Journal of Information Technology, Software Engineering and Computer Science
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume 4, Number 4, October 2026
DOI : **https://doi.org/10.58602/itsecs.v4i4.405**
Page 153-171

Deteksi Alzheimer Berbasis *Random Forest* dengan *Feature Selection* dan Analisis SHAP pada Data Klinis

Nirwana Hendrastuty[1*]✉, Marta Santra Wijaya[1]✉, Debby Alita[2]✉
[1,2]Sistem Informasi, Universitas Teknokrat Indonesia, Indonesia
[2]Informatika, Universitas Teknokrat Indonesia, Indonesia

Abstrak: Penyakit Alzheimer merupakan gangguan neurodegeneratif progresif yang menjadi penyebab utama demensia, namun deteksi dininya di Indonesia masih terkendala oleh mahalnya metode pencitraan otak seperti MRI dan PET Scan. *Machine learning* menawarkan alternatif deteksi dini berbasis data klinis rutin yang lebih terjangkau dan mudah diakses. Penelitian ini bertujuan mengembangkan model deteksi dini Alzheimer menggunakan algoritma *Random Forest* yang dioptimalkan melalui seleksi fitur dan disertai analisis interpretabilitas SHAP (*SHapley Additive exPlanations*) pada data klinis non-pencitraan. Dataset yang digunakan bersumber dari Kaggle, terdiri dari 2.149 data pasien dengan 32 fitur prediktor setelah pra-pemrosesan, mencakup aspek demografis, medis, kognitif, dan fungsional. Ketidakseimbangan kelas pada data latih (1.111 pasien sehat berbanding 608 pasien Alzheimer) ditangani menggunakan SMOTE. Seleksi fitur dilakukan melalui kombinasi *Random Forest Feature Importance* dan RFECV (*Recursive Feature Elimination with Cross-Validation*) dengan skema *5-fold Stratified Cross-Validation*, sementara optimasi *hyperparameter* dilakukan menggunakan *GridSearchCV* terhadap 72 kombinasi parameter. Hasil penelitian menunjukkan bahwa model *Random Forest* menghasilkan performa yang sangat baik dengan *accuracy* 94,19%, *precision* 94,41%, *recall* 88,82%, *F1-Score* 91,53%, dan *ROC-AUC* 0,938. Hasil RFECV secara objektif memvalidasi bahwa seluruh 32 fitur tetap relevan bagi model, berbeda dari pendekatan pemangkasan fitur sepihak pada penelitian acuan. Analisis SHAP mengungkap bahwa *Functional Assessment*, ADL, MMSE, *Memory Complaints*, dan *Behavioral Problems* merupakan lima fitur paling dominan, dengan pola hubungan berbentuk ambang batas pada *Functional Assessment* dan ADL, serta interpretasi transparan pada level individu pasien melalui *waterfall plot*. Penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan sistem pendukung keputusan medis untuk deteksi dini Alzheimer yang akurat sekaligus interpretatif.
Kata Kunci: Alzheimer; *Random Forest; Feature Selection*; RFECV; SHAP; *Machine Learning*;

Abstract: Alzheimer's disease is a progressive neurodegenerative disorder and the leading cause of dementia, yet its early detection in Indonesia remains constrained by the high cost of brain imaging methods such as MRI and PET Scan. Machine learning offers an affordable and accessible alternative for early detection based on routine clinical data. This study aims to develop an early Alzheimer's detection model using a *Random Forest* algorithm optimized through feature selection and equipped with SHAP (SHapley Additive exPlanations) interpretability analysis on non-imaging clinical data. The dataset was obtained from Kaggle, consisting of 2,149 patient records with 32 predictor features after preprocessing,

Nirwana Hendrastuty: *Penulis Korespondensi

Journal of Information Technology, Software Engineering and Computer Science
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume 4, Number 4, October 2026
DOI : **https://doi.org/10.58602/itsecs.v4i4.405**
Page 153-171

covering demographic, medical, cognitive, and functional aspects. Class imbalance in the training data (1,111 healthy patients versus 608 Alzheimer's patients) was addressed using SMOTE. Feature selection was performed by combining *Random Forest* Feature Importance with RFECV (Recursive Feature Elimination with Cross-Validation) using a 5-fold Stratified Cross-Validation scheme, while hyperparameter optimization was carried out using GridSearchCV across 72 parameter combinations. The results show that the *Random Forest* model achieved excellent performance with an accuracy of 94.19%, precision of 94.41%, recall of 88.82%, F1-Score of 91.53%, and ROC-AUC of 0.938. The RFECV results objectively validated that all 32 features remained relevant to the model, in contrast to the one-sided feature-pruning approach used in the reference study. SHAP analysis revealed that Functional Assessment, ADL, MMSE, Memory Complaints, and Behavioral Problems were the five most dominant features, with a threshold-like relationship pattern observed for Functional Assessment and ADL, as well as transparent patient-level interpretation through waterfall plots. This study is expected to contribute to the development of an accurate and interpretable medical decision-support system for the early detection of Alzheimer's disease.
Keywords: Alzheimer; Random Forest; Feature Selection; RFECV; SHAP; Machine Learning;

1. PENDAHULUAN

Penyakit Alzheimer merupakan gangguan neurodegeneratif progresif yang menjadi penyebab utama demensia, ditandai dengan penurunan fungsi kognitif, gangguan memori, serta hilangnya kemampuan menjalankan *activities of daily living* (ADL) secara mandiri. *World Health Organization* memproyeksikan jumlah penderita demensia di dunia akan meningkat tajam seiring pertumbuhan populasi lanjut usia global dalam beberapa dekade mendatang [1]. Di Indonesia sendiri, Kementerian Kesehatan RI mencatat sekitar 1,2 juta jiwa mengalami demensia dengan Alzheimer sebagai penyebab utama pada 60–70% kasus [2], dan proyeksi berbasis model regional memperkirakan angka ini meningkat menjadi sekitar 1,9 juta jiwa pada tahun 2030 dan 3,9 juta jiwa pada tahun 2050 [24]. Kondisi ini menjadikan deteksi dini Alzheimer sebagai isu kesehatan yang mendesak untuk terus dikembangkan solusinya.

Deteksi dini Alzheimer di Indonesia masih terkendala oleh keterbatasan metode diagnosis konvensional. Pemeriksaan *Mini-Mental State Examination* (MMSE) dan evaluasi klinis oleh tenaga medis spesialis bersifat subjektif dan memakan waktu, sementara metode pencitraan seperti *Magnetic Resonance Imaging* (MRI) atau *Positron Emission Tomography* (PET Scan) memerlukan biaya tinggi dan infrastruktur yang tidak tersedia merata, terutama di fasilitas kesehatan tingkat pertama di wilayah pedesaan. Akibatnya, banyak pasien lanjut usia baru terdiagnosis pada tahap lanjut ketika intervensi medis sudah kurang efektif. Oleh karena itu, dibutuhkan pendekatan deteksi dini yang akurat, efisien, dan dapat diakses secara luas berbasis data klinis rutin tanpa bergantung pada modalitas pencitraan otak yang mahal.

Machine learning menawarkan solusi potensial atas permasalahan tersebut, dengan algoritma *Random Forest* yang banyak dipilih karena sifatnya yang robust terhadap noise, mampu menangani data berdimensi tinggi, serta memiliki mekanisme ensemble yang menurunkan risiko overfitting. Beberapa penelitian terdahulu dalam lima tahun terakhir menunjukkan perkembangan pada topik ini. Yuni, Oktavianingrum, dan Aksa [3] menerapkan *Random Forest* dengan penyeimbangan kelas SMOTE dan seleksi fitur SelectKBest pada dataset klinis Alzheimer dari Kaggle berisi 2.149 data pasien dengan 35 fitur, memperoleh akurasi 94% dengan Functional Assessment, ADL, dan MMSE sebagai

Nirwana Hendrastuty: *Penulis Korespondensi

Journal of Information Technology, Software Engineering and Computer Science
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume 4, Number 4, October 2026
DOI : **https://doi.org/10.58602/itsecs.v4i4.405**
Page 153-171

fitur paling dominan, namun menggunakan metode seleksi fitur berbasis filter yang statis tanpa penjelasan kontribusi fitur pada level individu pasien. Rohini dan Surendran [4] mengembangkan model weighted hybrid *Random Forest* berbasis data pencitraan MRI ADNI dengan akurasi sekitar 93%, tetapi bergantung pada infrastruktur pencitraan yang mahal. Fadhila, Ulinnuha, dan Yuliati [5] menerapkan *Random Forest* dengan seleksi fitur Information Gain dan *hyperparameter tuning* namun belum menyertakan analisis interpretabilitas model pasca-pelatihan. Akbar dan Rahmaddeni [6] membandingkan beberapa algoritma *machine learning* untuk memprediksi Alzheimer namun berfokus pada perbandingan akurasi antar-algoritma tanpa mengkaji signifikansi fitur, sedangkan Sari dan Fatah [7] mengklasifikasikan Alzheimer menggunakan *Decision Tree* dengan performa lebih rendah dan tanpa penanganan ketidakseimbangan kelas.

