

## STUDI KOMPARATIF ALGORITMA KLASIFIKASI K-NEAREST NEIGHBOR (KNN) DAN SUPPORT VECTOR MACHINE (SVM) UNTUK DETEKSI DINI ALZHEIMER

M. Rian Hidayah, Trisna Oditia, Gaozhan Muhammad Zabran,

Ken Ditha Tania, Ahmad Rifai, Dedy Kurniawan

Sistem Informasi, Universitas Sriwijaya

Jalan Palembang-Prabumulih KM 32, Kec. Indralaya, Kabupaten Ogan Ilir, 30662, Indonesia

ryanhidayah369@gmail.com

### ABSTRAK

Penyakit *Alzheimer* merupakan gangguan neurodegeneratif yang memerlukan deteksi dini untuk meningkatkan efektivitas penanganan. Permasalahan dalam penelitian ini adalah menentukan algoritma klasifikasi yang paling optimal untuk mendeteksi *Alzheimer* berdasarkan data klinis. Penelitian ini bertujuan untuk membandingkan performa algoritma *K-Nearest Neighbors (KNN)* dan *Support Vector Machine (SVM)*. Metode yang digunakan meliputi *preprocessing* data, *feature selection*, pelatihan model, serta evaluasi menggunakan *10-fold cross validation* dan *confusion matrix* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa *SVM* memiliki performa lebih baik dengan *accuracy* 84,45%, *precision* 80,59%, *recall* 73,29%, dan *F1-score* 76,62%, dibandingkan *KNN* yang memiliki nilai lebih rendah pada seluruh metrik. Dengan demikian, *SVM* lebih efektif dalam mendeteksi *Alzheimer*, terutama dalam meminimalkan kesalahan *false negative*.

**Kata kunci :** *Alzheimer, KNN, SVM, machine learning, klasifikasi, deteksi dini*

### 1. PENDAHULUAN

Penyakit *Alzheimer* merupakan gangguan neurodegeneratif progresif yang menjadi penyebab utama demensia di seluruh dunia. Organisasi Kesehatan Dunia (*WHO*) melaporkan bahwa lebih dari 55 juta orang hidup dengan demensia secara global, dengan penyakit *Alzheimer* menyumbang sekitar 60–70% dari seluruh kasus tersebut [1]. Deteksi dini menjadi sangat penting karena intervensi pada tahap awal dapat membantu memperlambat penurunan fungsi kognitif serta meningkatkan kualitas hidup pasien [2]. Oleh karena itu, berbagai pendekatan berbasis teknologi informasi mulai dikembangkan untuk mendukung proses diagnosis dan prediksi penyakit secara lebih cepat dan akurat.

Dalam beberapa tahun terakhir, penerapan *machine learning* dalam bidang kesehatan mengalami perkembangan pesat, termasuk untuk mendukung deteksi dini penyakit *Alzheimer* menggunakan data klinis, *biomarker*, maupun rekam medis elektronik [3], [4]. Teknik *machine learning* mampu mengekstraksi pola tersembunyi dari data kesehatan sehingga dapat membantu tenaga medis dalam melakukan klasifikasi risiko penyakit secara lebih objektif dan berbasis data [5]. Pendekatan ini juga banyak digunakan dalam pengembangan sistem pendukung keputusan klinis (*Clinical Decision Support System/CDSS*).

Sejumlah penelitian menunjukkan bahwa algoritma *supervised learning* klasik seperti *K-Nearest Neighbors (KNN)* dan *Support Vector Machine (SVM)* memiliki performa yang kompetitif dalam analisis data klinis berukuran kecil hingga menengah [6]. *SVM* dikenal memiliki kemampuan generalisasi yang baik serta efektif dalam memodelkan hubungan nonlinear

melalui penggunaan fungsi *kernel* [7], sedangkan *KNN* memiliki keunggulan dalam kesederhanaan implementasi dan pendekatan berbasis kedekatan jarak antar data [8].

Berbagai studi komparatif dalam bidang *machine learning* menunjukkan bahwa evaluasi beberapa algoritma klasifikasi secara bersamaan dapat memberikan gambaran performa yang lebih komprehensif terhadap suatu dataset. Penelitian terdahulu menunjukkan bahwa perbandingan beberapa algoritma klasifikasi seperti *Naive Bayes*, *SVM*, *KNN*, *Decision Tree*, dan *Random Forest* dapat memberikan wawasan yang lebih luas terkait kelebihan dan keterbatasan masing-masing metode dalam proses analisis data [9]. Pendekatan komparatif serupa juga diterapkan dalam berbagai penelitian untuk meningkatkan akurasi klasifikasi pada berbagai domain, termasuk bidang kesehatan [10].

