

Classification of Primary Biliary Cirrhosis Using Supervised Machine Learning Algorithms

submitted by

Ivy Dey

Class Roll No: 002210503025

Exam Roll: MCA244024

Master of Computer Application

under the supervision of

Prof. (Dr.) Susmita Ghosh (De)

in the

Department of Computer Science and Engineering

Faculty of Engineering and Technology

Jadavpur University

2024

CERTIFICATE OF RECOMMENDATION

To Whom It May Concern

I hereby recommend that the project report entitled “Classification of Primary Biliary Cirrhosis Using Supervised Machine Learning Algorithms” has been carried out by Ivy Dey (Reg. No.: 163617 of 2022-23, Roll No.: 002210503025, Exam Roll: MCA244024), under my guidance and supervision and be accepted in partial fulfilment of the requirement for the degree of Master of COMPUTER APPLICATION in DEPARTMENT of COMPUTER SCIENCE and ENGINEERING, JADAVPUR UNIVERSITY during the academic year 2023-2024.

Prof. (Dr.) Susmita Ghosh (De)
Project Supervisor
Dept. of Computer Science and
Engineering
Jadavpur University

Prof. (Dr.) Nandini Mukhopadhyay
Head of the Department
Dept. of Computer Science & Engineering
Jadavpur University

Prof. Dipak Laha
Dean, Faculty Council
of Engineering & Technology
Jadavpur University, Kolkata-032

CERTIFICATE OF APPROVAL

This is to certify that the project report entitled “**Classification of Primary Biliary Cirrhosis Using Supervised Machine Learning Algorithms**” is a bonafide record of work carried out by IVY DEY, in fulfillment of the requirements for the award of the degree of Master of Computer Application in the Department of Computer Science and Engineering, Jadavpur University. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein but approve the project report only for the purpose for which it has been submitted.

Signature of Examiner 1

Date :

Signature of Examiner 2

Date :

Faculty of Engineering and Technology
Jadavpur University

**DECLARATION OF COMPLIANCE WITH
ACADEMIC ETHICS**

I hereby declare that this project report entitled "Classification of Primary Biliary Cirrhosis Using Supervised Machine Learning Algorithms" is an outcome of my genuine and independent work, as part of my Master of Computer Application program.

All information in this document has been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, all materials and results that are not original to this work have been fully cited and referenced.

Ivy Dey

Registration No.: 163617 of 2022-23

Examination Roll No.: MCA244024

Thesis Title :-Classification of Primary Biliary Cirrhosis Using Supervised Machine Learning Algorithms

Signature with Date

ACKNOWLEDGEMENT

With my most sincere and gratitude, I would like to thank Prof. (Dr.) Susmita Ghosh (De), Department of Computer Science & Engineering, my supervisor, for her overwhelming support throughout the duration of the project. Her motivation always gave me the required inputs and momentum to continue with my work, without which the project work would not have taken its current shape. Her valuable suggestions and numerous discussions have always inspired new ways of thinking. I feel deeply honored that I got this opportunity to work under her.

I would like to express my sincere thanks to all my teachers for providing sound knowledge base and cooperation.

I would like to thank all the faculty members of the Department of Computer Science & Engineering of Jadavpur University for their continuous support.

Last, but not the least, I would like to thank my batch mates for staying by my side when I needed them.

Ivy Dey

Class Roll No.: 002210503025

Examination Roll No.: MCA244024

Department of Computer Science & Engineering
Jadavpur University

Index

<u>Chapters</u>	<u>Page No.</u>
1. INTRODUCTION	01
1.1 SCOPE OF THE REPORT	01
1.2 ORGANIZATION OF THE REPORT	01-02
2. LITERATURE SURVEY	03-04
3. METHODOLOGY	05
3.1 INTRODUCTION	05
3.2 CLASSIFIERS USED	05
3.3 WORKING PRINCIPLE OF KNN	06
3.3.1 HOW IT WORKS?	06
3.3.2 ALGORITHM OF KNN	06-07
3.3.3 FLOWCHART OF KNN	08
3.3.4 ADVANTAGES OF KNN	09
3.3.5 DISADVANTAGES OF KNN	09
3.4 WORKING PRINCIPLE OF SVM	09
3.4.1 HOW IT WORKS?	09
3.4.2 ALGORITHM OF SVM	09-11
3.4.3 FLOWCHART OF SVM	12
3.4.4 ADVANTAGES OF SVM	12-13
3.4.5 DISADVANTAGES OF SVM	13
3.5 WORKING PRINCIPLE OF DECISION TREE	13
3.5.1 HOW IT WORKS?	13
3.5.2 ALGORITHM OF DECISION TREE	14
3.5.3 FLOWCHART OF DECISION TREE	15
3.5.4 ADVANTAGES OF DECISION TREE	16
3.5.5 DISADVANTAGES OF DECISION TREE	16
4. RESULTS AND ANALYSIS	17
4.1 INTRODUCTION	17
4.2 DATASET USED	17-23
4.3 PARAMETERS USED	23
4.4 PERFORMANCE METRICS	24
4.5 PREPROCESSING	25
4.6 RESULTS AND ANALYSIS	25-29
4.7 CLASSIFICATION REPORT	29-38
4.8 CONFUSION MATRIX	
5 CONCLUSION	39
REFERENCES USED	40

LIST OF TABLES	PAGE NO.
RESULTS OF KNN (20 SIMULATIONS)	25
RESULTS OF SVM (20 SIMULATIONS)	25
RESULTS OF Decision Tree (20 SIMULATIONS)	26

LIST OF TABLES

LIST OF FIGURES	PAGE NO.
3.1 FLOWCHART OF KNN Classifier	08
3.2 FLOWCHART OF SVM Classifier	12
3.3 FLOWCHART OF Decision Tree Classifier	15
4.1 CLASSIFICATION REPORT OF KNN FOR TEST SIZE 0.1	26
4.2 CLASSIFICATION REPORT OF KNN FOR TEST SIZE 0.2	27
4.3 CLASSIFICATION REPORT OF KNN FOR TEST SIZE 0.3	27
4.4 CLASSIFICATION REPORT OF DT FOR TEST SIZE 0.1	27
4.5 CLASSIFICATION REPORT OF DT FOR TEST SIZE 0.2	28
4.6 CLASSIFICATION REPORT OF DT FOR TEST SIZE 0.3	28
4.7 CLASSIFICATION REPORT OF SVM FOR TEST SIZE 0.1	28
4.8 CLASSIFICATION REPORT OF SVM FOR TEST SIZE 0.2	29
4.9 CLASSIFICATION REPORT OF SVM FOR TEST SIZE 0.3	29
4.10 CONFUSION MATRIX FOR KNN FOR TEST SIZE 0.1	30
4.11 CONFUSION MATRIX FOR KNN FOR TEST SIZE 0.2	31
4.12 CONFUSION MATRIX FOR KNN FOR TEST SIZE 0.3	32
4.13 CONFUSION MATRIX FOR DT FOR TEST SIZE 0.1	33
4.14 CONFUSION MATRIX FOR DT FOR TEST SIZE 0.2	34
4.15 CONFUSION MATRIX FOR DT FOR TEST SIZE 0.3	35
4.16 CONFUSION MATRIX FOR SVM FOR TEST SIZE 0.1	36
4.17 CONFUSION MATRIX FOR SVM FOR TEST SIZE 0.2	37
4.18 CONFUSION MATRIX FOR SVM FOR TEST SIZE 0.3	38

LIST OF FIGURES

CHAPTER 1 : INTRODUCTION

Primary Biliary Cirrhosis (PBC) is a chronic autoimmune liver disease characterized by the gradual destruction of the bile ducts, leading to liver damage and potentially liver failure. Accurate survey of PBC patients is crucial for effective treatment and outlining the factors which can lead to worsening of patient's condition. In recent years, advancements in machine learning have provided promising tools for the medical field, offering novel methods for disease classification and prediction.

This project explores the classification of Primary Biliary Cirrhosis using supervised machine learning algorithms. [1] By leveraging various data-driven techniques, we aim to develop a robust and reliable model that can assist in the detection of status of patient and accurate classification of PBC. Our approach involves analyzing clinical and biochemical data, applying feature selection methods, and employing different supervised learning algorithms to identify patterns [1] and correlations that may not be apparent through traditional diagnostic method.

1.1 SCOPE OF THE REPORT:

This report deploys the application of various machine learning algorithms on the provided dataset to detect the status of patients suffering from Primary Biliary Cirrhosis. The scope of this report encompasses data collection, pre- processing, model training, and evaluation and the implementation of parameter tuning methods [2] to find the best parameter which suits the model. The report also details the development and reckoning the effectiveness of different machine learning models applied on the dataset. Along with that it compares the performance of different models trained on the provided dataset.

1.2 ORGANIZATION OF REPORT:

Chapter 1 contains Introduction to the project, followed by Scope of the Report and the Organization of the report.

Whereas, Chapter 2 contains Literature Survey where a basic survey has been done about the booming concern about Primary Biliary Cirrhosis among different age groups. This section comprises of a comprehensive study to predict who all are more likely to get attacked by the autoimmune disease. Along with that, it also consists a brief overview of different types of learnings and a sneak peak upon the different types of algorithms we have used in the project.

While Chapter 3 contains Methodology used, working principle, algorithm and flowchart for all the three algorithms we have used which are KNN, SVM, and Decision Tree.

While Chapter 4 contains Analysis of Results, Datasets used, Parameters considered for parameter tuning, different performance metrics, steps for data pre-processing, accuracy calculation , plotting of confusion matrix and accuracy curve.

Finally Chapter 5 consists of Conclusion, Future Scope of the project, Comparison of our results with recent seminal research works on the same dataset and references and citations used

CHAPTER 2: LITERATURE SURVEY

Primary biliary cirrhosis (PBC) is an autoimmune disease where the bile ducts in the liver are slowly destroyed, leading to cirrhosis. It primarily affects women, with a prevalence rate of ninety percent, typically occurring between the ages of forty and sixty [3]. Symptoms include pain in the right upper quadrant, fatigue, nausea, and loss of appetite, progressing to more severe symptoms such as abdominal swelling, jaundice, and mental confusion as the disease advances.

Supervised learning algorithms can play a crucial role in the classification of PBC stages. By utilizing datasets containing information about patient demographics, symptoms, and laboratory

tests, these algorithms can learn patterns and relationships within the data to accurately classify patients into different stages of PBC. Algorithms such as decision trees, k-nearest neighbors, support vector machines can be trained on labeled data to predict the stage of PBC based on various input features.

For instance, a decision tree algorithm could learn to classify patients into different stages based on a combination of symptoms and laboratory test results. By analyzing a large dataset of patients with known PBC stages, the algorithm can identify which combination of features best predicts each stage of the disease. Similarly, support vector machine models can estimate the probability of a patient being in a particular stage of PBC based on their input features.

Supervised learning algorithms offer a data-driven approach to classify PBC stages, providing clinicians with valuable insights for diagnosis and treatment planning. By accurately predicting the stage of PBC, healthcare providers can intervene early to slow disease progression and improve patient outcomes.

In a study P. Aryan [4] used algorithms such as Logistic Regression and Random Forest performed better than the other algorithms. ROC-AUC score for all the models is above 0.50. The best performing model i.e. logistic regression has an accuracy of 0.55 and roc-auc score of 0.74. The second-best performing model is random forest with an accuracy of 0.49 and roc-auc score of 0.72. Models such as gradient boosting machines and naïve Bayes classifier also performed well with an accuracy of 0.50 and 0.35 respectively and a roc-auc score of 0.68 for both.

Somaya Hashem [6] the research makes use of several Machine Learning techniques, such as Decision Tree, Genetic Algorithm, Particle Swarm Optimization, and Multi-linear Regression. The accuracy and area under the ROC curve (AUROC) are used to evaluate the models. Results AUROC scores range from 0.73 to 0.76, and their levels of accuracy range from 66.3% to 84.4%.

Literature study shows the wide applicability of machine learning algorithms in liver disease diagnosis. Artificial neural networks (ANN) with supervised and unsupervised learning structures have enhanced the results.

CHAPTER 3: METHODOLOGY

3.1 INTRODUCTION

Supervised learning, also known as supervised machine learning, is a subcategory of machine learning and artificial intelligence. This subcategory is defined by its use of labelled data sets to train algorithms that to classify data or predict outcomes accurately.

As input data is fed into the model, it adjusts its weights until the model has been fitted appropriately, which occurs as part of the cross validation [7] process. Consequently, the accuracy is calculated, followed by calculation of precision, recall, etc.

3.2 CLASSIFIERS USED

3.2.1

Support Vector Machine

SVM works by mapping data to a high-dimensional feature space so that data points can be categorized, even when the data are not otherwise linearly separable. A separator between the categories is found, then the data are transformed in such a way that the separator could be drawn as a hyperplane.

3.2.2

K- Nearest Neighbors:

KNN classifier is a machine learning algorithm used for classification and regression problems. It works by finding the K nearest points in the training dataset and uses their class to predict the class or value of a new data point.[1]

3.2.3

Decision Tree:

Decision tree learning employs a divide and conquer strategy by conducting a greedy search to identify the optimal split points within a tree. This process of splitting is then repeated in a top-down, recursive manner until all, or the majority of records have been classified under specific class labels. The attribute that maximizes information gain is chosen as the splitting criterion for building the decision tree. Information gain is used in both classification and regression decision trees. In classification, entropy is used as a measure of impurity, while in regression, variance is used as a measure of impurity.[1]

3.3 Working Principle of KNN:

3.3.1 How does KNN works?

The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other. KNN which uses proximity to make classifications or predictions about the grouping of an individual data points. The corresponding algorithm is given below and the flowchart is in Fig. 3.1

3.3.2 ALGORITHM:

Step-1: Select the number K of the neighbors

Step-2: Calculate the Euclidean distance of K number of neighbors.

Distance Calculation: When a new input needs to be classified or predicted, the algorithm calculates the distance between this test instance and all instances in the training dataset. Common distance metrics include:

Euclidean distance: the distance between two points (x_1, y_1) and (x_2, y_2) is $d = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]}$.

Manhattan distance: Distance = $|x_1 - x_2| + |y_1 - y_2|$

Minkowski distance: Generalized form of Euclidean and Manhattan distances

□

Step-3: Take the K nearest neighbors as per the calculated Euclidean distance.

Step-4: Among these k neighbors, count the number of the data points in each category.

Step-5: Assign the new data points to that category for which the number of the neighbor is maximum.

