



Klasifikasi Multi-Class Penyakit Jantung Dengan SMOTE dan Pearson's Correlation menggunakan MLP

Rahmad Firdaus¹, Desti Mualfah², Julian Silvia Hasanah³

Email: ¹rahmadfirdaus@umri.ac.id, ²destimualfah@umri.ac.id, ³190401221@student.umri.ac.id

¹²³Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Muhammadiyah Riau

Diterima: 27 Maret 2023 | Direvisi: 05 April 2023 | Disetujui: 30 April 2023

©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Muhammadiyah Riau, Indonesia

Abstrak

Penyakit jantung merupakan penyakit paling mematikan didunia. Laporan WHO tahun 2019 menyebutkan penyakit jantung sebagai penyebab kematian tertinggi didunia dengan persentase 16% dari jumlah kematian atau 8.9 juta kematian. Tingginya kematian yang disebabkan oleh penyakit ini terjadi karena penyakit ini biasanya timbul tanpa adanya gejala sehingga sulit untuk diketahui sejak dini oleh penderita. Salah satu cara untuk mengatasi permasalahan tersebut adalah dengan pemanfaatan metode klasifikasi. Hasil klasifikasi *multi-class* pada penelitian sebelumnya dengan dataset dan metode yang sama masih terbilang rendah yang salah satunya disebabkan oleh adanya *imbalance data*. Untuk itu dibutuhkan teknik *balancing data* serta *feature selection (FS)* untuk melihat pengaruh *imbalance data* dan pengaruh korelasi fitur pada klasifikasi *multi-class*. Pada penelitian ini menggunakan *Synthetic Minority Over-sampling Technique (SMOTE)* untuk *balancing data* dan *Pearson's Correlation (PS)* untuk memilih fitur dengan korelasi yang baik pada klasifikasi penyakit jantung menggunakan metode *Multi-Layer Perceptron (MLP)*. Penelitian ini dengan *MLP+Pearson's Correlation* hanya mendapatkan akurasi tertinggi sebesar 63.33%. Akurasi tertinggi dengan penggabungan metode *MLP+SMOTE+Pearson's Correlation* sebesar 92.5%. Serta akurasi tertinggi pada penelitian ini didapat menggunakan metode *MLP+SMOTE* yaitu sebesar 97.5%.

Kata kunci: penyakit jantung, SMOTE, pearson's correlation, multi-layer perceptron

Multi-class classification of heart disease SMOTE and pearson's correlation using MLP

Abstract

Heart disease is the deadliest disease in the world. The 2019 WHO report stated that heart disease is the highest cause of death in the world with a percentage of 16% of the total number of deaths or 8.9 million deaths. The high mortality caused by this disease occurs because this disease usually occurs without any symptoms so it is difficult to be detected, early sufferers. One way to overcome these problems is to use the classification method. The results of a multi-class classification in previous studies with the same dataset and method are still relatively low, one of which is caused by the imbalance of data. For this reason, data balancing techniques and feature selection (FS) are needed to see the effect of imbalanced data and the effect of feature correlation on multi-class classification. This study uses Synthetic Minority Over-sampling Technique (SMOTE) to balance data and Pearson's Correlation (PS) to select features with a good correlation in the classification of heart disease using the Multi-Layer Perceptron method (MLP). This study with MLP + Pearson's Correlation only got the highest score of 63.33%. The highest accuracy with the MLP + SMOTE + Pearson's Correlation method is 92.5%. And the highest accuracy in this study was obtained using the MLP + SMOTE method, which was 97.5%.