Terkait aspek seleksi fitur, tinjauan literatur sistematis oleh Priyatno dan Widiyaningtyas [8] menemukan bahwa metode RF-RFE merupakan pendekatan embedded yang paling banyak digunakan setelah SVM-RFE di bidang kesehatan, namun integrasi *Recursive Feature Elimination with Cross-Validation* (RFECV) dengan interpretasi model *post-hoc* masih jarang dieksplorasi bersamaan. Hal ini didukung oleh Awad dan Fraihat [9] yang membuktikan RFECV mampu menghasilkan subset fitur lebih optimal dibandingkan seleksi fitur statis. Di sisi interpretabilitas, Ghasemi dkk. [10] mengonfirmasi bahwa SHAP merupakan teknik *Explainable Artificial Intelligence* (XAI) model-agnostic yang paling banyak digunakan dalam riset medis, dan Hidayat, Nugroho, dan Suprianto [11] telah membuktikan efektivitas kombinasi *Random Forest* dan SHAP untuk interpretasi risiko stroke, namun pendekatan serupa belum banyak diterapkan pada kasus Alzheimer berbasis data klinis non-pencitraan. Gholampour [12] turut menunjukkan bahwa validitas *feature importance* pasca-SMOTE [13] perlu divalidasi secara cermat. Berdasarkan uraian tersebut, teridentifikasi tiga kesenjangan penelitian utama. Pertama penelitian klasifikasi Alzheimer non-pencitraan yang ada [3], [5]–[7] masih mengandalkan seleksi fitur berbasis filter statis dan belum mengoptimalkan pemilihan fitur secara wrapper. Kedua, pendekatan berbasis pencitraan otak [4] memberikan akurasi tinggi namun sulit diskalakan karena keterbatasan biaya dan infrastruktur dan ketiga integrasi seleksi fitur dengan interpretasi SHAP yang terbukti efektif pada domain medis lain [10], [11], [14] masih minim dieksplorasi khusus pada data klinis Alzheimer.

Untuk menjawab kesenjangan tersebut, penelitian ini mengusulkan model deteksi dini Alzheimer berbasis *Random Forest* yang dioptimalkan melalui kombinasi *Random Forest* Feature Importance dan RFECV dengan skema 5-fold Stratified *Cross-Validation*, dilengkapi optimasi *hyperparameter* menggunakan *GridSearchCV*, serta dilengkapi analisis interpretabilitas SHAP untuk menjelaskan kontribusi setiap fitur klinis terhadap hasil prediksi baik secara global maupun pada level individu pasien. Kebaruan penelitian ini terletak pada penggunaan pendekatan seleksi fitur *wrapper* (RFECV) yang divalidasi silang, alih-alih metode filter statis seperti pada penelitian acuan [3], serta integrasi langsung dengan interpretasi SHAP yang belum banyak diterapkan pada data klinis Alzheimer non-pencitraan.

Penelitian ini bertujuan mengembangkan model deteksi dini Alzheimer berbasis *Random Forest* yang dioptimalkan melalui tahapan feature selection pada data klinis non-pencitraan yang mudah diakses, serta menyertakan analisis SHAP untuk memberikan interpretasi transparan atas kontribusi setiap fitur klinis terhadap hasil prediksi. Kontribusi penelitian ini meliputi: (1) evaluasi performa klasifikasi *Random Forest* dengan seleksi fitur pada dataset klinis Alzheimer; (2) identifikasi fitur klinis paling relevan dalam mendeteksi risiko Alzheimer; dan (3) penyediaan interpretasi model berbasis SHAP sebagai dasar pendukung keputusan medis yang dapat dipercaya oleh tenaga klinis dalam proses skrining dini. Tahapan penelitian secara rinci dijelaskan pada bagian Metode Penelitian berikut.

Nirwana Hendrastuty: *Penulis Korespondensi

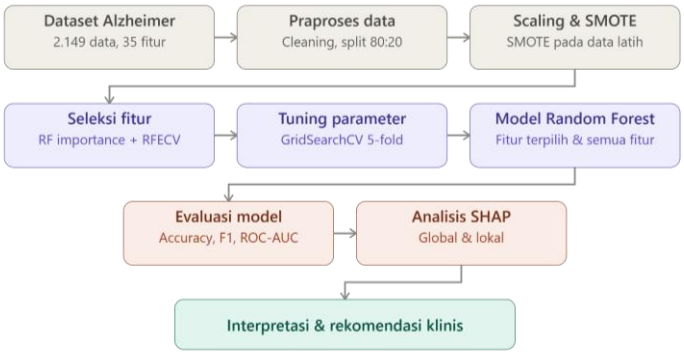
Journal of Information Technology, Software Engineering and Computer Science
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume 4, Number 4, October 2026
DOI : https://doi.org/10.58602/itsecs.v4i4.405
Page 153-171

2. METODE PENELITIAN

Pada bagian ini dijelaskan kontribusi metodologis penelitian dalam mengembangkan model deteksi dini Alzheimer, mencakup tahapan penelitian, dataset yang digunakan, pra-pemrosesan data, algoritma *Random Forest*, seleksi fitur, *optimasi hyperparameter*, skema evaluasi, hingga analisis interpretabilitas SHAP.

2.1. Tahapan Penelitian

Penelitian ini dilaksanakan secara sistematis melalui beberapa tahapan utama yang saling berkaitan, meliputi pengumpulan data, pra-pemrosesan, penanganan ketidakseimbangan kelas, seleksi fitur, pemodelan *Random Forest*, optimasi *hyperparameter* menggunakan *GridSearchCV*, evaluasi performa model, serta interpretasi model menggunakan analisis SHAP. Seluruh proses komputasi dijalankan menggunakan bahasa pemrograman *Python* dengan memanfaatkan *library scikit-learn*, *imbalanced-learn*, dan shap pada platform *Google Colaboratory*. Gambar 1 menyajikan tahapan penelitian secara keseluruhan.



Gambar 1. Tahapan Penelitian

Gambar 1 menunjukkan bahwa proses penelitian berjalan secara berurutan, dimulai dari pengumpulan dataset, dilanjutkan dengan pembersihan data dan *exploratory data analysis* (EDA), pembagian data latih-uji secara *stratified*, standardisasi fitur dan penyeimbangan kelas melalui SMOTE, seleksi fitur menggunakan kombinasi *Random Forest Importance* dan RFECV, optimasi *hyperparameter* melalui *GridSearchCV*, pelatihan model *Random Forest* final, evaluasi performa menggunakan lima metrik, hingga interpretasi model menggunakan SHAP. Alur bertahap ini memastikan setiap komponen metodologi diterapkan secara konsisten dan dapat divalidasi ulang.

2.2. Dataset Penelitian

Dataset yang digunakan dalam penelitian ini adalah *Alzheimer's Disease* Dataset, sebuah dataset publik yang diperoleh dari platform Kaggle, identik dengan yang digunakan pada penelitian acuan [3]. Dataset ini memuat 2.149 catatan data pasien yang mencakup atribut demografis, riwayat medis, pengukuran klinis, serta penilaian kognitif dan fungsional, terdiri dari 35 kolom: identitas pasien (*PatientID*), 32 fitur prediktor di antaranya usia (*Age*), jenis kelamin (*Gender*), etnisitas (*Ethnicity*), tingkat pendidikan (*EducationLevel*), indeks massa tubuh (BMI), riwayat keluarga Alzheimer (*FamilyHistoryAlzheimers*), skor kognitif (MMSE), penilaian fungsional (*FunctionalAssessment*), aktivitas hidup sehari-hari (ADL), keluhan memori (*MemoryComplaints*), dan masalah perilaku (*BehavioralProblems*) variabel target Diagnosis (0 = tidak Alzheimer, 1 = Alzheimer), serta kolom *DoctorInCharge* yang bernilai

Nirwana Hendrastuty: *Penulis Korespondensi

Journal of Information Technology, Software Engineering and Computer Science
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume 4, Number 4, October 2026
DOI : **https://doi.org/10.58602/itsecs.v4i4.405**
Page 153-171

konstan. Distribusi awal dataset menunjukkan ketidakseimbangan kelas, dengan proporsi 64,6% untuk kelas tidak Alzheimer dan 35,4% untuk kelas Alzheimer.

Secara spesifik, dataset ini dipublikasikan oleh Rabie El Kharoua di platform Kaggle pada tahun 2024 dengan tautan resmi https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset/data. Berdasarkan metadata resmi pada platform tersebut, dataset ini dikategorikan sebagai dataset sintetis (*synthetic*), artinya label Diagnosis dibentuk secara terprogram untuk mensimulasikan pola hubungan klinis yang realistis antara atribut demografis, riwayat medis, serta penilaian kognitif dan fungsional terhadap status Alzheimer, dan bukan berasal dari catatan rekam medis pasien di lapangan secara langsung. Karakteristik ini perlu menjadi catatan transparansi bagi pembaca dalam menginterpretasikan validitas eksternal model, meskipun pola hubungan antar-fitur yang dihasilkan tetap relevan untuk tujuan pengembangan dan evaluasi metodologis model machine learning.

Pemilihan dataset klinis non-pencitraan ini memberikan keuntungan tersendiri dibandingkan dataset berbasis pencitraan otak (MRI/PET) yang mahal dan sulit diakses secara luas [4], karena atribut yang digunakan bersifat rutin dan mudah diperoleh pada pemeriksaan kesehatan primer. Kombinasi atribut demografis, medis, dan kognitif fungsional menjadi dasar pertimbangan dalam menerapkan seleksi fitur, mengingat tidak seluruh atribut secara klinis memberikan kontribusi setara terhadap diagnosis Alzheimer.