3.3 FLOWCHART

SHOWN ON NEXT PAGE

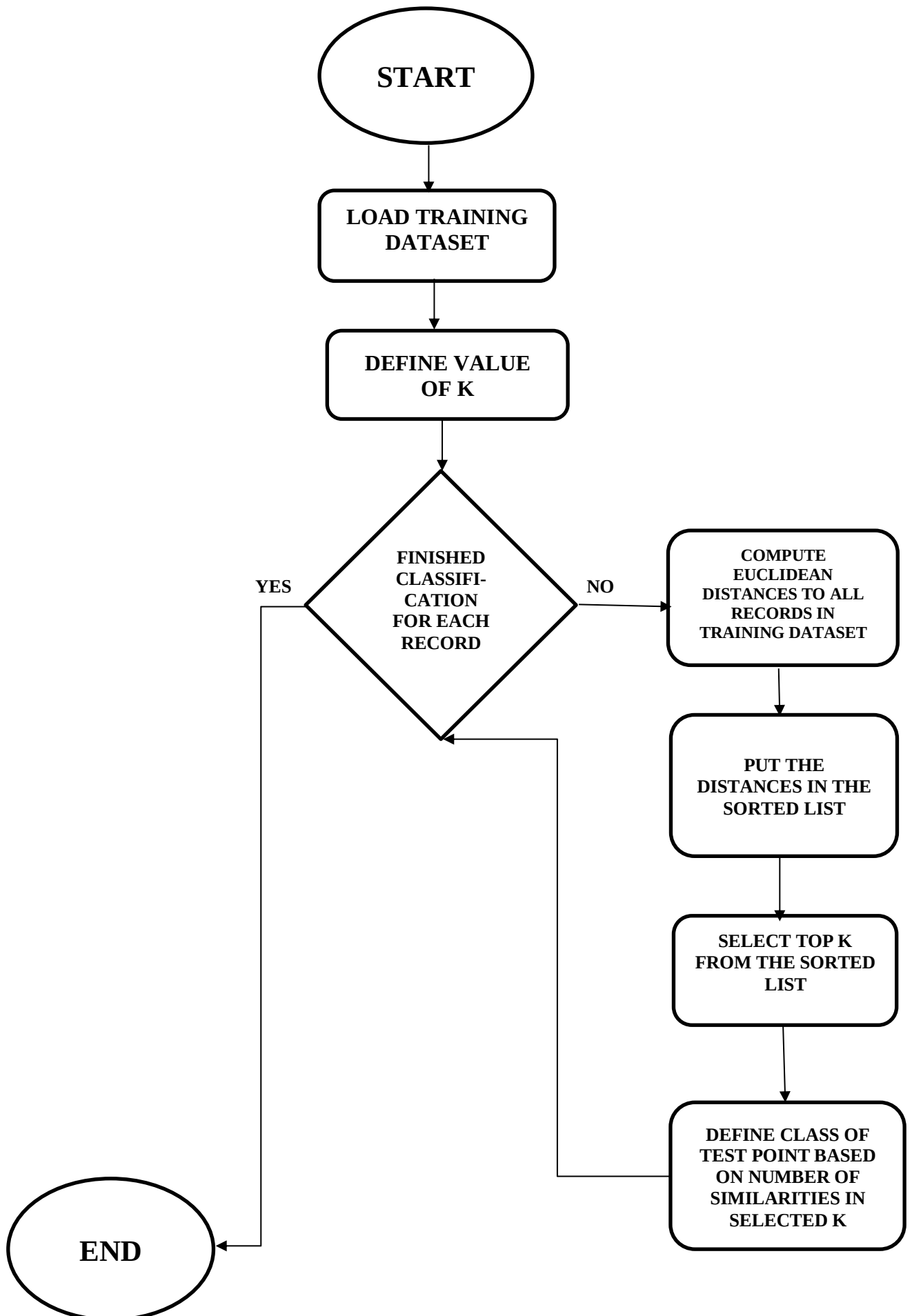


Fig. 3.1: FLOWCHART OF KNN CLASSIFIER

3.4.VANTAGES:

Advantages of KNN Algorithm:

- Simplicity: Easy to understand and implement.
- No Training: No explicit training phase, which can be advantageous for quickly updating the model with new data.
- Robust: It is robust to the noisy training data. It can be more effective if the training data is large.

3.4.2 DISADVANTAGES:

Disadvantages of KNN Algorithm:

- Computationally Intensive: It needs a lot of computing power because it has to calculate the distance between the new data point and all the stored data points every time it makes a prediction.
- Memory Intensive: It requires a lot of data storage since it has to keep all the training data.
- Curse of Dimensionality: This implies the algorithm faces a hard time classifying the data points properly when the dimensionality is too high.

3.4 WORKING PRINCIPLE OF SVM:

3.4.1 HOW DOES SVM WORKS?

SVM works by mapping data to a high-dimensional feature space so that data points can be categorized, even when the data are not otherwise linearly separable. The corresponding algorithm is given below and the flowchart is shown in Fig. 3.2:

3.4.2 SVM ALGORITHM:

1.Prepare the Data:

- Collect the training data consisting of input features and corresponding labels.
- Optionally, normalize or standardize the data to ensure all features contribute equally to the distance calculations

2.Choose the Kernel Function

- Select a kernel function depending on the nature of the data:
- Linear Kernel : For linearly separable data.
- Polynomial Kernel: For data where the relationship between features is polynomial.
- Radial Basis Function (RBF) Kernel/Gaussian Kernel: For data where the decision boundary is non-linear.
- Sigmoid Kernel: Similar to neural networks with a certain type of activation function

3. KERNEL FUNCTION:

1) Linear :

The linear kernel works really well when there are a lot of features, and text classification problems have a lot of features. Linear kernel functions are faster than most of the others and have fewer parameters to optimize.

Here's the function that defines the linear kernel:

$$f(X) = w^T * X + b$$

In this equation, w is the weight vector that you want to minimize, X is the data that you're trying to classify, and b is the linear coefficient estimated from the training data. This equation defines the decision boundary that the SVM returns.

2) Polynomial

The polynomial kernel isn't used in practice very often because it isn't as computationally efficient as other kernels and its predictions aren't as accurate.

Here's the function for a polynomial kernel:

$$f(X1, X2) = (a + X1^T * X2) ^ b$$

This is one of the more simple polynomial kernel equations you can use. $f(X1, X2)$ represents the polynomial decision boundary that will separate our data. $X1$ and $X2$ represent our data.

3) Gaussian Radial Basis Function (RBF)

One of the most powerful and commonly used kernels in SVMs. Usually the choice for non-linear data.

Here's the equation for an RBF kernel:

$$f(X1, X2) = \exp(-\gamma * \|X1 - X2\|^2)$$

In this equation, γ specifies how much a single training point has on the other data points around it. $\|X1 - X2\|$ is the dot product between your features.

4) Sigmoid

More useful in neural networks than in support vector machines, but there are occasional specific use cases.

Here's the function for a sigmoid kernel:

$$f(X, y) = \tanh(\alpha * X^T * y + C)$$

In this function, α is a weight vector and C is an offset value to account for some misclassification of data that can happen.

4) Construct the Decision Boundary:

- The decision function for classification is:

$$-f(x) = \sum_{i=1 \text{ to } n} \alpha_i * y_i * K(x_i, x) + b f(x) \text{ i.e, } i = (1 \text{ to } n) \sum \alpha_i * y_i * K(x_i, x) + b$$

For a new data point x , the class is determined by the sign of $f(x)$

5) Training :

-Split the data into training and validation sets.

-Train the SVM model using the training set.

-Use cross-validation to tune hyperparameters such as the regularization parameter C and kernel parameters (e.g., the width γ for the RBF kernel).

- Evaluate the performance of the model using appropriate metrics such as accuracy, precision, recall, F1 score, or ROC-AUC, depending on the problem.

-Once the model is trained and validated, use it to make predictions on new data points by applying the decision function $f(x)$

3.4.3 Flowchart:

SHOWN ON NEXT PAGE

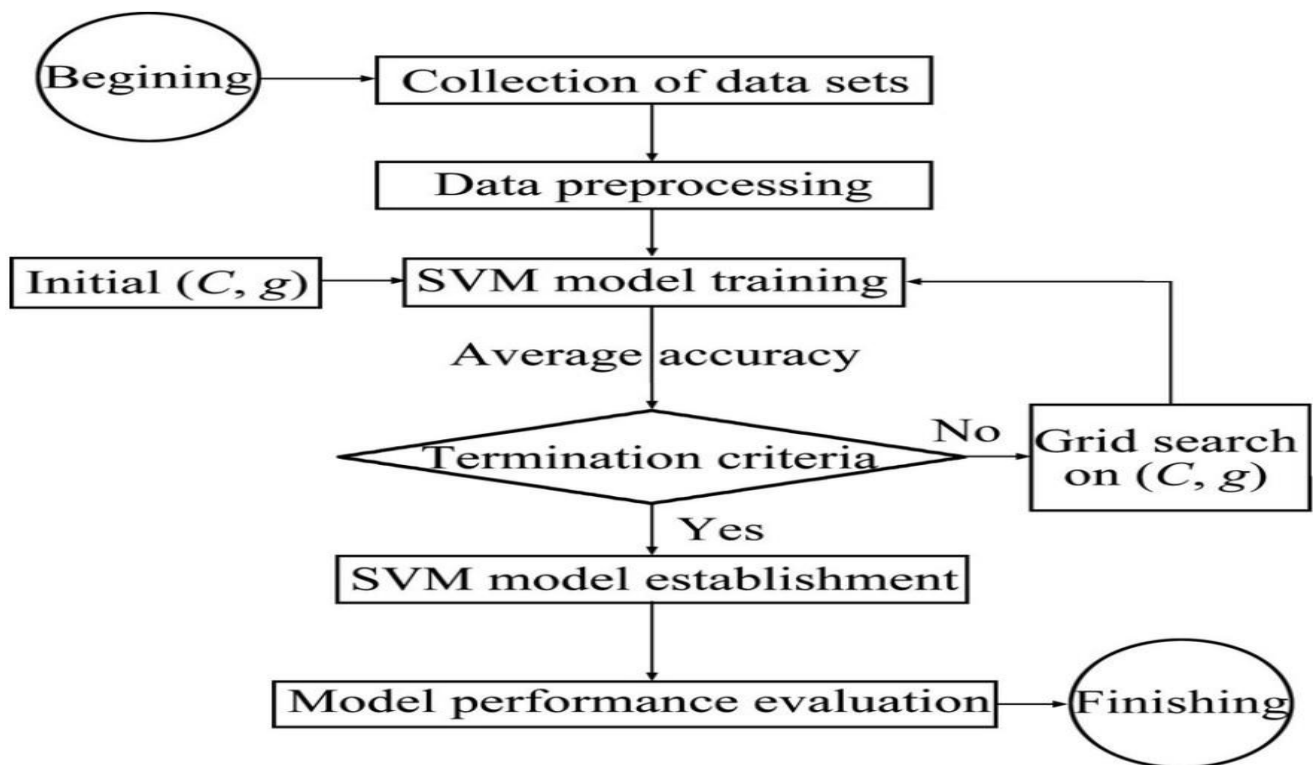


Fig. 3.2 : FLOWCHART OF SVM CLASSIFIER

3.4.4 ADVANTAGES:

Advantages of SVM Algorithm:

1. Effective in High-Dimensional Spaces:

SVM works well in cases where the number of dimensions is greater than the number of samples.

2. Memory Efficient:

SVM uses a subset of training points (support vectors) in the decision function, making it memory efficient.

3. Versatile with Different Kernel Functions:

The ability to use custom kernel functions allows SVM to adapt to different types of data distributions (linear, non-linear, polynomial, radial basis function, etc.).

4. Clear Margin of Separation:

SVM provides a clear margin of separation between classes, which can improve the generalization of the classifier.

5. Robust to Overfitting:

Proper regularization parameters (like C) help prevent overfitting, especially in high-dimensional spaces.

3.4.5 DISADVANTAGES:

Disadvantages of SVM:

1. Computationally expensive: SVMs can be computationally expensive for large datasets, as the algorithm requires solving a quadratic optimization problem.

2. Choice of kernel: The choice of kernel can greatly affect the performance of an SVM, and it can be difficult to determine the best kernel for a given dataset.

3. Sensitivity to the choice of parameters: SVMs can be sensitive to the choice of parameters, such as the regularization parameter, and it can be difficult to determine the optimal parameter values for a given dataset.

4. Not suitable for large datasets with many features: SVMs can be very slow and can consume a lot of memory when the dataset has many features.

3.5 WORKING PRINCIPLE OF DECISION TREE:

3.5.1 How Decision tree works?

Decision tree learning employs a divide and conquer strategy by conducting a greedy search to identify the optimal split points within a tree. This process of splitting is then repeated in a top-down, recursive manner until all, or the majority of records have been classified under specific class labels. The attribute that maximizes information gain is chosen as the splitting criterion for building the decision tree. [1] The corresponding algorithm is shown and the flowchart is depicted below

3.5.2 Algorithm of Decision Tree:

1. Selecting the Best Attribute: The process starts at the root node and involves selecting the best attribute to split the data. This selection is based on a metric such as:

Gini Index: Measures the impurity or purity of a split. Lower Gini indicates a better split.

• Entropy: Used in information gain, it measures the disorder in the dataset. Lower entropy after a split indicates a better split.

• Information Gain: The reduction in entropy after a dataset is split on an attribute. Higher information gain indicates a better attribute.

• Chi-square: Used for categorical data to measure the independence between attributes.

2. Splitting: Based on the best attribute, the data is split into subsets. This process is repeated recursively for each subset.

3. Stopping Criteria: The recursion stops when:

• All data points in a node belong to the same class.

• There are no more attributes to split.

• A pre-defined condition like maximum depth of tree is reached.

4. Pruning: This is an optional step to avoid overfitting. Pruning reduces the size of the tree by removing sections that provide little power in classifying instances.

The range of value Gini Impurity can have is between 0 to 0.5. The lesser the Gini Impurity, the better the split is.

A Gini Impurity of 0 denotes a pure node and 0.5 denotes a most impure node.

The Gini Impurity formula is: $1 - (p_1)^2 - (p_2)^2$, where p_1 and p_2 represent the probabilities of the two classes in a binary classification problem. For multi-classification it ranges from 1 to n .

The range of values Entropy can have is between 0 to 1.

Entropy of 0 denotes a pure node and 1 denotes most impure node (where we have 50–50 split of 'Yes' and 'No').

Use the formula for entropy: $\text{Entropy} = -\sum_{i=1}^c p_i \times \log_{\text{base}_2}(p_i)$

Where p_i is the proportion of data points belonging to class i and c is the number of classes.