Keywords: heart disease, SMOTE, Pearson's Correlation, multi-layer perceptron

1. PENDAHULUAN

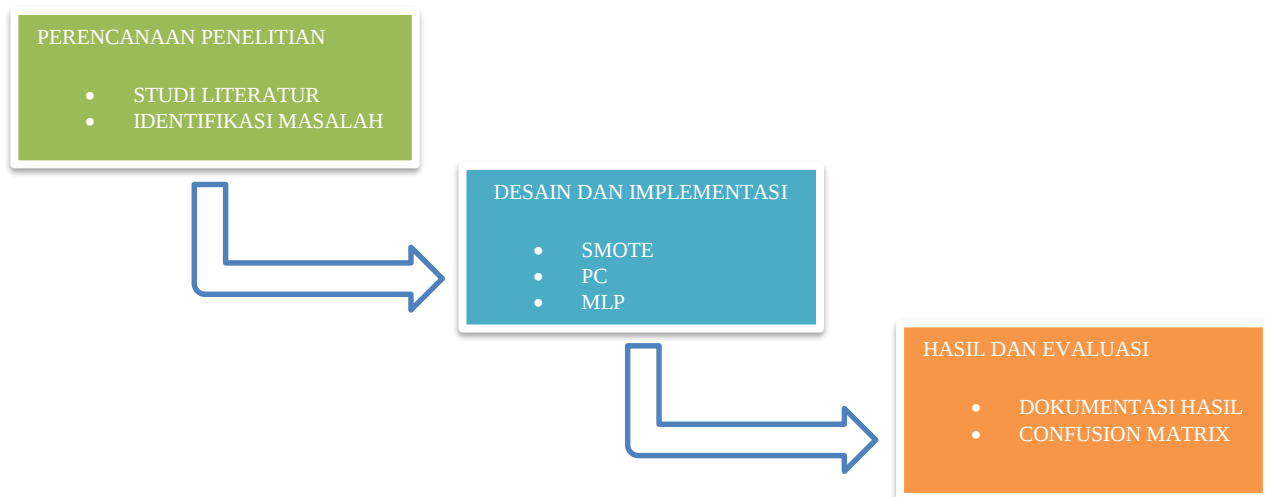
Penyakit jantung atau *heart disease* tercatat sebagai penyakit yang paling mematikan di dunia dan menduduki peringkat kedua sebagai penyakit yang paling mematikan di Indonesia [1]. Penyakit ini timbul karena adanya penyempitan pada lumen pembuluh darah arteri koroner akibat dari adanya plak atau yang disebut aterosklerosis [2]. Laporan WHO tahun 2019 menyebutkan penyakit jantung sebagai penyebab kematian tertinggi di dunia dengan persentase 16% dari jumlah kematian atau 8.9 juta kematian. Di Indonesia, Berdasarkan data Riset Kesehatan Dasar (Riskesdas) tahun 2018, setidaknya 15 dari 1000 orang, atau sekitar 2.784.064 orang di Indonesia menderita penyakit jantung [3] dengan presentase kematian sebesar 12,9 % dari penyebab kematian di Indonesia [4].

Penyakit jantung merupakan penyakit mematikan yang termasuk ke dalam Penyakit Tidak Menular (PTM) dimana penyakit tersebut tidak disebabkan oleh infeksi kuman serta biasanya muncul tanpa gejala awal sehingga penyakit ini sulit diidentifikasi atau disadari sejak dini oleh masyarakat umum [5] Maka dari itu, sangat penting mengetahui gejala Penyakit tersebut sejak dini agar dapat melakukan penanganan serta pengobatan dengan tepat. Untuk mengetahui penyakit tersebut sejak dini dapat dilakukan diagnosis terhadap hasil *medical check up* yang dimiliki pasien demi meminimalisir resiko penyakit yang lebih parah lagi. Salah satu cara untuk mengatasi permasalahan tersebut adalah dengan pemanfaatan metode klasifikasi. Klasifikasi merupakan proses menemukan model suatu data untuk memasukkan data tersebut kedalam kelas tertentu dari kelas yang ada [6]. Banyak metode yang dapat digunakan untuk melakukan klasifikasi, namun dalam penelitian ini menggunakan metode *Multi-Layer Perceptron*. Penelitian sebelumnya mengenai klasifikasi pada penyakit jantung (2 kelas dan 5 kelas) menunjukkan bahwa SVM dan MLP dapat memberikan akurasi lebih dari 90%, tetapi pada klasifikasi *multi-class* kedua algoritma tersebut hanya mendapat akurasi kurang dari 70% yaitu 68,86% pada MLP dan 59,01% pada SVM [7]. Klasifikasi *multi-class* pada penelitian lain juga mendapatkan akurasi yang rendah dimana akurasi tertinggi diperoleh metode MLP yaitu hanya sebesar 58,44% [8]. Pada klasifikasi *multi-class* lainnya (3 kelas), klasifikasi menghasilkan akurasi yang tinggi dengan menggunakan teknik SMOTE dan *feature selected* dimana akurasi tertinggi pada penelitian ini dengan menggunakan teknik SMOTE diperoleh metode *Multilayer perceptron* (MLP) sebesar 99.35% dan 94.44% pada metode *Logistic Regression* dengan menggunakan teknik SMOTE dan *Feature Selected* [9] Beberapa penelitian lain menggunakan teknik SMOTE untuk mengatasi imbalance data juga menghasilkan akurasi diatas 90%, seperti penelitian SMOTE untuk deteksi penyakit stroke menghasilkan akurasi sebesar 91% [10]

Dari uraian diatas, maka penelitian ini menggunakan metode *Multi-Layer Perceptron* dalam klasifikasi penyakit jantung dengan 5 kelas untuk mengetahui akurasi dalam pengklasifikasian tersebut. Teknik balanced data (SMOTE) dan *feature selection* (Pearson's Correlation) juga akan digunakan untuk meningkatkan akurasi pada klasifikasi *multi-class*.

