Hepatitis Disease Classification Analysis Using the Support Vector Machine (SVM) Method

Amanda Novia Syaputri[a1], Rahmatul Munawardah[a2], Ulysa Siregar[b3], Yesi Dwi Pratiwi[b4], Evta Indra[c5]

[a]Faculty of Technology and Computer Science, Universitas Prima Indonesia, Indonesia,
e-mail: [1]mandanovia5@gmail.com, [2]rahmamunawardah2003@gmail.com,
[3]ulysa.siregar02@gmail.com, [4]yesidwipratiwi48@gmail.com, [5]evtaindra@unprimdn.ac.id

Abstrak

Penyakit hepatitis, peradangan hati yang disebabkan oleh konsumsi alkohol berlebih maupun gangguan autoimun, menjadi permasalahan kesehatan global dengan peningkatan jumlah kematian. Berdasarkan Laporan Global Hepatitis 2024 WHO, hepatitis menjadi penyebab kematian menular tertinggi kedua, dengan 1,3 juta kasus pada 2022. Dari jumlah tersebut, 83% disebabkan oleh hepatitis B dan 17% oleh hepatitis C. Deteksi dini dan klasifikasi yang akurat sangat penting untuk meningkatkan efektivitas pengobatan. Penelitian ini menerapkan metode Support Vector Machine (SVM) untuk mengklasifikasikan pasien hepatitis. Model SVM yang diterapkan pada dataset penderita hepatitis, menghasilkan akurasi tinggi sebesar 90,50% setelah melalui 50 iterasi. Meskipun demikian, terdapat 8 kasus False Negative (pasien hidup diprediksi meninggal) dan 11 kasus False Positive (pasien meinggal diprediksi hidup). Disarankan untuk eksplorasi parameter SVM seperti kernel, nilai C, dan gamma. Serta mempertimbangkan metode ensemble seperti Bagging atau Boosting untuk meningkatkan akurasi dan mengurangi variansi model, khususnya pada dataset yang tidak seimbang atau kompleks.

Kata kunci: *Hepatitis, Klasifikasi, Support Vector Machine, Penyakit*

Abstract

Hepatitis, an inflammation of the liver caused by excessive alcohol consumption and autoimmune disorders, is a global health problem with an increasing number of deaths. Based on the WHO Global Hepatitis Report 2024, hepatitis is the second highest cause of infectious death, with 1.3 million cases in 2022. Of these, 83% were caused by hepatitis B and 17% by hepatitis C. Early detection and accurate classification are essential to improve treatment effectiveness. This study applied the Support Vector Machine (SVM) method to classify hepatitis patients. The SVM model applied to the hepatitis dataset, produced a high accuracy of 90.50% after going through 50 iterations. However, there were 8 cases of False Negative (living patients predicted to die) and 11 cases of False Positives (patients who died predicted to live). It is recommended for exploration of SVM parameters such as kernel, C value, and gamma. As well as considering ensemble methods such as Bagging or Boosting to improve accuracy and reduce model variance, especially on unbalanced or complex datasets.

Keywords: *Hepatitis, Classification, Support Vector Machine, Diseases*

1. Introduction

Hepatitis is a disease characterized by inflammation of the liver and can be triggered by various causes, such as excessive alcohol consumption and autoimmune disorders [1][2][3]. The diagnosis process is usually carried out through a number of examinations, especially through blood tests to detect the presence of the virus [4]. There are five main types of hepatitis viruses, namely hepatitis A, B, C, D, and E (VHA to VHE) [5][6]. People with hepatitis generally experience symptoms such as fever, nausea, diarrhea, weakness, and pain in the right abdominal area. In some patients, symptoms such as yellowing of the eyes and skin are

encountered and flu symptoms may appear. However, in children the symptoms of hepatitis are difficult to detect [7]. Hepatitis can be prevented by getting vaccinated against hepatitis and living a clean and healthy lifestyle as recommended by governments and global health institutions [8]

Transmission of hepatitis A to E disease can spread due to unclean and unhealthy lifestyle [9] Therefore, the implementation of a clean lifestyle such as maintaining food hygiene, the use of sterile medical equipment, and vaccination is an important step to avoid transmission.

Based on the 2024 Global Hepatitis Report released by WHO, hepatitis is now the infectious disease with the second highest mortality rate globally. It was recorded that the number of deaths due to the hepatitis virus increased from 1.1 million in 2019 to 1.3 million cases in 2022 [10]. This fact shows that hepatitis is a global health problem because it has the potential to develop into a more serious condition if not detected and treated early [11].

