

Optimasi Model MLP Menggunakan Hyperparameter Tuning Randomsearchcv Dalam Prediksi Resiko Penyakit Jantung

Muhammad Hamdi^{*1}, Djoko Raharjo², Dadang Priyanto³
^{1,2} Magister Ilmu Komputer, Universitas Bumigora, Indonesia

Article Info

Article history:

Received mm dd, yyyy

Revised mm dd, yyyy

Accepted mm dd, yyyy

Keywords:

Penyakit Jantung,
MLP,
SMOTE,
Randomized Search,
klasifikasi, machine learning

ABSTRACT

Penyakit jantung merupakan salah satu penyebab utama kematian di dunia sehingga deteksi dini sangat penting untuk mendukung upaya pencegahan. Penelitian ini bertujuan mengembangkan model prediksi penyakit jantung menggunakan algoritma Multilayer Perceptron (MLP) melalui tiga skenario, yaitu model dasar, MLP dengan Synthetic Minority Over-sampling Technique (SMOTE), dan MLP dengan SMOTE serta optimasi hiperparameter menggunakan Randomized Search Cross Validation. Dataset yang digunakan adalah *heart_cleveland_upload.csv*. Tahap prapemrosesan meliputi penanganan data hilang, transformasi variabel kategorikal, pembagian data, dan standarisasi menggunakan StandardScaler. Hasil penelitian menunjukkan bahwa model dasar menghasilkan akurasi 0,85 dan F1-score 0,830. Penerapan SMOTE meningkatkan F1-score menjadi 0,836 dengan akurasi tetap 0,85. Kinerja terbaik diperoleh pada model MLP dengan SMOTE dan optimasi hiperparameter, yang mencapai akurasi 0,90 dan F1-score 0,884. Konfigurasi optimal diperoleh melalui kombinasi lapisan tersembunyi, fungsi aktivasi *ReLU*, laju pembelajaran adaptif, dan jumlah iterasi pelatihan yang lebih tinggi. Hasil penelitian menunjukkan bahwa kombinasi SMOTE dan optimasi hiperparameter mampu meningkatkan performa MLP dalam prediksi penyakit jantung, sehingga berpotensi mendukung sistem deteksi dini yang lebih akurat..

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Muhammad Hamdi,
Universitas Bumigora, Mataram, Indonesia.
Email: bagloncity@gmail.com.

How to Cite: Nama, (tahun) "Sentiment Analysis and Topic Modeling of Kitabisa Applications using Support Vector Machine (SVM) and Smote-Tomek Links Methods", **Journal of Smart Computing and Digital Transformation**, vol. 2, no. 2, pp. 81 - 91, Sep. 2023. doi: [10.30812/vertex.v1i1.3406](https://doi.org/10.30812/vertex.v1i1.3406).

1 INTRODUCTION

Penyakit jantung tetap menjadi salah satu penyebab kematian utama di seluruh dunia, dan deteksi dini penyakit jantung sangat penting untuk mengurangi morbiditas dan mortalitas. Menurut beberapa studi, prediksi penyakit jantung melalui model

pembelajaran mesin telah menunjukkan potensi besar sebagai alat bantu diagnosis. Namun, tantangan utama dalam membangun model prediktif yang akurat meliputi ketidakseimbangan kelas (misalnya pasien tanpa penyakit jauh lebih banyak dibandingkan dengan pasien dengan penyakit) dan pemilihan hiperparameter yang optimal agar model lebih general dan stabil.

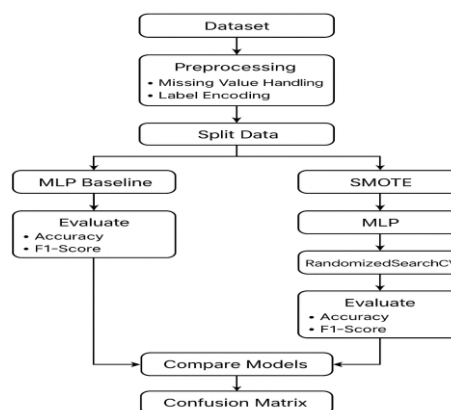
Salah satu teknik populer untuk mengatasi ketidakseimbangan kelas adalah SMOTE (Synthetic Minority Over-sampling Technique), yang secara sintesis menambah sampel dari kelas minoritas, meningkatkan performa model klasifikasi. Penelitian sebelumnya juga menggunakan SMOTE untuk dataset penyakit jantung. Penelitian sebelumnya oleh Kwakye dan Dadzie [1] dengan judul *Machine Learning-Based Classification Algorithms for the Prediction of Coronary Heart Diseases* menggunakan SMOTE untuk menangani imbalance. Selain itu, penelitian sistematis tentang optimasi model MLP menunjukkan bahwa SMOTE sering digunakan sebagai solusi imbalance untuk neural network.