2. METODE PENELITIAN

Berikut merupakan tahapan penelitian dapat dilihat pada gambar.1 berikut :



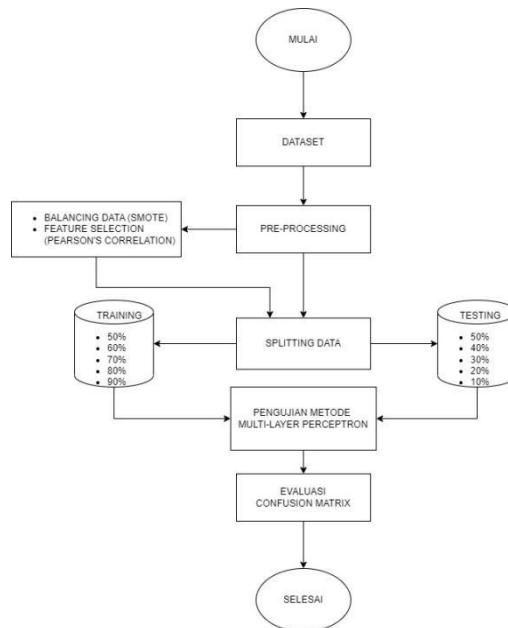
Gambar 1. Kerangka Penelitian

2.1. Perencanaan Penelitian

Merupakan tahap membaca dan mengumpulkan literatur melalui buku, jurnal maupun internet untuk mendapatkan teori-teori yang digunakan serta penelitian sebelumnya sebagai bahan referensi dalam penelitian ini.

2.2. Desain Dan Implementasi

Pada tahapan desain dan implementasi ini menjelaskan alur sistem untuk klasifikasi pada gambar ini seperti yang tertera pada gambar 2 :



Gambar 2. Flowchart Penelitian

2.2.1 Dataset

Dataset yang digunakan diambil dari website UCI yaitu *Heart Disease Data Set* yang dapat diunduh melalui link: dibawah ini <https://archive.ics.uci.edu/ml/datasets/heart+Disease>. Dataset terdiri dari 303 data dan 14 atribut dimana 13 atribut yang menjadi fitur dan satu atribut berfungsi sebagai *output* atau variabel *predictor*. Kelas pada variabel *predictor* terdiri dari 5 kelas (0,1,2,3,4) dimana kelas 0 menunjukkan tidak adanya penyakit jantung dan kelas 1 sampai 4 menunjukkan tingkat keparahan. Tabel 1 mendeskripsikan atribut yang terdapat pada dataset sebagai berikut :

Tabel 1. Atribut Pada Dataset

No	Nama Atribut	Deskripsi	Tipe Data
1	Age	Usia Pasien Dalam Satuan Tahun	Continuous
2	Sex	Jenis kelamin pasien. 1 = Laki-Laki 0 = Perempuan	Categorical
3	Cp	Jenis nyeri dada. Nilai 1 = Angina tipikal Nilai 2 = Angina atipikal Nilai 3 = Nyeri non angina Nilai 4 = Asimtomatik	Categorical
4	Trestbps	Tekanan darah diukur dalam mm/Hg.	Continuous
5	Chol	Kolesterol serum diukur dalam mg/dl.	Continuous
6	Fbs	Gula darah pasien. Jika lebih besar dari 120 mg/dl nilai atributnya adalah 1, selain itu nilai atribut adalah 0.	Continuous
7	Restecg	Hasil elektrokardiografi Nilai 0 = Normal Nilai 1 = memiliki kelainan gelombang ST-T (gelombang T inversi dan/atau elevasi atau depresi ST > 0,05 mV) Nilai 2 = menunjukkan kemungkinan atau pasti hipertrofi ventrikel kiri menurut kriteria Ester.	Categorical
8	Thalach	Detak jantung maksimal yang dicapai pasien.	
9	Exang	Latihan diinduksi angina. Nilai 1 = ya, Nilai 0 = tidak	Categorical
10	Oldpeak	Depresi ST yang diinduksi oleh latihan relatif terhadap set.	
11	Slope	Bentuk atau Ukuran Slope Nilai 1 = Miring ke atas Nilai 2 = Datar Nilai 3 = Miring ke bawah	Categorical

12	Ca	Jumlah pembuluh darah besar (0-3) yang diwarnai dengan fluoroskopi. Nilai atribut dapat berupa 0 hingga 3.	Categorical
13	Thal	Mewakili detak jantung pasien. Nilai 3 = Normal Nilai 6 = Cacat tetap Nilai 7 = Cacat reversibel	Categorical
14	Heart	Mewakili diagnosis penyakit jantung terdiri dari 5 kelas (0,1,2,3,4) dimana kelas 0 menunjukkan tidak adanya penyakit jantung dan kelas 1 sampai 4 menunjukkan tingkat keparahan.	Categorical

2.2.2 Pre-processing

Preprocessing Pengolahan data merupakan tahap mengolah dataset yang telah dikumpulkan sebelum digunakan pada proses klasifikasi. Pengolahan data atau *preprocessing* bertujuan untuk membersihkan data mentah menjadi data yang siap digunakan dalam diagnosis penyakit jantung.

2.2.3 Pearson's Correlation

Teknik mengurangi fitur yang tidak relevan untuk meningkatkan akurasi.

2.2.4 SMOTE

Merupakan metode *over-sampling* dimana data pada kelas minoritas diperbanyak dengan menggunakan data sintetik yang berasal dari replikasi data pada kelas minoritas.

2.2.5 Splitting Data

Membagi dataset menjadi data *training* dan data *testing* yang sebelumnya telah diberi label dengan pembagian 50:50, 60:40, 70:30, 80:20 dan 90:10.

2.2.6 Klasifikasi

Penelitian ini menggunakan metode *Multi-Layer Perceptron* yang terdiri dari 3 layer yaitu *input layer*, *hidden layer* dan *output layer*[11].