Classification problems are often encountered in daily life, for example in the process of diagnosing diseases in the medical world. Classification is a prediction method that produces an output in the form of discrete values, with the aim of building a Decision function f(x) that can predict data with a high level of accuracy [12]. Classification is an important process in the medical world, especially in identifying and diagnosing diseases quickly and accurately.

Therefore, effective methods are needed to classify hepatitis disease in making more informed medical decisions. Technological developments in the field of artificial intelligence have opened up new opportunities in the medical world, especially in disease analysis. A wide variety of machine learning models are available to understand and classify data across different fields [13]. The methods that are widely used in the artificial intelligence-based disease classification approach are the Machine Learning method such as Random forest, Decision Tree, Naïve Bayes, K-Nearest Neighbor, and Support Vector Machine (SVM).

Therefore, in this study, we applied the SVM method for the analysis of hepatitis disease classification in order to improve the accuracy of diagnosis, thereby assisting medical personnel in decision-making [14]. This research is expected to contribute to the health sector through the development of a more accurate and efficient diagnosis system, as well as being able to compare the SVM method with other methods.

2. Research Method

2.1 Data Used

The data used are 20 variables that will be used to classify hepatitis using the SVM method, namely demographic data (age and sex), medical symptoms (steroids, antivirals, fatigue, malaise, anorexia, liver_big, liver_firm, spleen_palpaple, spiders, ascites, varices), laboratory test results (bilirubin, alk_phosphate, sgot, albumin, protime), and classification determinants (histology and class).

2.2 Data Processing

In processing data for the classification of hepatitis disease in this study, a systematic workflow is needed so that the data used can produce optimal model accuracy. The stages of the data processing flow used in this study are as follows:

Figure 1 Data Processing Flow

2.3 Data Acquisition

Data collection is an essential initial stage in research where researchers access relevant data for analysis. In this study, data were obtained from open source [15] and included 155 data field entries with 20 attributes, which contained information related to patients diagnosed with hepatitis. The appearance of the data structure can be seen in the image below.

```
df = pd.read_csv('/content/drive/MyDrive/Dataset/Hepatitis/Dataset Hepatitis.csv',delimiter=";")
df.head(1000)
```

	age	sex	steroid	antivirals	fatigue	malaise	anorexia	liver_big	liver_firm	spleen_palpable	spiders	ascites	varices	bilirubin	alk_phosphate	sgot
0	30	male	False	False	False	False	False	False	False	False	False	False	False	1.0	85.0	18.0
1	50	female	False	False	True	False	False	False	False	False	False	False	False	9.0	135.0	42.0
2	78	female	True	False	True	False	False	True	False	False	False	False	False	7.0	96.0	32.0
3	31	female	True	True	False	False	False	True	False	False	False	False	False	7.0	46.0	52.0
4	34	female	True	False	False	False	False	True	False	False	False	False	False	1.0	NaN	200.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
150	46	female	True	False	True	True	True	True	False	False	True	True	True	76.0	NaN	242.0
151	44	female	True	False	True	False	False	True	True	False	False	False	False	9.0	126.0	142.0
152	61	female	False	False	True	True	False	False	True	False	True	False	False	8.0	75.0	20.0
153	53	male	False	False	True	False	False	True	False	True	True	False	True	15.0	81.0	19.0
154	43	female	True	False	True	False	False	True	False	True	True	True	False	12.0	100.0	19.0

155 rows × 20 columns

Figure 2 Hepatitis Classification Dataset

2.4 Preprocessing Data

The data that has been obtained is then processed through the preprocessing stage, which is the process of preparing data before being used in the development and training of machine learning models. This stage aims to improve the quality of the data so that the model can work optimally. The important steps in this stage are missing values, converting categorical data into numerical forms, viewing NaN data and replacing it with Mean so that it can be recognized by machine learning algorithms. The preprocessing process is carried out through several stages as follows:

a. Missing Values

At this stage, the dataset is checked for each instance of missing data. After performing the missing data analysis, it was determined that no data was lost. So, the process proceeds to the next stage, which involves the removal of duplicate entries. The missing values can be seen in the following image:

```
df.isna().sum()
```

	0		0
age	0	spiders	0
sex	0	ascites	0
steroid	0	varices	0
antivirals	0	bilirubin	0
fatigue	0	alk_phosphate	0
malaise	0	sgot	0
anorexia	0	albumin	0
liver_big	0	protime	0
liver_firm	0	histology	0
spleen_palpable	0	class	0

