Methotika : Jurnal Ilmiah Teknik Informatika
E-ISSN : 2776 – 5792
Vol. 6, No. 1, April 2026, pp. 128-134
https://ejurnal.methodist.ac.id/index.php/methotika/

Penerapan Random Forest Sebagai Sistem Pendukung Keputusan Diagnosis Dini Penyakit Tiroid

Putri Maria Cornelia Purba[1], Darwis Manalu[2], Samuel Manurung[3]
Fakultas Ilmu Komputer, Universitas Methodist Indonesia

Info Artikel

Histori Artikel:

Received, Feb 21 , 2026
Revised, Mar 17, 2026
Accepted, April 08, 2026

Keywords:

Random Forest,
Klasifikasi,
Penyakit Tiroid,
Machine Learning,
Diagnosis Medis.

ABSTRAK

Penyakit tiroid merupakan salah satu gangguan endokrin yang memengaruhi metabolisme tubuh dan memerlukan diagnosis cepat serta akurat. Penelitian ini bertujuan mengimplementasikan algoritma Random Forest dalam mengklasifikasikan tingkat keparahan penyakit tiroid (low, medium, high) berdasarkan data pasien dari platform Kaggle yang berjumlah 383 sampel. Algoritma Random Forest dipilih karena kemampuannya dalam mengelola data dengan dimensi yang tinggi serta menghasilkan prediksi yang stabil dan akurat. Tahapan penelitian meliputi pengumpulan data, preprocessing, pemilihan fitur relevan, implementasi algoritma, serta evaluasi model menggunakan bahasa pemrograman Python. Model dievaluasi dengan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa algoritma Random Forest mampu mengklasifikasikan tingkat keparahan penyakit tiroid dengan tingkat akurasi sebesar 82%, presisi 76%, recall 63%, dan F1-score 68%. Metode ini dapat menjadi alternatif efektif sebagai sistem pendukung keputusan bagi tenaga medis dalam diagnosis dini penyakit tiroid.

Penulis Koresponden:

Putri Maria Cornelia Purba,
Fakultas Ilmu Komputer,
Universitas Methodist Indonesia, Medan,
Jl. Hang Tuah No.8, Medan - Sumatera Utara.
Email: putrimariacorneliapurba@gmail.com

1. PENDAHULUAN

Gangguan pada kelenjar tiroid—yang mencakup hipotiroidisme, hipertiroidisme, hingga keganasan tiroid—telah menjelma sebagai salah satu persoalan kesehatan publik dengan dampak signifikan di berbagai belahan dunia. Data dari World Health Organization (WHO) mencatat bahwa tidak kurang dari 200 juta individu di seluruh penjuru dunia hidup dengan berbagai bentuk disfungsi tiroid[1]. Kondisi ini turut berdampak besar di Indonesia, di mana lebih dari 10 juta warganya tercatat telah didiagnosis menderita penyakit tiroid, menjadikan penatalaksanaan yang tepat sasaran sebagai kebutuhan strategis dalam sistem pelayanan kesehatan nasional. Besarnya beban penyakit ini menegaskan bahwa gangguan tiroid telah

bergeser dari sekadar masalah kesehatan individu menjadi isu kesehatan masyarakat yang memerlukan perhatian lintas sektor, baik di ranah medis maupun teknologi.

Kondisi ini membutuhkan kewaspadaan khusus sebab manifestasi klinisnya yang tidak spesifik kerap mengakibatkan penundaan dalam proses penegakan diagnosis. Sebagai organ endokrin dengan ukuran terbesar di dalam tubuh manusia, kelenjar tiroid bertanggung jawab memproduksi hormon T3 dan T4 yang memegang peranan krusial dalam regulasi metabolisme, tumbuh kembang, serta berbagai fungsi fisiologis organ secara menyeluruh [1]. Apabila kelenjar ini mengalami disfungsi, konsekuensinya bersifat multisistemik dan berpotensi menurunkan kualitas hidup penderita secara bermakna, mencakup gangguan pada proses metabolisme, fluktuasi berat badan yang tidak normal, rasa lelah berkepanjangan, hingga komplikasi serius pada sistem kardiovaskular.