Berdasarkan latar belakang tersebut, permasalahan dalam penelitian ini adalah bagaimana menentukan algoritma klasifikasi yang paling optimal antara *K-Nearest Neighbors (KNN)* dan *Support Vector Machine (SVM)* dalam mendeteksi dini penyakit *Alzheimer* berdasarkan data klinis [11]. Tujuan penelitian ini adalah membandingkan performa kedua algoritma tersebut menggunakan metrik evaluasi yang komprehensif. Adapun rencana penyelesaian masalah dilakukan melalui tahapan *preprocessing* data, *feature selection*, pelatihan model, serta evaluasi menggunakan teknik *10-fold cross validation* dan *confusion matrix* untuk memperoleh hasil yang akurat dan objektif.

## 2. TINJAUAN PUSTAKA

### 2.1. Penyakit Alzheimer dan Pentingnya Deteksi Dini

Penyakit *Alzheimer* merupakan gangguan neurodegeneratif yang menyebabkan penurunan fungsi kognitif secara bertahap dan menjadi penyebab utama demensia. Deteksi dini penyakit ini sangat penting karena dapat membantu memperlambat progresivitas penyakit melalui intervensi medis yang tepat [1], [2]. Seiring berkembangnya teknologi, berbagai pendekatan berbasis data mulai digunakan untuk mendukung proses diagnosis yang lebih cepat dan akurat, khususnya melalui pemanfaatan *machine learning* dalam analisis data kesehatan [12].

### 2.2. Machine Learning dalam Deteksi Alzheimer

Penerapan *machine learning* dalam bidang kesehatan, khususnya untuk deteksi *Alzheimer*, telah berkembang pesat dalam beberapa tahun terakhir. Berbagai penelitian menunjukkan bahwa algoritma *machine learning* mampu mengolah data klinis, biomarker, maupun data citra medis untuk mengidentifikasi pola yang tidak terlihat secara langsung oleh manusia [13]. Selain itu, *machine learning* juga memungkinkan pengembangan sistem klasifikasi yang dapat membedakan kondisi normal, *mild cognitive impairment (MCI)*, dan *Alzheimer* secara otomatis [14].

Studi terbaru menunjukkan bahwa pendekatan berbasis *machine learning* dan *deep learning* mampu meningkatkan akurasi deteksi *Alzheimer* secara signifikan, terutama ketika dikombinasikan dengan teknik pemrosesan data dan ekstraksi fitur yang tepat [13]. Namun demikian, model *deep learning* umumnya membutuhkan data dalam jumlah besar dan sumber daya komputasi tinggi, sehingga tidak selalu optimal untuk dataset klinis berukuran kecil atau menengah [12].

### 2.3. Algoritma K-Nearest Neighbors (KNN)

*K-Nearest Neighbors (KNN)* merupakan algoritma *supervised learning* yang bekerja dengan mengklasifikasikan data berdasarkan kedekatan jarak dengan data lain dalam ruang fitur. *KNN* dikenal karena kesederhanaannya serta kemampuannya dalam menangani data nonlinear tanpa memerlukan proses pelatihan model yang kompleks [15].

Dalam konteks deteksi *Alzheimer*, *KNN* sering digunakan sebagai model dasar (*baseline model*) karena kemudahan implementasi dan interpretasinya. Namun, performa *KNN* sangat dipengaruhi oleh pemilihan nilai *k*, skala fitur, serta ukuran dataset. Pada beberapa penelitian, *KNN* menunjukkan performa yang cukup baik pada data klinis, meskipun cenderung memiliki keterbatasan dalam menangani dataset dengan dimensi tinggi atau jumlah data besar [16].

### 2.4. Algoritma Support Vector Machine (SVM)

*Support Vector Machine (SVM)* merupakan algoritma klasifikasi yang bekerja dengan mencari *hyperplane* optimal untuk memisahkan data ke dalam kelas yang berbeda. *SVM* memiliki keunggulan dalam menangani data berdimensi tinggi serta mampu memodelkan hubungan nonlinear melalui penggunaan fungsi *kernel* [13].