3.5.3 FLOWCHART:

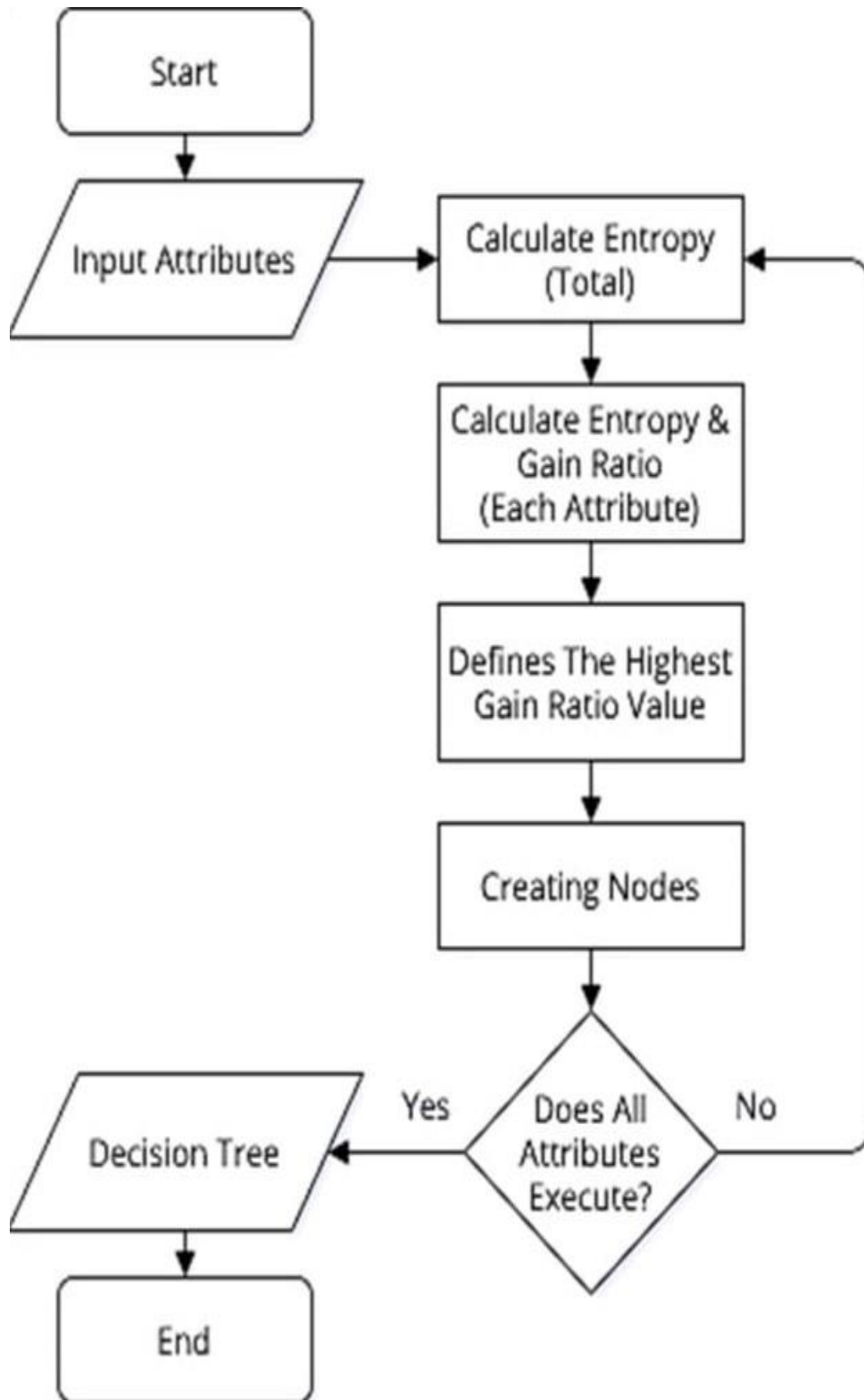


Fig 3.3 : FLOWCHART OF DECISION TREE CLASSIFIER

3.5.4 ADVANTAGES:

1. **Simplicity and Interpretability:** Decision trees are easy to understand and interpret. The visual representation closely mirrors human decision-making processes.
2. **Versatility:** Can be used for both classification and regression
3. **No Need for Feature Scaling:** Decision trees do not require normalization or scaling of the data.
4. **Handles Non-linear Relationships:** Capable of capturing non-linear relationships between features and target variables.

3.5.5 DISADVANTAGES:

1. **Overfitting:** Decision trees can easily overfit the training data, especially if they are deep with many nodes.
2. **Instability:** Small variations in the data can result in a completely different tree being generated.
3. **Bias towards Features with More Levels:** Features with more levels can dominate the tree structure.

CHAPTER 4 : RESULTS AND ANALYSIS

4.1 Introduction:

In this section, we explore the outcomes of our classification model evaluations, complemented by visualizations to better understand their performance. We analyze three different machine learning classifiers: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Decision Tree (DT).[2]

For each classifier, we carefully assess their accuracy scores and F1 scores [2], providing a clear view of their predictive capabilities. Additionally, we delve into their performance details using confusion matrices and classification reports. These metrics reveal the models' proficiency in correctly classifying and distinguishing between various classes.

To make these insights more accessible, we use heat maps to visualize the confusion matrices. These graphical representations offer a clearer picture of the models' accuracy and their tendencies to misclassify certain classes. By combining quantitative metrics with visual aids, we can make more informed decisions about which models are best suited for our dataset and problem.

4.2 DATASET USED –

The data provided is sourced from a Mayo Clinic study on primary biliary cirrhosis (PBC) [7] of the liver carried out from 1974 to 1984. During 1974 to 1984, 424 PBC patients referred to the Mayo Clinic qualified for the randomized placebo-controlled trial testing the drug D-penicillamine. Of these, the initial 312 patients took part in the trial and have mostly comprehensive data. The remaining 112 patients didn't join the clinical trial but agreed to record basic metrics and undergo survival tracking. Six of these patients were soon untraceable after their diagnosis, leaving data for 106 of these individuals in addition to the 312 who were part of the randomized trial.

1. ID: unique identifier
2. N_Days: number of days between registration and the earlier of death, transplantation, or study analysis time in July 1986
3. Status: status of the patient C (censored), CL (censored due to liver tx), or D (death)
4. Drug: type of drug D-penicillamine or placebo
5. Age: age in [days]
6. Sex: M (male) or F (female)
7. Ascites: presence of ascites N (No) or Y (Yes)
8. Hepatomegaly: presence of hepatomegaly N (No) or Y (Yes)
9. Spiders: presence of spiders N (No) or Y (Yes)
10. Edema: presence of edema N (no edema and no diuretic therapy for edema), S (edema present without diuretics, or edema resolved by diuretics), or Y (edema despite diuretic therapy)
11. Bilirubin: serum bilirubin in [mg/dl]
12. Cholesterol: serum cholesterol in [mg/dl]
13. Albumin: albumin in [gm/dl]

14. Copper: urine copper in [ug/day]
15. Alk_Phos: alkaline phosphatase in [U/liter]
16. SGOT: SGOT in [U/ml]
17. Triglycerides: triglycerides in [mg/dl]
18. Platelets: platelets per cubic [ml/1000]
19. Prothrombin: prothrombin time in seconds [s]
20. Stage: histologic stage of disease (1, 2, 3, or 4)

ID	N_Days	Status	Drug	Age	Sex	Ascites	Hepatome	Spiders	Edema	Bilirubin	Cholesterc	Albumin	Copper	Alk_Phos	SGOT	Tryglicerid	Platelets	Prothromt	Stage
1	400	D	D-penicilla	21464	F	Y	Y	Y	Y	14.5	261	2.6	156	1718	137.95	172	190	12.2	4
2	4500	C	D-penicilla	20617	F	N	Y	Y	N	1.1	302	4.14	54	7394.8	113.52	88	221	10.6	3
3	1012	D	D-penicilla	25594	M	N	N	N	S	1.4	176	3.48	210	516	96.1	55	151	12	4
4	1925	D	D-penicilla	19994	F	N	Y	Y	S	1.8	244	2.54	64	6121.8	60.63	92	183	10.3	4
5	1504	CL	Placebo	13918	F	N	Y	Y	N	3.4	279	3.53	143	671	113.15	72	136	10.9	3
6	2503	D	Placebo	24201	F	N	Y	N	N	0.8	248	3.98	50	944	93	63	NA	11	3
7	1832	C	Placebo	20284	F	N	Y	N	N	1	322	4.09	52	824	60.45	213	204	9.7	3
8	2466	D	Placebo	19379	F	N	N	N	N	0.3	280	4	52	4651.2	28.38	189	373	11	3
9	2400	D	D-penicilla	15526	F	N	N	Y	N	3.2	562	3.08	79	2276	144.15	88	251	11	2
10	51	D	Placebo	25772	F	Y	N	Y	Y	12.6	200	2.74	140	918	147.25	143	302	11.5	4
11	3762	D	Placebo	19619	F	N	Y	Y	N	1.4	259	4.16	46	1104	79.05	79	258	12	4
12	304	D	Placebo	21600	F	N	N	Y	N	3.6	236	3.52	94	591	82.15	95	71	13.6	4
13	3577	C	Placebo	16688	F	N	N	N	N	0.7	281	3.85	40	1181	88.35	130	244	10.6	3
14	1217	D	Placebo	20535	M	Y	Y	N	Y	0.8	NA	2.27	43	728	71	NA	156	11	4
15	3584	D	D-penicilla	23612	F	N	N	N	N	0.8	231	3.87	173	9009.8	127.71	96	295	11	3
16	3672	C	Placebo	14772	F	N	N	N	N	0.7	204	3.66	28	685	72.85	58	198	10.8	3
17	769	D	Placebo	19060	F	N	Y	N	N	2.7	274	3.15	159	1533	117.8	128	224	10.5	4
18	131	D	D-penicilla	19698	F	N	Y	Y	Y	11.4	178	2.8	588	961	280.55	200	283	12.4	4
19	4232	C	D-penicilla	18102	F	N	Y	N	S	0.7	235	3.56	39	1881	93	123	209	11	3
20	1356	D	Placebo	21898	F	N	Y	N	N	5.1	374	3.51	140	1919	122.45	135	322	13	4
21	3445	C	Placebo	23445	M	N	Y	Y	N	0.6	252	3.83	41	843	65.1	83	336	11.4	4
22	673	D	D-penicilla	20555	F	N	N	Y	N	3.4	271	3.63	464	1376	120.9	55	173	11.6	4
23	264	D	Placebo	20442	F	Y	Y	Y	Y	17.4	395	2.94	558	6064.8	227.04	191	214	11.7	4
24	4079	D	D-penicilla	16261	M	N	Y	N	N	2.1	456	4	124	5719	221.88	230	70	9.9	2
25	4127	C	Placebo	16463	F	N	N	N	N	0.7	298	4.1	40	661	106.95	66	324	11.3	2

26	1444	D	Placebo	19002	F	N	Y	Y	N	5.2	1128	3.68	53	3228	165.85	166	421	9.9	3
27	77	D	Placebo	19884	F	Y	Y	Y	S	21.6	175	3.31	221	3697.4	101.91	168	80	12	4
28	549	D	Placebo	16417	F	Y	Y	Y	Y	17.2	222	3.23	209	1975	189.1	195	144	13	4
29	4509	C	Placebo	23331	F	N	N	N	N	0.7	370	3.78	24	5833	73.53	86	390	10.6	2
30	321	D	Placebo	15116	F	N	Y	Y	N	3.6	260	2.54	172	7277	121.26	158	124	11	4
31	3839	D	Placebo	15177	F	N	Y	N	N	4.7	296	3.44	114	9933.2	206.4	101	195	10.3	2
32	4523	C	Placebo	19722	F	N	Y	N	N	1.8	262	3.34	101	7277	82.56	158	286	10.6	4
33	3170	D	Placebo	18731	F	N	N	N	N	0.8	210	3.19	82	1592	218.55	113	180	12	3
34	3933	C	D-penicilla	19015	F	N	N	N	N	0.8	364	3.7	37	1840	170.5	64	273	10.5	2
35	2847	D	Placebo	17758	F	N	N	N	N	1.2	314	3.2	201	12258.8	72.24	151	431	10.6	3
36	3611	C	Placebo	20604	F	N	N	N	N	0.3	172	3.39	18	558	71.3	96	311	10.6	2
37	223	D	D-penicilla	22546	F	Y	Y	N	Y	7.1	334	3.01	150	6931.2	180.6	118	102	12	4
38	3244	D	Placebo	13378	F	N	Y	Y	N	3.3	383	3.53	102	1234	137.95	87	234	11	4
39	2297	D	D-penicilla	20232	F	N	Y	N	N	0.7	282	3	52	9066.8	72.24	111	563	10.6	4
40	4467	C	D-penicilla	17046	F	N	N	N	N	1.3	NA	3.34	105	11046.6	104.49	NA	358	11	4
41	1350	D	D-penicilla	12285	F	N	Y	N	N	6.8	NA	3.26	96	1215	151.9	NA	226	11.7	4
42	4453	C	Placebo	12307	F	N	Y	Y	N	2.1	NA	3.54	122	8778	56.76	NA	344	11	4
43	4556	C	D-penicilla	17850	F	N	N	N	N	1.1	361	3.64	36	5430.2	67.08	89	203	10.6	2
44	3428	D	Placebo	13727	F	N	Y	Y	Y	3.3	299	3.55	131	1029	119.35	50	199	11.7	3
45	4025	C	Placebo	15265	F	N	N	N	N	0.6	NA	3.93	19	1826	71.3	NA	474	10.9	2
46	2256	D	D-penicilla	16728	F	N	Y	N	N	5.7	482	2.84	161	11552	136.74	165	518	12.7	3
47	2576	C	Placebo	17323	F	N	N	N	N	0.5	316	3.65	68	1716	187.55	71	356	9.8	3
48	4427	C	Placebo	17947	M	N	N	N	N	1.9	259	3.7	281	10396.8	188.34	178	214	11	3
49	708	D	Placebo	22336	F	N	Y	N	N	0.8	NA	3.82	58	678	97.65	NA	233	11	4
50	2598	D	D-penicilla	19544	F	N	Y	N	N	1.1	257	3.36	43	1080	106.95	73	128	10.6	4
51	3853	D	Placebo	19025	F	N	N	N	N	0.8	276	3.6	54	4332	99.33	143	273	10.6	2

