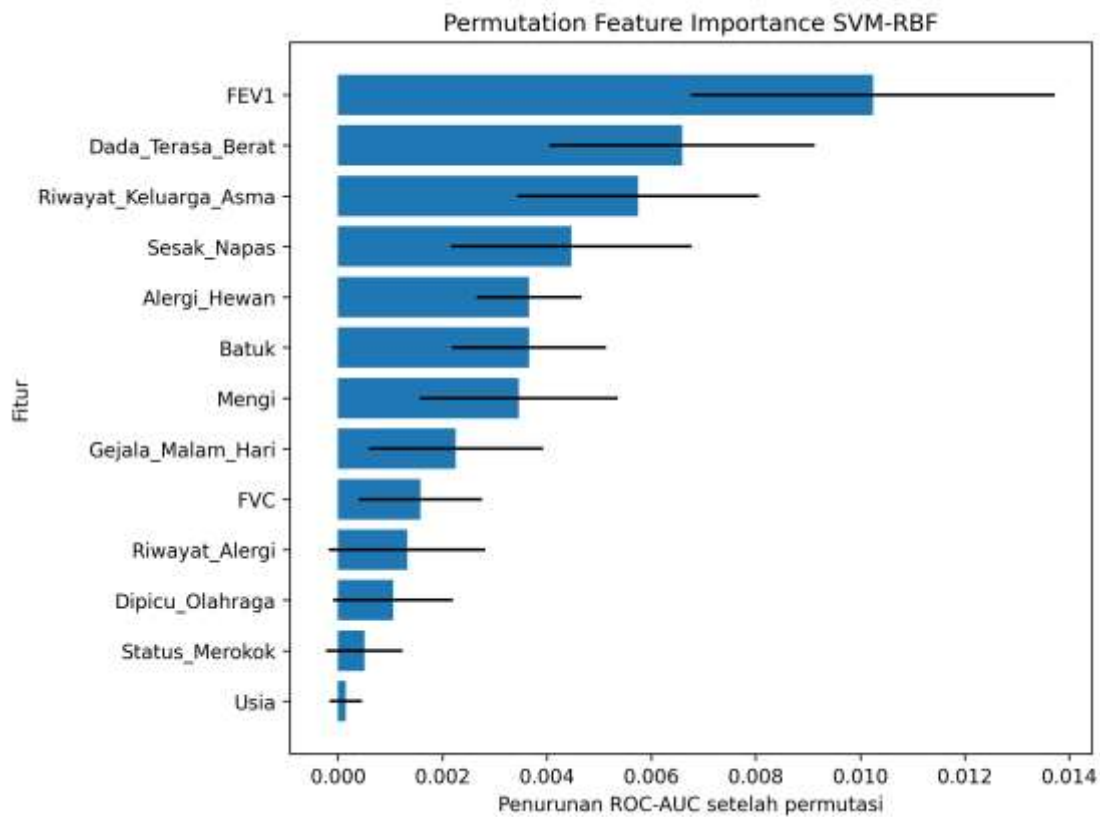
Gambar 2. Kurva ROC Model SVM-RBF Teroptimasi



Permutation feature importance dengan ROC-AUC sebagai scoring function menunjukkan FEV1 sebagai fitur dengan kontribusi prediktif terbesar (mean decrease ROC-AUC=0,0102), diikuti dada terasa berat (0,0066), riwayat keluarga asma (0,0057), sesak napas (0,0045), serta alergi hewan (0,0037). Hasil ini menunjukkan bahwa keputusan model dibentuk oleh kombinasi

fungsi paru dan informasi klinis. Nilai importance merupakan ukuran prediktif dan tidak boleh ditafsirkan sebagai hubungan kausal.

Gambar 3. Permutation Feature Importance Model SVM-RBF



E. Analisis Kesalahan Klasifikasi

Error analysis pada tujuh kasus misklasifikasi menunjukkan enam false negative dan satu false positive. Seluruh false negative tidak memiliki fitur Sesak_Napas dan Gejala_Malam_Hari; hanya satu dari enam memiliki Mengi dan satu dari enam memiliki Dipicu_Olahraga. Rata-rata FEV1 pada false negative adalah 77,62, lebih tinggi daripada true positive 66,09; rata-rata FVC false negative 91,82, juga lebih tinggi daripada true positive 75,83. Pola tersebut membuat sejumlah pasien berlabel asma memiliki representasi fitur yang lebih menyerupai kelas tidak asma. Temuan ini bersifat deskriptif dan harus ditafsirkan bersama keterbatasan provenance label serta unit fungsi paru.