2.2.7 Evaluasi

Pada evaluasi, hasil pengklasifikasian di evaluasi dengan *Confusion Matrix*. Hasil pengklasifikasian di cocokkan dengan data sebenarnya.S

2.3. Hasil Dan Evaluasi

Pada tahapan ini menjelaskan hasil dari penelitian yang telah dilakukan seperti dokumentasi berupa *screenshot* atau laporan dari *output* penelitian. Selanjutnya dilakukan evaluasi menggunakan *confusion matrix* untuk melihat perbandingan hasil klasifikasi yang dilakukan oleh sistem (model) dengan hasil klasifikasi sebenarnya yang berbentuk tabel matriks yang menggambarkan kinerja model klasifikasi pada serangkaian data uji yang nilai sebenarnya diketahui. Hasil dari *confusion matrix* berupa nilai *accuracy*, *precision*, *recall* dan *f1-Score* terhadap model yang digunakan pada *Heart Disease Data Set*.

3. HASIL DAN PEMBAHASAN

Bab ini menguraikan tentang tahapan lebih lanjut, dimulai dari tahapan *data preparation*, *preprocessing*, *balancing data*, *feature selection*, *splitting data*, desain dan implementasi, dan evaluasi.

3.1. Data Preparation

Mengimport beberapa *library* yang diperlukan sebelum mengimport dataset yang akan digunakan. Adapun *library* yang digunakan sebagai berikut:

```

1 import matplotlib.pyplot as plt
2 import pandas as pd
3 import numpy as np
4
5 from sklearn.neural_network import MLPClassifier
6 from sklearn.model_selection import train_test_split
7 from sklearn.metrics import accuracy_score, confusion_matrix, precision_score, recall_score, f1_score
8 from sklearn.preprocessing import StandardScaler
9 from imblearn.over_sampling import SMOTE

```

Gambar 3. Library Yang Diperlukan

3.2. Data Preprocessing

Pada tahap ini, *preprocessing* yang dilakukan adalah *handling missing value* dan melakukan standarisasi pada data. *Missing value* dapat dilihat pada gambar dibawah ini :

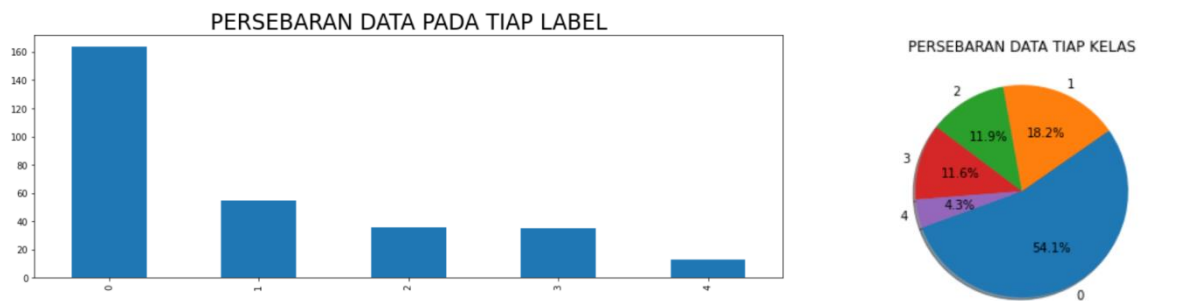
```
Age      0
Sex      0
Cp       0
Trestbps 0
Chol     0
Fbs      0
Restecg  0
Thalach  0
Exang    0
Oldpeak  0
Slope    0
Ca       4
Thal     2
Heart    0
dtype: int64
```

Gambar 4. Missing Value

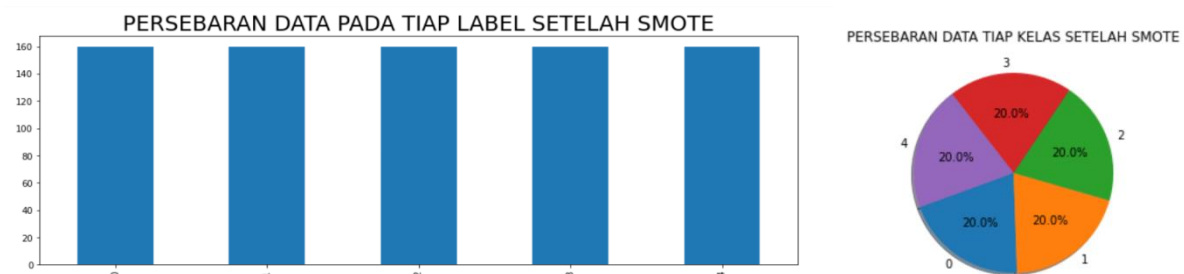
Dalam penelitian ini, 6 *missing value* tersebut diselesaikan dengan metode *drop* (menghapus baris yang memiliki *missing value*). Selanjutnya data distandarisasi agar perbedaan nilai/skala pada setiap atribut tidak terlalu jauh.