Figure 3 Missing Handling

b. Converting Categorical Data to Numerical Data

This process is carried out by utilizing the LabelEncoder from the scikit-learn library with the aim of converting categorical data, such as text or true/false values, into numerical data that can be understood by machine learning algorithms.

age	sex	steroid	antivirals	fatigue	malaise	anorexia	liver_big	liver_firm	spleen_palpable	spiders	ascites	varices	bilirubin	alk_phosphate	sgot	albumin	protime	histology	class
30	1	0	0	0	0	0	0	0	0	0	0	0	1.0	85.0	18.0	4.0	NaN	0	1
50	0	0	0	1	0	0	0	0	0	0	0	0	9.0	135.0	42.0	35.0	NaN	0	1
78	0	1	0	1	0	0	1	0	0	0	0	0	7.0	96.0	32.0	4.0	NaN	0	1
31	0	1	1	0	0	0	1	0	0	0	0	0	7.0	46.0	52.0	4.0	80.0	0	1
34	0	1	0	0	0	0	1	0	0	0	0	0	1.0	NaN	200.0	4.0	NaN	0	1
34	0	1	0	0	0	0	1	0	0	0	0	0	9.0	95.0	28.0	4.0	75.0	0	1
51	0	0	0	1	0	1	1	0	1	1	0	0	NaN	NaN	NaN	NaN	NaN	0	0
23	0	1	0	0	0	0	1	0	0	0	0	0	1.0	NaN	NaN	NaN	NaN	0	1
39	0	1	0	1	0	0	1	1	0	0	0	0	7.0	NaN	48.0	44.0	NaN	0	1
30	0	1	0	0	0	0	1	0	0	0	0	0	1.0	NaN	120.0	39.0	NaN	0	1
39	0	0	1	0	0	0	0	1	0	0	0	0	13.0	78.0	30.0	44.0	85.0	0	1
32	0	1	1	1	0	0	1	1	0	1	0	0	1.0	59.0	249.0	37.0	54.0	0	1
41	0	1	1	1	0	0	1	1	0	0	0	0	9.0	81.0	60.0	39.0	52.0	0	1
30	0	1	0	1	0	0	1	1	0	0	0	0	22.0	57.0	144.0	49.0	78.0	0	1
47	0	0	1	0	0	0	1	0	0	0	0	0	NaN	NaN	60.0	NaN	NaN	0	1
38	0	0	0	1	1	1	1	0	0	0	1	0	2.0	72.0	89.0	29.0	46.0	0	1
66	0	1	0	1	0	0	1	0	0	0	0	0	12.0	102.0	53.0	43.0	NaN	0	1
40	0	0	0	1	0	0	1	1	0	0	0	0	6.0	62.0	166.0	4.0	63.0	0	1
38	0	1	0	0	0	0	1	0	0	0	0	0	7.0	53.0	42.0	41.0	85.0	1	1
38	0	0	1	0	0	0	0	1	0	0	0	0	7.0	70.0	28.0	42.0	62.0	0	1
22	1	1	1	1	0	0	1	0	0	0	0	0	9.0	48.0	20.0	42.0	64.0	0	1
27	0	1	0	1	1	1	0	1	1	1	0	0	12.0	133.0	98.0	41.0	39.0	0	1
31	0	1	0	0	0	0	1	0	0	0	0	0	1.0	85.0	20.0	4.0	100.0	0	1
42	0	1	0	0	0	0	1	0	0	0	0	0	9.0	60.0	63.0	47.0	47.0	0	1
25	1	0	1	0	0	0	1	0	0	0	0	0	4.0	45.0	18.0	43.0	70.0	0	1

Figure 4 Converting Categorical Data to Numerical Data

c. Viewing NaN Data and Replacing With Mean (Numeric Contuniu)
The purpose of Seeing NaN Data and Replacing with Mean is to clean the data by filling in the missing values in the numerical column using the average value in the data for further analysis:

```
df.fillna(df.mean(), inplace = True)
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 155 entries, 0 to 154
Data columns (total 20 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   age              155 non-null    int64
 1   sex              155 non-null    int64
 2   steroid          155 non-null    int64
 3   antivirals       155 non-null    int64
 4   fatigue          155 non-null    int64
 5   malaise          155 non-null    int64
 6   anorexia         155 non-null    int64
 7   liver_big        155 non-null    int64
 8   liver_firm       155 non-null    int64
 9   spleen_palpable  155 non-null    int64
 10  spiders          155 non-null    int64
 11  ascites          155 non-null    int64
 12  varices          155 non-null    int64
 13  bilirubin        155 non-null    float64
 14  alk_phosphate    155 non-null    float64
 15  sgot             155 non-null    float64
 16  albumin          155 non-null    float64
 17  protime          155 non-null    float64
 18  histology        155 non-null    int64
 19  class            155 non-null    int64
```