Pengelompokan pasien tiroid ke dalam strata risiko yang akurat merupakan langkah fundamental dalam menentukan strategi terapi yang paling tepat. Sejumlah faktor klinis telah terbukti berkontribusi terhadap derajat keparahan penyakit, antara lain: jenis kelamin, mengingat perempuan memperlihatkan kerentanan yang lebih tinggi terhadap keganasan tiroid; riwayat merokok yang secara paradoksal justru dikaitkan dengan penurunan insidensi kanker tiroid; paparan radioterapi pada regio kepala dan leher yang secara bermakna meningkatkan risiko transformasi maligna; serta kondisi fungsi tiroid abnormal seperti hipotiroidisme dan hipertiroidisme yang berkorelasi kuat dengan progresi penyakit. Di samping itu, temuan klinis berupa pembesaran kelenjar limfa servikal (adenopati), karakteristik histopatologi, dan pola pertumbuhan nodul (fokalitas) turut berperan sebagai penanda prognostik yang krusial bagi penentuan perjalanan penyakit.

Proses penegakan diagnosis penyakit tiroid secara tradisional masih sangat mengandalkan kompetensi dokter spesialis dalam menafsirkan serangkaian hasil pemeriksaan, mulai dari uji laboratorium, evaluasi fisik, hingga analisis biopsi jaringan. Prosedur ini tidak hanya menyita waktu yang cukup lama, namun juga berpotensi menimbulkan variabilitas penilaian antarpemeriksa. Penundaan diagnosis dapat memperparah kondisi klinis pasien, khususnya pada kasus keganasan tiroid yang bersifat progresif dan memerlukan intervensi segera. Situasi ini mendorong kebutuhan mendesak akan pengembangan sistem berbasis komputer yang dapat mendukung pengambilan keputusan medis—membantu klinisi dalam mengelompokkan pasien berdasarkan tingkat risiko penyakit tiroid dengan cara yang lebih cepat, terstandarisasi, dan objektif.

Di tengah pesatnya perkembangan teknologi informasi, penerapan kecerdasan buatan khususnya machine learning di sektor pelayanan kesehatan telah membuka cakrawala baru dalam upaya peningkatan ketepatan dan kecepatan diagnosis klinis. Salah satu pendekatan yang telah berhasil divalidasi dalam analisis data medis berskala besar adalah Random Forest [2]. Algoritma ini bekerja dengan prinsip ensemble learning melalui teknik bagging, di mana sejumlah pohon keputusan dibangun secara paralel dan independen, lalu hasil prediksi masing-masing pohon diintegrasikan lewat mekanisme pemungutan suara mayoritas guna menghasilkan output yang lebih andal dan konsisten [3]. Kelebihan utama metode ini terletak pada kapasitasnya mengelola fitur berdimensi tinggi, resistensinya terhadap fenomena overfitting, serta kemampuannya mengukur kontribusi relatif setiap variabel prediktor—aspek yang sangat bermanfaat dalam konteks interpretasi model klinis.

Beberapa penelitian terdahulu telah menerapkan Random Forest dalam berbagai domain klasifikasi medis dan menunjukkan hasil yang menjanjikan. Sebayang et al. (2023) berhasil mencapai akurasi 95,67% dalam klasifikasi data kesehatan mental, Hidayat et al. (2023) memperoleh akurasi 92% pada klasifikasi penyakit jantung, dan Fadli & Saputra (2023) mencapai akurasi 93,6% pada prediksi stroke menggunakan Random Forest [6][7]. Namun,

penelitian yang secara khusus mengklasifikasikan tingkat keparahan penyakit tiroid ke dalam tiga kategori risiko (low, medium, high) menggunakan pendekatan ini masih sangat terbatas, sehingga menjadi celah penelitian yang perlu diisi.