Tabel 8. Karakteristik Ringkas Kasus Misklasifikasi (profil Mengi/Sesak/Gejala Malam)

ID	Jenis	Usia	Mengi/Sesak/ Malam	FEV1	FVC	Decision score
P0041	FN	24	0/0/0	79.9	95.3	-1.541
P0681	FN	11	0/0/0	62.0	90.2	-0.934
P0705	FN	72	0/0/0	70.1	82.6	-0.699
P0233	FN	38	1/0/0	84.5	101.9	-0.525
P0860	FN	76	0/0/0	81.9	90.1	-0.352
P1029	FN	64	0/0/0	87.3	90.8	-0.207
P0905	FP	37	1/0/0	73.8	65.7	1.142

F. Implikasi, Quality Control, dan Keterbatasan Penelitian

Model berpotensi menjadi komponen Clinical Decision Support System untuk membantu klasifikasi awal berbasis rekam medis. Namun, keluaran sistem tidak memiliki kewenangan menetapkan diagnosis secara mandiri dan tidak boleh menggantikan anamnesis, pemeriksaan fisik, spirometri yang tervalidasi, maupun pertimbangan profesional dokter. Keputusan diagnosis dan tata laksana akhir tetap berada pada tenaga medis yang berwenang. Verifikasi manual fungsi keputusan dipertahankan hanya sebagai quality control implementasi dan dipindahkan dari fokus pembahasan utama.

Keterbatasan penelitian mencakup sumber data dari satu fasilitas kesehatan, distribusi kelas 449 tidak asma versus 638 asma, ketidakterediaan metadata unit FEV1/FVC, dan provenance rinci label Diagnosis_Asma yang tidak tersedia pada dataset analisis. Walaupun nested cross-validation

mengurangi ketergantungan terhadap satu holdout split, generalisasi ke populasi lain belum terbukti. Penelitian berikutnya perlu melibatkan data multisenter, validasi prospektif, harmonisasi definisi label oleh klinisi berwenang, pencatatan spirometri dengan unit dan reference values yang jelas, evaluasi class weighting/cost-sensitive threshold untuk menekan false negative, serta external validation sebelum implementasi klinis.

CONCLUSION

SVM kernel RBF mampu mengklasifikasikan status asma berdasarkan kombinasi 13 fitur klinis rekam medis dengan performa diskriminatif tinggi. Optimasi Grid Search menghasilkan $C=30$ dan $\gamma=0,01$. Pada final holdout test set, model memperoleh akurasi 96,79%, presisi 99,19%, sensitivitas 95,31%, spesifisitas 98,89%, F1-score 97,21%, MCC 0,935, dan ROC-AUC 0,9949. Nested stratified 5-fold cross-validation menghasilkan akurasi $95,21 \pm 1,56\%$ dan ROC-AUC $0,9898 \pm 0,0049$, sehingga menunjukkan bahwa performa relatif stabil dan tidak hanya bergantung pada satu pembagian data.

Interpretabilitas model menunjukkan bahwa FEV1, dada terasa berat, riwayat keluarga asma, dan sesak napas merupakan fitur dengan kontribusi prediktif utama. Error analysis memperlihatkan bahwa false negative cenderung memiliki lebih sedikit gejala respirasi khas dan nilai fungsi paru yang relatif lebih tinggi, sehingga patient safety tetap menjadi perhatian meskipun metrik agregat sangat tinggi. Keterbatasan data satu pusat, provenance label diagnosis, metadata spirometri, dan ketidakseimbangan kelas membatasi generalisasi. Pengembangan selanjutnya perlu menggunakan data multisenter, validasi prospektif dan eksternal, standardisasi label dan spirometri, serta strategi cost-sensitive untuk menekan false negative. Model harus digunakan sebagai Clinical Decision Support System dan tidak boleh menggantikan keputusan diagnosis profesional dokter.