Tabel 1. Contoh Data *Alzheimer's Disease*

Age	Gender	BMI	MMSE	Functional Assessment	ADL	Memory Complaints	Behavioral Problems	Diagnosis
73	0	22,93	21,46	6,52	1,73	0	0	0
89	0	26,83	20,61	7,12	2,59	0	0	0
73	0	17,80	7,36	5,90	7,12	0	0	0
74	1	33,80	13,99	8,97	6,48	0	1	0
89	0	20,72	13,52	6,05	0,01	0	0	0
86	1	30,63	27,52	5,51	9,02	0	0	0
75	0	18,78	10,14	3,40	4,52	0	0	1
78	1	28,87	7,85	4,51	1,94	1	0	1

Tabel 1 menyajikan contoh data yang merepresentasikan variasi karakteristik pasien, mulai dari kombinasi usia dan skor kognitif (MMSE) hingga variasi skor fungsional (Functional Assessment, ADL) yang berkaitan dengan status Diagnosis. Terlihat bahwa dua pasien dengan skor Functional Assessment rendah (3,40 dan 4,51) serta ADL rendah (4,52 dan 1,94) terdiagnosis Alzheimer, sejalan dengan temuan penelitian acuan bahwa kedua fitur tersebut merupakan prediktor dominan [3].

2.3. Pra-Pemrosesan Data

Tahap Tahap pra-pemrosesan data dilakukan untuk memastikan kualitas dan konsistensi data sebelum memasuki proses pemodelan. Terdapat lima langkah yang dilaksanakan secara berurutan.

a. Pemeriksaan kualitas data. Pemeriksaan dilakukan terhadap nilai hilang (*missing value*) dan data duplikat, hasilnya tidak ditemukan satu pun nilai hilang maupun baris duplikat pada dataset.
b. Penghapusan kolom tidak informatif. Kolom *PatientID* (identitas unik) dan *DoctorInCharge* (bernilai konstan "*XXXConfid*") dibuang karena tidak memberikan informasi prediktif, menyisakan 32 fitur.
c. Pembagian data latih dan uji. Data dibagi dengan rasio 80:20 secara *stratified* untuk menjaga proporsi kelas, menghasilkan 1.719 data latih dan 430 data uji.

Nirwana Hendrastuty: *Penulis Korespondensi

Journal of Information Technology, Software Engineering and Computer Science
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume 4, Number 4, October 2026
DOI : **https://doi.org/10.58602/itsecs.v4i4.405**
Page 153-171

d. Standardisasi fitur numerik. Seluruh fitur numerik distandardisasi menggunakan *StandardScaler* agar berada pada skala yang seragam, sekaligus mempermudah interpretasi nilai SHAP pada tahap akhir.
e. Penanganan ketidakseimbangan kelas. Mengingat distribusi kelas pada data latih tidak seimbang (1.111 pasien tidak Alzheimer berbanding 608 pasien Alzheimer), teknik SMOTE [13] diterapkan hanya pada data latih untuk membangkitkan sampel sintetis pada kelas minoritas hingga distribusi menjadi seimbang (1.111:1.111), sekaligus mencegah kebocoran informasi (*data leakage*) ke data uji.

Kelima tahapan pra-pemrosesan tersebut dilakukan secara berurutan agar kualitas data yang masuk ke tahap pemodelan benar-benar terjaga. Standardisasi menjaga agar fitur dengan rentang nilai besar seperti kolesterol tidak mendominasi proses pembelajaran, sementara SMOTE memastikan model tidak bias terhadap kelas mayoritas. Penelitian klasifikasi kesehatan lain di Indonesia yang menerapkan SMOTE sebelum penelitian turut menegaskan bahwa langkah ini memberikan kontribusi nyata terhadap peningkatan sensitivitas model dalam mengenali kelas minoritas [13], meskipun validitas klinisnya tetap perlu dicermati melalui interpretasi model pasca-pelatihan [12].

2.4. Random Forest

Random Forest adalah algoritma ensemble yang membangun T pohon keputusan secara paralel menggunakan teknik *bootstrap* sampling dan pemilihan fitur secara acak pada setiap pemisahan node [15]. Prediksi akhir ditentukan melalui mekanisme *majority voting*, dirumuskan pada (1).

$$\hat{y} = \arg\max_{k} \sum_{t=1}^{T} I\left(h_t(x) = k\right) \qquad (1)$$

Keterangan: T adalah jumlah pohon (T=200, hasil *GridSearchCV*), h_t(x) adalah prediksi pohon ke-t, $k \in \{0,1\}$ adalah label kelas (tidak Alzheimer, Alzheimer), dan $I(\cdot)$ adalah fungsi indikator bernilai 1 apabila prediksi pohon ke-t sama dengan kelas k, dan 0 jika sebaliknya. Mekanisme *bootstrap* sampling memungkinkan setiap pohon dilatih pada subset data acak dengan pengembalian, sehingga korelasi antar-pohon relatif rendah dan model tahan terhadap overfitting [15].

2.5. Seleksi Fitur

Dua pendekatan dikombinasikan agar hasil seleksi fitur kuat secara metodologis [14]. Pertama, peringkat awal fitur dihitung menggunakan *Gini Importance* (*Mean Decrease in Impurity*) sebagaimana Persamaan (2):

$$Imp(x_j) = \sum_{t=1}^{T} \sum_{n \in N_t(x_j)} p\,(n)\,\Delta\text{Gini}(n) \qquad (2)$$

Keterangan: $N_t(x_j)$ = himpunan *node* pada pohon ke-t yang menggunakan fitur x_j sebagai kriteria pemisah, $p(n)$ = proporsi sampel yang mencapai *node* n, dan $\Delta Gini(n)$ = penurunan impuritas Gini akibat pemisahan pada *node* tersebut. Kedua, RFECV (*Recursive Feature Elimination with Cross-Validation*) [7] diterapkan dengan estimator *Random Forest* (*n_estimators*=200) dan skema *5-fold Stratified Cross-Validation*, mengeliminasi fitur dengan kontribusi terlemah secara bertahap dan memilih jumlah fitur yang memaksimalkan skor rata-rata F1 pada validasi silang, sehingga jumlah fitur optimal ditentukan secara otomatis, bukan berdasarkan ambang batas sepihak. Pendekatan *wrapper* ini sejalan dengan tinjauan sistematis yang menemukan bahwa kombinasi

Nirwana Hendrastuty: *Penulis Korespondensi

Journal of Information Technology, Software Engineering and Computer Science
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume 4, Number 4, October 2026
DOI : **https://doi.org/10.58602/itsecs.v4i4.405**
Page 153-171

Random Forest dan RFE (RF-RFE) merupakan metode paling banyak diterapkan pada domain kesehatan [6].

2.6. Optimasi Hyperparameter dengan GridSearchCV

GridSearchCV merupakan metode pencarian *hyperparameter* yang bekerja dengan mengevaluasi secara menyeluruh setiap kombinasi nilai parameter dalam sebuah *grid* pencarian [14], [15], dirumuskan pada Persamaan (3):

$$\theta^* = \arg\max_{\theta \in \Theta} \frac{1}{K} \sum_{k=1}^{K} F\,1(RF_\theta;\, D_{val}^{(k)}) \qquad (3)$$

Keterangan: θ = kombinasi *hyperparameter* dalam ruang pencarian Θ, K=5 (jumlah *fold*), dan *F1(RF_θ; D_val^(k))* = skor F1 model dengan parameter θ pada subset validasi *fold* ke-*k*. Skor F1 dipilih sebagai target optimasi karena lebih representatif dibanding *accuracy* untuk data yang tidak seimbang. Tabel 2 menyajikan parameter beserta rentang nilai yang diuji.

Tabel 2. Rentang *Hyperparameter* yang Diuji pada *GridSearchCV*

Hyperparameter	Rentang Nilai yang Diuji
n_estimators	200, 300, 500
max_depth	None, 10, 20
min_samples_split	2, 5
min_samples_leaf	1, 2
max_features	sqrt, log2

Rentang nilai pada Tabel 2 menghasilkan 72 kombinasi (3×3×2×2×2), diuji melalui 360 proses pelatihan (72 kombinasi × 5 *fold*) menggunakan *5-fold Stratified Cross-Validation*, memastikan proses *GridSearchCV* mengeksplorasi ruang parameter secara memadai sebelum konfigurasi terbaik ditentukan.

2.7. Evaluasi Model

Performa model dievaluasi pada data uji yang belum pernah dilihat model menggunakan *Confusion Matrix* serta lima metrik kinerja: (1) *accuracy* mengukur proporsi prediksi benar dari keseluruhan data uji; (2) *precision* menggambarkan ketepatan prediksi kelas positif (Alzheimer); (3) *recall* mengukur kemampuan model mendeteksi seluruh sampel Alzheimer; (4) *F1-score* menyeimbangkan *precision* dan *recall* dalam satu nilai harmonik; dan (5) *ROC-AUC* mengukur kemampuan model membedakan kedua kelas di berbagai ambang keputusan, yang relevan untuk data tidak seimbang karena tidak terpengaruh oleh distribusi proporsi kelas [10]. Sebagai pembanding, model *Random Forest* dengan parameter terbaik yang sama juga dilatih menggunakan seluruh fitur (tanpa seleksi RFECV) untuk membuktikan secara kuantitatif kontribusi tahap seleksi fitur terhadap efisiensi dan performa model.