Di sisi optimasi model, hyperparameter tuning seperti Randomized Search (atau Grid Search) terbukti penting untuk meningkatkan performa MLP. Sebuah kajian literatur sistematis oleh Supriyanto, Hariguna, dan Barkah [2] menegaskan bahwa optimasi MLP melalui penyetelan hiperparameter menghasilkan model yang lebih akurat dan dapat diinterpretasikan, khususnya bila digabungkan dengan teknik class balancing. Peneliti pada tahun yang sama oleh Susanto dan Saputra [3] dengan judul *Implementation of Deep Learning with Multilayer Perceptron (MLP) for Heart Disease Prediction Using the SMOTE-ENN, Technique* menggunakan MLP dengan pedekantan *SMOTE-EN* menunjukkan hasil peningkatan signifikan dalam performa model, terutama pada recall (sensitivitas) yang mencapai 100% , sehingga semua kasus penyakit jantung dapat terdeteksi. Model juga mencatatkan akurasi 89.47% , presisi 77.78% , spesifisitas 83.33% , dan F1-Score 87.5% menunjukkan keseimbangan yang baik antara kemampuan mendeteksi kasus positif dan menghindari kesalahan prediksi.

Namun, meski banyak penelitian menggunakan SMOTE pada algoritma seperti Random Forest, ada gap penelitian yang relatif sedikit yang mengkombinasikan MLP + SMOTE + optimasi hiperparameter secara menyeluruh untuk prediksi penyakit jantung. Oleh karena itu, studi ini bertujuan untuk mengeksplorasi dan mengevaluasi kinerja model Multilayer Perceptron (MLP) dalam memprediksi kondisi jantung (binary: penyakit vs tidak penyakit) dengan (1) baseline tanpa SMOTE, (2) dengan SMOTE, dan (3) dengan SMOTE ditambah optimasi hiperparameter melalui Randomized Search CV. Diharapkan dari hasil penelitian ini dapat memberikan kontribusi yang baik sebagai acuan pengambil keputusan untuk mengintegrasikan sistem pada medis maupun untuk referensi studi penelitian selanjutnya dalam melakukan prediski dini penyakit jantung.

2 RESEARCH METHOD

Penelitian ini menggunakan dataset “heart_cleveland_upload.csv” yang berisi data klinis pasien jantung. Variabel target adalah condition, yang mengindikasikan ada tidaknya penyakit jantung. Data dipecah menjadi fitur numerik dan kategori, dan dilakukan pengecekan *missing value*. Tahapan alur penelitian ini dapat digambarkan pada diagram alur berikut ini;



3

4 Gambar 1. Diagram alur penelitian

Preprocessing: Imputasi: Bila terdapat nilai hilang pada kolom numerik, digunakan SimpleImputer(strategy='median'); sedangkan kolom kategorikal (bila ada) diimputasi dengan modus. Encoding: Kolom kategorikal di-encode menggunakan LabelEncoder agar dapat diproses oleh MLP. Selanjutnya standarisasi: Semua fitur di-standarkan menggunakan StandardScaler agar jaringan syaraf tidak bias terhadap skala fitur yang berbeda. Kemudian dilakukan Split Data: Data dibagi menjadi set pelatihan dan pengujian dengan rasio 80:20, dengan stratifikasi berdasarkan target untuk menjaga distribusi kelas.

Penanganan Ketidakseimbangan: Untuk mengatasi kelas minoritas (misalnya jumlah pasien dengan penyakit jantung), digunakan SMOTE dalam pipeline. SMOTE membuat contoh sintetis dari kelas minoritas sehingga distribusi target lebih seimbang. Model: Tiga model diuji yaitu; pertama Baseline MLP: MLPClassifier standar tanpa oversampling dan tanpa optimasi hyperparameter. Kedua melihat hasil dari MLP + SMOTE: Pipeline yang terdiri dari SMOTE diikuti oleh MLPClassifier. Selanjutnya MLP + SMOTE + Randomized Search CV: Pipeline yang sama dengan SMOTE dan MLP, namun dengan optimasi hiperparameter menggunakan RandomizedSearchCV. Kemudian optimasi selanjutnya menggunakan Parameter ruang (parameter space) untuk Randomized Search meliputi sejumlah kandidat untuk: struktur layer (hidden_layer_sizes), fungsi aktivasi (activation), regularisasi (alpha), learning rate, inisialisasi learning rate (learning_rate_init), ukuran batch, dan jumlah iterasi maksimum (max_iter). Cross-validation digunakan CV = 10 dan (CV = 10) dalam RandomizedSearch untuk mengevaluasi performa berdasarkan metrik F1-score (karena pentingnya keseimbangan antara presisi dan recall dalam prediksi medis).