Berdasarkan latar belakang dan permasalahan tersebut, penelitian ini bertujuan untuk menerapkan algoritma Random Forest dalam mengklasifikasikan tingkat keparahan penyakit tiroid (low, medium, high) dengan memanfaatkan data pasien dari platform Kaggle yang mencakup 383 sampel serta sembilan fitur klinis yang relevan. Model dikembangkan menggunakan bahasa pemrograman Python dan dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem pendukung keputusan klinis yang efektif bagi tenaga medis dalam mendeteksi dini dan menentukan tingkat risiko penyakit tiroid.

2. METODE PENELITIAN

2.1 Random Forest

Random Forest merupakan salah satu varian algoritma ensemble yang mengadopsi teknik Bootstrap Aggregating (Bagging) dengan cara mengonstruksi sejumlah besar pohon keputusan yang masing-masing dilatih secara terpisah dan independen [4]. Pada proses pembentukannya, tiap pohon mendapatkan data pelatihan yang berbeda melalui pengambilan sampel acak dengan pengembalian, sementara subset fitur juga dipilih secara stokastik pada setiap proses pemisahan node. Hasil prediksi final ditentukan berdasarkan agregasi suara terbanyak dari seluruh pohon yang telah dibangun. Mekanisme ini terbukti mampu menekan nilai bias sekaligus varians secara simultan dibandingkan penggunaan pohon keputusan tunggal, sehingga model yang dihasilkan memiliki generalisasi yang lebih baik dan lebih stabil [5].

Penentuan node pada setiap pohon menggunakan perhitungan entropy dan information gain dengan rumus:

$$Entropy(S) = -\sum \rho i \times log2(\rho i)$$

$$Gain(S, A) = Entropy(S) - \sum |Si| / |S| \times Entropy(Si)$$

Di mana S adalah kumpulan data, A adalah atribut yang diuji, ρi adalah probabilitas kemunculan kelas i, dan |Si| adalah jumlah kasus dalam kelas i dari atribut A [4].

2.2 Dataset dan Fitur Penelitian

Dataset yang digunakan diperoleh dari platform Kaggle (https://www.kaggle.com/datasets/jainaru/thyroid-disease-data) dan terdiri dari 383 sampel pasien penyakit tiroid. Terdapat sembilan fitur utama yang digunakan, yaitu: jenis kelamin (Gender), kebiasaan merokok (Smoking), riwayat merokok (Hx Smoking), riwayat radioterapi (Hx Radiotherapy), fungsi tiroid (Thyroid Function), pemeriksaan fisik (Physical Examination), adenopati (Adenopathy), jenis patologi (Pathology), dan fokalitas (Focality). Label klasifikasi terdiri dari tiga kategori: Low, Medium, dan High.

2.3 Tinjauan Pustaka

Sejumlah penelitian terdahulu yang menjadi acuan dalam penelitian ini dirangkum dalam Tabel 1.

Tabel 1. Tinjauan Literatur Penelitian Terkait

Peneliti	Judul / Topik	Hasil
Sebayang et al. (2023)	Klasifikasi data kesehatan mental menggunakan Random Forest	Akurasi 95,67%

Hidayat et al. (2023)	Klasifikasi penyakit jantung dengan Random Forest Classifier	Akurasi 92%
Widya Kayohana (2024)	Klasifikasi penyakit hati dengan Random Forest dan KNN	Random Forest: 89%, KNN: 84%
Junus et al. (2023)	Deteksi dini diabetes menggunakan SVM dan Random Forest	Random Forest: 98,08%, SVM: 91,03%
Fadli & Saputra (2023)	Klasifikasi risiko stroke menggunakan Random Forest	Akurasi 93,6%, F1-score 93,7%

2.4 Tahapan Penelitian

Tahapan penelitian meliputi: (1) Pengumpulan data dari Kaggle; (2) Preprocessing data mencakup encoding data kategorik menjadi numerik menggunakan LabelEncoder dari scikit-learn; (3) Pembagian data menjadi 80% data latih dan 20% data uji; (4) Implementasi algoritma Random Forest menggunakan Python dengan 100 pohon dan kedalaman maksimum 5; serta (5) Evaluasi model menggunakan metrik akurasi, presisi, recall, dan F1-score melalui confusion matrix.