Dalam penelitian terkait *Alzheimer*, *SVM* banyak digunakan karena kemampuannya dalam menghasilkan model dengan generalisasi yang baik. Beberapa studi menunjukkan bahwa *SVM* mampu mencapai tingkat akurasi yang tinggi dalam klasifikasi *Alzheimer*, terutama ketika dikombinasikan dengan teknik seleksi fitur atau ekstraksi fitur yang efektif [17]. Namun, performa *SVM* juga bergantung pada pemilihan parameter seperti jenis *kernel* dan nilai regularisasi.

### 2.5. Penelitian Terdahulu

Beberapa penelitian sebelumnya telah mengkaji penggunaan *machine learning* dalam deteksi penyakit *Alzheimer*. Penelitian oleh Khalid [16] menunjukkan bahwa algoritma *machine learning* mampu memberikan performa yang baik dalam klasifikasi *Alzheimer*, terutama ketika dikombinasikan dengan teknik *preprocessing* yang tepat. Selanjutnya, Li [17] menemukan bahwa *Support Vector Machine (SVM)* mampu mencapai tingkat akurasi yang tinggi dalam diagnosis *Alzheimer* berbasis data klinis.

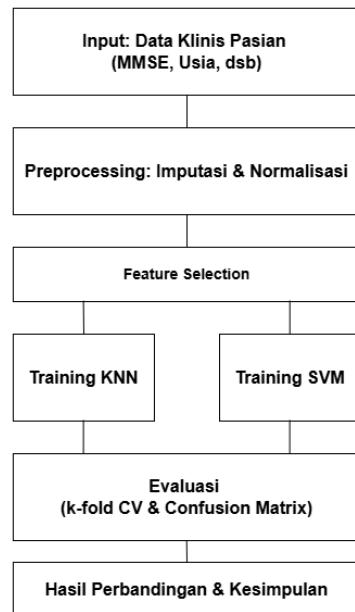
Di sisi lain, penelitian oleh Mokadem [18] membandingkan beberapa algoritma *machine learning* dan menunjukkan bahwa algoritma *ensemble* seperti *Random Forest* memiliki performa tinggi, namun dengan kompleksitas yang lebih besar. Penelitian lain oleh Aldila dan Supriyono [6] juga menunjukkan bahwa algoritma klasik seperti *KNN* dan *SVM* tetap kompetitif dalam dataset berukuran kecil hingga menengah.

Berdasarkan penelitian-penelitian tersebut, terlihat bahwa belum terdapat kesimpulan yang konsisten mengenai algoritma terbaik, sehingga diperlukan analisis komparatif lebih lanjut, khususnya antara *KNN* dan *SVM* pada dataset *Alzheimer* berbasis tabular.

## 3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental komparatif untuk mengevaluasi performa dua algoritma klasifikasi. Alur Metode penelitian dapat dilihat pada gambar 1.

Tahapan penelitian meliputi pengumpulan data yang didapat dari *Kaggle*, *preprocessing* data, *feature selection*, pelatihan model *KNN* dan *SVM*, serta evaluasi performa model menggunakan *10-fold cross validation* dan *confusion matrix*, lalu analisis perbandingan dari hasil yang didapatkan. Proses ini bertujuan untuk menghasilkan dan menentukan model klasifikasi yang optimal.



Gambar 1. Alur penelitian

### 3.1. Sumber Data

Dataset yang digunakan diperoleh dari *Kaggle*, *Alzheimer's Disease Dataset* oleh *Rabie El Kharoua* [11]. Dataset berformat CSV dengan 2.149 observasi dan 35 atribut klinis, setiap observasi merepresentasikan satu pasien dengan rentang *id* dari 4751 hingga 6900. Struktur dataset meliputi variabel demografis, skor *MMSE*, riwayat penyakit, *biomarker*, dan indikator perilaku.

Variabel target dalam dataset ini adalah diagnosis, yang menunjukkan status *Alzheimer*, dengan nilai 0 untuk *Non-Alzheimer* dan 1 untuk *Alzheimer*. Dataset bersifat tabular dan sintesis, serta dirancang untuk keperluan penelitian dan pengembangan model *machine learning*. Dengan karakteristik tersebut, dataset ini sangat sesuai untuk pendekatan *supervised learning*, khususnya dalam klasifikasi penyakit.