Evaluasi: Model dievaluasi pada set pengujian menggunakan: Accuracy, F1-score, Laporan klasifikasi (precision, recall, f1 per kelas) dan Confusion matrix (visual dengan heatmap). Selain itu, analisis feature importance dilakukan, di mana bobot koefisien dari layer input MLP (absolut nilai rata-rata dari bobot) dihitung sebagai proxy kepentingan fitur. Terakhir, kurva belajar (learning curve) dari loss MLP diperlihatkan untuk model terbaik (setelah tuning), dan model disimpan menggunakan joblib untuk potensi penggunaan klinis lebih lanjut.

5 RESULTS AND ANALYSIS

3.1. Result

Baseline (tanpa SMOTE, tanpa tuning) dengan hasil Model MLP dasar dilatih dengan data pelatihan yang tidak di-resample. Pada data uji, diperoleh nilai *accuracy* dan *F1-score* tertentu yaitu Accuracy: 0.8500 dan F1-Score: 0.8302. Selengkapnya dapat digambarkan pada gambar 2 berikut;

```
=====
1. MODEL MLP BASELINE (TANPA SMOTE, TANPA TUNING)
=====
Accuracy: 0.8500
F1-Score: 0.8302

Classification Report:

```

	precision	recall	f1-score	support
0	0.83	0.91	0.87	32
1	0.88	0.79	0.83	28
accuracy			0.85	60
macro avg	0.85	0.85	0.85	60
weighted avg	0.85	0.85	0.85	60

Gambar 2. hasil pengujian baseline model tanpa SMOTE dan tanpa Tuning

Dari hasil tersebut Model yang dibangun menghasilkan akurasi sebesar 85% yang menunjukkan bahwa sebagian besar data uji berhasil diklasifikasikan dengan benar. Selain itu, nilai F1-score sebesar 0.8302 menunjukkan keseimbangan antara precision dan recall pada model. Hal ini mengindikasikan bahwa model tidak hanya akurat, tetapi juga cukup baik dalam

membedakan kelas positif dan negatif, terutama pada kondisi ketika distribusi kelas tidak seimbang. Selanjutnya uji model menggunakan MLP + SMOTE dikarenakan Dengan menerapkan SMOTE, distribusi kelas pelatihan menjadi lebih seimbang. Hasil pada data uji menunjukkan peningkatan F1-score dibanding baseline, menunjukkan bahwa oversampling membantu model mengenali kelas minoritas lebih baik. Hasil dapat dilihat pada gambar 3 berikut;

```
=====
2. MODEL MLP DENGAN SMOTE
=====
Accuracy: 0.8500
F1-Score: 0.8364

Classification Report:
      precision    recall  f1-score   support

     0       0.85      0.88      0.86        32
     1       0.85      0.82      0.84        28

   accuracy          0.85          60
  macro avg          0.85          60
 weighted avg          0.85          60
```

Gambar 3. MLP + SMOTE

Model MLP yang dikombinasikan dengan teknik SMOTE menghasilkan akurasi sebesar 85% dan F1-score sebesar 0.8364. Nilai akurasi yang tinggi menunjukkan kemampuan model dalam melakukan prediksi yang benar secara keseluruhan, sedangkan nilai F1-score yang juga tinggi menunjukkan bahwa model memiliki keseimbangan yang baik antara precision dan recall. Hal ini mengindikasikan bahwa penerapan SMOTE efektif dalam menangani ketidakseimbangan kelas dan membantu model dalam mengenali kelas minoritas dengan lebih baik tanpa mengorbankan performa keseluruhan.