Figure 5 Replace with Mean (Numeric Example)

d. Resampling Data
After data cleansing, a significant imbalance was found between class 1 (123 samples) and class 0 (32 samples). This imbalance has the potential to result in a machine learning model that is biased towards the majority class. To address this and improve performance in minority classes, an oversampling technique was applied to balance the data to 500 samples for each class.

2.5 Exploratory Data Analysis

This stage includes the initial stage of exploratory data analysis (EDA), which requires examining the distribution pattern of hepatitis cases. The data exploration carried out is as follows:

a. Correlation Variable
This heatmap is used to show the linear relationships (correlations) between pairs of variables (features) in the dataset. Correlation is expressed in values between -1 to 1.

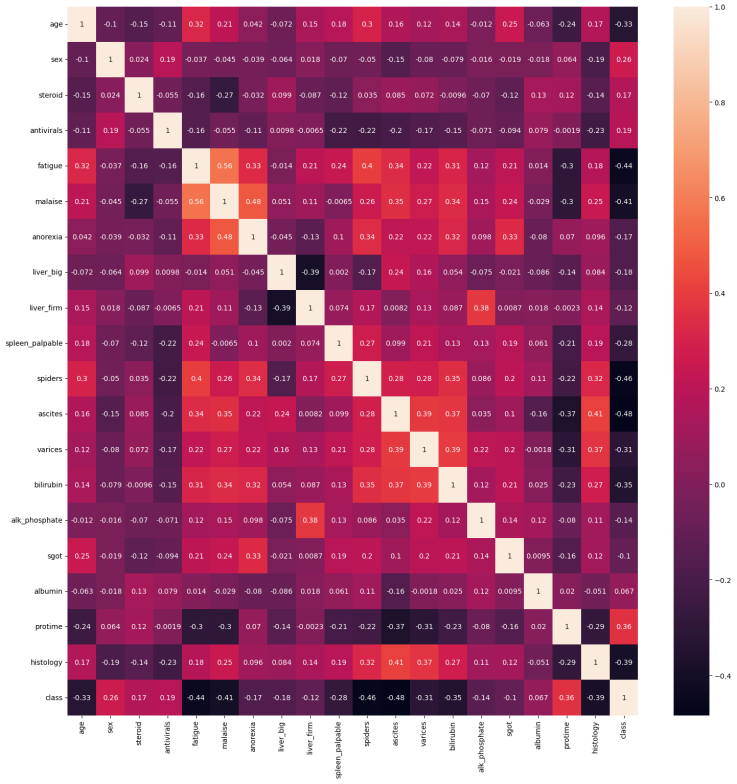

Figure 6 Data Correlation

The first cluster that stands out the most is systemic symptom syndrome, in which fatigue shows a very strong positive correlation with malaise (r = 0.56), reflecting how these two symptoms are manifestations of the same systemic decompensation. This association was then strengthened by the correlation of fatigue with anorexia (r = 0.33) and malaise with anorexia (r = 0.48), forming a consistent triad of systemic symptoms.

b. Age and Class Column Visualization

To understand the data in more detail, a visualization of the distribution of age variables in the hepatitis patient dataset was carried out both in general and based on the classification of the patient's survival status.