3. HASIL DAN PEMBAHASAN

3.1 Preprocessing Data

Proses preprocessing dilakukan terhadap 383 sampel data pasien penyakit tiroid yang diperoleh dari platform Kaggle. Dataset ini mencakup sembilan fitur klinis, yaitu: Gender, Smoking, Hx Smoking, Hx Radiotherapy, Thyroid Function, Physical Examination, Adenopathy, Pathology, dan Focality, dengan label klasifikasi berupa tingkat risiko (Low, Medium, High). Distribusi kelas pada dataset menunjukkan ketidakseimbangan yang cukup signifikan, di mana kelas Low mendominasi dengan jumlah sampel terbesar, diikuti kelas Medium, dan kelas High dengan jumlah sampel yang paling sedikit. Ketidakseimbangan distribusi ini menjadi salah satu faktor yang memengaruhi performa model, terutama pada kelas minoritas.

Karena seluruh data awal bersifat kategorik, dilakukan proses encoding menggunakan LabelEncoder dari library scikit-learn untuk mengubah setiap variabel kategorik menjadi representasi numerik yang dapat diproses oleh algoritma machine learning. Hasil encoding untuk setiap variabel adalah sebagai berikut: (1) Gender: Female = 0, Male = 1; (2) Smoking dan Hx Smoking: No = 0, Yes = 1; (3) Hx Radiotherapy: No = 0, Yes = 1; (4) Thyroid Function: Clinical Hyperthyroidism = 0, Clinical Hypothyroidism = 1, Euthyroid = 2, Subclinical Hyperthyroidism = 3, Subclinical Hypothyroidism = 4; (5) Adenopathy: Bilateral = 0, Extensive = 1, Left = 2, No = 3, Right = 5; (6) Pathology: Follicular = 0, Hurthel cell = 1, Micropapillary = 2, Papillary = 3; (7) Focality: Multi-Focal = 0, Uni-Focal = 1; dan (8) variabel target Risk: High = 0, Low = 1, Medium = 2. Setelah encoding, data dibagi menjadi 80% data latih (306 sampel) dan 20% data uji (77 sampel) untuk keperluan pelatihan dan evaluasi model.

3.2 Pembentukan Pohon Keputusan

Pada setiap pohon dalam Random Forest, terdapat tiga tahapan inti yang dijalankan secara sekuensial. Tahap pertama adalah bootstrap sampling, yaitu pengambilan sampel secara acak dengan pengembalian dari data latih sehingga masing-masing pohon mendapatkan komposisi data pelatihan yang unik. Tahap kedua adalah seleksi fitur acak sebesar $\sqrt{9} = 3$ variabel dari keseluruhan 9 fitur yang tersedia di setiap titik percabangan. Tahap ketiga adalah pembangunan struktur pohon keputusan berdasarkan fitur-fitur terpilih

menggunakan kalkulasi nilai entropi dan information gain sebagai kriteria pemisahan. Berdasarkan perhitungan manual pada subset 30 sampel dengan konfigurasi 5 pohon berkedalaman 5, diperoleh nilai entropi root node sebesar 1,555. Dalam simulasi tersebut, pohon pertama mengidentifikasi fitur Smoking sebagai root node dengan information gain 0,184; pohon kedua memilih Adenopathy (0,299); pohon ketiga memilih Pathology (0,524); pohon keempat memilih Gender (0,260); dan pohon kelima kembali memilih Adenopathy (0,650). Penetapan prediksi akhir dilakukan melalui agregasi hasil voting dari keseluruhan pohon. Pada implementasi skala penuh menggunakan Python dengan 100 pohon dan 383 sampel data, fitur Adenopathy, Pathology, Smoking, dan Gender secara konsisten muncul sebagai prediktor dominan—mengindikasikan bahwa pembesaran kelenjar limfa, karakteristik patologi jaringan, kebiasaan merokok, dan faktor jenis kelamin merupakan determinan paling berpengaruh dalam klasifikasi tingkat keparahan penyakit tiroid.