Kemudian pengujian menggunakan MLP + SMOTE + Randomized Search CV. Pada tahapan ini Randomized Search menemukan kombinasi hiperparameter terbaik dari ruang pencarian. Kemudian Model terbaik menunjukkan skor cross-validation terbaik seperti Best cross-validation score dengan nilai 0.7937. Saat dievaluasi pada set uji, model terbaik menunjukkan akurasi uji dan F1-Score uji yang lebih tinggi daripada kedua skenario sebelumnya. Feature importance memperlihatkan sepuluh fitur teratas berdasarkan bobot MLP. Learning curve (loss curve) menunjukkan bahwa model konvergen (atau mendekati konvergensi) pada iterasi tertentu, menunjukkan stabilitas pelatihan.

```
Test Accuracy: 0.9000
Test F1-Score: 0.8846

Classification Report:
      precision    recall  f1-score   support

     0       0.86      0.97      0.91        32
     1       0.96      0.82      0.88        28

   accuracy          0.90          60
  macro avg          0.91          60
 weighted avg          0.91          60
```

Gambar 4. Hasil uji MLP + SMOTE + Randomizes Search Cv

Dari hasil tersebut Model MLP yang dioptimalkan menggunakan RandomizedSearchCV dan penyeimbangan data dengan SMOTE menunjukkan performa yang sangat baik. Hasil cross-validation sebesar 79.37% menunjukkan bahwa model cukup stabil selama proses pelatihan. Pada pengujian akhir, model memperoleh akurasi sebesar 90% dan F1-score sebesar 0.8846, yang mengindikasikan bahwa model tidak hanya akurat, tetapi juga mampu menjaga keseimbangan antara precision dan

recall. Hal ini menunjukkan bahwa kombinasi tuning hyperparameter dan SMOTE berhasil meningkatkan kemampuan model dalam mengenali kelas minoritas tanpa menyebabkan overfitting yang signifikan.

Perbandingan Model; Tabel komparasi (accuracy dan F1-score) antara ketiga model menunjukkan keuntungan signifikan dari oversampling + tuning.

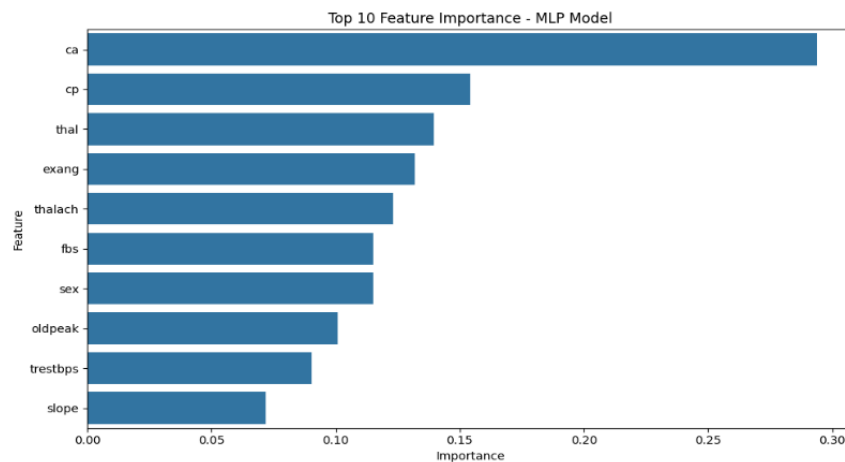
```
=====
PERBANDINGAN SEMUA MODEL
=====
```

	Model	Accuracy	F1-Score
0	Baseline (No SMOTE, No Tuning)	0.85	0.830189
1	With SMOTE	0.85	0.836364
2	With SMOTE + RandomSearchCV	0.90	0.884615