### 3.2. Preprocessing Data

*Preprocessing* dilakukan untuk meningkatkan kualitas model sesuai praktik standar *data mining* modern:

- Penghapusan atribut yang tidak relevan (*PatientID*, *DoctorInCharge*).
- Label *Encoding* untuk fitur kategorikal.
- Normalisasi menggunakan *StandardScaler* karena *KNN* dan *SVM* sensitif terhadap skala fitur.

Normalisasi sangat penting karena *KNN* berbasis perhitungan jarak *Euclidean* [8]. Sedangkan *SVM* dengan kernel *RBF* juga bergantung pada jarak antar titik data dalam ruang fitur [7].

### 3.3. Feature Selection

Tahap *feature selection* dilakukan untuk memilih atribut yang paling relevan terhadap variabel target

serta mengurangi dimensi data. Dari total 35 atribut awal, dilakukan proses seleksi fitur dengan cara menghilangkan fitur yang tidak relevan, dan memilih fitur yang memiliki hubungan signifikan terhadap diagnosis *Alzheimer* menggunakan pendekatan statistik.

Tujuan dari tahap ini adalah untuk mengurangi *noise* pada data, meningkatkan akurasi model, dan mempercepat proses komputasi. Hasil dari *feature selection* adalah subset fitur optimal yang kemudian digunakan dalam proses pelatihan model *KNN* dan *SVM*.

### 3.4. K-Nearest Neighbors (KNN)

*KNN* mengklasifikasikan data berdasarkan mayoritas label tetangga terdekat.

Parameter yang digunakan:

- $k = 5$
- Distance metric: *Euclidean*

Menggunakan rumus jarak *euclidean*:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan:

- $d(x, y)$  = jarak antara dua data (data uji dan data latih)
- $x_i$  = nilai fitur ke- $i$  pada data pertama
- $y_i$  = nilai fitur ke- $i$  pada data kedua
- $n$  = jumlah total fitur

Rumus jarak *euclidean* menghitung jarak antara dua titik dalam ruang fitur. Dalam *KNN*, jarak ini digunakan untuk menentukan tetangga terdekat yang akan menentukan hasil klasifikasi.

### 3.5. Support Vector Machine (SVM)

*SVM* mencari *hyperplane* optimal untuk memisahkan kelas.

Parameter yang digunakan:

- Kernel: *RBF*
- $C = 1$
- $\gamma = \text{textscales}$

Menggunakan rumus kernel *RBF*:

$$K(x, x') = \exp(-\gamma \|x - x'\|^2) \quad (2)$$

Keterangan:

- $K(x, x')$  = nilai kernel (kemiripan antara dua data)
- $x, x'$  = dua data dalam ruang fitur
- $\|x - x'\|^2$  = jarak kuadrat antara dua data
- $\gamma$  = parameter yang mengontrol tingkat pengaruh jarak

Kernel *RBF* mengubah data ke ruang berdimensi lebih tinggi sehingga pola nonlinear dapat dipisahkan. Nilai *kernel* yang besar menunjukkan bahwa dua data memiliki kemiripan tinggi.

### 3.6. Evaluasi Model

Evaluasi model pertama menggunakan *10-fold cross validation*, di mana data dibagi menjadi 10

bagian dan diuji secara bergantian untuk memastikan hasil stabil.

Selanjutnya, untuk mengukur performa kedua model digunakan empat metrik utama dari *confusion matrix*:

- Accuracy:** Persentase total prediksi yang benar.
- Precision:** Ketepatan model dalam memprediksi kelas positif.
- Recall (Sensitivity):** Kemampuan model mengidentifikasi seluruh pasien yang benar-benar sakit. Metrik ini sangat diprioritaskan dalam konteks medis untuk meminimalisir *false negatives*.
- F1-Score:** Rata-rata harmonik antara *precision* dan *recall* untuk memberikan gambaran performa model secara seimbang.

#### 4. HASIL DAN PEMBAHASAN

##### 4.1. Analisis menggunakan Cross Validation

Hasil pengujian menggunakan teknik *10-fold cross validation* dapat dilihat pada Tabel 1.