51	3853 D	Placebo	19025 F	N	N	N	N	0.8	276	3.6	54	4332	99.33	143	273	10.6	2
52	2386 D	D-penicilla	18460 M	N	N	N	N	6	614	3.7	158	5084.4	206.4	93	362	10.6	1
53	1000 D	D-penicilla	24621 F	N	Y	N	N	2.6	NA	3.1	94	6456.2	56.76	NA	214	11	4
54	1434 D	D-penicilla	14317 F	Y	Y	Y	Y	1.3	288	3.4	262	5487.2	73.53	125	254	11	4
55	1360 D	D-penicilla	24020 M	N	N	N	N	1.8	416	3.94	121	10165	79.98	219	213	11	3
56	1847 D	Placebo	12279 F	N	Y	Y	N	1.1	498	3.8	88	13862.4	95.46	319	365	10.6	2
57	3282 D	D-penicilla	19567 F	N	Y	N	S	2.3	260	3.18	231	11320.2	105.78	94	216	12.4	3
58	4459 C	D-penicilla	16279 M	N	N	N	N	0.7	242	4.08	73	5890	56.76	118	NA	10.6	1
59	2224 D	D-penicilla	14754 F	N	Y	Y	N	0.8	329	3.5	49	7622.8	126.42	124	321	10.6	3
60	4365 C	D-penicilla	21324 F	N	N	N	N	0.9	604	3.4	82	876	71.3	58	228	10.3	3
61	4256 C	Placebo	16034 M	N	N	N	N	0.6	216	3.94	28	601	60.45	188	211	13	1
62	3090 D	Placebo	22173 F	Y	N	N	N	1.3	302	2.75	58	1523	43.4	112	329	13.2	4
63	859 D	Placebo	17031 F	N	N	Y	Y	22.5	932	3.12	95	5396	244.9	133	165	11.6	3
64	1487 D	Placebo	22977 F	N	Y	N	N	2.1	373	3.5	52	1009	150.35	188	178	11	3
65	3992 C	D-penicilla	14684 F	N	N	N	N	1.2	256	3.6	74	724	141.05	108	430	10	1
66	4191 D	D-penicilla	16967 M	N	Y	N	N	1.4	427	3.7	105	1909	182.9	171	123	11	3
67	2769 D	Placebo	18733 F	N	N	N	N	1.1	466	3.91	84	1787	328.6	185	261	10	3
68	4039 C	D-penicilla	11912 F	N	N	N	N	0.7	174	4.09	58	642	71.3	46	203	10.6	3
69	1170 D	D-penicilla	18021 F	N	Y	Y	S	20	652	3.46	159	3292	215.45	184	227	12.4	3
70	3458 C	D-penicilla	20600 F	N	N	N	N	0.6	NA	4.64	20	666	54.25	NA	265	10.6	2
71	4196 C	Placebo	17841 F	N	Y	N	N	1.2	258	3.57	79	2201	120.9	76	410	11.5	4
72	4184 C	Placebo	11868 F	N	N	N	N	0.5	320	3.54	51	1243	122.45	80	225	10	3
73	4190 C	Placebo	14060 F	N	N	N	N	0.7	132	3.6	17	423	49.6	56	265	11	1
74	1827 D	D-penicilla	18964 F	N	Y	Y	N	8.4	558	3.99	280	967	89.9	309	278	11	4
75	1191 D	D-penicilla	15895 F	Y	Y	Y	S	17.1	674	2.53	207	2078	182.9	598	268	11.5	4
76	71 D	D-penicilla	18972 F	N	Y	Y	S	12.2	394	3.08	111	2132	155	243	165	11.6	4

76	71 D	D-penicilla	18972 F	N	Y	Y	S	12.2	394	3.08	111	2132	155	243	165	11.6	4
77	326 D	Placebo	18199 F	N	Y	Y	S	6.6	244	3.41	199	1819	170.5	91	132	12.1	3
78	1690 D	D-penicilla	17512 F	N	Y	N	N	6.3	436	3.02	75	2176	170.5	104	236	10.6	4
79	3707 C	D-penicilla	16990 F	N	Y	N	N	0.8	315	4.24	13	1637	170.5	70	426	10.9	3
80	890 D	Placebo	24622 M	N	Y	N	N	7.2	247	3.72	269	1303	176.7	91	360	11.2	4
81	2540 D	D-penicilla	23107 F	N	Y	Y	N	14.4	448	3.65	34	1218	60.45	318	385	11.7	4
82	3574 D	D-penicilla	24585 F	N	N	N	N	4.5	472	4.09	154	1580	117.8	272	412	11.1	3
83	4050 C	D-penicilla	20459 F	N	Y	N	S	1.3	250	3.5	48	1138	71.3	100	81	12.9	4
84	4032 C	Placebo	20392 F	N	N	N	N	0.4	263	3.76	29	1345	137.95	74	181	11.2	3
85	3358 D	Placebo	17246 F	N	Y	N	N	2.1	262	3.48	58	2045	89.9	84	225	11.5	4
86	1657 D	D-penicilla	19270 F	N	Y	Y	N	5	1600	3.21	75	2656	82.15	174	181	10.9	3
87	198 D	D-penicilla	13616 F	N	N	N	N	1.1	345	4.4	75	1860	218.55	72	447	10.7	3
88	2452 C	Placebo	15119 F	N	N	N	S	0.6	296	4.06	37	1032	80.6	83	442	12	3
89	1741 D	D-penicilla	19155 F	N	Y	N	N	2	408	3.65	50	1083	110.05	98	200	11.4	2
90	2689 D	D-penicilla	12227 M	N	N	N	N	1.6	660	4.22	94	1857	151.9	155	337	11	2
91	460 D	Placebo	16658 F	N	Y	Y	S	5	325	3.47	110	2460	246.45	56	430	11.9	4
92	388 D	D-penicilla	28018 F	Y	N	N	Y	1.4	206	3.13	36	1626	86.8	70	145	12.2	4
93	3913 C	D-penicilla	13344 F	N	N	N	N	1.3	353	3.67	73	2039	232.5	68	380	11.1	2
94	750 D	D-penicilla	19693 F	N	Y	Y	N	3.2	201	3.11	178	1212	159.65	69	188	11.8	4
95	130 D	Placebo	16944 F	Y	Y	Y	Y	17.4	NA	2.64	182	559	119.35	NA	401	11.7	2
96	3850 C	D-penicilla	17841 F	N	N	N	N	1	NA	3.7	33	1258	99.2	NA	338	10.4	3
97	611 D	Placebo	26259 M	N	Y	N	S	2	420	3.26	62	3196	77.5	91	344	11.4	3
98	3823 C	D-penicilla	10550 F	N	N	N	N	1	239	3.77	77	1877	97.65	101	312	10.2	1
99	3820 C	Placebo	17703 M	N	N	N	N	1.8	460	3.35	148	1472	108.5	118	172	10.2	2
100	552 D	Placebo	18799 M	N	Y	N	N	2.3	178	3	145	746	178.25	122	119	12	4
101	3581 C	Placebo	16418 F	N	N	N	N	0.9	400	3.6	31	1689	164.3	166	327	10.4	3

108	2583 D	D-penicilla	18393 F	N	N	N	N	0.4	127	3.5	14	1062	49.6	84	334	10.3	2
109	2504 C	Placebo	16094 F	N	N	N	N	0.5	120	3.61	53	804	110.05	52	271	10.6	3
110	2105 D	D-penicilla	14212 F	N	Y	Y	N	1.9	486	3.54	74	1052	108.5	109	141	10.9	3
111	2350 CL	D-penicilla	15031 F	N	N	N	N	5.5	528	4.18	77	2404	172.05	78	467	10.7	3
112	3445 D	Placebo	20256 F	N	Y	Y	N	2	267	3.67	89	754	196.85	90	136	11.8	4
113	980 D	D-penicilla	18713 F	N	Y	Y	N	6.7	374	3.74	103	979	128.65	100	266	11.1	4
114	3395 D	Placebo	19295 M	N	N	N	N	3.2	259	4.3	208	1040	110.05	78	268	11.7	3
115	3422 C	Placebo	15574 F	N	N	Y	N	0.7	303	4.19	81	1584	111.6	156	307	10.3	3
116	3336 C	D-penicilla	22306 F	N	N	Y	S	3	458	3.63	74	1588	106.95	382	438	9.9	3
117	1083 D	D-penicilla	18137 F	N	Y	Y	N	6.5	950	3.11	111	2374	170.5	149	354	11	4
118	2288 D	D-penicilla	17844 F	N	Y	N	N	3.5	390	3.3	67	878	137.95	93	207	10.2	3
119	515 D	D-penicilla	19817 F	N	N	Y	N	0.6	636	3.83	129	944	97.65	114	306	9.5	3
120	2033 CL	D-penicilla	12839 M	N	N	N	N	3.5	325	3.98	444	766	130.2	210	344	10.6	3
121	191 D	Placebo	24803 M	Y	Y	N	Y	1.3	151	3.08	73	1112	46.5	49	213	13.2	4
122	3297 C	D-penicilla	20248 F	N	N	N	N	0.6	298	4.13	29	758	65.1	85	256	10.7	3
123	971 D	D-penicilla	16736 F	N	Y	Y	Y	5.1	NA	3.23	18	790	179.8	NA	104	13	4
124	3069 C	D-penicilla	19318 M	N	Y	N	N	0.6	251	3.9	25	681	57.35	107	182	10.8	4
125	2468 CL	Placebo	17233 F	N	Y	N	N	1.3	316	3.51	75	1162	147.25	137	238	10	4
126	824 D	D-penicilla	19577 F	Y	Y	Y	N	1.2	269	3.12	NA	1441	165.85	68	166	11.1	4
127	3255 C	Placebo	16109 F	N	N	N	N	0.5	268	4.08	9	1174	86.8	95	453	10	2
128	1037 D	D-penicilla	15322 F	N	Y	Y	N	16.2	NA	2.89	42	1828	299.15	NA	123	12.6	4
129	3239 C	D-penicilla	23235 F	N	Y	N	N	0.9	420	3.87	30	1009	57.35	232	NA	9.7	3
130	1413 D	Placebo	16154 F	N	Y	Y	N	17.4	1775	3.43	205	2065	165.85	97	418	11.5	3
131	850 D	Placebo	22646 F	N	Y	Y	N	2.8	242	3.8	74	614	136.4	104	121	13.2	4
132	2944 C	D-penicilla	14812 F	N	N	N	N	1.9	448	3.83	60	1052	127.1	175	181	9.8	3
133	2796 D	Placebo	22881 M	N	N	N	N	1.5	331	3.95	13	577	128.65	99	165	10.1	4

133	2796 D	Placebo	22881 M	N	N	N	N	1.5	331	3.95	13	577	128.65	99	165	10.1	4
134	3149 C	Placebo	15463 F	N	N	N	N	0.7	578	3.67	35	1353	127.1	105	427	10.7	2
135	3150 C	D-penicilla	15694 F	N	N	N	N	0.4	263	3.57	123	836	74.4	121	445	11	2
136	3098 C	D-penicilla	20440 F	N	N	N	N	0.8	263	3.35	27	1636	116.25	69	206	9.8	2
137	2990 C	D-penicilla	22960 F	N	N	N	N	1.1	399	3.6	79	3472	155	152	344	10.1	2
138	1297 D	D-penicilla	18719 M	N	Y	N	N	7.3	426	3.93	262	2424	145.7	218	252	10.5	3
139	2106 C	Placebo	17080 F	N	Y	N	N	1.1	328	3.31	159	1260	94.55	134	142	11.6	4
140	3059 C	D-penicilla	19751 F	N	Y	N	N	1.1	290	4.09	38	2120	186	146	318	10	3
141	3050 C	D-penicilla	17180 F	N	N	N	N	0.9	346	3.77	59	794	125.55	56	336	10.6	2
142	2419 D	Placebo	20354 F	N	Y	N	N	1	364	3.48	20	720	134.85	88	283	9.9	2
143	786 D	Placebo	16839 F	N	Y	N	N	2.9	332	3.6	86	1492	134.85	103	277	11	4
144	943 D	Placebo	19098 F	N	Y	N	S	28	556	3.26	152	3896	198.4	171	335	10	3
145	2976 C	Placebo	18701 F	N	N	Y	N	0.7	309	3.84	96	858	41.85	106	253	11.4	3
146	2615 C	Placebo	12369 F	N	N	N	S	1.2 NA		3.89	58	1284	173.6 NA		239	9.4	3
147	2995 C	D-penicilla	27398 F	N	N	N	S	1.2	288	3.37	32	791	57.35	114	213	10.7	2
148	1427 D	Placebo	11273 F	N	Y	N	N	7.2	1015	3.26	247	3836	198.4	280	330	9.8	3
149	762 D	D-penicilla	22574 M	N	Y	Y	S	3	257	3.79	290	1664	102.3	112	140	9.9	4
150	2891 C	Placebo	12779 F	N	N	Y	N	1 NA		3.63	57	1536	134.85 NA		233	10	1
151	2870 C	D-penicilla	20104 F	N	N	N	N	0.9	460	3.03	57	721	85.25	174	301	9.4	2
152	1152 D	D-penicilla	25546 M	N	Y	N	N	2.3	586	3.01	243	2276	114.7	126	339	10.9	3
153	2863 C	D-penicilla	18118 F	N	N	N	N	0.5	217	3.85	68	453	54.25	68	270	11.1	1
154	140 D	D-penicilla	25340 M	N	N	Y	Y	2.4	168	2.56	225	1056	120.9	75	108	14.1	3
155	2666 C	Placebo	15909 F	N	Y	Y	S	0.6	220	3.35	57	1620	153.45	80	311	11.2	4
156	853 D	Placebo	21699 F	N	Y	N	N	25.5	358	3.52	219	2468	201.5	205	151	11.5	2
157	2835 C	Placebo	17809 F	N	N	N	N	0.6	286	3.42	34	1868	77.5	206	487	10	2
158	2475 CL	D-penicilla	13329 F	N	N	N	N	3.4	450	3.37	32	1408	116.25	118	313	11.2	2
cirrhosis																	