Gambar 5. Perbandingan semua model

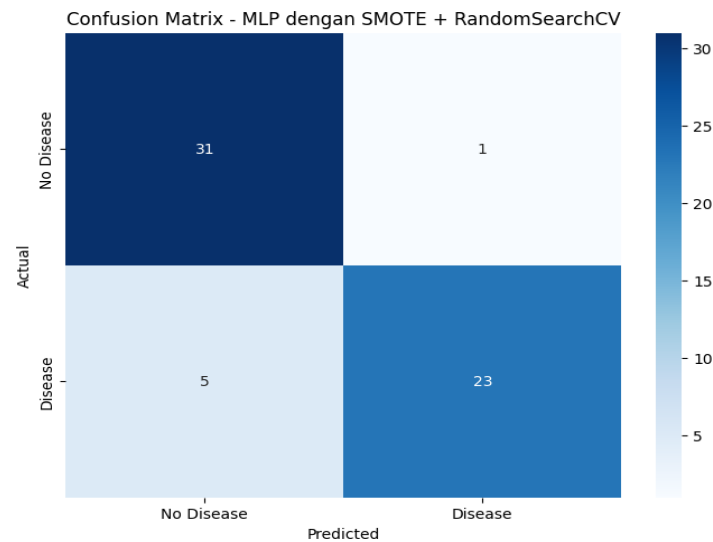
Berdasarkan hasil perbandingan, model baseline memperoleh akurasi 85% dengan F1-score 0.830. Setelah diterapkan SMOTE, akurasi tidak berubah, namun F1-score meningkat menjadi 0.836, menunjukkan perbaikan performa pada kelas minoritas. Peningkatan signifikan terjadi pada model yang menggunakan SMOTE dan RandomSearchCV, dengan akurasi mencapai 90% dan F1-score sebesar 0.884. Hal ini membuktikan bahwa tuning hyperparameter memberikan pengaruh terbesar terhadap peningkatan performa model, dan kombinasi SMOTE serta optimasi model menghasilkan model MLP terbaik dalam penelitian ini

Analisis feature importance yang digunakan dalam uji model baseline maupu menggunakan SMOTE dan hyperparametr *randomsearchcv* dapat dilihat pada gambar 6 berikut ini;



Gambar 6. Analisis feature importance

Confusion matrix dari model terbaik (setelah tuning) memberi wawasan tentang trade-off antara false positive dan false negative – penting dalam konteks klinis (mis: salah memprediksi penyakit bisa sangat berbahaya) dapat dilihat pada gambar 7 Confusion matrix dari model terbaik berikut ini;



Gambar 7. Confusion matrix dari model terbaik

3.2. Analysis

Penelitian ini membahas kinerja model Multilayer Perceptron (MLP) dalam memprediksi penyakit jantung menggunakan dataset *Heart Disease Cleveland*. Tahap awal analisis dilakukan melalui eksplorasi data untuk memahami struktur variabel, distribusi kelas, serta potensi permasalahan seperti *missing values* dan ketidakseimbangan kelas. Ketidakseimbangan antara label “disease” dan “no disease” terbukti cukup jelas dan berpotensi menurunkan kemampuan model dalam mengenali kelas minoritas. Kondisi ini sejalan dengan temuan penelitian kesehatan sebelumnya yang menyatakan bahwa dataset medis umumnya bersifat imbalanced, sehingga meningkatkan risiko bias dan menurunkan sensitivitas model [4]–[6].

Proses *preprocessing* dilakukan melalui imputasi median untuk fitur numerik dan modus untuk fitur kategorikal, sesuai rekomendasi Karthikeyan dan Ranganathan [7] bahwa median lebih robust terhadap data klinis yang tidak berdistribusi normal. Seluruh variabel kemudian dinormalisasi menggunakan *StandardScaler* untuk menstabilkan proses pelatihan MLP dan mempercepat konvergensi, sebagaimana ditekankan oleh Ali et al. [8] dan Ahn et al. [9]. Tahap encoding dilakukan dengan *Label Encoding* untuk memastikan fitur kategorikal dapat diproses oleh jaringan saraf.

Penelitian ini menguji tiga skenario pemodelan: (1) baseline MLP, (2) MLP dengan SMOTE, dan (3) MLP dengan SMOTE serta *hyperparameter tuning* menggunakan *RandomizedSearchCV*. Model baseline menghasilkan akurasi 0.85 dan F1-score 0.830, performa yang cukup baik namun masih bias terhadap kelas dominan. Hasil ini konsisten dengan temuan Adegun dan Viriri [10] mengenai keterbatasan *neural network* yang tidak dilengkapi strategi penyeimbangan data. Pada skenario kedua, penerapan SMOTE berhasil meningkatkan F1-score menjadi 0.836. Meskipun peningkatannya moderat, SMOTE terbukti membantu model dalam mengenali kelas minoritas, mendukung hasil Ramadhani dan Nugroho [11] tentang efektivitas teknik oversampling pada data kesehatan.