2.8. Analisis SHAP (*SHapley Additive exPlanations*)

Untuk mengatasi keterbatasan interpretabilitas model *black-box* pada penelitian acuan sebelumnya [3], penelitian ini menyertakan analisis SHAP menggunakan *TreeExplainer* [14],[15]. Nilai kontribusi setiap fitur terhadap prediksi dihitung berdasarkan nilai *Shapley* dari teori permainan kooperatif, dirumuskan pada Persamaan (4):

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(|F|-|S|-1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \qquad (4)$$

Nirwana Hendrastuty: *Penulis Korespondensi

Journal of Information Technology, Software Engineering and Computer Science
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume 4, Number 4, October 2026
DOI : **https://doi.org/10.58602/itsecs.v4i4.405**
Page 153-171

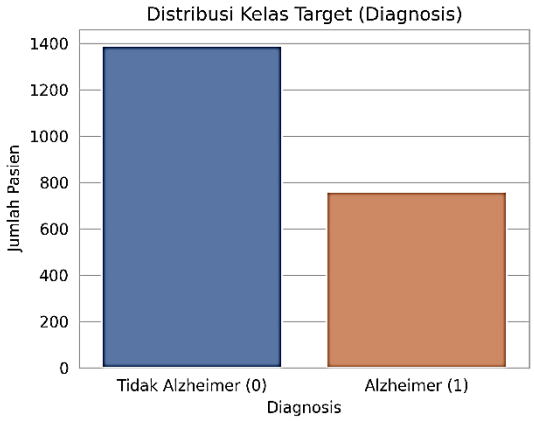
Keterangan: φ_i = nilai SHAP untuk fitur ke-*i*, *F* = himpunan seluruh fitur, *S* = *subset* fitur yang tidak memuat fitur *i*, dan *f(S)* = prediksi model menggunakan subset fitur *S*. Analisis mencakup tiga tingkat interpretasi pertama interpretasi global melalui *bar plot* dan *beeswarm summary plot* untuk mengetahui fitur paling berpengaruh secara keseluruhan, kedua *dependence plot* pada tiga fitur teratas untuk melihat arah pengaruh nilai fitur terhadap prediksi, dan ketiga interpretasi lokal melalui *waterfall plot* yang menjelaskan kontribusi setiap fitur terhadap prediksi satu pasien tertentu, aspek penting untuk mendukung kepercayaan tenaga klinis terhadap rekomendasi model secara individual[8], [9], bukan hanya secara *agregat*.

3. HASIL DAN PEMBAHASAN

Dataset *Alzheimer's Disease* yang digunakan pada penelitian ini terdiri dari 2.149 data awal dengan 35 kolom, menyusut menjadi 32 fitur prediktor setelah kolom *PatientID* dan *DoctorInCharge* dibuang. Pemeriksaan kualitas data tidak menemukan nilai hilang maupun baris duplikat. Bagian ini memaparkan hasil eksplorasi data, seleksi fitur, optimasi *hyperparameter*, evaluasi model pada data uji, analisis interpretabilitas SHAP, serta perbandingan kuantitatif dengan penelitian acuan [3] dan studi terkait lainnya.

3.1. Analisis Eksplorasi Data (EDA)

Distribusi kelas target menunjukkan ketidakseimbangan, dengan 1.389 pasien tidak Alzheimer (64,6%) dan 760 pasien Alzheimer (35,4%) (Gambar 2), menjadi dasar penerapan SMOTE pada tahap pra-pemrosesan.

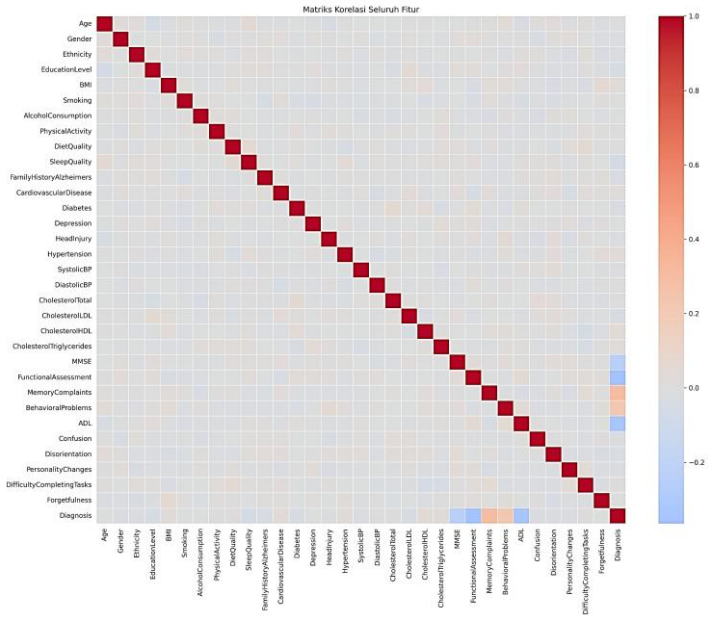


Gambar 2. Distribusi Kelas Target (Diagnosis)

Analisis korelasi Pearson antar seluruh 32 fitur terhadap satu sama lain (Gambar 3) menunjukkan bahwa mayoritas fitur klinis tidak saling berkorelasi kuat (nilai mendekati nol), kecuali kelompok fitur kognitif-fungsional (MMSE, *FunctionalAssessment*, ADL) yang menunjukkan korelasi negatif moderat dengan Diagnosis, serta *MemoryComplaints* dan *BehavioralProblems* yang berkorelasi positif dengan Diagnosis.

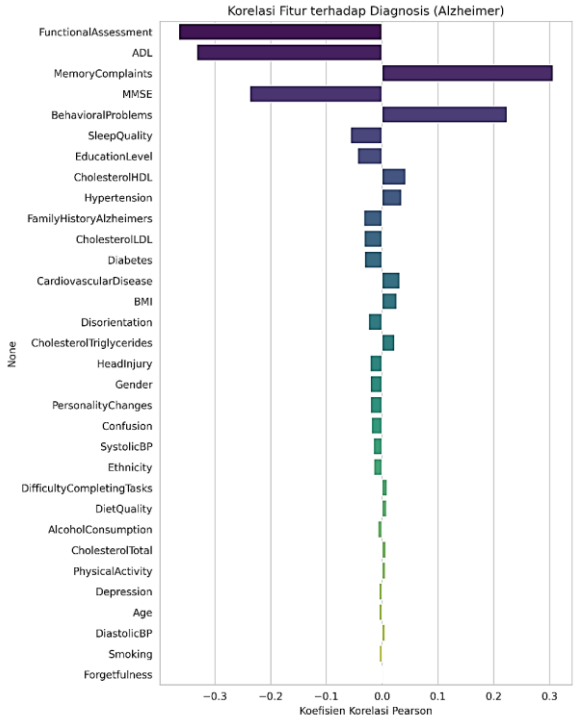
Nirwana Hendrastuty: *Penulis Korespondensi

Journal of Information Technology, Software Engineering and Computer Science
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume 4, Number 4, October 2026
DOI : **https://doi.org/10.58602/itsecs.v4i4.405**
Page 153-171



Gambar 3. Matriks Korelasi Seluruh Fitur

Pola korelasi ini penting secara metodologis: nilai korelasi antar-fitur prediktor yang secara umum rendah (multikolinearitas minimal) mendukung validitas interpretasi *Gini Importance* dan SHAP pada tahap selanjutnya, karena kontribusi setiap fitur terhadap prediksi tidak saling "tumpang tindih" akibat redundansi informasi antar-fitur.

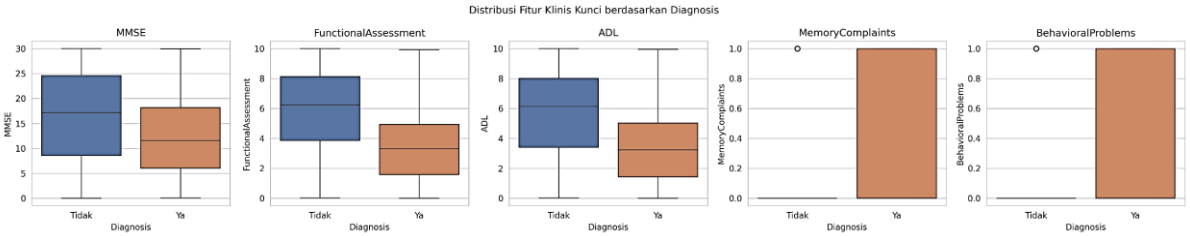


Gambar 4. Korelasi Fitur terhadap Diagnosis (Alzheimer)

Nirwana Hendrastuty: *Penulis Korespondensi

Journal of Information Technology, Software Engineering and Computer Science
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume 4, Number 4, October 2026
DOI : **https://doi.org/10.58602/itsecs.v4i4.405**
Page 153-171

Gambar 4 memperkuat hasil Gambar 3: *FunctionalAssessment* menunjukkan korelasi negatif terkuat terhadap Diagnosis (sekitar −0,36), diikuti ADL (sekitar −0,33) dan MMSE (sekitar −0,24) — mengonfirmasi bahwa penurunan ketiga skor tersebut berasosiasi dengan meningkatnya risiko Alzheimer. Sebaliknya, *MemoryComplaints* (sekitar +0,31) dan *BehavioralProblems* (sekitar +0,22) menunjukkan korelasi positif terkuat, mengindikasikan bahwa kehadiran keluhan memori dan masalah perilaku berasosiasi dengan diagnosis positif. Seluruh fitur klinis dan gaya hidup lainnya (kolesterol, tekanan darah, BMI, usia, dsb.) menunjukkan korelasi sangat lemah (|r| < 0,05) terhadap target konsisten dengan hierarki *Gini Importance* yang akan ditunjukkan pada Gambar 6.