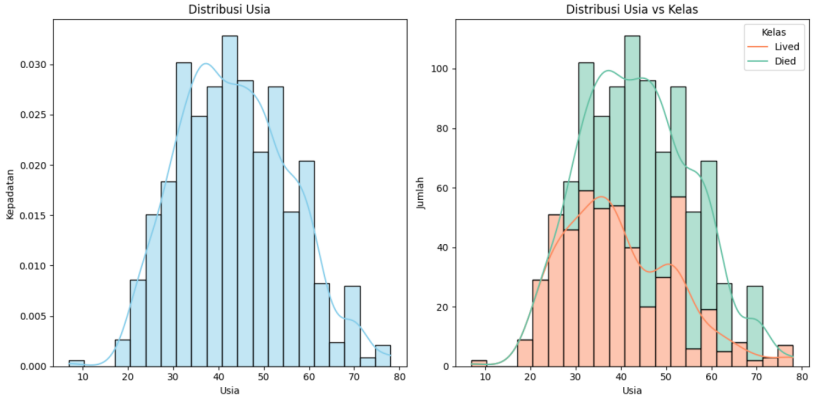

Figure 7 Visualization of Age and Class Columns

The image above shows the age distribution of hepatitis patients. The left graph shows that the majority of patients are 30-4- years old. The right graph compares the age distribution by patient status: orange (survival) is concentrated at 30-40 years of age, while bluish-green (death) is more widespread and tends to increase at age over 40, especially in old age. This indicates age as an important factor that affects the chances of survival.

2.6 Data Splitting

The data sharing process is carried out by separating the dataset into two main parts, namely training data and test data. Training data is used to build and train machine learning models to recognize patterns from available data. Meanwhile, test data serves to evaluate the performance of the trained model, to ensure that the model is able to accurately predict new data that has never been seen before.

a. Creating a Determinant Variable (X)
In making the determinant variable (X) 9 columns are used which can be seen in the image below:

```
X.head(1000)
```

	age	sex	steroid	antivirals	fatigue	malaise	anorexia	liver_big	liver_firm	spleen_palpable	spiders	ascites	varices	bilirubin	alk_phosphate	sgot	
0	34	0	0	0	1	1	0	1	1	1	0	1	1	11.375839	105.325397	85.89404	3
1	58	1	1	0	1	0	0	1	1	0	1	0	0	14.000000	175.000000	55.00000	2
2	36	0	0	0	1	1	1	0	1	0	1	0	1	17.000000	295.000000	60.00000	2
3	42	0	1	0	0	0	0	1	0	0	0	0	0	9.000000	60.000000	63.00000	4
4	35	0	0	0	1	0	0	1	1	1	1	1	0	15.000000	138.000000	58.00000	2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
995	70	0	0	0	1	1	1	1	0	1	1	1	1	17.000000	109.000000	528.00000	2
996	57	0	0	0	1	1	0	1	0	0	1	1	0	46.000000	82.000000	55.00000	3
997	41	0	1	1	1	0	0	1	1	0	0	0	0	9.000000	81.000000	60.00000	3
998	38	0	0	0	0	0	0	1	1	0	0	0	0	4.000000	243.000000	49.00000	3
999	54	0	0	0	1	1	0	1	1	1	0	1	0	39.000000	120.000000	28.00000	3

1000 rows × 19 columns

Figure 8 Determinant Variables (X)

b. Creating a Target Variable (Y)
In creating the target variable (Y) 1 column is used which can be seen in the image below:

	class
0	1
1	1
2	1
3	1
4	0
...	...
995	0
996	0
997	1
998	0
999	0

1000 rows × 1 columns

Figure 9 Target Variable (Y)

c. Data Splitting
The data is divided into 2 parts, namely 80% training data and 20% test data with random state 3 settings using the Skcitlearn library.

3. Literature Study

SVM is known as one of the algorithms in machine learning that is effective in solving classification and regression problems [16], SVM's job is to find the best hyperplane that separates data from two classes with maximum margins [17] SVM has various types of kernels that can be adjusted to the characteristics of the data, including: linear, polynomial, radial base function (RBF), and sigmoid kernel [18]. This research is included in the quantitative category, which uses numerical data to be analyzed statistically to produce objective conclusions [19].

Previous research has been conducted by [20], Comparative Analysis of Hepatitis Disease Prediction Classification using K-Nearest Neighbor, Naïve Bayes and Neural Network Algorithms. This analysis used 20 attributes and was carried out through three experiments. The K-Nearest Neighbor algorithm produced the highest accuracy of 93% with an error rate of 7%, compared to the Neural Network algorithm of 82.97% with an error accuracy rate of 17.03 and Naïve Bayes of 76.92 with an error rate of 23.01%.

By utilizing proper data processing techniques, such as preprocessing, oversampling, and parameter tuning such as kernel and C, SVM performance can be significantly improved. Therefore, SVM is seen as a very potential method in the development of decision support systems for rapid and accurate diagnosis of hepatitis.

4. Result and Discussion

Data processing is carried out using the Support Vector Machine method using the python programming language through the Google Collaboratory platform.