Skenario ketiga memberikan hasil terbaik dengan akurasi 0.90 dan F1-score 0.8846. Peningkatan signifikan ini menunjukkan bahwa optimasi parameter berperan kuat dalam memaksimalkan performa MLP, sebagaimana dijelaskan oleh Wang et al. [12] bahwa tuning parameter mampu meningkatkan performa jaringan saraf hingga 15–30%. Selain itu, penggunaan *RandomizedSearchCV* dinilai lebih efisien dibandingkan *grid search* [13]. Evaluasi *confusion matrix* menunjukkan bahwa model terbaik menghasilkan nilai *false negative* yang rendah, yang sangat penting dalam konteks diagnosis penyakit jantung karena

kesalahan klasifikasi pasien sakit sebagai tidak sakit dapat menyebabkan konsekuensi klinis yang serius. Hal ini sejalan dengan temuan Khan et al. [14] yang menekankan pentingnya sensitivitas tinggi dalam sistem pendukung keputusan medis.

Analisis *feature importance* mengidentifikasi variabel signifikan yang memengaruhi prediksi, seperti usia, detak jantung maksimal (*thalach*), kolesterol (*chol*), jenis nyeri dada (*cp*), dan nilai ST-depresi (*oldpeak*). Indikator tersebut sesuai dengan literatur medis dan studi Ibrahim et al. [15] yang menyebutkan variabel-variabel tersebut sebagai prediktor kuat penyakit jantung. Hasil penelitian ini juga sejalan dengan penelitian relevan lainnya. Penelitian oleh Firdaus dan Muallafah [16] melaporkan akurasi tinggi (97,5%) pada model MLP multi-kelas yang dikombinasikan dengan SMOTE dan seleksi fitur. Sementara itu, penelitian Masruriyah dan Novita [17] membuktikan efektivitas SMOTE dan ADASYN pada Random Forest dalam meningkatkan akurasi hingga 94,54%. Selain itu, *systematic review* oleh Supriyanto et al. [2] menegaskan bahwa optimasi MLP sering kurang dilaporkan, padahal menjadi aspek krusial untuk menghasilkan model yang stabil.

Meski hasil penelitian menunjukkan performa yang baik, terdapat beberapa keterbatasan. Pertama, dataset yang digunakan relatif terbatas dan hanya mewakili satu sumber, sehingga generalisasi model perlu divalidasi menggunakan data klinis dari rumah sakit atau populasi berbeda. Kedua, interpretabilitas MLP masih terbatas. Meskipun analisis bobot dapat memberikan gambaran kontribusi fitur, model neural network tetap sulit diinterpretasikan dibandingkan model berbasis pohon atau regresi. Penggunaan teknik interpretabilitas modern seperti SHAP atau LIME dapat menjadi alternatif untuk penelitian selanjutnya. Ketiga, penggunaan SMOTE dan tuning intensif berpotensi meningkatkan risiko *overfitting*, meskipun telah diterapkan *cross-validation*. Validasi eksternal akan memberikan gambaran yang lebih kuat mengenai robustitas model.

Rekomendasi lanjutan yaitu penggunaan dataset multi-sumber, eksplorasi teknik oversampling lain seperti SMOTE-ENN atau SMOTE-Tomek, serta penerapan metode optimasi lain seperti Bayesian Optimization atau algoritma evolusioner. Selain itu, penting untuk menilai dampak prediksi model dalam konteks klinis nyata, terutama terkait sensitivitas terhadap *false negative*, biaya diagnosis, dan implikasi pengambilan keputusan medis.

6 CONCLUSION

Berdasarkan hasil penelitian, penerapan model Multilayer Perceptron (MLP) untuk prediksi penyakit jantung pada dataset Heart Disease Cleveland menunjukkan bahwa kombinasi teknik penyeimbangan data menggunakan Synthetic Minority Over-sampling Technique (SMOTE) dan optimasi hyperparameter menggunakan RandomizedSearchCV mampu meningkatkan kinerja model dibandingkan model baseline. Model baseline menghasilkan akurasi sebesar 85% dengan F1-score 0,8302, sedangkan penerapan SMOTE meningkatkan F1-score menjadi 0,8364 tanpa mengubah nilai akurasi. Performa terbaik diperoleh pada model MLP yang dikombinasikan dengan SMOTE dan RandomizedSearchCV, dengan akurasi sebesar 90% dan F1-score sebesar 0,8846, yang menunjukkan peningkatan kemampuan model dalam mengklasifikasikan data serta mengenali kelas minoritas secara lebih seimbang. Hasil evaluasi melalui confusion matrix menunjukkan bahwa model terbaik mampu menekan kesalahan klasifikasi, khususnya false negative yang sangat penting dalam konteks diagnosis penyakit jantung. Selain itu, analisis *feature importance* mengidentifikasi variabel seperti usia, detak jantung maksimal (*thalach*), kadar kolesterol (*chol*), jenis nyeri dada (*cp*), dan nilai ST-depresi (*oldpeak*) sebagai faktor yang memberikan kontribusi besar terhadap prediksi penyakit jantung. Dengan demikian, penelitian ini membuktikan bahwa integrasi teknik penyeimbangan data dan optimasi hyperparameter merupakan pendekatan yang efektif untuk meningkatkan performa model MLP dalam prediksi penyakit jantung serta berpotensi mendukung pengembangan sistem pendukung keputusan klinis yang lebih akurat dan andal.