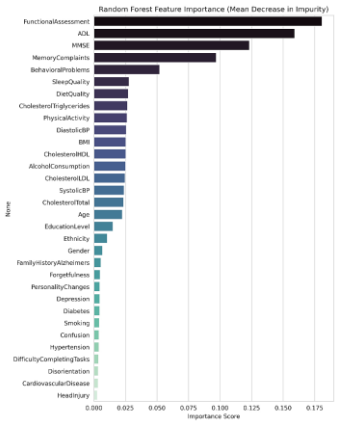


Gambar 5. Distribusi Fitur Klinis Kunci berdasarkan Diagnosis

Gambar 5 memberikan gambaran visual yang sangat jelas. *Boxplot* MMSE, *FunctionalAssessment*, dan ADL menunjukkan median yang jauh lebih rendah pada kelompok Alzheimer ("Ya") dibandingkan kelompok sehat ("Tidak"), dengan rentang interkuartil yang hampir tidak tumpang tindih untuk ADL dan *FunctionalAssessment* mengonfirmasi kekuatan diskriminatif kedua fitur ini secara visual. Yang lebih mencolok, *MemoryComplaints* dan *BehavioralProblems* menunjukkan pola hampir *biner* terpisah sempurna: hampir seluruh pasien pada kelompok "Ya" bernilai 1 sementara kelompok "Tidak" bernilai 0, dengan hanya segelintir kasus pencilan (*outlier*) pada arah sebaliknya. Pola pemisahan yang sangat tajam ini menjelaskan mengapa kedua fitur tersebut memperoleh skor SHAP dan *Gini Importance* yang tinggi meskipun bertipe *biner*.

3.2. Hasil Seleksi Fitur

Peringkat awal fitur menggunakan *Random Forest Feature Importance* (Gambar 6) menunjukkan bahwa *FunctionalAssessment* (0,1809), ADL (0,1591), dan MMSE (0,1232) merupakan tiga fitur paling dominan, diikuti *MemoryComplaints* (0,0969) dan *BehavioralProblems* (0,0523) kelima fitur ini bersama-sama menyumbang lebih dari 62% dari total *importance*, sementara 27 fitur sisanya masing-masing menyumbang kurang dari 3%.



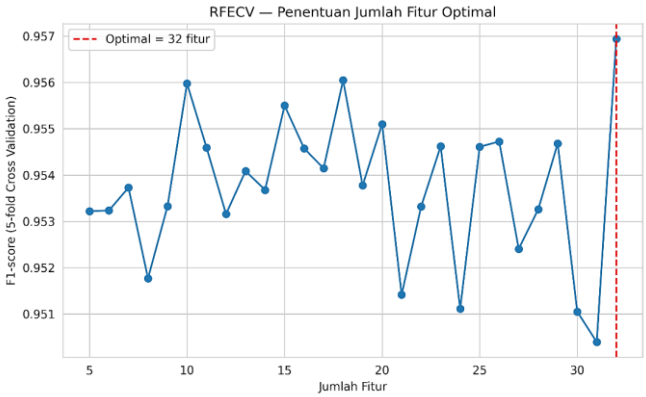
Gambar 6. *Random Forest Feature Importance (Mean Decrease in Impurity)*

Nirwana Hendrastuty: *Penulis Korespondensi

Journal of Information Technology, Software Engineering and Computer Science
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume 4, Number 4, October 2026
DOI : **https://doi.org/10.58602/itsecs.v4i4.405**
Page 153-171

Kelima fitur teratas seluruhnya berasal dari kategori penilaian kognitif dan fungsional, konsisten dengan temuan penelitian acuan [3] yang juga menempatkan *Functional Assessment*, ADL, dan MMSE sebagai fitur paling dominan, serta selaras dengan temuan Soladoye dkk. [16] dan Jahan dkk. [17] yang juga melaporkan skor kognitif-fungsional sebagai prediktor paling berpengaruh pada model *Random Forest* untuk Alzheimer berbasis data klinis multimodal.

Hasil RFECV (Gambar 7) menunjukkan temuan yang menarik: skor F1 validasi silang bervariasi pada kisaran 0,951–0,956 di sepanjang jumlah fitur 5 hingga 31 tanpa pola kenaikan yang monoton, dan baru mencapai puncaknya (0,9569) tepat pada 32 fitur (seluruh fitur dipertahankan).



Gambar 7. Kurva RFECV Penentuan Jumlah Fitur Optimal

Pola ini berbeda dari pendekatan penelitian acuan [3] yang memangkas fitur menjadi 10 secara sepihak menggunakan SelectKBest; hasil RFECV justru memvalidasi secara objektif melalui cross-validation sejalan dengan prinsip seleksi fitur berbasis validasi yang ditekankan oleh Guyon dan Elisseeff [18], bahwa seluruh 32 fitur termasuk atribut gaya hidup dan klinis yang tampak "lemah" pada Gambar 6 tetap relevan bagi model. Dengan demikian, RFECV pada penelitian ini berfungsi sebagai mekanisme validasi terhadap kelengkapan dan relevansi fitur, bukan sebagai metode reduksi dimensi.

Table 3. Parameter Optimal Hasil *GridSearchCV*

Hyperparameter	Nilai Optimal
n_estimators	200
max_depth	20
min_samples_split	2
min_samples_leaf	1
max_features	sqrt
F1-Score CV terbaik	0,9574

Kedalaman pohon sedang (20 level) terbukti cukup untuk menangkap pola data klinis tanpa risiko *overfitting*, sementara pemilihan fitur acak berbasis akar kuadrat (*sqrt*) pada setiap pemisahan *node* menjaga keragaman antar pohon sesuai prinsip *ensemble* Breiman [15], dan konsisten dengan praktik optimasi *hyperparameter* berbasis *GridSearchCV* pada studi klasifikasi kesehatan Indonesia lainnya [19], [20].

Perbedaan antara skor F1 hasil *cross-validation* (0,9574) dan F1 pada data uji (0,9153) sebesar 0,0421 perlu dicermati lebih lanjut. Skor F1 CV dihitung pada lipatan (*fold*) yang seluruhnya berasal dari data latih yang telah diseimbangkan melalui SMOTE (proporsi kelas 1.111:1.111), sedangkan data uji tetap mempertahankan distribusi kelas asli yang tidak seimbang (278:152, atau sekitar 65:35). Perbedaan distribusi kelas antara validasi lipat

Nirwana Hendrastuty: *Penulis Korespondensi

Journal of Information Technology, Software Engineering and Computer Science
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume 4, Number 4, October 2026
DOI : **https://doi.org/10.58602/itsecs.v4i4.405**
Page 153-171

dan data uji inilah yang menjadi penyebab utama gap tersebut, bukan indikasi overfitting yang serius, mengingat kedua skor tetap berada pada rentang tinggi (>0,91) dan konsisten dengan proporsi false negative yang teramati pada *Confusion Matrix* (Gambar 8). Dengan kata lain, performa model pada kondisi kelas seimbang selama *cross-validation* sedikit lebih optimistis dibandingkan performa pada kondisi dunia nyata yang tidak seimbang, sebuah temuan yang wajar dan penting untuk transparansi evaluasi model klasifikasi pada data medis tidak seimbang.

3.3. Perbandingan Performa: Semua Fitur vs Fitur Terpilih (RFECV)

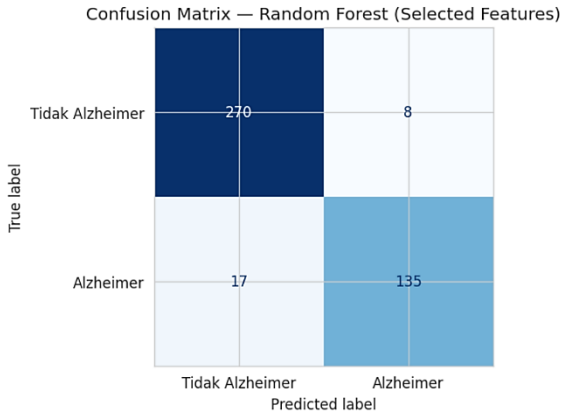
Table 4. Perbandingan Performa Semua Fitur vs Fitur Terpilih (RFECV)

Metrik	Semua Fitur (32)	Fitur Terpilih RFECV (32)
Accuracy	0,9419	0,9419
Precision	0,9441	0,9441
Recall	0,8882	0,8882
F1-Score	0,9153	0,9153
ROC-AUC	0,9380	0,9380
PR-AUC	0,9194	0,9194

Karena RFECV mempertahankan seluruh 32 fitur sebagai konfigurasi optimal, performa kedua kondisi identik secara numerik. Meskipun tidak terjadi reduksi dimensi, RFECV tetap berfungsi sebagai mekanisme validasi objektif yang membuktikan tidak ada subset fitur lebih kecil yang mampu menyamai performa model dengan fitur lengkap sebuah keunggulan metodologis dibandingkan pendekatan *SelectKBest* statis pada penelitian acuan [3]. Dengan kata lain, RFECV pada penelitian ini difungsikan sebagai konfirmasi relevansi fitur (*feature relevance validation*), bukan sebagai metode pengurangan jumlah fitur (*dimensionality reduction*).