REFERENCES

- [1] M. A. Tosca, I. Schiavetti, G. Ciprandi, et al., "The challenge of asthma control grading in clinical practice," *Respiratory Investigation*, vol. 59, no. 5, pp. 683–685, 2021, doi: 10.1016/j.resinv.2021.06.001.
- [2] G. Guarnieri et al., "Asthma severity: the patient's point of view," *European Clinical Respiratory Journal*, vol. 11, no. 1, Art. no. 2381307, 2024, doi: 10.1080/20018525.2024.2381307.
- [3] Y. Y. Sio and F. T. Chew, "Risk factors of asthma in the Asian population: a systematic review and meta-analysis," *Journal of Physiological Anthropology*, vol. 40, Art. no. 22, 2021, doi: 10.1186/s40101-021-00273-x.
- [4] T. Sarwar, S. Seifollahi, J. Chan, X. Zhang, V. Aksakalli, I. Hudson, K. Verspoor, and L. Cavedon, "The secondary use of electronic health records for data mining: data characteristics and challenges," *ACM Computing Surveys*, vol. 55, no. 2, Art. no. 33, pp. 1–40, 2022, doi: 10.1145/3490234.
- [5] U. M. Butt, S. Letchmunan, M. Ali, F. H. Hassan, A. Baqir, and H. H. R. Sherazi, "Machine learning based diabetes classification and prediction for healthcare applications," *Journal of Healthcare Engineering*, vol. 2021, Art. no. 9930985, 2021, doi: 10.1155/2021/9930985.
- [6] R. Guido, S. Ferrisi, D. Lofaro, and D. Conforti, "An overview on the advancements of Support Vector Machine models in healthcare applications: a review," *Information*, vol. 15, no. 4, Art. no. 235, 2024, doi: 10.3390/info15040235.
- [7] X. Zhou et al., "Support vector machine deep mining of electronic medical records to predict the prognosis of severe acute myocardial infarction," *Frontiers in Physiology*, vol. 13, Art. no. 991990, 2022, doi: 10.3389/fphys.2022.991990.
- [8] H. Joo, D. Lee, S. H. Lee, Y. K. Kim, and C. K. Rhee, "Increasing the accuracy of the asthma diagnosis using an operational definition for asthma and a machine learning method," *BMC Pulmonary Medicine*, vol. 23, Art. no. 196, 2023, doi: 10.1186/s12890-023-02479-4.
- [9] M. Moloney et al., "Validation of adult asthma case definitions for primary care sentinel surveillance," *Allergy, Asthma & Clinical Immunology*, vol. 19, Art. no. 95, 2023, doi: 10.1186/s13223-023-00854-8.

- [10] M. M. Ahsan, M. A. P. Mahmud, P. K. Saha, K. D. Gupta, and Z. Siddique, "Effect of data scaling methods on machine learning algorithms and model performance," *Technologies*, vol. 9, no. 3, Art. no. 52, 2021, doi: 10.3390/technologies9030052.
- [11] I. Izonin, R. Tkachenko, N. Shakhovska, B. Ilchyshyn, and K. K. Singh, "A two-step data normalization approach for improving classification accuracy in the medical diagnosis domain," *Mathematics*, vol. 10, no. 11, Art. no. 1942, 2022, doi: 10.3390/math10111942.
- [12] A. Razaque et al., "Improved Support Vector Machine enabled Radial Basis Function and linear variants for remote sensing image classification," *Sensors*, vol. 21, no. 13, Art. no. 4431, 2021, doi: 10.3390/s21134431.
- [13] D. Chicco, N. Tötsch, and G. Jurman, "The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation," *BioData Mining*, vol. 14, Art. no. 13, 2021, doi: 10.1186/s13040-021-00244-z.
- [14] N. Budhathoki, R. Bhandari, S. Bashyal, and C. Lee, "Predicting asthma using imbalanced data modeling techniques: evidence from 2019 Michigan BRFSS data," *PLOS ONE*, vol. 18, no. 12, Art. no. e0295427, 2023, doi: 10.1371/journal.pone.0295427.
- [15] O. K. Gargari et al., "Assessing the diagnostic accuracy of machine learning algorithms for identification of asthma in United States adults based on NHANES dataset," *Scientific Reports*, vol. 15, Art. no. 4537, 2025, doi: 10.1038/s41598-025-88345-1.
- [16] H. Tang, G. Yang, X. Gu, H. Mao, and H. Cao, "A support vector machine-based tool for rapid pediatric asthma exacerbation risk assessment: development and nursing application," *Frontiers in Pediatrics*, vol. 13, Art. no. 1660895, 2025, doi: 10.3389/fped.2025.1660895.
- [17] Global Initiative for Asthma (GINA), *Global Strategy for Asthma Management and Prevention*, 2026. [Online]. Available: <https://ginasthma.org/2026-gina-strategy-report/>.
- [18] J. Landeo-Gutierrez et al., "Leveraging Electronic Health Records for Guideline-Based Asthma Documentation," *Journal of Allergy and Clinical Immunology: In Practice*, vol. 11, no. 3, pp. 855–862.e4, 2023, doi: 10.1016/j.jaip.2022.11.032.
- [19] S. Sohn, C.-I. Wi, Y. J. Juhn, and H. Liu, "Analysis of clinical variations in asthma care documented in electronic health records between staff and resident physicians," *Studies in Health Technology and Informatics*, vol. 245, pp. 1170–1174, 2017, doi: 10.3233/978-1-61499-830-3-1170.