3.3. Implementasi SMOTE

Pada tahap ini, dilakukan *oversampling* dimana data kelas minoritas direplikasi menggunakan data sintetik yang berasal dari reproduksi data kelas minoritas agar setara atau seimbang dengan kelas mayoritas seperti gambar dibawah ini :



Gambar 5. Persebaran Data Tanpa SMOTE



Gambar 6. Persebaran Data Dengan SMOTE

3.4. Implementasi Pearson's Correlation

Pearson's correlation digunakan untuk melihat fitur mana yang memiliki korelasi tertinggi yang dijadikan dasar dalam *feature selection*. Korelasi pada setiap fitur dapat dilihat pada matrix berikut:

	Age	Sex	Cp	Trestbps	Chol	Fbs	Restecg	Thalach	Exang	Oldpeak	Slope	Ca	Thal	Heart
Age	1.000000	-0.031793	0.157435	0.301518	0.102521	0.117414	0.197580	-0.301098	0.098628	0.146102	0.157387	0.436737	0.069477	0.301902
Sex	-0.031793	1.000000	0.025857	-0.140445	-0.174285	-0.018454	0.018716	-0.077662	0.141221	0.134557	0.029308	0.074951	0.295782	0.203943
Cp	0.157435	0.025857	1.000000	0.041819	0.048424	-0.063869	0.172502	-0.356087	0.351751	0.222719	0.201917	0.254743	0.317650	0.422208
Trestbps	0.301518	-0.140445	0.041819	1.000000	0.044737	0.208126	0.171202	-0.134614	0.045975	0.140786	0.145338	0.103484	0.139753	0.220387
Chol	0.102521	-0.174285	0.048424	0.044737	1.000000	0.084215	0.142002	0.040198	0.061992	-0.052047	-0.083034	0.157225	-0.034658	0.009120
Fbs	0.117414	-0.018454	-0.063869	0.208126	0.084215	1.000000	0.046898	0.037458	-0.006968	-0.064093	-0.002298	0.134261	0.041349	0.015925
Restecg	0.197580	0.018716	0.172502	0.171202	0.142002	0.046898	1.000000	-0.111287	0.027953	0.070821	0.087529	0.229411	0.032726	0.309440
Thalach	-0.301098	-0.077662	-0.356087	-0.134614	0.040198	0.037458	-0.111287	1.000000	-0.362468	-0.344584	-0.440461	-0.246964	-0.269828	-0.397750
Exang	0.098628	0.141221	0.351751	0.045975	0.061992	-0.006968	0.027953	-0.362468	1.000000	0.213816	0.223657	0.043823	0.233087	0.242651
Oldpeak	0.146102	0.134557	0.222719	0.140786	-0.052047	-0.064093	0.070821	-0.344584	0.213816	1.000000	0.633943	0.259598	0.357477	0.486218
Slope	0.157387	0.029308	0.201917	0.145338	-0.083034	-0.002298	0.087529	-0.440461	0.223657	0.633943	1.000000	0.144326	0.323074	0.426868
Ca	0.436737	0.074951	0.254743	0.103484	0.157225	0.134261	0.229411	-0.246964	0.043823	0.259598	0.144326	1.000000	0.192721	0.522926
Thal	0.069477	0.295782	0.317650	0.139753	-0.034658	0.041349	0.032726	-0.269828	0.233087	0.357477	0.323074	0.192721	1.000000	0.469104
Heart	0.301902	0.203943	0.422208	0.220387	0.009120	0.015925	0.309440	-0.397750	0.242651	0.486218	0.426868	0.522926	0.469104	1.000000

Gambar 7. Matrix Korelasi

3.4. Implentasi *Multi-Layer Perceptron*

Klasifikasi menggunakan *MLP* dengan 3 *hidden layer* dimana setiap *hidden layer* memiliki 50 *neurons* atau *nodes*, learning rate sebesar 0.0016 dan jumlah *iterasi maksimal* adalah 1000 *iterasi*.

3.5. Hasil Dan Evaluasi

Berikut tabel hasil dari klasifikasi yang telah dilakukan dan hasil tersebut sesuai dengan hasil evaluasi menggunakan *confussion matrix*, dimana pada setiap tabel terdiri dari 5 *splitting data* (50:50, 60:40, 70:30, 80:20, 90:10), *accuracy*, *precession*, *recall* dan *f-score*.

Tabel 2. *Original Dataset*

<i>Splitting Data</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Score</i>
50:50	53.02%	48.94%	53.02%	50.61%
60:40	50.42%	45.21%	50.42%	47.4%
70:30	51.11%	45.69%	51.11%	48.09%
80:20	50.0%	45.25%	50.0%	47.46%
90:10	60.0%	53.15%	60.0%	56.37%

Tabel 2 merupakan hasil dari klasifikasi menggunakan metode *Multi-Layer Perceptron* (Tanpa *SMOTE* dan *PC*).

Tabel 3. *Dataset Dengan SMOTE*

<i>Splitting Data</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Score</i>
50:50	84.75%	84.96%	84.75%	84.7%
60:40	87.5%	87.77%	87.5%	87.36%
70:30	90.0%	90.07%	90.0%	89.94%
80:20	91.88%	91.95%	91.88%	91.76%
90:10	97.5%	97.56%	97.5%	97.48%

Tabel 3 merupakan hasil dari klasifikasi menggunakan metode *Multi-Layer Perceptron* dan dengan penerapan *SMOTE*.