Tabel 1. Metrik Evaluasi

| Metrik Evaluasi | KNN    | SVM    |
|-----------------|--------|--------|
| Accuracy        | 0.7212 | 0.8445 |
| Precision       | 0.6560 | 0.8059 |
| Recall          | 0.4461 | 0.7329 |
| F1-Score        | 0.5303 | 0.7662 |

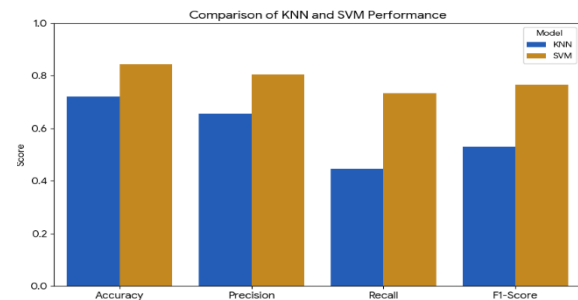
Tabel 1. menampilkan hasil evaluasi performa algoritma *K-Nearest Neighbors (KNN)* dan *Support Vector Machine (SVM)* berdasarkan empat metrik utama, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Berdasarkan hasil pengujian menggunakan teknik *10-fold cross validation*, terlihat bahwa model SVM secara konsisten menunjukkan performa yang lebih unggul dibandingkan KNN pada seluruh metrik yang digunakan.

Pada metrik *accuracy*, SVM memperoleh nilai sebesar 0,8445 atau 84,45%, sedangkan KNN hanya mencapai 0,7212 atau 72,12%. Hal ini menunjukkan bahwa secara keseluruhan SVM mampu menghasilkan prediksi yang lebih tepat dibandingkan KNN. Dari sisi *precision*, SVM juga menunjukkan kinerja yang lebih baik dengan nilai 0,8059 dibandingkan KNN sebesar 0,6560, yang berarti bahwa prediksi positif yang dihasilkan oleh SVM lebih akurat dan memiliki tingkat kesalahan yang lebih rendah.

Perbedaan yang paling signifikan terlihat pada metrik *recall*. SVM memperoleh nilai *recall* sebesar 0,7329, sedangkan KNN hanya mencapai 0,4461. Nilai ini menunjukkan bahwa SVM jauh lebih mampu dalam mendeteksi pasien yang benar-benar menderita *Alzheimer*. Dalam konteks medis, kemampuan ini sangat penting karena kesalahan dalam mendeteksi pasien yang sebenarnya sakit (*false negative*) dapat berdampak serius terhadap penanganan penyakit. Oleh karena itu, keunggulan SVM pada metrik *recall* menjadi faktor penting dalam menilai efektivitas model.

Selain itu, pada metrik *F1-score* yang merupakan kombinasi antara *precision* dan *recall*, SVM kembali menunjukkan performa yang lebih baik dengan nilai 0,7662 dibandingkan KNN sebesar 0,5303. Hal ini mengindikasikan bahwa SVM memiliki keseimbangan yang lebih baik antara ketepatan prediksi dan kemampuan dalam mendeteksi kasus positif.

Visualisasi perbandingan performa kedua model juga dapat dilihat pada Gambar 2.

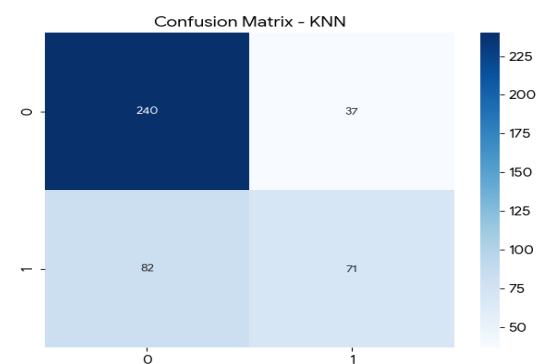


Gambar 2. Grafik batang perbandingan performa KNN dan SVM

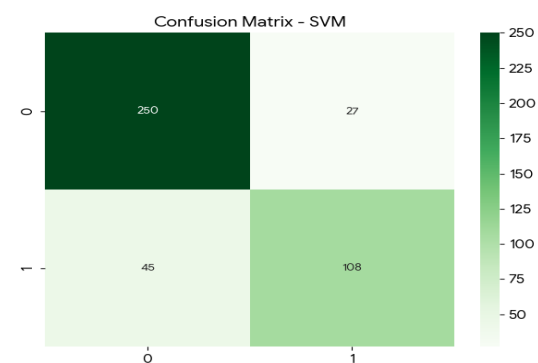
Secara keseluruhan, hasil pada Gambar 2. menunjukkan bahwa SVM lebih unggul dibandingkan KNN dalam mengklasifikasikan data klinis untuk deteksi dini *Alzheimer*.