3.3 Evaluasi Model

Evaluasi model dilakukan menggunakan data uji (77 sampel, 20% dari total dataset) terhadap model Random Forest yang dilatih dengan seluruh 383 data, 100 pohon keputusan, dan kedalaman maksimum 5. Proses evaluasi menggunakan confusion matrix untuk mengukur performa klasifikasi pada tiga kelas sekaligus. Hasil confusion matrix dapat dilihat pada Tabel 2.

Tabel 2. Confusion Matrix Hasil Klasifikasi

Kelas Sebenarnya \ Prediksi	Low	Medium	High
Low	53	2	0
Medium	8	7	1
High	2	1	3

Dari confusion matrix pada Tabel 2 terlihat bahwa model paling baik mengklasifikasikan kelas Low (53 benar dari 55), namun masih mengalami kesalahan pada kelas Medium dan High. Hasil perhitungan selengkapnya ditampilkan pada Tabel 3.

Tabel 3. Hasil Evaluasi Metrik Klasifikasi

Kelas / Metrik	Precision	Recall	F1-Score
Low (0)	0.84	0.96	0.90
Medium (1)	0.70	0.44	0.54
High (2)	0.75	0.50	0.60
Accuracy	-	-	0.82
Macro Avg	0.76	0.63	0.68

Berdasarkan Tabel 3, model Random Forest menghasilkan akurasi sebesar 82%, dengan macro average precision 76%, recall 63%, dan F1-score 68%. Analisis per kelas menunjukkan perbedaan performa yang cukup mencolok. Kelas Low memperoleh hasil terbaik dengan precision 0,84, recall 0,96, dan F1-score 0,90. Tingginya nilai recall pada kelas ini (0,96) menunjukkan bahwa model mampu mendeteksi 53 dari 55 sampel kelas Low dengan benar, yang berarti hanya 2 sampel yang salah diklasifikasikan sebagai Medium. Kinerja unggul pada kelas Low ini secara langsung berkontribusi pada tingginya akurasi keseluruhan model.

Sebaliknya, kelas Medium dan High menunjukkan performa yang lebih rendah. Kelas Medium memperoleh precision 0,70, recall 0,44, dan F1-score 0,54. Rendahnya recall (0,44)

mengindikasikan bahwa model hanya mampu mengidentifikasi 7 dari 16 sampel Medium yang sebenarnya dengan benar, sementara 8 sampel Medium salah diklasifikasikan sebagai Low dan 1 sampel sebagai High. Kelas High mendapatkan precision 0,75, recall 0,50, dan F1-score 0,60. Dari 6 sampel High yang ada di data uji, hanya 3 yang berhasil terdeteksi dengan benar, sedangkan 2 sampel salah terklasifikasi sebagai Low dan 1 sebagai Medium. Rendahnya performa pada kelas Medium dan High ini secara utama disebabkan oleh ketidakseimbangan distribusi data (class imbalance) pada dataset, di mana model cenderung bias terhadap kelas mayoritas (Low).

Perlu dicatat bahwa dalam konteks medis, nilai recall pada kelas High menjadi sangat kritis, karena kegagalan mendeteksi pasien dengan risiko tinggi (false negative) dapat berdampak serius bagi keselamatan pasien. Nilai recall 0,50 pada kelas High berarti setengah dari pasien berisiko tinggi tidak teridentifikasi oleh model, yang merupakan kelemahan yang perlu dikembangkan lebih lanjut pada penelitian selanjutnya, misalnya melalui pengaplikasian metode SMOTE (Synthetic Minority Over-sampling Technique)atau pengaturan class weight pada algoritma.