4.1 Modelling Support Vector Machine

In the Loop the data is divided into training data (X_train, y_train) and test data (X_tes, y_test) using train_test_split. test_size=0.2 indicates that 20% of the data is used for testing. Random_state=i ensures a different data sharing in each iteration, so the model is trained with diverse data variations. From the Looping experiment, the maximum accuracy value of 90.50% was obtained on the 20th iteration, for the accuracy value obtained from 50 experiments can be seen in the following table:

Table 1 Accuracy Values

Looping 1: Accuracy: 0.86		Looping 26: Accuracy: 0.835
Looping 2: Accuracy: 0.825		Looping 27: Accuracy: 0.885
Looping 3: Accuracy: 0.885		Looping 28: Accuracy: 0.81
Looping 4: Accuracy: 0.885		Looping 29: Accuracy: 0.845
Looping 5: Accuracy: 0.84		Looping 30: Accuracy: 0.845
Looping 6: Accuracy: 0.875		Looping 31: Accuracy: 0.855
Looping 7: Accuracy: 0.865		Looping 32: Accuracy: 0.86
Looping 8: Accuracy: 0.805		Looping 33: Accuracy: 0.845
Looping 9: Accuracy: 0.825		Looping 34: Accuracy: 0.865
Looping 10: Accuracy: 0.835		Looping 35: Accuracy: 0.85
Looping 11: Accuracy: 0.89		Looping 36: Accuracy: 0.825
Looping 12: Accuracy: 0.835		Looping 37: Accuracy: 0.885
Looping 13: Accuracy: 0.885		Looping 38: Accuracy: 0.835
Looping 14: Accuracy: 0.865		Looping 39: Accuracy: 0.825
Looping 15: Accuracy: 0.835		Looping 40: Accuracy: 0.85
Looping 16: Accuracy: 0.875		Looping 41: Accuracy: 0.845
Looping 17: Accuracy: 0.85		Looping 42: Accuracy: 0.835
Looping 18: Accuracy: 0.835		Looping 43: Accuracy: 0.85
Looping 19: Accuracy: 0.825		Looping 44: Accuracy: 0.875

Looping 20: Accuracy: 0.905		Looping 45: Accuracy: 0.85
Looping 21: Accuracy: 0.9		Looping 46: Accuracy: 0.845
Looping 22: Accuracy: 0.865		Looping 47: Accuracy: 0.885
Looping 23: Accuracy: 0.8		Looping 48: Accuracy: 0.85
Looping 24: Accuracy: 0.87		Looping 49: Accuracy: 0.85
Looping 25: Accuracy: 0.855		Looping 50: Accuracy: 0.885

To make it easier to see the accuracy value of each loop, the accuracy value data is visualized using a line diagram that we can see in the image below:

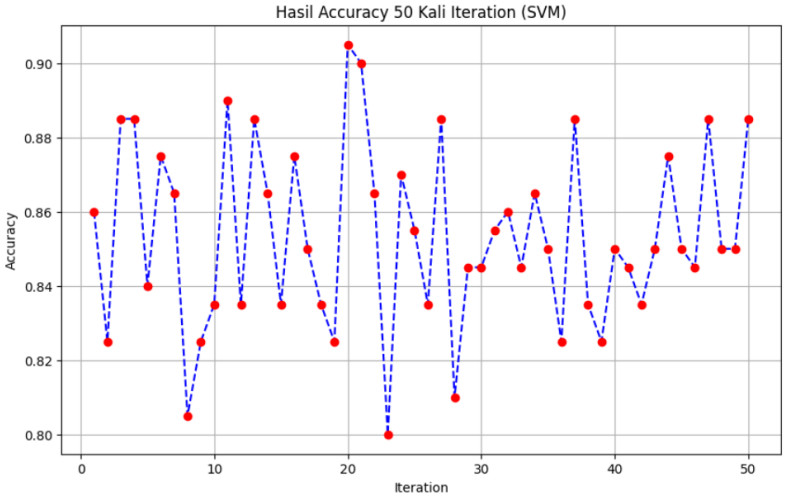

Figure 10 Accuracy Value Line Diagram

4.2 Model Evaluation

The confusion matrix image shown shows the performance evaluation of the Support Vector Machine (SVM) classification model in identifying the condition of hepatitis patients, with two target classes, namely "Lived" and "Died". This matrix presents the number of correct or false predictions made by the model against the test data. From the total predictions, the model managed to classify 98 patients who were really alive (True Positive) and 83 patients who were really dead (True Negative) precisely.