3.4. Evaluasi Model dan Confusion Matrix

Model *Random Forest* final dievaluasi pada 430 data uji, menghasilkan *accuracy* 94,19%, *precision* 94,41%, *recall* 88,82%, *F1-score* 91,53%, *ROC-AUC* 0,938, dan *PR-AUC* 0,919.



Gambar 8. *Confusion Matrix* Model *Random Forest* pada Data Uji (n=430)

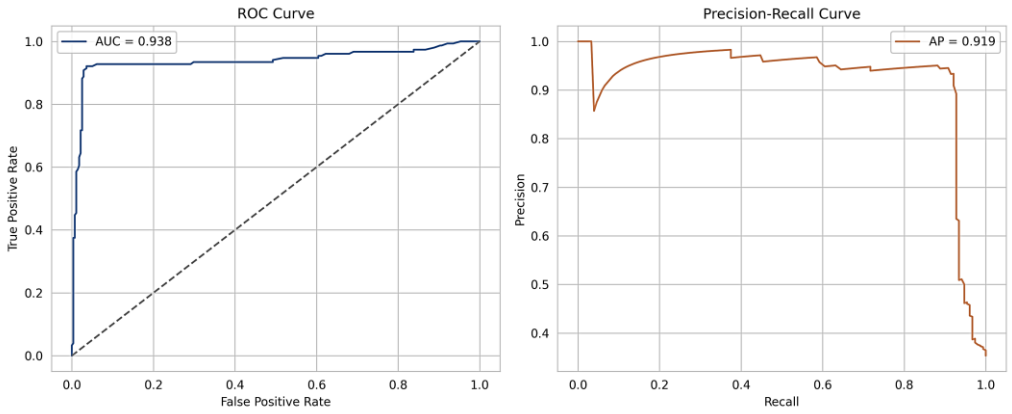
Dari 278 pasien tidak Alzheimer, model berhasil mengidentifikasi 270 dengan benar (97,1%) dan hanya salah mengklasifikasikan 8 pasien sebagai Alzheimer (*false positive*). Dari 152 pasien Alzheimer, model berhasil mendeteksi 135 dengan benar (88,8%) namun melewatkan 17 kasus (*false negative*, 11,2%). Jumlah *false negative* yang lebih tinggi dibandingkan *false positive* konsisten dengan nilai *recall* yang lebih rendah dari *precision*, dan menjadi catatan penting bagi pengembangan sistem skrining klinis: dalam konteks

Nirwana Hendrastuty: *Penulis Korespondensi

Journal of Information Technology, Software Engineering and Computer Science
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume 4, Number 4, October 2026
DOI : **https://doi.org/10.58602/itsecs.v4i4.405**
Page 153-171

deteksi dini penyakit, melewatkan kasus positif (*false negative*) umumnya menimbulkan konsekuensi lebih besar dibandingkan kesalahan positif palsu, sehingga ambang keputusan (*threshold*) 0,5 dapat dipertimbangkan untuk digeser guna menekan *false negative*.

3.5. Kurva ROC dan Precision-Recall

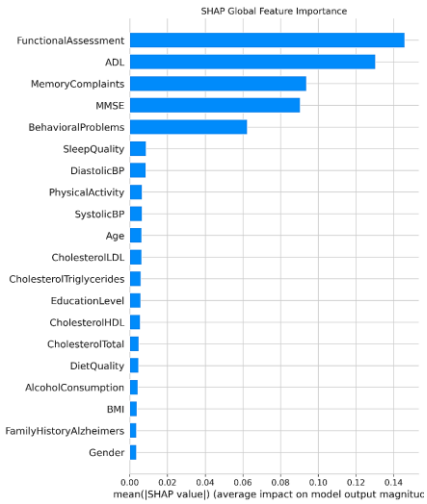


Gambar 9. Kurva ROC dan Precision-Recall Model Random Forest

Kurva ROC pada Gambar 9 naik tajam mendekati pojok kiri atas grafik dengan AUC 0,938, yang menurut standar interpretasi AUC [21] tergolong *excellent classifier* (AUC > 0,90). Namun, mengingat data uji ini tetap tidak seimbang (278:152), kurva *Precision-Recall* di sisi kanan Gambar 9 memberikan gambaran yang lebih informatif secara statistik dibandingkan ROC untuk kasus seperti ini [22]. *Average Precision* (AP) sebesar 0,919 dengan presisi yang tetap tinggi (>0,90) hingga *recall* mencapai sekitar 0,60, sebelum menurun tajam pada *recall* di atas 0,90 — pola umum pada data medis tidak seimbang di mana mendeteksi kasus-kasus positif paling sulit memerlukan *trade-off* presisi yang lebih besar [12].

3.6. Analisis SHAP (*SHapley Additive exPlanations*)

Analisis interpretabilitas global menggunakan SHAP *TreeExplainer* menghasilkan peringkat kepentingan fitur berbasis rata-rata nilai SHAP absolut (Gambar 10), yang selaras dengan peringkat *Gini Importance* pada Gambar 6.

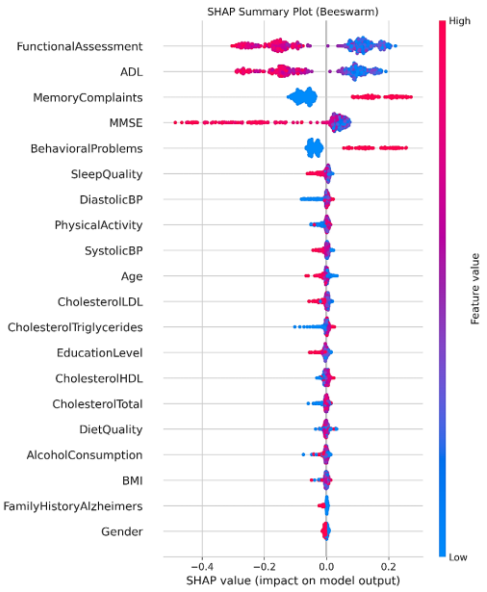


Gambar 10. SHAP Global *Feature Importance*

Nirwana Hendrastuty: *Penulis Korespondensi

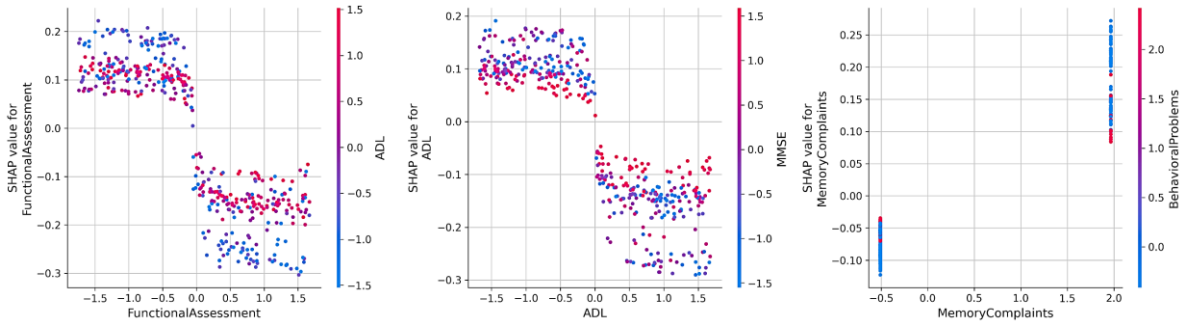
Journal of Information Technology, Software Engineering and Computer Science
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume 4, Number 4, October 2026
DOI : **https://doi.org/10.58602/itsecs.v4i4.405**
Page 153-171

Menariknya, pada Gambar 10 urutan posisi ketiga dan keempat sedikit bertukar dibandingkan Gambar 6: SHAP menempatkan *MemoryComplaints* di atas MMSE, sementara *Gini Importance* menempatkan MMSE di atas *MemoryComplaints*. Perbedaan kecil ini wajar terjadi karena kedua metode mengukur kontribusi fitur dengan mekanisme berbeda. *Gini Importance* mengukur penurunan impuritas rata-rata di seluruh pohon, sedangkan SHAP mengukur kontribusi marjinal aktual terhadap setiap prediksi individual [17]-[22] namun kesesuaian kelima fitur teratas pada kedua metode tetap menjadi bukti kuat bahwa hierarki fitur ini bukan artefak metodologis semata.



Gambar 11. SHAP *Summary Plot* (*Beeswarm*)

Pola warna pada Gambar 11 memperlihatkan arah pengaruh setiap fitur: titik biru (nilai fitur rendah) pada *FunctionalAssessment* dan ADL tersebar di sisi kanan (mendorong prediksi ke arah Alzheimer), sementara titik merah (nilai tinggi) berkumpul di sisi kiri menegaskan bahwa penurunan skor fungsional dan kemampuan hidup sehari-hari meningkatkan risiko prediksi Alzheimer, sesuai pemahaman klinis. Untuk *MemoryComplaints* dan *BehavioralProblems*, titik merah (nilai=1, keluhan ada) sepenuhnya terpisah di sisi kanan (SHAP positif) sementara titik biru (nilai=0) berkumpul di sisi kiri mengonfirmasi pola pemisahan hampir *biner* yang telah teramati pada Gambar 5.