##### 4.2. Analisis menggunakan Confusion Matrix

Hasil pengujian *confusion matrix* dapat dilihat pada Gambar 3. dan Gambar 4.



Gambar 3. Confusion matrix KNN



Gambar 4. Confusion matrix SVM

Pada Gambar 3. *confusion matrix KNN*, terlihat bahwa model berhasil mengklasifikasikan 240 data sebagai *true negative (TN)*, yaitu pasien *non-Alzheimer* yang diprediksi dengan benar. Namun, terdapat 37 *false positive (FP)*, yaitu pasien *non-Alzheimer* yang salah diprediksi sebagai *Alzheimer*. Untuk kelas positif, *KNN* hanya mampu mendeteksi 71 *true positive (TP)*, yaitu pasien *Alzheimer* yang diprediksi dengan benar, tetapi memiliki 82 *false negative (FN)*, yaitu pasien *Alzheimer* yang tidak terdeteksi oleh model. Jumlah *false negative* yang cukup besar ini menunjukkan bahwa *KNN* memiliki kelemahan dalam mendeteksi pasien yang benar-benar menderita *Alzheimer*.

Sementara itu, pada Gambar 4. *confusion matrix SVM* terlihat peningkatan performa yang cukup signifikan. Model ini mampu mengklasifikasikan 250 *true negative (TN)*, lebih tinggi dibandingkan *KNN*, serta menurunkan jumlah *false positive (FP)* menjadi 27. Pada kelas positif, *SVM* berhasil mendeteksi 108 *true positive (TP)*, yang jauh lebih tinggi dibandingkan *KNN*, serta menurunkan jumlah *false negative (FN)* menjadi 45. Penurunan *false negative* ini menunjukkan bahwa *SVM* jauh lebih baik dalam mengidentifikasi pasien *Alzheimer*.

Perbandingan kedua *confusion matrix* tersebut menunjukkan bahwa *SVM* tidak hanya lebih akurat secara keseluruhan, tetapi juga lebih seimbang dalam menangani kedua kelas. Terutama dalam konteks medis, kemampuan *SVM* dalam mengurangi jumlah *false negative* menjadi keunggulan utama, karena kesalahan dalam bentuk pasien sakit yang tidak terdeteksi dapat berdampak serius terhadap penanganan dan pengobatan [19].

## 5. KESIMPULAN DAN SARAN

Penelitian ini berhasil membandingkan performa algoritma *Support Vector Machine (SVM)* dan *K-Nearest Neighbors (KNN)* untuk deteksi dini penyakit *Alzheimer*. Hasil eksperimen menggunakan *10-fold cross validation* menunjukkan bahwa *SVM* dengan *kernel Radial Basis Function (RBF)* memiliki performa yang jauh lebih unggul dibandingkan *KNN* di seluruh metrik evaluasi, dengan tingkat akurasi mencapai 84,45% dan *recall* sebesar 73,29%. Keunggulan *SVM* ini sangat krusial dalam konteks medis karena nilai *recall* yang tinggi efektif untuk meminimalisir kesalahan deteksi pada pasien yang sebenarnya sakit (*false negatives*). Keunggulan *SVM* dalam menangani hubungan nonlinear antar fitur menjadikannya lebih efektif dalam memodelkan kompleksitas data klinis dan mendukung pengambilan keputusan yang lebih akurat. Adapun saran untuk penelitian selanjutnya adalah mengembangkan metode penanganan ketidakseimbangan data seperti *SMOTE*, mengeksplorasi algoritma lain seperti *Random Forest* atau *XGBoost*, serta melakukan optimasi parameter model untuk meningkatkan performa, sekaligus mempertimbangkan penggunaan dataset yang lebih

besar dan beragam agar hasil penelitian menjadi lebih *generalizable* dan aplikatif di dunia nyata.