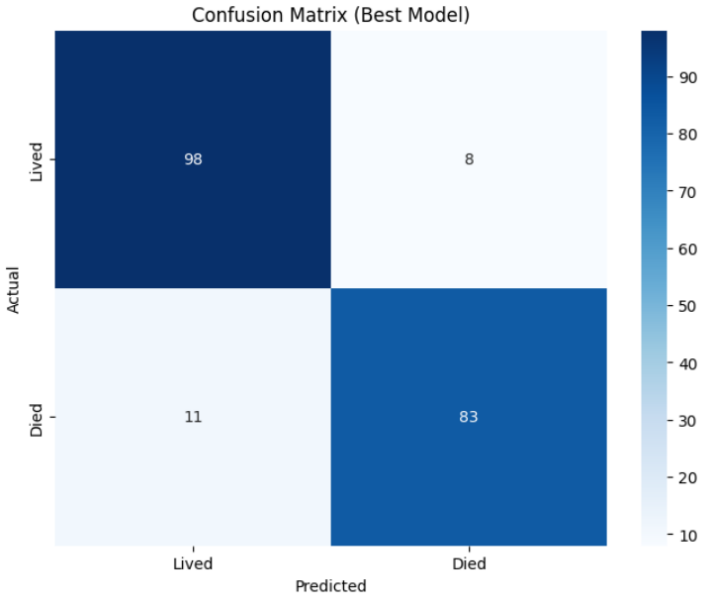

Figure 11 Confusion Matrix algoritma Support Vector Machine

In the picture above, 8 cases where patients who are actually alive are predicted to die (False Negative), and 11 cases of patients who actually die are predicted to be alive (False Negative). These results show that the model has a high degree of accuracy in identifying cases of hepatitis death, but there are still few errors in distinguishing between surviving patients. This visualization provides a clear picture of the distribution of predictive results and is an important tool in understanding the strengths and weaknesses of the SVM model in the context of hepatitis diagnosis.

5. Conclusion

The application of the Support Vector Machine (SVM) method on the dataset of hepatitis patients showed very positive results. The SVM model used was able to achieve a high level of accuracy, which was 90.50% after going through a series of experiments of 50 iterations. This achievement confirms that this model is quite reliable in classifying the condition of hepatitis patients both who survive and die. This optimal performance suggests that the SVM algorithm can be used effectively in data-driven diagnostic processes to support medical decision-making against hepatitis disease.

References

[1] J. E. Aurelia, Z. Rustam, I. Wirasati, S. Hartini, and G. S. Saragih, “Hepatitis classification using support vector machines and random forest,” *IAES Int. J. Artif. Intell.*, vol. 10, no. 2, pp. 446–451, 2021, doi: 10.11591/IJAI.V10.I2.PP446-451.

[2] A. Damayanti and G. Testiana, “PENERAPAN DATA MINING UNTUK PREDIKSI PENYAKIT HEPATITIS C MENGGUNAKAN ALGORITMA NAÏVE BAYES,” *J. Manaj. Inform. Jayakarta*, vol. 3, no. 2, pp. 177–186, 2023, doi: 10.52362/jmijayakarta.v3i2.1098.

[3] B. A. Herlambang, A. G. Prathama, and R. R. Waliyansyah, “Aplikasi Pendeteksi Dini Penyakit Hepatitis Menggunakan Mesin Inferensi Forward Chaining Berbasis Android,” *J. Appl. Sci. Technol.*, vol. 1, no. 02, p. 1, 2021, doi: 10.30659/jast.1.02.1-12.

[4] M. A. G. Hazber, E. M. Senan, and H. S. Alrashidi, “Feature Engineering Methods for Analyzing Blood Samples for Early Diagnosis of Hepatitis Using Machine Learning Approaches,” *C. - Comput. Model. Eng. Sci.*, vol. 142, no. 3, pp. 3229–3254, Mar. 2025, doi: 10.32604/CMES.2025.062302.

[5] M. D. S. Kusuma and I. A. P. Wulandari, “Perilaku Hidup Bersih dan Sehat untuk Pencegahan Penyakit Hepatitis di Panti Asuhan,” *J. Kreat. Pengabdi. Kpd. Masy.*, vol. 7, no. 4, pp. 1856–1866, Apr. 2024, doi: 10.33024/jkpm.v7i4.13312.

[6] D. W. Dede Widiyanto and A.Sidiq Purnomo, “Diagnosa Penyakit Hepatitis pada RSUD

Bunut Menggunakan Sistem Pakar dengan Metode Certainty Factor," *JEKIN - J. Tek. Inform.*, vol. 4, no. 3, pp. 650–662, 2024, doi: 10.58794/jekin.v4i3.893.