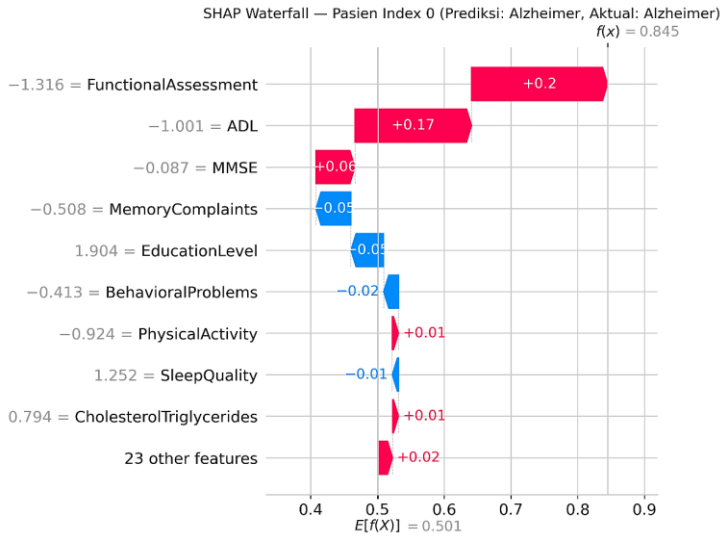


Gambar 12. SHAP Dependence Plot Tiga Fitur Teratas

Nirwana Hendrastuty: *Penulis Korespondensi

Journal of Information Technology, Software Engineering and Computer Science
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume 4, Number 4, October 2026
DOI : **https://doi.org/10.58602/itsecs.v4i4.405**
Page 153-171

Gambar 12 mengungkap pola yang menarik dan tidak linear: baik *FunctionalAssessment* maupun ADL menunjukkan hubungan berbentuk ambang batas (*threshold effect*) nilai SHAP relatif stabil di kisaran +0,1 hingga +0,2 ketika nilai fitur (terstandardisasi) berada di bawah nol, kemudian turun tajam ke kisaran −0,1 hingga −0,3 begitu nilai fitur melewati nol, dan relatif datar setelahnya. Pola ini menunjukkan bahwa model *Random Forest* menangkap suatu titik potong klinis yang tajam pada kedua skor tersebut, bukan hubungan linear proporsional sebuah wawasan yang tidak dapat diperoleh dari nilai *Gini Importance* semata dan menjadi salah satu nilai tambah utama analisis SHAP dibandingkan penelitian acuan [3]. Panel ketiga menegaskan pola biner pada *MemoryComplaints*: nilai SHAP terkonsentrasi negatif (∼−0,05 hingga −0,12) saat *MemoryComplaints*=0 dan positif (∼+0,08 hingga +0,27) saat *MemoryComplaints*=1, tanpa nilai antara.



Gambar 13. SHAP *Waterfall* Contoh Interpretasi Individual Pasien

Gambar 13 mendemonstrasikan interpretasi lokal untuk satu pasien uji (indeks-0) yang diprediksi Alzheimer (prediksi benar, *f(x)*=0,845 dari baseline *E[f(X)]*=0,501). Skor *FunctionalAssessment* yang rendah (nilai terstandardisasi −1,316) memberikan kontribusi terbesar (+0,20) mendorong prediksi ke arah Alzheimer, diikuti ADL rendah (−1,001, kontribusi +0,17) dan MMSE rendah (−0,087, kontribusi +0,06). Menariknya, *MemoryComplaints* pasien ini justru bernilai rendah (−0,508, artinya tidak ada keluhan memori) dan memberikan kontribusi negatif kecil (−0,05) yang sedikit menurunkan probabilitas namun tidak cukup mengimbangi kontribusi besar dari *FunctionalAssessment* dan ADL. Jenis penjelasan per-pasien seperti ini sepenuhnya tidak tersedia pada penelitian acuan [3] yang hanya melaporkan *feature importance* agregat, dan merupakan kontribusi utama penelitian ini terhadap transparansi model sejalan dengan urgensi XAI pada domain medis yang ditegaskan oleh Ghasemi dkk. [10] dan diterapkan secara praktis oleh Hidayat dkk. [11] pada risiko stroke, Soladoye dkk. [16], serta Jahan dkk. [17] pada Alzheimer berbasis data multimodal.

3.7. Perbandingan dengan Penelitian Acuan dan Studi Terkait

Perbandingan menyeluruh antara hasil penelitian ini dan penelitian sebelumnya [3] *accuracy* penelitian ini sedikit lebih tinggi (+0,19%), namun *recall*, *F1-Score*, dan *ROC-AUC* justru sedikit lebih rendah dibandingkan penelitian acuan [3]. Perbedaan ini kemungkinan disebabkan oleh perbedaan metrik pelaporan penelitian acuan melaporkan *precision*/*recall* rata-rata tertimbang dari kedua kelas, sementara penelitian ini melaporkan metrik untuk kelas positif (Alzheimer) secara spesifik, yang secara inheren lebih ketat

Nirwana Hendrastuty: *Penulis Korespondensi

Journal of Information Technology, Software Engineering and Computer Science
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume 4, Number 4, October 2026
DOI : **https://doi.org/10.58602/itsecs.v4i4.405**
Page 153-171

karena kelas Alzheimer adalah kelas minoritas dengan tantangan klasifikasi lebih tinggi [12]. Temuan ini menegaskan kontribusi metodologis penelitian ini bukan terletak pada peningkatan angka akurasi semata, melainkan pada: (1) validasi objektif seleksi fitur melalui RFECV dibandingkan pemangkasan sepihak menjadi 10 fitur pada [3]; dan (2) penyediaan interpretasi SHAP yang transparan pada level individu pasien, sepenuhnya tidak tersedia pada [3].

Dibandingkan pendekatan berbasis pencitraan MRI oleh Rohini dan Surendran [23] (akurasi 93% dengan >60 fitur struktural otak), penelitian ini menunjukkan bahwa data klinis non-pencitraan yang jauh lebih murah dan mudah diakses mampu mencapai akurasi setara atau lebih tinggi (94,19%), memperkuat kelayakan deteksi dini berbasis data klinis rutin untuk skrining awal di fasilitas kesehatan primer. Dibandingkan Fadhila dkk. [5] dan Akbar serta Rahmaddeni [6] yang menerapkan optimasi tanpa interpretabilitas model, serta Sari dan Fatah [7] yang berfokus pada *Decision Tree* tanpa penanganan ketidakseimbangan kelas, penelitian ini melengkapi celah tersebut melalui kombinasi seleksi fitur tervalidasi silang dan interpretasi SHAP dua tingkat (global-lokal).

Dibandingkan studi internasional terbaru yang juga menggabungkan Random Forest dan SHAP pada data klinis Alzheimer multimodal, Soladoye dkk. [16] melaporkan akurasi 95% dan AUC 98% menggunakan *Random Forest* dengan *backward elimination* dan optimasi *ant colony*, sementara Jahan dkk. [17] melaporkan bahwa *Random Forest* menjadi model dengan performa terbaik di antara sembilan algoritma yang diuji pada data multimodal (klinis, MRI, psikologis) dengan SHAP sebagai kerangka interpretasi. Performa penelitian ini (94,19%) berada pada rentang yang sebanding dengan kedua studi tersebut meskipun hanya menggunakan data klinis non-pencitraan tanpa fitur MRI atau data psikologis tambahan yang menegaskan bahwa fitur kognitif-fungsional rutin (MMSE, *FunctionalAssessment*, ADL) sudah cukup informatif untuk skrining awal, sejalan dengan kesimpulan kedua studi internasional tersebut yang juga menempatkan skor kognitif sebagai penentu utama prediksi.

Temuan bahwa RFECV mempertahankan seluruh fitur juga memberikan implikasi praktis penting: berbeda dengan asumsi umum bahwa seleksi fitur selalu mereduksi kompleksitas model, hasil ini menunjukkan bahwa pada dataset dengan jumlah fitur yang sudah relatif ringkas (32 fitur) dan telah melalui pra-pemrosesan yang baik, kontribusi utama seleksi fitur bergeser dari *reduksi dimensi* menjadi *validasi relevansi* sebuah nuansa yang jarang dilaporkan pada studi klasifikasi kesehatan di Indonesia yang cenderung berasumsi seleksi fitur selalu menghasilkan subset lebih kecil [5], [6]. Dengan demikian, secara keseluruhan, penelitian ini memberikan bukti empiris bahwa model *Random Forest* dengan seleksi fitur tervalidasi silang dan interpretasi SHAP mampu menghasilkan sistem deteksi dini Alzheimer yang tidak hanya akurat, tetapi juga transparan dan dapat dipertanggungjawabkan secara klinis memenuhi kesenjangan penelitian yang diidentifikasi pada bagian Pendahuluan.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan model deteksi dini Alzheimer berbasis algoritma *Random Forest* yang dioptimalkan melalui kombinasi seleksi fitur (RFECV), penyetelan *hyperparameter (GridSearchCV),* dan interpretasi model menggunakan SHAP pada dataset klinis non-pencitraan berisi 2.149 data pasien dengan 32 fitur prediktor. Model final mencapai performa yang sangat baik pada data uji, dengan *accuracy* 94,19%, *precision* 94,41%, *recall* 88,82%, *F1-Score* 91,53%, ROC-AUC 0,938, dan PR-AUC 0,919.