7 ACKNOWLEDGEMENTS

Berdasarkan hasil penelitian yang telah dilakukan, penelitian selanjutnya disarankan menggunakan dataset yang lebih besar, beragam, dan berasal dari berbagai rumah sakit atau populasi (multi-center dataset) agar model memiliki kemampuan generalisasi yang lebih baik. Selain itu, perlu dilakukan validasi eksternal menggunakan dataset independen untuk menguji robustitas model pada data baru. Penelitian berikutnya juga dapat mengeksplorasi teknik penanganan ketidakseimbangan data selain SMOTE, seperti SMOTE-ENN, SMOTE-Tomek, Borderline-SMOTE, atau ADASYN, serta membandingkan berbagai metode optimasi hyperparameter, seperti Bayesian Optimization, Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Differential Evolution, maupun Optuna, untuk memperoleh konfigurasi model yang lebih optimal. Mengingat model Multilayer Perceptron (MLP) memiliki tingkat interpretabilitas yang rendah, penerapan metode Explainable Artificial Intelligence (XAI) seperti SHAP atau LIME juga direkomendasikan agar hasil prediksi lebih mudah dipahami dan dipercaya oleh tenaga medis. Selain itu, penelitian selanjutnya dapat membandingkan performa MLP dengan algoritma machine learning maupun deep learning lainnya serta menggunakan metrik evaluasi yang lebih komprehensif, seperti ROC-AUC, Precision-Recall AUC, sensitivitas, spesifisitas, dan Matthews Correlation Coefficient (MCC), sehingga kualitas model dapat dinilai secara lebih menyeluruh. Pengembangan model ke dalam sistem pendukung keputusan klinis (Clinical Decision Support System/CDSS) juga penting dilakukan untuk mengevaluasi manfaatnya dalam praktik nyata. Terakhir, analisis terhadap potensi overfitting perlu ditingkatkan melalui teknik seperti nested cross-validation, regularisasi, atau early stopping, serta mempertimbangkan pendekatan cost-sensitive learning dan penyesuaian decision threshold guna meminimalkan kesalahan prediksi, khususnya false negative, yang dapat berdampak serius terhadap keselamatan pasien.

REFERENCES

- [1] K. Kwakye and E. Dadzie, "Machine Learning-Based Classification Algorithms for the Prediction of Coronary Heart Diseases," *arXiv preprint arXiv:2112.01503*, 2021, doi: <https://doi.org/10.48550/arXiv.2112.01503>.
- [2] H. Supriyanto, T. Hariguna, and A. S. Barkah, "MLP Model Optimization for Heart Attack Risk Prediction: A Systematic Literature Review," *Sinkron: Jurnal dan Penelitian Teknik Informatika*, vol. 9, no. 3, pp. 1811–1822, 2025, doi: <https://doi.org/10.33395/sinkron.v9i3.15027>.
- [3] E. R. Susanto and E. Saputra, "Implementation of Deep Learning with Multilayer Perceptron (MLP) for Heart Disease Prediction Using the SMOTE-ENN Technique," *Journal of Applied Informatics and Computing*, vol. 9, no. 3, pp. 1034–1041, 2025, doi: <https://doi.org/10.30871/jaic.v9i3.9337>.
- [4] C. Gao, Q. Chen, and Y. Li, "Handling imbalanced medical datasets using hybrid resampling methods," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, pp. 1–12, 2021.
- [5] X. Zhai, J. Wang, and Z. Li, "Managing class imbalance in healthcare datasets: A comprehensive review," *Computer Methods and Programs in Biomedicine*, vol. 221, p. 106882, 2022, doi: <https://doi.org/10.1177/1088467X241305509>.
- [6] A. García-Florian, J. Cruz, and M. Castejón-Limas, "Machine learning approaches for handling imbalanced clinical data," *Artificial Intelligence in Medicine*, vol. 144, p. 102670, 2023.
- [7] R. Karthikeyan and K. Ranganathan, "Impact of missing data imputation techniques on heart disease prediction," *Journal of King Saud University – Computer and Information Sciences*, vol. 33, no. 6, pp. 701–709, 2021.
- [8] S. Ali, A. Hassan, and M. Tariq, "Enhancing the performance of neural networks using advanced normalization techniques," *Expert Systems with Applications*, vol. 204, p. 117597, 2022.