## DAFTAR PUSTAKA

- [1] World Health Organization, "Dementia," Fact Sheet, World Health Organization, Geneva, Switzerland, Mar. 2025. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/dementia>
- [2] S. G. Singh, D. Das, U. Barman, and M. J. Saikia, "Early Alzheimer's Disease Detection: A Review of Machine Learning Techniques for Forecasting Transition from Mild Cognitive Impairment," *Diagnostics*, vol. 14, no. 16, p. 1759, 2024.
- [3] G. Battineni, N. Chintalapudi, and F. Amenta, "Machine learning driven by magnetic resonance imaging for the classification of Alzheimer disease progression: Systematic review and meta-analysis," *JMIR Aging*, vol. 7, 2024, Art. no. e59370.
- [4] G. Yan, S. Wu, Q. Qian, et al., "Early Diagnosis of Alzheimer's Disease Using Blood Biomarkers and Machine Learning," *BMC Medical Informatics and Decision Making*, 2025.
- [5] S. Demir and H. Selvitopi, "Early Diagnosis of Alzheimer's Disease Using Machine Learning Methods," *Procedia Computer Science*, vol. 258, pp. 107–117, 2025.
- [6] A. S. Aldila and L. Supriyono, "Evaluating Logistic Regression, SVM, KNN, and Ensemble Models for Accurate Heart Disease Risk Prediction," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 11, no. 3, pp. 949–957, 2026.
- [7] V. Piccialli and M. Sciandrone, "Nonlinear optimization and support vector machines," *Annals of Operations Research*, vol. 314, pp. 15–47, 2022.
- [8] B. Wang and S. Zhang, "A new locally adaptive K-nearest centroid neighbor classification based on the average distance," *Connection Science*, vol. 34, no. 1, pp. 2084–2107, 2022.
- [9] M. Al Fachrozi and K. D. Tania, "Comparison of Naive Bayes, SVM, K-NN, Decision Tree, and Random Forest in Sentiment Analysis Based on SeaBank Application Aspects," *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. XII, no. 1, pp. 69–80, 2025.
- [10] R. Rayadhani, W. Aldora, and M. Rahardi, "Comparative Analysis of Random Forest, SVM, and Naive Bayes for Cardiovascular Disease Prediction," *Journal of Applied Informatics and Computing*, vol. 9, no. 6, pp. 3234–3243, 2025.
- [11] R. El Kharoua, "Alzheimer's Disease Dataset," Kaggle, 2024. [Online]. Available: <https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset>
- [12] O. A. Dara, J. M. Lopez-Guede, H. I. Raheem, J. Rahebi, E. Zulueta, and U. Fernandez-Gamiz, "Alzheimer's disease diagnosis using machine

- learning: a survey,” *Applied Sciences*, vol. 13, no. 14, p. 8298, 2023.
- [13] S. Mohsen, “Alzheimer’s disease detection using deep learning and machine learning: a review,” *Artificial Intelligence Review*, vol. 58, no. 9, p. 262, 2025.
- [14] M. K. U. Ahamed, R. Hossen, B. K. Paul, M. Hasan, W. H. Al-Arashi, M. Kazi, and M. A. Talukder, “A hybrid filtering and deep learning approach for early Alzheimer’s disease identification,” *Scientific Reports*, vol. 15, no. 1, p. 27694, 2025.
- [15] Y. Manzali, K. A. Barry, R. Flouchi, Y. Balouki, and M. Elfar, “A feature weighted K-nearest neighbor algorithm based on association rules,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 15, no. 7, pp. 2995–3008, 2024.
- [16] M. Z. Khalid, N. Iqbal, B. Ali, J. S. U. Rahman, S. Iqbal, L. Almudaimiegh, Z. Y. Hamd, and A. Gareeballah, “Detection and classification of Alzheimer’s disease using deep and machine learning,” *Tomography*, vol. 12, no. 1, p. 4, 2026.
- [17] M. Li, H. Liu, Y. Li, Z. Wang, Y. Yuan, and H. Dai, “Intelligent diagnosis of Alzheimer’s disease based on machine learning,” *Proceedings of the 2023 4th International Symposium on Artificial Intelligence for Medicine Science*, pp. 456–462, Oct. 2023.
- [18] Z. Mokadem, M. Djerioui, B. Attallah, and Y. Brik, “A comparison of machine learning algorithms for predicting Alzheimer’s disease using neuropsychological data,” *Science, Engineering and Technology*, vol. 5, no. 1, pp. 177–191, 2025.
- [19] P. Iacobescu, V. Marina, C. Anghel, and A. D. Anghel, “Evaluating binary classifiers for cardiovascular disease prediction: enhancing early diagnostic capabilities,” *Journal of Cardiovascular Development and Disease*, vol. 11, no. 12, p. 396, 2024.