Tabel 4. *Dataset Dengan Pearson's Correlation* (3 Fitur)

<i>Splitting Data</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Score</i>
50:50	51.01%	45.79%	51.01%	46.96%
60:40	52.1%	47.55%	52.1%	46.64%
70:30	60.0%	53.36%	60.0%	54.61%
80:20	50.0%	42.89%	50.0%	46.06%
90:10	63.33%	58.22%	63.33%	60.29%

Tabel 4 merupakan hasil dari klasifikasi menggunakan metode *Multi-Layer Perceptron* dengan 3 fitur (penerapan *PC*).

Tabel 5. *Dataset Dengan Pearson's Correlation* (5 Fitur)

<i>Splitting Data</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Score</i>
50:50	52.35%	48.01%	52.35%	49.93%
60:40	48.74%	43.75%	48.74%	46.01%
70:30	55.56%	47.57%	55.56%	50.34%
80:20	53.33%	46.75%	53.33%	49.74%
90:10	53.33%	50.94%	53.33%	51.24%

Tabel 5 merupakan hasil dari klasifikasi menggunakan metode *Multi-Layer Perceptron* dengan 5 fitur (penerapan *PC*).

Tabel 6. *Dataset Dengan Pearson's Correlation* (6 Fitur)

<i>Splitting Data</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Score</i>
-----------------------	-----------------	------------------	---------------	----------------

50:50	51.01%	48.82%	51.01%	48.8%
60:40	50.42%	44.83%	50.42%	47.0%
70:30	54.44%	46.49%	54.44%	49.77%
80:20	56.67%	46.96%	56.67%	51.32%
90:10	60.0%	52.47%	60.0%	54.4%

Tabel 6 merupakan hasil dari klasifikasi menggunakan metode Multi-Layer Perceptron dengan 6 fitur (penerapan PC).

Tabel 7, Dataset Dengan *Pearson's Correlation* (7 Fitur)

<i>Splitting Data</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Score</i>
50:50	49.66%	46.57%	49.66%	47.95%
60:40	48.74%	46.68%	48.74%	47.47%
70:30	52.22%	49.07%	52.22%	50.08%
80:20	53.33%	50.77%	53.33%	51.76%
90:10	46.67%	46.67%	46.67%	46.49%

Tabel 7 merupakan hasil dari klasifikasi menggunakan metode Multi-Layer Perceptron dengan 7 fitur (penerapan PC).

Tabel 8. Dataset Dengan *Pearson's Correlation* (10 Fitur)

<i>Splitting Data</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Score</i>
50:50	50.34%	44.71%	50.34%	47.1%
60:40	45.38%	38.7%	45.38%	41.56%
70:30	56.67%	52.43%	56.67%	54.34%
80:20	51.67%	46.63%	51.67%	48.94%
90:10	60.0%	58.75%	60.0%	59.33%

Tabel 8 merupakan hasil dari klasifikasi menggunakan metode Multi-Layer Perceptron dengan 10 fitur (penerapan PC).

Tabel 9. Dataset Dengan *Pearson's Correlation* (12 Fitur)

<i>Splitting Data</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Score</i>
50:50	51.01%	47.1%	51.01%	48.77%
60:40	52.1%	46.5%	52.1%	48.51%
70:30	52.22%	46.44%	52.22%	48.53%
80:20	51.67%	49.25%	51.67%	50.37%
90:10	50.0%	48.28%	50.0%	48.48%

Tabel 9 merupakan hasil dari klasifikasi menggunakan metode Multi-Layer Perceptron dengan 12 fitur (penerapan PC).

Tabel 10. Dataset Dengan *SMOTE* Dan *Pearson's Correlation* (3 Fitur)

<i>Splitting Data</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Score</i>
50:50	58.25%	59.26%	58.25%	58.31%
60:40	58.75%	58.86%	58.75%	58.64%
70:30	56.67%	57.53%	56.67%	56.48%
80:20	56.88%	58.43%	56.88%	56.68%
90:10	60.0%	62.83%	60.0%	60.41%

Tabel 4.10 merupakan hasil dari klasifikasi menggunakan metode MLP dengan 3 fitur (penerapan PC) dan dengan penerapan *SMOTE*.

Tabel 11. Dataset Dengan *SMOTE* Dan *Pearson's Correlation* (5 Fitur)

<i>Splitting Data</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Score</i>
50:50	64.0%	64.74%	64.0%	64.12%
60:40	62.81%	64.07%	62.81%	62.77%
70:30	64.58%	64.77%	64.58%	64.52%
80:20	71.88%	71.36%	71.88%	70.79%
90:10	72.5%	71.76%	72.5%	71.65%

Tabel 11 merupakan hasil dari klasifikasi menggunakan metode Multi-Layer Perceptron dengan 5 fitur (penerapan *PC*) dan dengan penerapan *SMOTE* .