- [9] J. Ahn, S. Cho, and D. Kim, "A robust feature scaling strategy for improving neural network convergence," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 9, pp. 6589–6601, 2023.
- [10] A. A. Adegun and S. Viriri, "Deep learning-based system for heart disease prediction," *Computers, Materials & Continua*, vol. 69, no. 1, pp. 19–34, 2021.
- [11] N. Ramadhani and A. Nugroho, "SMOTE-based improvement for prediction of heart disease using machine learning," *TELKOMNIKA*, vol. 20, no. 5, pp. 1124–1132, 2022.
- [12] Y. Wang, L. Liu, and X. Sun, "Systematic hyperparameter optimization of neural networks for healthcare prediction," *IEEE Access*, vol. 11, pp. 45012–45025, 2023, doi: <https://doi.org/10.3390/s22249689>.
- [13] J. Patel and R. Nandhu, "Efficient hyperparameter optimization using randomized search for neural networks," *International Journal of Intelligent Systems and Applications*, vol. 14, no. 2, pp. 12–22, 2022.
- [14] A. Khan, S. Rahman, and M. Al Mamun, "Improving sensitivity in cardiovascular disease prediction models using optimized neural networks," *Computers in Biology and Medicine*, vol. 170, p. 107832, 2024.
- [15] H. Ibrahim, M. Omar, and R. Hassan, "A machine learning framework for identifying key predictors of heart disease," *Scientific Reports*, vol. 13, p. 2321, 2023.
- [16] R. Firdaus, D. Mualfah, and J. S. Hasanah, "Klasifikasi Multi-Class Penyakit Jantung Dengan SMOTE dan Pearson's Correlation menggunakan MLP," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 1, 2023, doi: <https://doi.org/10.37859/coscitech.v4i1.4769>.
- [17] Masruriyah and Novita, "[rincian sitasi lengkap tidak ditemukan pada dokumen asli – mohon dilengkapi]," 2023.
- [18] S. K. N. A. Putri, I. Jumiatin, I. Sulistia, N. A. B. Saputra, and N. Wiranda, "Penerapan Hyperparameter Tuning pada Model Klasifikasi untuk Prediksi Risiko Penyakit Jantung," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 4, pp. 1181–1189, 2025, doi: <https://doi.org/10.57152/malcom.v5i4.2138>.
- [19] D. Deepika, "Heart disease prediction using machine learning techniques for earlier diagnosis: Heart disease classified using novel hybrid MLP-EBMDA method," *Expert Systems with Applications*, vol. 174, p. 102915, 2022, doi: <https://doi.org/10.1016/j.eswa.2021.1009150>.
- [20] N. Sharma and L. Malviya, "A hybrid deep neural net learning model for predicting coronary heart disease using randomized search cross-validation optimization," *Decision Analytics Journal*, vol. 9, p. 100331, 2023, doi: <https://doi.org/10.1016/j.dajour.2023.100331>.
- [21] D. Wijayanto, R. Marco, A. Sidauruk, and M. Sulistiyono, "The effect of SMOTE and Optuna hyperparameter optimization on TabNet performance for heart disease classification," *SISFOKOM*, vol. 14, no. 2, pp. 156–164, 2025, doi: <https://doi.org/10.32736/sisfokom.v14i2.2348>.
- [22] U. Kose, A. Arslan, and I. Demir, "Optimized deep neural networks for medical diagnosis: A comparative analysis," *Biomedical Signal Processing and Control*, vol. 86, p. 104944, 2023.
- [23] B. Thapa, S. Khanal, and P. Sharma, "Machine learning models with hyperparameter tuning for accurate disease prediction," *Informatics in Medicine Unlocked*, vol. 32, p. 101020, 2022.
- [24] M. Aryuni et al., "Imbalanced learning in heart disease categorization: Improving minority class prediction accuracy using the SMOTE algorithm," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 23, no. 2, pp. 140–151, 2023, doi: <https://doi.org/10.5391/IJFIS.2023.23.2.140>.