159	1536 D	Placebo	16714 M	N	N	N	N	2.5	317	3.46	217	714	130.2	140	207	10.1	3
160	2772 C	Placebo	20955 F	N	N	N	N	0.6	217	3.62	13	414	75.95	119	224	10.5	3
161	2797 C	Placebo	15612 F	N	N	N	N	2.3	502	3.56	4	964	120.9	180	269	9.6	2
162	186 D	Placebo	21483 F	N	Y	Y	N	3.2	260	3.19	91	815	127.1	101	160	12	4
163	2055 D	D-penicilla	19540 F	N	N	N	N	0.3	233	4.08	20	622	66.65	68	358	9.9	3
164	264 D	Placebo	15857 F	N	Y	Y	S	8.5 NA		3.34	161	1428	181.35 NA		88	13.3	4
165	1077 D	D-penicilla	19470 M	N	Y	N	N	4	196	3.45	80	2496	133.3	142	212	11.3	4
166	2721 C	Placebo	15105 F	N	Y	N	N	5.7	1480	3.26	84	1960	457.25	108	213	9.5	2
167	1682 D	D-penicilla	22265 M	N	Y	N	N	0.9	376	3.86	200	1015	83.7	154	238	10.3	4
168	2713 C	Placebo	17442 F	N	Y	N	N	0.4	257	3.8	44	842	97.65	110 NA		9.2	2
169	1212 D	Placebo	12963 F	N	N	N	N	1.3	408	4.22	67	1387	142.6	137	295	10.1	3
170	2692 C	D-penicilla	17774 F	N	N	N	N	1.2	390	3.61	32	1509	88.35	52	263	9	3
171	2574 C	D-penicilla	19237 F	N	N	N	N	0.5 NA		4.52	31	784	74.4 NA		361	10.1	3
172	2301 C	Placebo	18215 F	N	N	Y	N	1.3	205	3.34	65	1031	91.45	126	217	9.8	3
173	2657 C	D-penicilla	11058 F	N	Y	Y	N	3	236	3.42	76	1403	89.9	86	493	9.8	2
174	2644 C	D-penicilla	20296 F	N	N	N	N	0.5 NA		3.85	63	663	79.05 NA		311	9.7	1
175	2624 C	Placebo	19049 F	N	N	N	N	0.8	283	3.8	152	718	108.5	168	340	10.1	3
176	1492 D	D-penicilla	15198 F	N	Y	Y	N	3.2 NA		3.56	77	1790	139.5 NA		149	10.1	4
177	2609 C	Placebo	20254 F	N	N	N	N	0.9	258	4.01	49	559	43.4	133	277	10.4	2
178	2580 C	D-penicilla	25569 F	N	N	N	N	0.6 NA		4.08	51	665	74.4 NA		325	10.2	4
179	2573 C	Placebo	16050 F	N	Y	N	N	1.8	396	3.83	39	2148	102.3	133	278	9.9	4
180	2563 C	Placebo	15548 F	N	N	N	N	4.7	478	4.38	44	1629	237.15	76	175	10.4	3
181	2556 C	D-penicilla	16279 F	N	Y	Y	N	1.4	248	3.58	63	554	75.95	106	79	10.3	4
182	2555 C	D-penicilla	20799 F	N	Y	N	N	0.6 NA		3.69	161	674	26.35 NA		539	9.9	2
183	2241 CL	Placebo	14705 F	N	N	N	N	0.5	201	3.73	44	1345	54.25	145	445	10.1	2
184	974 D	Placebo	13736 F	N	Y	N	N	11	674	3.55	358	2412	167.4	140	471	9.8	3

185	2527 C	D-penicilla	17664 F	N	N	N	N	0.8	256	3.54	42	1132	74.4	94	192	10.5	3
186	1576 D	D-penicilla	25873 F	N	N	Y	S	2	225	3.53	51	933	69.75	62	200	12.7	3
187	733 D	Placebo	13073 F	N	Y	N	N	14	808	3.43	251	2870	153.45	137	268	11.5	3
188	2332 C	D-penicilla	22873 F	N	Y	N	N	0.7	187	3.48	41	654	120.9	98	164	11	4
189	2456 C	Placebo	18499 F	N	Y	N	N	1.3	360	3.63	52	1812	97.65	164	256	9.9	3
190	2504 C	D-penicilla	19916 F	N	N	Y	N	2.3 NA		3.93	24	1828	133.3 NA		327	10.2	2
191	216 D	Placebo	19246 F	Y	Y	Y	N	24.5	1092	3.35	233	3740	147.25	432	399	15.2	4
192	2443 C	D-penicilla	19256 F	N	Y	N	N	0.9	308	3.69	67	696	51.15	101	344	9.8	4
193	797 D	Placebo	20736 F	N	N	N	N	10.8	932	3.19	267	2184	161.2	157	382	10.4	4
194	2449 C	D-penicilla	16216 F	N	N	N	N	1.5	293	4.3	50	975	125.55	56	336	9.1	2
195	2330 C	D-penicilla	10795 F	N	Y	N	N	3.7	347	3.9	76	2544	221.65	90	129	11.5	4
196	2363 C	D-penicilla	20834 F	N	Y	Y	N	1.4	226	3.36	13	810	72.85	62	117	11.6	4
197	2365 C	D-penicilla	16300 F	N	N	N	N	0.6	266	3.97	25	1164	102.3	102	201	10.1	2
198	2357 C	Placebo	13075 F	N	N	Y	N	0.7	286	2.9	38	1692	141.05	90	381	9.6	2
199	1592 C	D-penicilla	14872 F	N	N	N	N	2.1	392	3.43	52	1395	184.45	194	328	10.2	3
200	2318 C	Placebo	11773 F	N	N	Y	N	4.7	236	3.55	112	1391	137.95	114	332	9.9	3
201	2294 C	Placebo	15009 F	N	Y	N	N	0.6	235	3.2	26	1758	106.95	67	228	10.8	4
202	2272 C	D-penicilla	22514 F	N	N	N	N	0.5	223	3.8	15	1044	80.6	89	514	10	2
203	2221 C	Placebo	13535 F	N	Y	N	N	0.5	149	4.04	227	598	52.7	57	166	9.9	2
204	2090 D	Placebo	22857 F	N	N	N	N	0.7	255	3.74	23	1024	77.5	58	281	10.2	3
205	2081 D	D-penicilla	17889 F	Y	N	N	N	2.5	382	3.55	108	1516	238.7 NA		126	10.3	3
206	2255 C	D-penicilla	22642 F	N	N	N	N	0.6	213	4.07	12	5300	57.35	68	240	11	1
207	2171 C	D-penicilla	26580 F	N	N	N	S	0.6 NA		3.33	14	733	85.25 NA		259	10.1	4
208	904 D	D-penicilla	22388 F	N	Y	N	N	3.9	396	3.2	58	1440	153.45	131	156	10	4
209	2216 C	Placebo	19221 F	N	Y	Y	N	0.7	252	4.01	11	1210	72.85	58	309	9.5	2
210	2224 C	Placebo	18176 M	N	Y	N	N	0.9	346	3.37	81	1098	122.45	90	298	10	2

212	2176 C	Placebo	17263 F	N	N	N	N	1.2	232	3.98	11	1074	100.75	99	223	9.9	3
213	2178 C	D-penicilla	18337 F	N	N	Y	N	0.5	400	3.4	9	1134	96.1	55	356	10.2	3
214	1786 D	Placebo	25329 F	N	Y	N	N	0.9	404	3.43	34	1866	79.05	224	236	9.9	3
215	1080 D	Placebo	15037 F	N	N	N	N	5.9	1276	3.85	141	1204	203.05	157	216	10.7	3
216	2168 C	D-penicilla	21610 F	N	N	N	N	0.5 NA		3.68	20	856	55.8 NA		146	10.4	3
217	790 D	Placebo	13178 F	N	Y	N	N	11.4	608	3.31	65	1790	151.9	210	298	10.8	4
218	2170 C	D-penicilla	12636 F	N	N	N	N	0.5 NA		3.89	29	897	66.65 NA		423	10.1	1
219	2157 C	Placebo	15601 F	N	N	N	N	1.6	215	4.17	67	936	134.85	85	176	9.6	3
220	1235 D	D-penicilla	23241 F	N	N	Y	N	3.8	426	3.22	96	2716	210.8	113	228	10.6	2
221	2050 C	Placebo	20684 F	N	Y	N	N	0.9	360	3.65	72	3186	94.55	154	269	9.7	4
222	597 D	Placebo	16898 F	N	Y	N	N	4.5	372	3.38	227	2310	167.4	135	240	12.4	3
223	334 D	D-penicilla	22369 F	Y	Y	N	Y	14.1	448	2.43	123	1833	134	155	210	11	4
224	1945 C	D-penicilla	14106 F	N	N	N	N	1	309	3.66	67	1214	158.1	101	309	9.7	3
225	2022 C	D-penicilla	14161 F	N	N	N	N	0.7	274	3.66	108	1065	88.35	135	251	10.1	2
226	1978 C	Placebo	20708 F	N	Y	N	N	0.5	223	3.7	39	884	75.95	104	231	9.6	3
227	999 D	D-penicilla	21532 M	N	N	N	N	2.3	316	3.35	172	1601	179.8	63	394	9.7	2
228	1967 C	Placebo	13486 F	N	N	N	N	0.7	215	3.35	41	645	93	74	165	9.6	3
229	348 D	D-penicilla	22797 F	Y	Y	N	S	4.5	191	3.05	200	1020	175.15	118	139	11.4	4
230	1979 C	Placebo	12641 F	N	Y	Y	N	3.3	302	3.41	51	310	83.7	44	95	11.5	4
231	1165 D	Placebo	21307 F	N	Y	Y	N	3.4	518	1.96	115	2250	203.05	90	190	10.7	4
232	1951 C	D-penicilla	18329 F	N	Y	N	N	0.4	267	3.02	47	1001	133.3	87	265	10.6	3
233	1932 C	D-penicilla	15591 F	N	Y	Y	N	0.9	514	3.06	412	2622	105.4	87	284	9.8	4
234	1776 C	Placebo	12557 F	N	N	N	N	0.9	578	3.35	78	976	116.25	177	322	11.2	2
235	1882 C	Placebo	12120 F	N	Y	N	N	13	1336	4.16	71	3510	209.25	111	338	11.9	3
236	1908 C	D-penicilla	14019 F	N	Y	Y	N	1.5	253	3.79	67	1006	139.5	106	341	9.7	3
237	1882 C	D-penicilla	21828 F	N	Y	N	N	1.6	442	2.95	105	820	85.25	108	181	10.1	3

237	1882 C	D-penicilla	21828 F	N	Y	N	N	1.6	442	2.95	105	820	85.25	108	181	10.1	3
238	1874 C	Placebo	24257 F	N	N	N	S	0.6	280	3.35 NA		1093	128.65	81	295	9.8	2
239	694 D	D-penicilla	17090 F	N	Y	Y	N	0.8	300	2.94	231	1794	130.2	99	319	11.2	4
240	1831 C	D-penicilla	20483 F	N	N	N	N	0.4	232	3.72	24	369	51.15	139	326	10.1	3
241	837 CL	Placebo	15112 F	N	Y	Y	N	4.4	316	3.62	308	1119	114.7	322	282	9.8	4
242	1810 C	D-penicilla	23585 F	N	Y	N	N	1.9	354	2.97	86	1553	196.85	152	277	9.9	3
243	930 D	Placebo	24650 F	N	Y	N	N	8	468	2.81	139	2009	198.4	139	233	10	4
244	1690 D	D-penicilla	16374 F	N	N	Y	N	3.9	350	3.22	121	1268	272.8	231	270	9.6	3
245	1790 C	Placebo	16718 F	N	Y	N	N	0.6	273	3.65	48	794	52.7	214	305	9.6	3
246	1435 CL	D-penicilla	12035 F	N	Y	N	N	2.1	387	3.77	63	1613	150.35	33	185	10.1	4
247	732 CL	D-penicilla	15056 F	N	Y	N	N	6.1	1712	2.83	89	3681	158.1	139	297	10	3
248	1785 C	Placebo	20241 F	N	Y	N	N	0.8	324	3.51	39	1237	66.65	146	371	10	3
249	1783 C	D-penicilla	17525 F	N	N	Y	N	1.3	242	3.2	35	1556	175.15	71	195	10.6	4
250	1769 C	Placebo	14899 F	N	Y	N	N	0.6	299	3.36	23	2769	220.1	85	303	10.9	4
251	1457 C	D-penicilla	20810 F	N	N	N	N	0.5	227	3.61	40	676	83	120	249	9.9	2
252	1770 C	D-penicilla	25006 F	N	Y	Y	N	1.1	246	3.35	116	924	113.15	90	317	10	4
253	1765 C	D-penicilla	28650 M	Y	Y	Y	N	7.1	243	3.03	380	983	158.1	154	97	11.2	4
254	737 CL	D-penicilla	14558 F	N	Y	Y	N	3.1	227	3.75	121	1136	110	91	264	10	3
255	1735 C	Placebo	12897 F	N	Y	Y	N	0.7	193	3.85	35	466	53	118	156	10.3	3
256	1701 C	D-penicilla	11485 F	N	N	N	N	1.1	336	3.74	48	823	84	108	242	9.7	3
257	1614 C	D-penicilla	21281 F	N	N	N	N	0.5	280	4.23	36	377	56	146	227	10.6	2
258	1702 C	D-penicilla	18806 F	N	N	N	N	1.1	414	3.44	80	1003	99	55	271	9.6	1
259	1615 C	Placebo	21904 F	N	Y	N	N	3.1	277	2.97	42	1110	125	126	221	9.8	3
260	1656 C	Placebo	27220 M	N	Y	N	N	5.6	232	3.59	188	1120	98	128	248	10.9	4
261	1677 C	Placebo	19126 F	N	Y	Y	N	3.2	375	3.14	129	857	89 NA		375	9.5	3
262	1666 C	Placebo	15628 F	N	Y	N	N	2.8	322	3.06	65	2562	91	209	231	9.5	3