Tabel 12. Dataset Dengan *SMOTE* Dan *Pearson's Correlation* (6 Fitur)

<i>Splitting Data</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Score</i>
50:50	69.25%	69.08%	69.25%	69.02%
60:40	71.88%	71.81%	71.88%	71.47%
70:30	73.33%	73.76%	73.33%	73.4%
80:20	75.62%	76.18%	75.62%	75.55%
90:10	82.5%	83.68%	82.5%	82.85%

Tabel 12 merupakan hasil dari klasifikasi menggunakan metode Multi-Layer Perceptron dengan 3 fitur (penerapan *PC*) dan dengan penerapan *SMOTE* .

Tabel 13. Dataset Dengan *SMOTE* Dan *Pearson's Correlation* (7 Fitur)

<i>Splitting Data</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Score</i>
50:50	72.5%	72.26%	72.5%	72.05%
60:40	77.5%	77.23%	77.5%	77.12%
70:30	81.25%	81.04%	81.25%	81.1%
80:20	83.12%	83.15%	83.12%	82.98%
90:10	86.25%	87.14%	86.25%	86.38%

Tabel 13 merupakan hasil dari klasifikasi menggunakan metode Multi-Layer Perceptron dengan 7 fitur (penerapan *PC*) dan dengan penerapan *SMOTE* .

Tabel 14. Dataset Dengan *SMOTE* Dan *Pearson's Correlation* (10 Fitur)

<i>Splitting Data</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Score</i>
50:50	78.0%	78.44%	78.0%	77.89%
60:40	82.81%	82.75%	82.81%	82.69%
70:30	84.17%	84.19%	84.17%	84.11%
80:20	87.5%	88.03%	87.5%	87.48%
90:10	91.25%	92.07%	91.25%	91.26%

Tabel 14 merupakan hasil dari klasifikasi menggunakan metode Multi-Layer Perceptron dengan 10 fitur (penerapan *PC*) dan dengan penerapan *SMOTE* .

Tabel 15. Dataset Dengan *SMOTE* Dan *Pearson's Correlation* (12 Fitur)

<i>Splitting Data</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Score</i>
50:50	83.0%	83.33%	83.0%	82.99%
60:40	87.19%	87.63%	87.19%	87.05%
70:30	87.08%	87.53%	87.08%	86.98%
80:20	90.62%	90.71%	90.62%	90.59%
90:10	92.5%	92.68%	92.5%	92.47%

Tabel 15 merupakan hasil dari klasifikasi menggunakan metode Multi-Layer Perceptron dengan 12 fitur (penerapan PC dan dengan penerapan SMOTE).

Dan berikut perbandingan hasil yang didapat pada penelitian ini (*MLP+SMOTE*, *MLP+Pearson's Correlation*, *MLP+ SMOTE+ Pearson's Correlation*) dengan penelitian sebelumnya dengan dataset yang sama (klasifikasi multi-class) :

Tabel 16. Perbandingan Hasil Dengan Penelitian Sebelumnya

Author	Year	Deskripsi/ Model (Used)	Accuracy
[7]	2019	MLP	68.86%
		SVM	59.01%
		RF + CDTL	89.30%
[12]	2020	RF	76.70%
		SVM	69.70%
		LM	63.90%
		DT	62,70%
		LR	55.26%
[13]	2021	Multinomial NB	53.95%
		Gaussian NB	38.16%
		Bernoulli NB	51.32%
		Linear SVC	53.95%
		DT	52.63%
		RF	55.26%
		KNN	57.9%
[14]	2021	DT + Feature Selection	67%
		KNN	48%
		LR	62%
		SVM	60%
		C5.0	54%
		RF	66%
[15]	2022	LR + SMOTE	62.67%
		SGD + SMOTE	63,20%
		ET + SMOTE	84.53%
		XGBoost + SMOTE	84.26%
		KNN + SMOTE	68
		SVM + SMOTE	84
This Study	2023	MLP + Pearson's Correlation	63.33%
		MLP + SMOTE + Pearson's Correlation	92.5%
		MLP + SMOTE	97.5%

4. KESIMPULAN

Berdasarkan hasil dan pembahasan pada penelitian klasifikasi multi-class pada penyakit jantung dengan SMOTE dan *pearson's correlation* menggunakan MLP, dapat diambil kesimpulan yaitu:

1. Proses klasifikasi yang dilakukan berupa melakukan *cleaning* dan standarisasi pada data, penerapan SMOTE untuk menyeimbangkan data, penerapan *Pearson's Correlation* untuk memilih fitur dengan korelasi yang tinggi, melakukan *splitting data* dan melakukan klasifikasi menggunakan MLP dengan 3 *hidden layer* dimana setiap *hidden layer* memiliki 50 *neurons* atau *nodes*, learning rate sebesar 0.0016 dan jumlah *iterasi maksimal* adalah 1000 *iterasi*.
2. *Imbalance data* menjadi faktor rendahnya akurasi pada klasifikasi *multi-class* serta seluruh fitur pada dataset ini dibutuhkan pada proses klasifikasi. Hal ini dibuktikan dengan klasifikasi menggunakan *MLP+Pearson's Correlation* yang hanya mendapatkan akurasi tertinggi sebesar 63.33%. Akurasi tertinggi dengan penggabungan metode *MLP+SMOTE+Pearson's*

Correlation sebesar 92.5%. Serta akurasi tertinggi pada penelitian ini didapat menggunakan metode *MLP+SMOTE* yaitu sebesar 97.5% dan hasil tersebut sesuai dengan hasil evaluasi menggunakan *confussion matrix*.