Hasil ini sejalan dengan penelitian Hidayat et al. (2023) yang mencapai akurasi 92% pada klasifikasi penyakit jantung menggunakan Random Forest, serta penelitian Fadli & Saputra (2023) yang memperoleh akurasi 93,6% pada prediksi stroke. Perbedaan akurasi yang diperoleh dalam penelitian ini (82%) dibandingkan studi-studi tersebut dapat disebabkan oleh beberapa faktor: (1) kompleksitas klasifikasi multi-kelas (tiga kelas) yang secara inheren lebih sulit daripada klasifikasi biner; (2) ketidakseimbangan distribusi kelas yang lebih ekstrem pada dataset penyakit tiroid; serta (3) ukuran dataset yang relatif kecil (383 sampel) yang membatasi kemampuan generalisasi model. Meskipun demikian, akurasi 82% yang dicapai penelitian ini tetap menunjukkan potensi yang signifikan sebagai sistem pendukung keputusan dalam konteks skrining awal penyakit tiroid, khususnya untuk mengidentifikasi pasien dengan risiko rendah yang mendominasi populasi. Perbandingan hasil penelitian ini dengan studi-studi terkait menunjukkan bahwa Random Forest secara konsisten memberikan performa yang kompetitif pada berbagai domain klasifikasi penyakit medis, memperkuat posisinya sebagai salah satu algoritma pilihan dalam aplikasi machine learning di bidang kesehatan.

4. KESIMPULAN

Algoritma Random Forest berhasil diterapkan dalam penelitian ini untuk mengklasifikasikan tingkat keparahan penyakit tiroid ke dalam tiga kategori: Low, Medium, dan High. Model yang dibangun menggunakan 383 data pasien dari Kaggle dengan bahasa pemrograman Python memperoleh nilai akurasi sebesar 82%, presisi 76%, recall 63%, dan F1-score 68%. Fitur-fitur yang paling berpengaruh dalam proses klasifikasi adalah Adenopathy, Pathology, Physical Examination, Smoking, dan Focality. Metode ini berpotensi menjadi sistem pendukung keputusan yang efektif bagi tenaga medis dalam diagnosis dini penyakit tiroid.

Untuk penelitian selanjutnya, disarankan untuk mengoptimalkan parameter model (jumlah pohon dan kedalaman), menerapkan teknik penanganan data tidak seimbang seperti SMOTE, serta mengeksplorasi algoritma lain sebagai perbandingan. Pengembangan aplikasi berbasis web yang mengintegrasikan model ini dengan sistem informasi medis juga direkomendasikan.

REFERENSI

[1] E. Safitri, D. Rofianto, S. Karnila, and H. Kurniawan, “Prediksi Kekambuhan Kanker Tiroid Menggunakan Algoritma Random Fores t,” pp. 178–184, 2025.

[2] T. Tambunan, M. Yohanna, and A. P. Silalahi, “Penerapan Metode Random Forest Dalam Mendeteksi Berita Hoax,” *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 7, no. 2, pp. 301–306, 2023, doi: 10.46880/jmika.vol7no2.pp301-306.

[3] L. M. Barus, H. G. Simanullang, and A. P. Silalahi, “Classification of English-Language Racial Hate Speech Using Random Forest ”,” *J. Intell. Comput. Inf. Syst.*, vol. 1, no. 1, pp. 1–17, 2025.

[4] A. P. Silalahi and H. G. Simanullang, *Decission Tree dan Penerapannya dengan Algoritma Iterative Dichotomizes versi 3 (ID3)*. 2025.

[5] M. Mahdikhani, “Exploring commonly used terms from online reviews in the fashion field to predict review helpfulness,” *Int. J. Inf. Manag. Data Insights*, vol. 3, no. 1, 2023, doi: 10.1016/j.jjimei.2023.100172.

[6] Hidayat, H., Sunyoto, A., & Al Fatta, H. (2023). Klasifikasi Penyakit Jantung Menggunakan Random Forest Classifier. Jurnal SISKOM-KB, 7(1), 31–40.

[7] Fadli, M., & Saputra, R. A. (2023). Klasifikasi Dan Evaluasi Performa Model Random Forest Untuk Prediksi Stroke. JT: Jurnal Teknik, 12(02), 72–80.