263	1301 CL	Placebo	12738 F	N	Y	Y	S	1.1	432	3.57	45	1406	190	77	248	11.4	4
264	1542 CL	Placebo	16122 F	N	Y	Y	N	3.4	356	3.12	188	1911	92	130	318	11.2	3
265	1084 CL	Placebo	16941 F	N	Y	N	N	3.5	348	3.2	121	938	120	146	296	10	4
266	1614 C	D-penicilla	20567 F	N	N	N	N	0.5	318	3.32	52	613	70	260	279	10.2	3
267	179 D	D-penicilla	25899 F	Y	Y	Y	Y	6.6	222	2.33	138	620	106	91	195	12.1	4
268	1191 D	D-penicilla	20233 F	Y	Y	N	S	6.4	344	2.75	16	834	82	179	149	11	4
269	1363 C	Placebo	16467 F	N	N	N	N	3.6	374	3.5	143	1428	188	44	151	10.1	2
270	1568 C	D-penicilla	9598 F	N	Y	Y	N	1	448	3.74	102	1128	71	117	228	10.2	3
271	1569 C	Placebo	18435 F	N	Y	N	N	1	321	3.5	94	955	111	177	289	9.7	3
272	1525 C	D-penicilla	14025 F	N	N	N	N	0.5	226	2.93	22	674	58	85	153	9.8	1
273	1558 C	Placebo	17320 F	N	N	Y	N	2.2	328	3.46	75	1677	87	116	202	9.6	3
274	1447 CL	D-penicilla	17525 F	N	N	N	N	1.6 NA		3.07	136	1995	128 NA		372	9.6	4
275	1349 C	D-penicilla	13995 F	N	N	N	N	2.2	572	3.77	77	2520	92	114	309	9.5	4
276	1481 C	D-penicilla	18302 F	N	N	N	N	1	219	3.85	67	640	145	108	95	10.7	2
277	1434 C	Placebo	12816 F	N	N	N	S	1	317	3.56	44	1636	84	111	394	9.8	3
278	1420 C	Placebo	11872 F	N	N	N	N	5.6	338	3.7	130	2139	185	193	215	9.9	4
279	1433 C	Placebo	20510 F	N	N	N	N	0.5	198	3.77	38	911	57	56	280	9.8	2
280	1412 C	D-penicilla	16858 F	N	N	N	N	1.6	325	3.69	69	2583	142	140	284	9.6	3
281	41 D	D-penicilla	24064 F	Y	N	N	Y	17.9	175	2.1	220	705	338	229	62	12.9	4
282	1455 C	Placebo	12398 F	N	Y	N	N	1.3	304	3.52	97	1622	71	169	255	9.5	4
283	1030 C	Placebo	22960 F	N	N	N	N	1.1	412	3.99	103	1293	91	113	422	9.6	4
284	1418 C	Placebo	17738 F	N	N	N	N	1.3	291	3.44	75	1082	85	195	251	9.5	3
285	1401 C	D-penicilla	16929 F	N	N	N	N	0.8	253	3.48	65	688	57	80	252	10	1
286	1408 C	D-penicilla	14191 F	N	Y	Y	N	2	310	3.36	70	1257	122	118	143	9.8	3
287	1234 C	D-penicilla	21421 F	N	N	Y	N	6.4	373	3.46	155	1768	120	151	258	10.1	4
288	1067 CL	Placebo	17874 F	N	Y	N	S	8.7	310	3.89	107	637	117	242	298	9.6	2

288	1067	CL	Placebo	17874	F	N	Y	N	S	8.7	310	3.89	107	637	117	242	298	9.6	2
289	799	D	D-penicilla	24681	M	N	Y	N	S	4	416	3.99	177	960	86	242	269	9.8	2
290	1363	C	D-penicilla	24101	F	N	N	N	N	1.4	294	3.57	33	722	93	69	283	9.8	3
291	901	CL	D-penicilla	14939	F	N	N	N	N	3.2	339	3.18	123	3336	205	84	304	9.9	4
292	1329	C	Placebo	18352	M	N	Y	N	N	8.6	546	3.73	84	1070	127	153	291	11.2	3
293	1320	C	Placebo	20891	F	N	Y	Y	Y	8.5	194	2.98	196	815	163	78	122	12.3	4
294	1302	C	D-penicilla	22111	M	N	Y	N	N	6.6	1000	3.07	88	3150	193	133	299	10.9	4
295	877	CL	D-penicilla	12912	M	N	N	N	N	2.4	646	3.83	102	855	127	194	306	10.3	3
296	1321	C	Placebo	11462	F	N	N	N	N	0.8	328	3.31	62	1105	137	95	293	10.9	4
297	533	CL	D-penicilla	20449	M	N	Y	N	N	1.2	275	3.43	100	1142	75	91	217	11.3	4
298	1300	C	Placebo	19258	F	N	Y	N	N	1.1	340	3.37	73	289	97	93	243	10.2	3
299	1293	C	D-penicilla	13913	F	N	N	N	N	2.4	342	3.76	90	1653	150	127	213	10.8	3
300	207	D	Placebo	21247	F	N	Y	N	N	5.2	NA	2.23	234	601	135	NA	206	12.3	4
301	1295	C	Placebo	16513	F	N	N	N	N	1	393	3.57	50	1307	74	103	295	10.5	4
302	1271	C	D-penicilla	13806	F	N	N	N	N	0.7	335	3.95	43	657	52	104	268	10.6	2
303	1250	C	Placebo	22156	F	N	Y	Y	N	1	372	3.25	108	1190	140	55	248	10.6	4
304	1230	C	D-penicilla	12979	F	N	N	N	N	0.5	219	3.93	22	663	45	75	246	10.8	3
305	1216	C	Placebo	15730	F	N	Y	Y	N	2.9	426	3.61	73	5184	288	144	275	10.6	3
306	1216	C	Placebo	20597	F	N	Y	N	N	0.6	239	3.45	31	1072	55	64	227	10.7	2
307	1149	C	Placebo	11167	F	N	N	N	N	0.8	273	3.56	52	1282	130	59	344	10.5	2
308	1153	C	D-penicilla	22347	F	N	Y	N	N	0.4	246	3.58	24	797	91	113	288	10.4	2
309	994	C	Placebo	21294	F	N	N	N	N	0.4	260	2.75	41	1166	70	82	231	10.8	2
310	939	C	D-penicilla	22767	F	N	N	N	N	1.7	434	3.35	39	1713	171	100	234	10.2	2
311	839	C	D-penicilla	13879	F	N	N	N	N	2	247	3.16	69	1050	117	88	335	10.5	2
312	788	C	Placebo	12109	F	N	N	Y	N	6.4	576	3.79	186	2115	136	149	200	10.8	2
313	4062	C	NA	21915	F	NA	NA	NA	N	0.7	NA	3.65	NA	NA	NA	NA	378	11	NA

314	3561	D	NA	23741	F	NA	NA	NA	S	1.4	NA	3.04	NA	NA	NA	NA	331	12.1	4
315	2844	C	NA	19724	F	NA	NA	NA	N	0.7	NA	4.03	NA	NA	NA	NA	226	9.8	4
316	2071	D	NA	27394	F	NA	NA	NA	S	0.7	NA	3.96	NA	NA	NA	NA	NA	11.3	4
317	3030	C	NA	22646	F	NA	NA	NA	N	0.8	NA	2.48	NA	NA	NA	NA	273	10	NA
318	1680	C	NA	15706	F	NA	NA	NA	N	0.7	NA	3.68	NA	NA	NA	NA	306	9.5	2
319	41	D	NA	16802	F	NA	NA	NA	N	5	NA	2.93	NA	NA	NA	NA	260	10.4	NA
320	2403	C	NA	16071	F	NA	NA	NA	S	0.4	NA	3.81	NA	NA	NA	NA	226	10.5	3
321	1170	C	NA	22280	M	NA	NA	NA	S	1.3	NA	3.41	NA	NA	NA	NA	259	10.9	4
322	2011	D	NA	23376	F	NA	NA	NA	N	1.1	NA	3.69	NA	NA	NA	NA	139	10.5	NA
323	3523	C	NA	14610	F	NA	NA	NA	N	0.6	NA	4.04	NA	NA	NA	NA	130	11.2	2
324	3468	C	NA	23011	F	NA	NA	NA	N	0.6	NA	3.94	NA	NA	NA	NA	234	11.5	2
325	4795	C	NA	12419	F	NA	NA	NA	N	1.8	NA	3.24	NA	NA	NA	NA	NA	18	2
326	1236	C	NA	18993	F	NA	NA	NA	N	1.5	NA	3.42	NA	NA	NA	NA	246	10.3	3
327	4214	C	NA	17897	F	NA	NA	NA	N	1.2	NA	3.99	NA	NA	NA	NA	NA	11.2	2
328	2111	D	NA	19724	F	NA	NA	NA	N	1	NA	3.6	NA	NA	NA	NA	NA	12.1	2
329	1462	D	NA	23011	F	NA	NA	NA	N	0.7	NA	3.4	NA	NA	NA	NA	371	10.1	4
330	1746	D	NA	19724	M	NA	NA	NA	N	3.5	NA	3.63	NA	NA	NA	NA	325	10.3	2
331	94	D	NA	16802	F	NA	NA	NA	S	3.1	NA	3.56	NA	NA	NA	NA	142	13.6	4
332	785	D	NA	19358	F	NA	NA	NA	N	12.6	NA	2.87	NA	NA	NA	NA	114	11.8	4
333	1518	D	NA	20454	F	NA	NA	NA	N	2.8	NA	3.92	NA	NA	NA	NA	NA	10.6	4
334	466	D	NA	20454	F	NA	NA	NA	N	7.1	NA	3.51	NA	NA	NA	NA	721	11.8	NA
335	3527	C	NA	20089	F	NA	NA	NA	N	0.6	NA	4.15	NA	NA	NA	NA	280	10.1	2
336	2635	C	NA	23741	F	NA	NA	NA	N	2.1	NA	3.34	NA	NA	NA	NA	155	10.1	4
337	2286	D	NA	20454	F	NA	NA	NA	N	1.8	NA	3.64	NA	NA	NA	NA	141	10	NA
338	791	D	NA	17167	F	NA	NA	NA	N	16	NA	3.42	NA	NA	NA	NA	475	13.8	2
339	3492	C	NA	21915	F	NA	NA	NA	N	0.6	NA	4.38	NA	NA	NA	NA	269	10.6	2

339	3492	C	NA	21915	F	NA	NA	NA	N	0.6	NA	4.38	NA	NA	NA	NA	269	10.6	2
340	3495	C	NA	19358	F	NA	NA	NA	N	5.4	NA	4.19	NA	NA	NA	NA	141	11.2	2
341	111	D	NA	19724	F	NA	NA	NA	N	9	NA	3.29	NA	NA	NA	NA	286	13.1	4
342	3231	C	NA	18263	F	NA	NA	NA	N	0.9	NA	4.01	NA	NA	NA	NA	244	10.5	3
343	625	D	NA	17532	F	NA	NA	NA	N	11.1	NA	2.84	NA	NA	NA	NA	NA	12.2	2
344	3157	C	NA	13149	F	NA	NA	NA	N	8.9	NA	3.76	NA	NA	NA	NA	209	10.6	3
345	3021	CL	NA	17532	F	NA	NA	NA	N	0.5	NA	3.76	NA	NA	NA	NA	388	10.1	2
346	559	D	NA	25568	F	NA	NA	NA	S	0.6	NA	3.81	NA	NA	NA	NA	160	11	4
347	2812	D	NA	18628	F	NA	NA	NA	N	3.4	NA	3.92	NA	NA	NA	NA	NA	9.3	2
348	2834	C	NA	18993	M	NA	NA	NA	N	0.9	NA	3.14	NA	NA	NA	NA	191	12.3	2
349	2855	C	NA	19724	F	NA	NA	NA	N	1.4	NA	3.82	NA	NA	NA	NA	249	10.3	2
350	662	D	NA	17532	F	NA	NA	NA	N	2.1	NA	4.1	NA	NA	NA	NA	200	9	3
351	727	D	NA	24107	F	NA	NA	NA	N	15	NA	3.4	NA	NA	NA	NA	150	11.1	4
352	2716	C	NA	19358	F	NA	NA	NA	N	0.6	NA	4.19	NA	NA	NA	NA	330	9.9	1
353	2698	C	NA	22646	F	NA	NA	NA	N	1.3	NA	3.4	NA	NA	NA	NA	167	10.6	4
354	990	D	NA	21550	F	NA	NA	NA	N	1.3	NA	3.12	NA	NA	NA	NA	125	9.6	2
355	2338	C	NA	14245	F	NA	NA	NA	N	1.6	NA	3.75	NA	NA	NA	NA	145	10.4	3
356	1616	D	NA	24472	F	NA	NA	NA	S	2.2	NA	3.26	NA	NA	NA	NA	171	11.1	4
357	2563	C	NA	21185	F	NA	NA	NA	N	3	NA	3.46	NA	NA	NA	NA	109	10.4	4
358	2537	C	NA	23376	F	NA	NA	NA	N	0.8	NA	3.49	NA	NA	NA	NA	314	10.3	3
359	2534	C	NA	16802	F	NA	NA	NA	N	0.8	NA	2.89	NA	NA	NA	NA	419	NA	1
360	778	D	NA	23376	F	NA	NA	NA	N	1.8	NA	3.15	NA	NA	NA	NA	183	10.4	4
361	617	CL	NA	14975	F	NA	NA	NA	N	5.5	NA	2.31	NA	NA	NA	NA	517	10.4	4
362	2267	CL	NA	17897	F	NA	NA	NA	N	18	NA	3.04	NA	NA	NA	NA	432	9.7	2
363	2249	C	NA	16071	F	NA	NA	NA	N	0.6	NA	3.5	NA	NA	NA	NA	150	9.9	3
364	359	D	NA	21550	F	NA	NA	NA	N	2.7	NA	3.35	NA	NA	NA	NA	142	11.5	4