Secara metodologis, RFECV mengonfirmasi bahwa seluruh 32 fitur klinis tetap relevan bagi model, sehingga berfungsi sebagai validasi objektif atas kelengkapan fitur dan bukan sebagai metode reduksi dimensi. Analisis *feature importance* dan SHAP secara konsisten mengidentifikasi *Functional Assessment*, ADL, MMSE, *Memory Complaints*, dan

Nirwana Hendrastuty: *Penulis Korespondensi

Journal of Information Technology, Software Engineering and Computer Science
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume 4, Number 4, October 2026
DOI : **https://doi.org/10.58602/itsecs.v4i4.405**
Page 153-171

Behavioral Problems sebagai lima fitur klinis paling berpengaruh, dengan interpretasi SHAP memberikan penjelasan yang transparan baik secara global maupun pada level individu pasien, sebuah aspek interpretabilitas yang menjadi kontribusi utama penelitian ini.

Penelitian ini memiliki keterbatasan, yaitu optimasi *hyperparameter* hanya menggunakan *GridSearchCV*, evaluasi hanya menggunakan satu skema pembagian data uji (80:20), dan cakupan penelitian terbatas pada algoritma *Random Forest* tunggal. Penelitian selanjutnya disarankan untuk membandingkan dengan algoritma ensemble lain serta mengintegrasikan data pencitraan sebagai pelengkap data klinis. Secara praktis, kerangka kerja yang dikembangkan berpotensi digunakan sebagai dasar sistem pendukung keputusan medis untuk *skrining* dini Alzheimer di fasilitas kesehatan primer.

PERNYATAAN KONTRIBUSI PENULIS (CREDIT)

Nirwana Hendrastuty: Penulisan – peninjauan dan penyuntingan. **Marta Santra Wijaya:** Konseptualisasi, Metodologi, Perangkat lunak, Validasi, Analisis formal, Investigasi, Kurasi data, Penulisan – draf awal, Penulisan – peninjauan dan penyuntingan, Visualisasi. **Debby Alita:** Penulisan – peninjauan dan penyuntingan.

PERNYATAAN KONFLIK KEPENTINGAN

Para penulis menyatakan bahwa tidak terdapat kepentingan finansial atau hubungan pribadi yang diketahui yang dapat memengaruhi hasil penelitian yang dilaporkan dalam artikel ini.

PERNYATAAN PENDANAAN

Penulis tidak menerima dukungan finansial untuk penelitian, kepenulisan, dan/atau publikasi artikel ini.

PERNYATAAN PENGGUNAAN AI GENERATIF DAN TEKNOLOGI BERBANTUAN AI DALAM PROSES PENULISAN

Para penulis menggunakan Claude Sonnet 5 untuk meningkatkan kejelasan penulisan artikel ini. Seluruh konten yang dibantu oleh AI telah ditinjau dan disunting oleh penulis, dan para penulis bertanggung jawab sepenuhnya atas isi akhir yang dipublikasikan.

REFERENCES

[1] "Dementia." Accessed: Aug. 31, 2026. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dementia
[2] "Direktorat Jenderal Kesehatan Lanjutan." Accessed: Aug. 31, 2026. [Online]. Available: https://keslan.kemkes.go.id/view_artikel/2819/mengenal-demensia-alzheimer
[3] S. P. Yuni, M. Oktavianingrum, and F. Aksa, "Analisis Feature Importance pada Penyakit Alzheimer Menggunakan Random Forest," vol. 02, no. 01, 2026.
[4] M. Rohini and D. Surendran, "Toward Alzheimer's disease classification through machine learning," *Soft Computing*, vol. 25, no. 4, pp. 2589–2597, Feb. 2021, doi: 10.1007/s00500-020-05292-x.
[5] R. N. Fadhila, N. Ulinnuha, and D. Yuliati, "Implementation of Random Forest Method with Information Gain Selection and Hyperparameter Tuning for Alzheimer's Disease Classification," *LKJITI*, vol. 16, no. 1, pp. 1–13, Oct. 2025, doi: 10.24843/LKJITI.2025.v16.i01.p01.

Nirwana Hendrastuty: *Penulis Korespondensi

Journal of Information Technology, Software Engineering and Computer Science
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume 4, Number 4, October 2026
DOI : **https://doi.org/10.58602/itsecs.v4i4.405**
Page 153-171

[6] Firman Akbar and Rahmaddeni, "Komparasi Algoritma Machine Learning Untuk Memprediksi Penyakit Alzheimer," *JKT*, vol. 8, no. 2, pp. 236–245, Nov. 2022, doi: 10.35143/jkt.v8i2.5713.
[7] Y. Sari and Z. Fatah, "Klasifikasi Penyakit Alzheimer Menggunakan Data Mining Decision Tree".
[8] A. M. Priyatno and T. Widiyaningtyas, "A SYSTEMATIC LITERATURE REVIEW: RECURSIVE FEATURE ELIMINATION ALGORITHMS," *jitk*, vol. 9, no. 2, pp. 196–207, Feb. 2024, doi: 10.33480/jitk.v9i2.5015.
[9] M. Awad and S. Fraihat, "Recursive Feature Elimination with Cross-Validation with Decision Tree: Feature Selection Method for Machine Learning-Based Intrusion Detection Systems," *JSAN*, vol. 12, no. 5, p. 67, Sep. 2023, doi: 10.3390/jsan12050067.
[10] A. Ghasemi, S. Hashtarkhani, D. L. Schwartz, and A. Shaban-Nejad, "Explainable artificial intelligence in breast cancer detection and risk prediction: A systematic scoping review," *Cancer Innovation*, vol. 3, no. 5, p. e136, Oct. 2024, doi: 10.1002/cai2.136.
[11] C. Hidayat, A. Nugroho, and A. Suprianto, "Prediksi Risiko Stroke Berdasarkan Faktor Klinis Menggunakan Random Forest Dengan Optimasi Threshold dan SHAP," *jacis*, vol. 6, no. 1, pp. 127–137, Apr. 2026, doi: 10.47134/jacis.v6i1.178.
[12] S. Gholampour, "Impact of Nature of Medical Data on Machine and Deep Learning for Imbalanced Datasets: Clinical Validity of SMOTE Is Questionable," *MAKE*, vol. 6, no. 2, pp. 827–841, Apr. 2024, doi: 10.3390/make6020039.
[13] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *jair*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.
[14] A. Yaqoob *et al.*, "SGA-Driven feature selection and random forest classification for enhanced breast cancer diagnosis: A comparative study," *Sci Rep*, vol. 15, no. 1, p. 10944, Mar. 2025, doi: 10.1038/s41598-025-95786-1.
[15] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
[16] A. A. Soladoye, N. Aderinto, D. Osho, and D. B. Olawade, "Explainable machine learning models for early Alzheimer's disease detection using multimodal clinical data," *International Journal of Medical Informatics*, vol. 204, p. 106093, Dec. 2025, doi: 10.1016/j.ijmedinf.2025.106093.
[17] S. Jahan *et al.*, "Explainable AI-based Alzheimer's prediction and management using multimodal data," *PLoS ONE*, vol. 18, no. 11, p. e0294253, Nov. 2023, doi: 10.1371/journal.pone.0294253.
[18] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection".
[19] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions".
[20] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nat Mach Intell*, vol. 2, no. 1, pp. 56–67, Jan. 2020, doi: 10.1038/s42256-019-0138-9.
[21] J. N. Mandrekar, "Receiver Operating Characteristic Curve in Diagnostic Test Assessment," *Journal of Thoracic Oncology*, vol. 5, no. 9, pp. 1315–1316, Sep. 2010, doi: 10.1097/JTO.0b013e3181ec173d.
[22] T. Saito and M. Rehmsmeier, "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets," *PLoS ONE*, vol. 10, no. 3, p. e0118432, Mar. 2015, doi: 10.1371/journal.pone.0118432.

Nirwana Hendrastuty: *Penulis Korespondensi

Journal of Information Technology, Software Engineering and Computer Science
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume 4, Number 4, October 2026
DOI : **https://doi.org/10.58602/itsecs.v4i4.405**
Page 153-171

[23] M. Rohini and D. Surendran, "Weighted Hybrid Random Forest Model for Significant Feature prediction in Alzheimer's Disease Stages," *Int J Comput Intell Syst*, vol. 18, no. 1, p. 53, Mar. 2025, doi: 10.1007/s44196-025-00780-0.

[24] P. A. Ong, F. R. Annisafitrie, N. Purnamasari, C. Calista, N. Sagita, Y. Sofiatin, and Y. Dikot, "Dementia Prevalence, Comorbidities, and Lifestyle Among Jatinangor Elders," Front. Neurol., vol. 12, p. 643480, Jul. 2021, doi: 10.3389/fneur.2021.643480.

Nirwana Hendrastuty: *Penulis Korespondensi