[7] A. Hajjah, Y. Nora Marlim, and R. Noratama Putri, "APLIKASI PENDETEKSI PENYAKIT HEPATITIS MENGGUNAKAN METODE NAÏVE BAYES," *JOISIE J. Inf. Syst. Informatics Eng.*, vol. 7, no. 1, pp. 155–164, 2023.

[8] A. W. Feliansyah and E. Purwanto, "Analisis faktor yang berhubungan dengan penyakit hepatitis di Indonesia," *Holistik J. Kesehat.*, vol. 18, no. 9, pp. 1131–1138, Nov. 2024, doi: 10.33024/hjk.v18i9.587.

[9] M. Zulham, D. Saripurna, and M. Z. Siambaton, "Aplikasi Diagnosa Penyakit Hepatitis dengan Menggunakan Metode Teorema Bayes dan Certainty Factor," *Blend Sains J. Tek.*, vol. 2, no. 1, pp. 1–15, Apr. 2023, doi: 10.56211/blendsains.v2i1.241.

[10] "WHO membunyikan peringatan tentang infeksi hepatitis virus yang merenggut 3500 nyawa setiap hari." https://www.who.int/news/item/09-04-2024-who-sounds-alarm-on-viral-hepatitis-infections-claiming-3500-lives-each-day (accessed Feb. 25, 2025).

[11] J. Ilmiah Matematika, S. Sofia Nurwananda, and R. Sulaiman, "APLIKASI HIMPUNAN FUZZY INTUISIONISTIK DALAM DIAGNOSA PENYAKIT HEPATITIS MENGGUNAKAN EXTENDED HAUSDORFF DISTANCE," *Mathunesa J. Ilm. Mat.*, vol. 10, no. 01, pp. 41–49, 2022.

[12] Jatnika Fahmi Idris, Rafid Ramadhani, and Muhammad Malik Mutoffar, "Klasifikasi Penyakit Kanker Paru Menggunakan Perbandingan Algoritma Machine Learning," *J. Media Akad.*, vol. 2, no. 2, pp. 1981–2000, 2024, doi: 10.62281/v2i2.145.

[13] R. Aprilia, "Implementasi Machine Learning Untuk Memprediksi Hepatitis Mortality Menggunakan Algoritma Decision Tree," *OKTAL J. Ilmu Komput. dan Sci.*, vol. 2, no. 4, pp. 1146–1150, 2023, [Online]. Available: https://journal.mediapublikasi.id/index.php/oktal

[14] W. D. Septiani, "OPTIMASI ALGORITMA C4.5 MENGGUNAKAN ALGORITMA GENETIKA UNTUK PREDIKSI PENYAKIT HEPATITIS," *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 59–64, Jul. 2020, doi: 10.33480/inti.v15i1.1421.

[15] "Hepatitis Data." https://www.kaggle.com/datasets/codebreaker619/hepatitis-data/data (accessed May 21, 2025).

[16] T. Sugihartono and R. R. C. Putra, "Penerapan Metode Support Vector Machine Dalam Classifikasi Ulasan Pengguna Aplikasi Mobile Jkn," *SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 7, no. 2, pp. 144–153, 2024, doi: 10.36080/skanika.v7i2.3193.

[17] S. Kenny Riva, S. Casi, and S. Randy Erfa, "Analisis Algoritma Support Vector Machine Dalam Klasifikasi Penyakit Stroke," *e-Proceeding Eng.*, vol. 9, no. 3, p. 922, 2022.

[18] V. RISKASARI YR, I. P. E. N. KENCANA, and I. K. G. SUKARSA, "Klasifikasi Penyakit Sirosis Menggunakan Support Vector Machine," *E-Jurnal Mat.*, vol. 12, no. 2, p. 87, 2023, doi: 10.24843/mtk.2023.v12.i02.p404.

[19] A. F. Djollong, "Teknik Pelaksanaan Penelitian Kuantitatif (Technique of Quantiative Research)," *Istiqra'*, vol. 2, no. 1, pp. 86–100, 2014.

[20] S. Sulastri, K. Hadiono, and M. T. Anwar, "Analisis Perbandingan Klasifikasi Prediksi Penyakit Hepatitis Dengan Menggunakan Algoritma K-Nearest Neighbor, Naïve Bayes Dan Neural Network," *Dinamik*, vol. 24, no. 2, pp. 82–91, 2020, doi: 10.35315/dinamik.v24i2.7867.