364	359 D	NA	21550 F	NA	NA	NA	N	2.7 NA	3.35 NA	NA	NA	NA	142	11.5	4
365	1925 C	NA	23011 F	NA	NA	NA	N	0.9 NA	3.58 NA	NA	NA	NA	224	10	3
366	249 D	NA	22280 F	NA	NA	NA	N	1.3 NA	3.01 NA	NA	NA	NA	223	10.7	3
367	2202 C	NA	23376 F	NA	NA	NA	N	1.1 NA	3.49 NA	NA	NA	NA	166	9.8	3
368	43 D	NA	17897 F	NA	NA	NA	N	13.8 NA	2.77 NA	NA	NA	NA	388 NA		4
369	1197 D	NA	15341 F	NA	NA	NA	N	4.4 NA	4.52 NA	NA	NA	NA	102	10.8	4
370	1095 D	NA	18263 F	NA	NA	NA	N	16 NA	3.36 NA	NA	NA	NA	384	10	3
371	489 D	NA	18628 F	NA	NA	NA	S	7.3 NA	3.52 NA	NA	NA	NA	265	11.1	1
372	2149 C	NA	13514 F	NA	NA	NA	N	0.6 NA	3.55 NA	NA	NA	NA	248	10.3	2
373	2103 C	NA	22646 F	NA	NA	NA	N	0.7 NA	3.29 NA	NA	NA	NA	190	9.8	2
374	1980 C	NA	18628 F	NA	NA	NA	N	0.7 NA	3.1 NA	NA	NA	NA	274	10.6	3
375	1347 CL	NA	18993 F	NA	NA	NA	N	1.7 NA	3.24 NA	NA	NA	NA	231	10.5	3
376	1478 D	NA	16071 M	NA	NA	NA	N	9.5 NA	3.63 NA	NA	NA	NA	292	10.2	3
377	1987 C	NA	12053 F	NA	NA	NA	N	2.2 NA	3.76 NA	NA	NA	NA	253	9.9	3
378	1168 D	NA	21915 F	NA	NA	NA	S	1.8 NA	3.62 NA	NA	NA	NA	225	9.9	2
379	597 D	NA	23011 F	NA	NA	NA	S	3.3 NA	2.73 NA	NA	NA	NA	224	11.1	4
380	1725 CL	NA	12053 F	NA	NA	NA	N	2.9 NA	4.08 NA	NA	NA	NA	418	10.5	3
381	1899 C	NA	14975 M	NA	NA	NA	N	1.7 NA	3.66 NA	NA	NA	NA	92	11	4
382	221 D	NA	18628 F	NA	NA	NA	N	14 NA	2.58 NA	NA	NA	NA	190	11.6	4
383	1022 CL	NA	13514 F	NA	NA	NA	S	0.8 NA	3 NA	NA	NA	NA	76	10.8	4
384	1639 C	NA	21550 F	NA	NA	NA	N	1.3 NA	3.4 NA	NA	NA	NA	243	9.7	1
385	1635 C	NA	20089 F	NA	NA	NA	N	0.7 NA	2.93 NA	NA	NA	NA	209	10.6	3
386	1654 C	NA	19724 M	NA	NA	NA	N	1.7 NA	2.38 NA	NA	NA	NA	166	9.8	3
387	1653 C	NA	17897 F	NA	NA	NA	S	13.6 NA	3 NA	NA	NA	NA	233	9.9	3
393	193 D	NA	18993 F	NA	NA	NA	S	0.7 NA	2.96 NA	NA	NA	NA	319	9.9	4
394	1367 C	NA	20819 F	NA	NA	NA	S	2 NA	3.07 NA	NA	NA	NA	80	12.1	4
395	1329 C	NA	13149 F	NA	NA	NA	N	1.4 NA	3.98 NA	NA	NA	NA	402	11	1
396	1343 C	NA	18263 F	NA	NA	NA	N	1.6 NA	3.48 NA	NA	NA	NA	277	10.2	2
397	1328 C	NA	23376 F	NA	NA	NA	N	0.5 NA	3.65 NA	NA	NA	NA	425	10.2	4
398	1375 C	NA	22646 F	NA	NA	NA	N	7.3 NA	3.49 NA	NA	NA	NA	189	10.9	4
399	1260 C	NA	15341 F	NA	NA	NA	N	8.1 NA	2.82 NA	NA	NA	NA	193	10.4	2
400	1223 C	NA	16071 F	NA	NA	NA	N	0.5 NA	3.34 NA	NA	NA	NA	258	10.6	2
401	935 D	NA	25202 F	NA	NA	NA	N	4.2 NA	3.19 NA	NA	NA	NA	120	11.1	4
402	943 C	NA	18993 F	NA	NA	NA	N	0.8 NA	3.01 NA	NA	NA	NA	256	10.6	3
403	1141 C	NA	24107 F	NA	NA	NA	N	2.5 NA	3.33 NA	NA	NA	NA	256	10.8	4
404	1092 C	NA	14610 F	NA	NA	NA	N	4.6 NA	3.6 NA	NA	NA	NA	337	10.4	3
405	1150 C	NA	18993 F	NA	NA	NA	N	1 NA	3.64 NA	NA	NA	NA	340	10.6	3
406	703 D	NA	16802 F	NA	NA	NA	N	4.5 NA	2.68 NA	NA	NA	NA	219	11.5	4
407	1129 C	NA	19724 M	NA	NA	NA	N	1.1 NA	3.69 NA	NA	NA	NA	220	10.8	3
408	1086 C	NA	18628 F	NA	NA	NA	S	1.9 NA	3.17 NA	NA	NA	NA	162	10.7	3
409	1067 C	NA	15706 F	NA	NA	NA	N	0.7 NA	3.73 NA	NA	NA	NA	214	10.8	3
410	1072 C	NA	14245 F	NA	NA	NA	N	1.5 NA	3.81 NA	NA	NA	NA	255	10.8	3
411	1119 C	NA	18628 F	NA	NA	NA	N	0.6 NA	3.57 NA	NA	NA	NA	286	10.6	3
412	1097 C	NA	24472 F	NA	NA	NA	N	1 NA	3.58 NA	NA	NA	NA	244	10.8	3
413	989 C	NA	12784 F	NA	NA	NA	N	0.7 NA	3.23 NA	NA	NA	NA	312	10.8	3
414	681 D	NA	24472 F	NA	NA	NA	N	1.2 NA	2.96 NA	NA	NA	NA	174	10.9	3
415	1103 C	NA	14245 F	NA	NA	NA	N	0.9 NA	3.83 NA	NA	NA	NA	180	11.2	4
416	1055 C	NA	20819 F	NA	NA	NA	N	1.6 NA	3.42 NA	NA	NA	NA	143	9.9	3
417	691 C	NA	21185 F	NA	NA	NA	N	0.8 NA	3.75 NA	NA	NA	NA	269	10.4	3
418	976 C	NA	19358 F	NA	NA	NA	N	0.7 NA	3.29 NA	NA	NA	NA	350	10.6	4

4.3 PARAMETERS USED:

FOR KNN :

The hyper parameter ‘k’ was tuned to find the optimal value. We consider value of k ranging from 1 to 50.

FOR SVM :

For SVM parameter tuning we have considered C VALUES as [0.1, 1, 10, 100] and KERNEL as [LINEAR, POLY, RBF, SIGMOID]

FOR DT:

For Decision Tree parameter tuning we have considered depth values ranging from 1 to 16

4.4 PERFORMANCE METRICS:

Accuracy: The accuracy metric is one of the simplest Classification metrics to implement, and it can be determined as the number of correct predictions to the total number of predictions.

Confusion Matrix: A confusion matrix is a tabular representation of prediction outcomes of any binary classifier, which is used to describe the performance of the classification model on a set of test data when true values are known. States of confusion matrix

1. **True Positive(TP):** In this case, the prediction outcome is true, and it is true in reality, also.
2. **True Negative(TN):** in this case, the prediction outcome is false, and it is false in reality, also.
3. **False Positive(FP):** In this case, prediction outcomes are true, but they are false in actuality.
4. **False Negative(FN):** In this case, predictions are false, and they are true in actuality.

Precision: The precision metric is used to overcome the limitation of Accuracy. The precision determines the proportion of positive prediction that was actually correct. It can be calculated as the True Positive or predictions that are actually true to the total positive predictions (True Positive and False Positive).

Recall or Sensitivity aims to calculate the proportion of actual positive that was identified incorrectly. It can be calculated as True Positive or predictions that are actually true to the total number of positives, either correctly predicted as positive or incorrectly predicted as negative (true Positive and false negative).

F-Score- F-score or F1 Score is a metric to evaluate a binary classification model on the basis of predictions that are made for the positive class. It is calculated with the help of Precision and Recall. It is a type of single score that represents both Precision and Recall. So, ***the F1 Score can be calculated as the harmonic mean of both precision and Recall, assigning equal weight to each of them.***

4.5 Preprocessing

In an ideal world, our dataset would be perfect and without any problems. Unfortunately, real-world data will always present some issues that need to be addressed. Our dataset consists of null values in the columns Drugs, Ascites, Cholesterol, Copper, Alk-Phos, SGOT, Tryglicerides, Platelets, Prothrombin, Stage, Drug, Hepatomegaly, and Spiders. We have filled the null values with modal values of that column. To normalize [2] data we have used MinMaxScaler. We have used pandas getdummies function for encoding categorical columns and label encoder for target column i.e., Status.

4.6 Results and Analysis

Maximum accuracy, Average Accuracy, and Standard Deviation obtained after 20 simulations for each classifiers (KNN, SVM, DT) are put into table:

TEST SIZES	MAXIMUM ACCURACY	MEAN ACCURACY	STANDARD DEVIATION
0.1	76.19	66.55	6.80
0.2	75	66.19	4.74
0.3	72.22	65.56	3.72

TABLE 4.1 : FOR RESULTS OF KNN (20 SIMULATIONS)

TEST SIZES	MAXIMUM ACCURACY	MEAN ACCURACY	STANDARD DEVIATION
0.1	70.83	78.57	5.20
0.2	78.57	70.71	4.28
0.3	76.19	70.67	3.33

TABLE 4.2 : FOR RESULTS OF DT (20 SIMULATIONS)

TEST SIZES	HIGHEST ACCURACY	MEAN ACCURACY	STANDARD DEVIATION
0.1	85.71	74.05	6.43
0.2	80.95	74.35	4.23
0.3	80.16	75.00	2.69

TABLE 4.3 : FOR RESULTS OF SVM (20 SIMULATIONS)

The Dataset is split into 90:10, 80:20, 70:30 where 90%, 80%, 70%, of data were used for training purposes while the remaining 10%, 20%, 30% were used for testing purposes.

4.7 CLASSIFICATION REPORT

The classification report is a tool used in machine learning to evaluate the performance of a classification model. It displays various metrics such as precision, recall, and F1 score for each class in the classification problem.

The precision is the ratio of true positives to the total number of predicted positives. It measures the accuracy of the positive predictions made by the model. The recall is the ratio of true positives to the total number of actual positives. It measures the ability of the model to correctly identify positive samples. The F1 score is the harmonic mean of precision and recall, and it provides a balanced measure of the model's performance.

	precision	recall	f1-score	support
0	0.66	0.83	0.73	23
1	0.00	0.00	0.00	3
2	0.69	0.56	0.62	16
accuracy			0.67	42
macro avg	0.45	0.46	0.45	42
weighted avg	0.62	0.67	0.64	42

Fig 4.1 CLASSIFICATION REPORT OF KNN FOR TEST SIZE 0.1

	precision	recall	f1-score	support
0	0.62	0.70	0.65	23
1	0.00	0.00	0.00	3
2	0.56	0.56	0.56	16
accuracy			0.60	42
macro avg	0.39	0.42	0.41	42
weighted avg	0.55	0.60	0.57	42

Fig 4.2 CLASSIFICATION REPORT OF KNN FOR TEST SIZE 0.2

	precision	recall	f1-score	support
0	0.64	0.79	0.71	70
1	0.00	0.00	0.00	8
2	0.58	0.46	0.51	48
accuracy			0.61	126
macro avg	0.41	0.41	0.41	126
weighted avg	0.58	0.61	0.59	126

FIG 4.3 CLASSIFICATION REPORT OF KNN FOR TEST SIZE 0.3

	precision	recall	f1-score	support
0	0.63	0.83	0.72	23
1	1.00	0.33	0.50	3
2	0.55	0.38	0.44	16
accuracy			0.62	42
macro avg	0.73	0.51	0.55	42
weighted avg	0.63	0.62	0.60	42

FIG 4.4 CLASSIFICATION REPORT OF DT FOR TEST SIZE 0.1

	precision	recall	f1-score	support
0	0.68	0.83	0.75	47
1	0.00	0.00	0.00	5
2	0.59	0.50	0.54	32
accuracy			0.65	84
macro avg	0.43	0.44	0.43	84
weighted avg	0.61	0.65	0.63	84

FIG 4.5 CLASSIFICATION REPORT OF DT FOR TEST SIZE 0.2

	precision	recall	f1-score	support
0	0.83	0.77	0.80	70
1	0.50	0.38	0.43	8
2	0.69	0.79	0.74	48
accuracy			0.75	126
macro avg	0.67	0.65	0.66	126
weighted avg	0.76	0.75	0.75	126

FIG 4.6 CLASSIFICATION REPORT OF DT FOR TEST SIZE 0.3

	precision	recall	f1-score	support
0	0.66	0.83	0.73	23
1	0.00	0.00	0.00	3
2	0.69	0.56	0.62	16
accuracy			0.67	42
macro avg	0.45	0.46	0.45	42
weighted avg	0.62	0.67	0.64	42

FIG 4.7 CLASSIFICATION REPORT OF SVM FOR TEST SIZE 0.1

	precision	recall	f1-score	support
0	0.67	0.89	0.76	47
1	0.00	0.00	0.00	5
2	0.76	0.50	0.60	32
accuracy			0.69	84
macro avg	0.48	0.46	0.46	84
weighted avg	0.66	0.69	0.66	84

FIG 4.8 CLASSIFICATION REPORT OF SVM FOR TEST SIZE 0.2

	precision	recall	f1-score	support
0	0.73	0.87	0.80	70
1	0.00	0.00	0.00	8
2	0.70	0.62	0.66	48
accuracy			0.72	126
macro avg	0.48	0.50	0.49	126
weighted avg	0.67	0.72	0.69	126

FIG 4.9 CLASSIFICATION REPORT OF SVM FOR TEST SIZE 0.3

4.7 CONFUSION MATRIX :

A **confusion matrix** is a matrix that summarizes the performance of a machine learning model on a set of test data. It is a means of displaying the number of accurate and inaccurate instances based on the model's predictions. It is often used to measure the performance of classification models, which aim to predict a categorical label for each input instance.

The matrix displays the number of instances produced by the model on the test data.

- **True positives (TP):** occur when the model accurately predicts a positive data point.
- **True negatives (TN):** occur when the model accurately predicts a negative data point.
- **False positives (FP):** occur when the model predicts a positive data point incorrectly.
- **False negatives (FN):** occur when the model mispredicts a negative data point.