DAFTAR PUSTAKA

- [1] R. Purbaningtyas, "Analisa Denyut Jantung Menggunakan Aplikasi Mobile Self Integrated BioInformatics System," *Jurnal Teknologi Informasi dan Terapan*, vol. 6, no. 2, pp. 67–74, Dec. 2019, doi: 10.25047/jtit.v6i2.113.
- [2] D. Selvia and A. Vradinatika, "Fungsi Tomat Sebagai Anti Aterosklerosis Dalam Pencegahan Penyakit Jantung Koroner," *Jurnal Kesehatan Pena Medika*, vol. 10, no. 1, 2020.
- [3] A. Nurmasani and Y. Pristyanto, "Algoritme Stacking Untuk Klasifikasi Penyakit Jantung Pada Dataset Imbalanced Class," *Jurnal Pseudocode*, vol. 8, no. 1, pp. 21–26, 2021, [Online]. Available: www.ejournal.unib.ac.id/index.php/pseudocode
- [4] Y. Dwi Hastuti, E. Dwi Mulyani, P. Sarjana Keperawatan, D. Ilmu Keperawatan, F. Kedokteran, and U. Diponegoro, "Kecemasan Pasien Dengan Penyakit Jantung Koroner Paska Percutaneous Coronary Intervention," *Jurnal Perawat Indonesia*, vol. 3, no. 3, pp. 167–174, 2019.
- [5] D. Rahayu *et al.*, "Deteksi Dini Penyakit Tidak Menular Pada Lansia," *Jurnal Peduli Masyarakat*, vol. 3, no. 1, pp. 91–96, 2021.
- [6] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *Jurnal Media Informatika Budidarma*, vol. 4, no. 2, pp. 437–444, Apr. 2020, doi: 10.30865/mib.v4i2.2080.
- [7] M. Nahiduzzaman, Md. J. Nayeem, Md. T. Ahmed, and Md. S. U. Zaman, "Prediction of Heart Disease Using Multi-Layer Perceptron Neural Network and Support Vector Machine," *International Conference on Electrical Information and Communication Technology (EICT)*, 2019.
- [8] I. Elujide, S. G. Fashoto, B. Fashoto, E. Mbunge, S. O. Folorunso, and J. O. Olamijuwon, "Application of deep and machine learning techniques for multi-label classification performance on psychotic disorder diseases," *Inform Med Unlocked*, vol. 23, Jan. 2021, doi: 10.1016/j.imu.2021.100545.
- [9] R. Vohra, A. Hussain, A. K. Dudyala, J. Pahareeya, and W. Khan, "Multi-class classification algorithms for the diagnosis of anemia in an outpatient clinical setting," *PLoS One*, vol. 17, no. 7 July, pp. 1–18, Jul. 2022, doi: 10.1371/journal.pone.0269685.
- [10] D. Muallafah, W. Fadila, and R. Firdaus, "Teknik SMOTE untuk Mengatasi Imbalance Data pada Deteksi Penyakit Stroke Menggunakan Algoritma Random Forest," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 2, pp. 107–113, Aug. 2022, doi: 10.37859/coscitech.v3i2.3912.
- [11] H. Mukhtar, R. Muhammad, T. Reny Medikawati, and Yoze Rizki, "Peramalan Kedatangan Wisatawan Mancanegara Ke Indonesia Menurut Kebangsaan Perbulannya Menggunakan Metode Multilayer Perceptron," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 2, no. 2, pp. 113–119, Dec. 2021, doi: 10.37859/coscitech.v2i2.3324.
- [12] G. Magesh and P. Swarnalatha, "Optimal feature selection through a cluster-based DT learning (CDTL) in heart disease prediction," *Evol Intell*, vol. 14, no. 2, pp. 583–593, Jun. 2020, doi: 10.1007/s12065-019-00336-0.
- [13] P. S. Patil, "Automated Heart Disease Recogniton System," *International Journal of Research in Engineering and Science (IJRES) ISSN*, vol. 9, no. 8, pp. 18–23, 2021, [Online]. Available: www.ijres.org18
- [14] S. Subhash Mandavkar, K. J. Somaiya, A. Sukhdev Jagdale, P. Pandurang Shinde, and V. Pinjarkar, "Heart diseases prediction system using classification and genetic algorithm," *International Conference on Advances in Science & Technology*, 2021, [Online]. Available: <https://ssrn.com/abstract=3866565>
- [15] A. Abdellatif, H. Abdellatef, J. Kanesan, C. O. Chow, J. H. Chuah, and H. M. Ghenni, "An Effective Heart Disease Detection and Severity Level Classification Model Using Machine Learning and Hyperparameter Optimization Methods," *IEEE Access*, vol. 10, pp. 79974–79985, 2022, doi: 10.1109/ACCESS.2022.3191669.