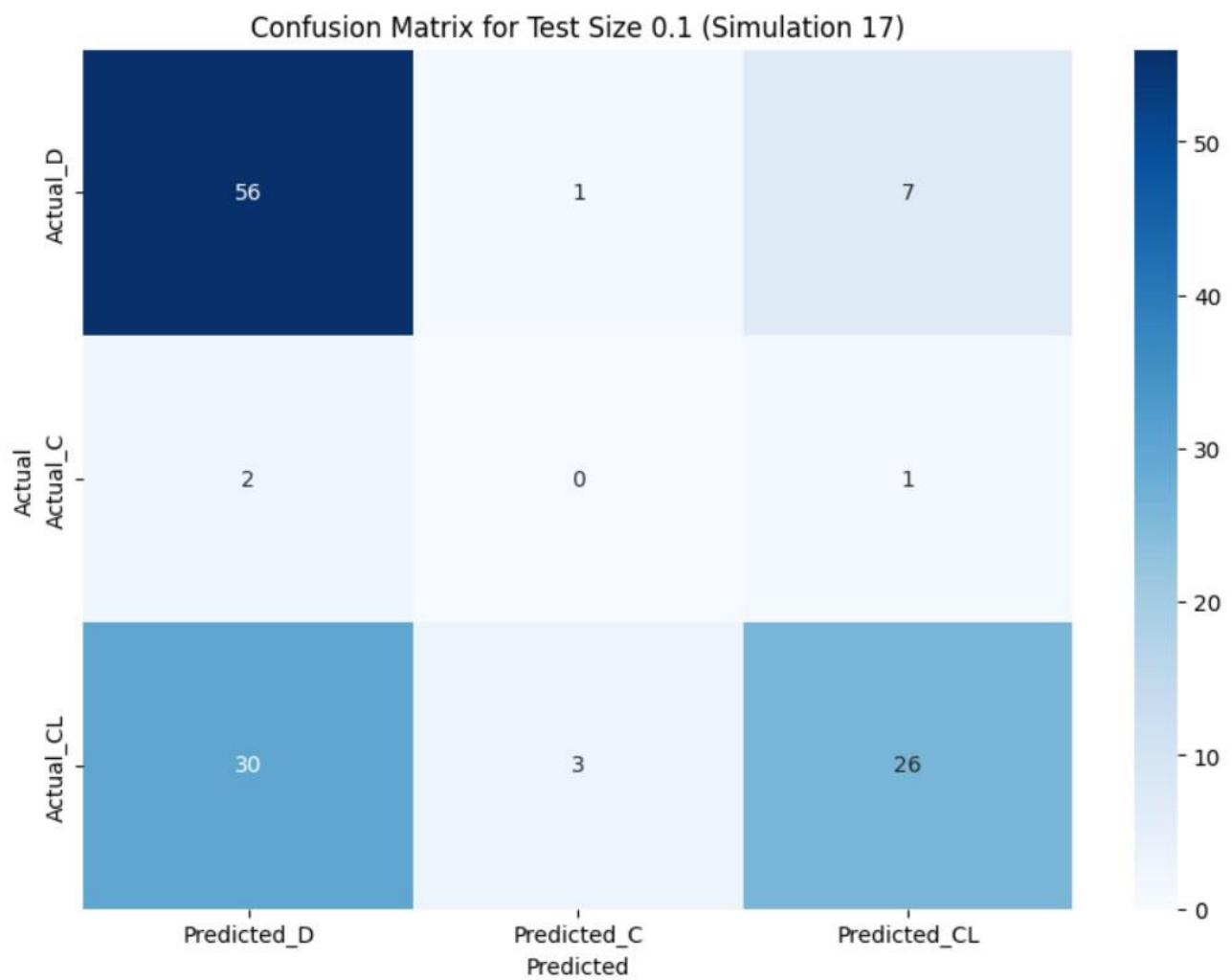


Fig 4.10 CONFUSION MATRIX FOR HIGHEST ACCURACY FOR TEST SIZE 0.1 FOR KNN

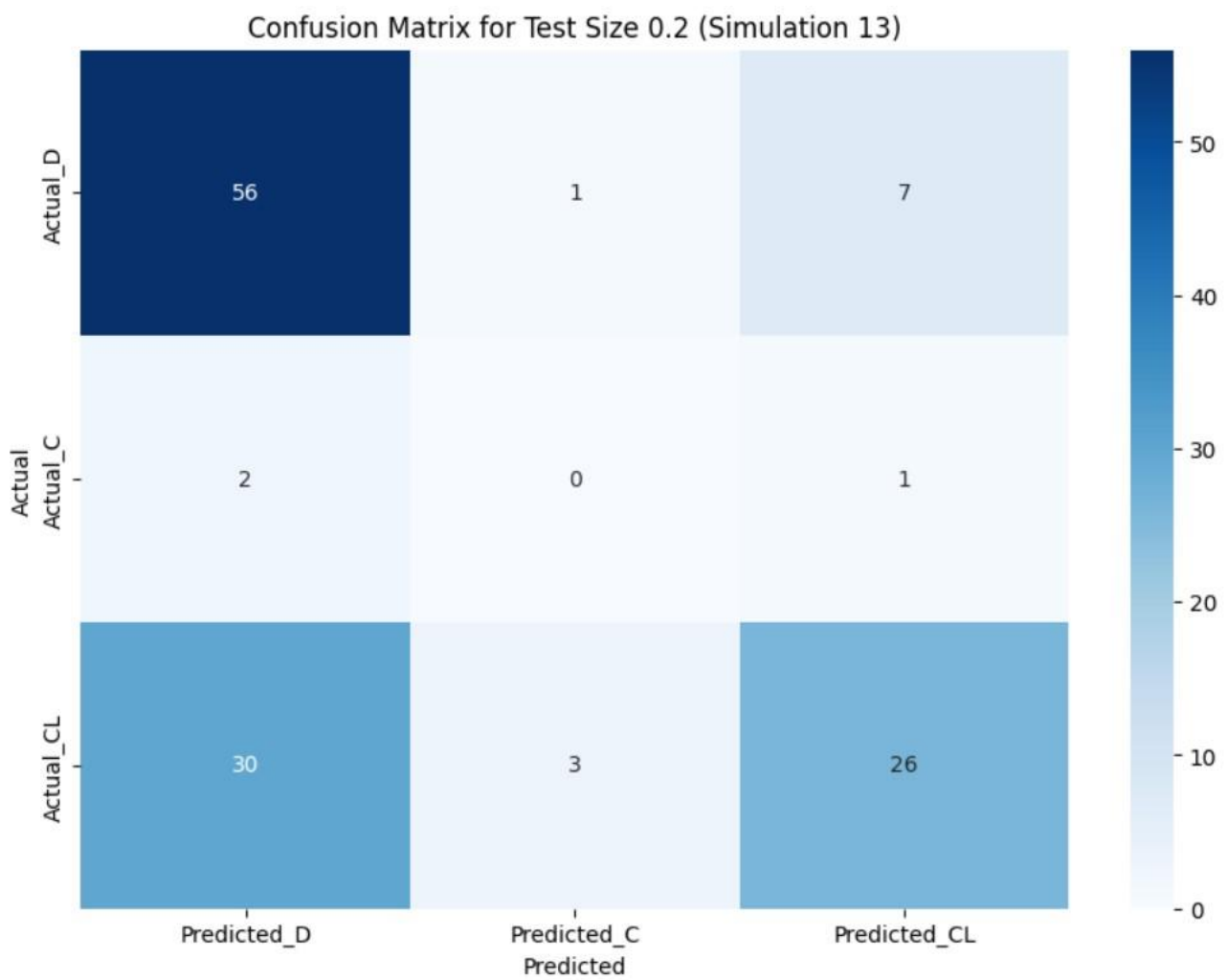


Fig 4.11 CONFUSION MATRIX FOR HIGHEST ACCURACY FOR TEST SIZE 0.2 FOR KNN

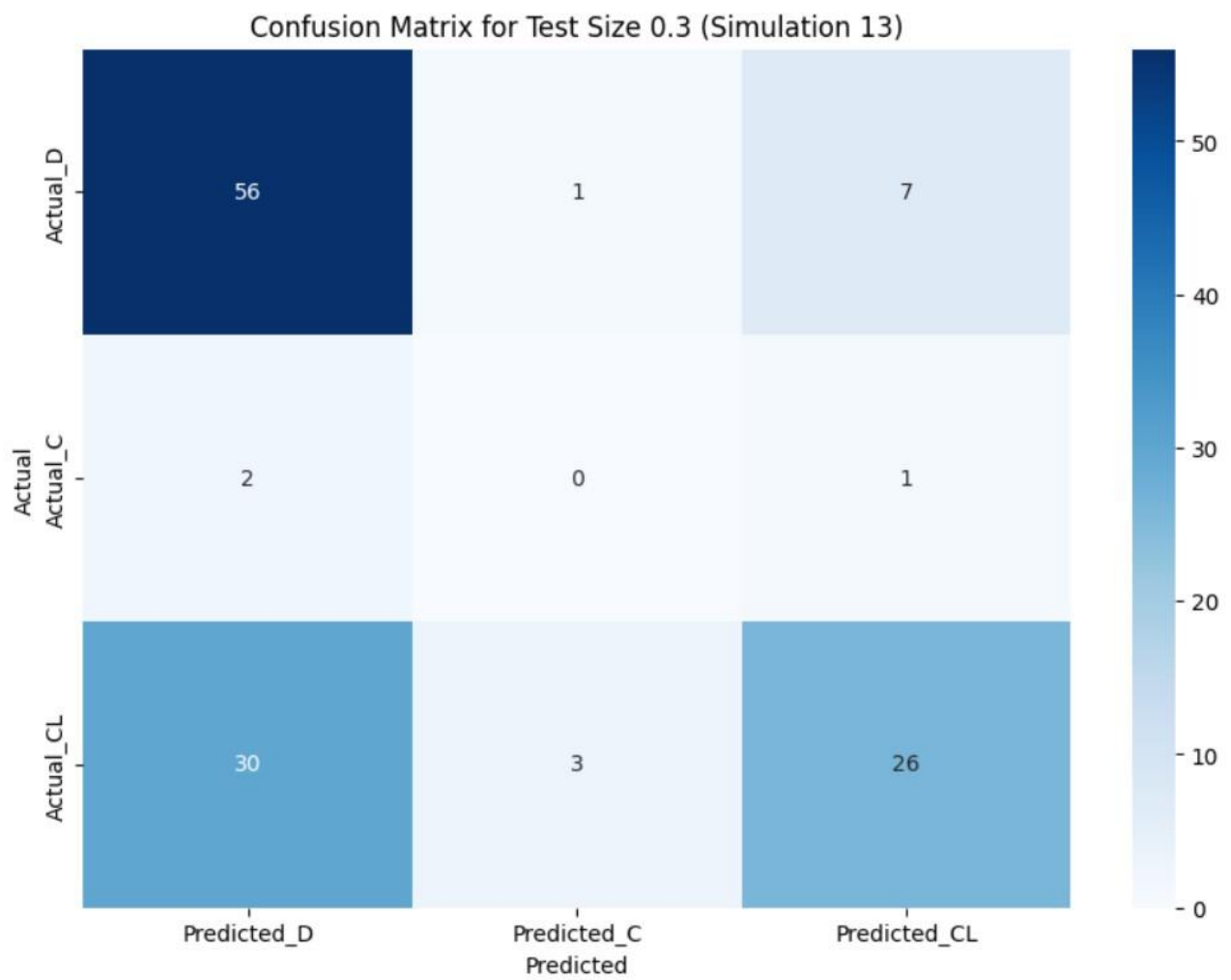


Fig 4.12 CONFUSION MATRIX FOR HIGHEST ACCURACY FOR TEST SIZE 0.3 FOR KNN

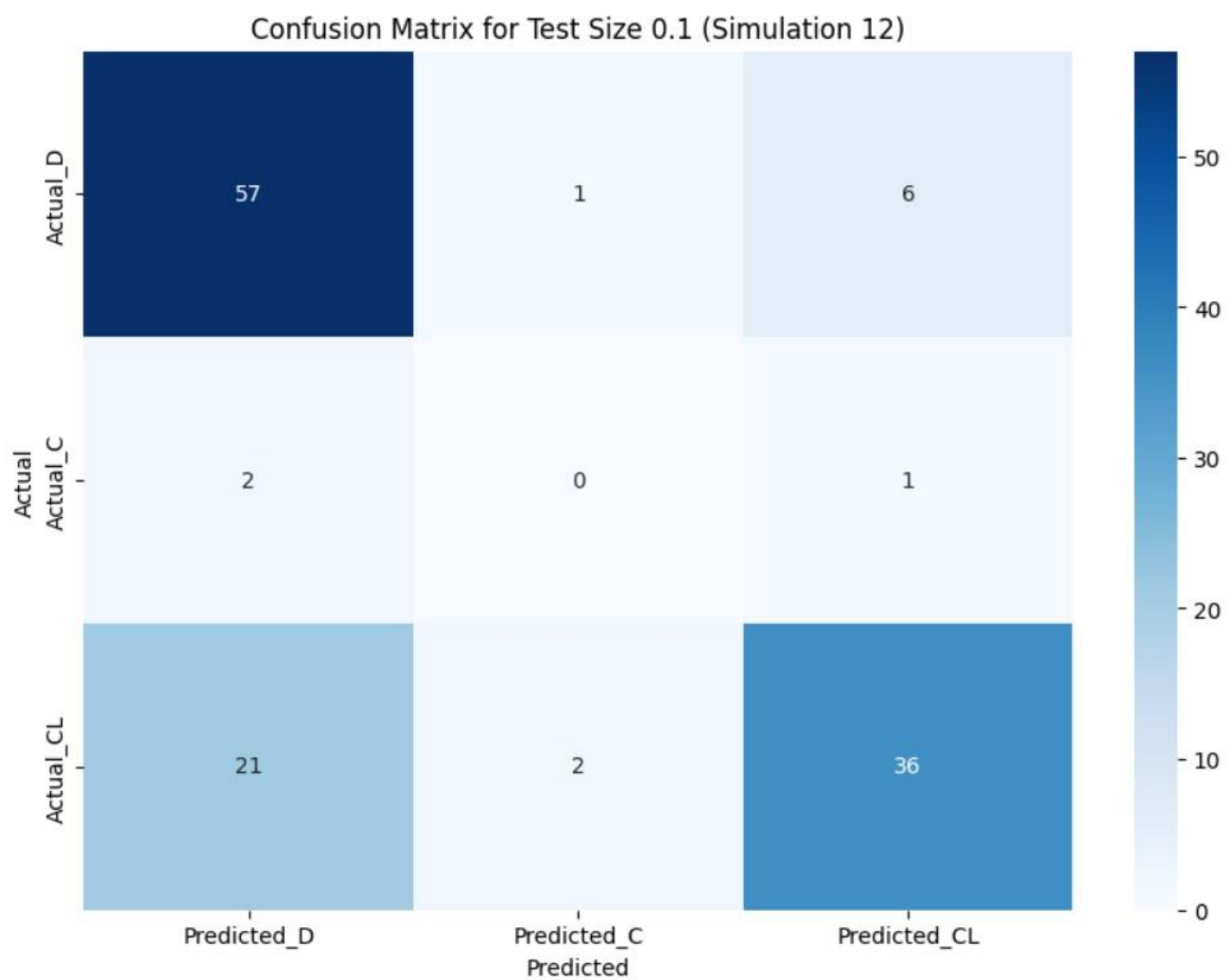


Fig 4.13 CONFUSION MATRIX FOR HIGHEST ACCURACY FOR TEST SIZE 0.1 FOR DT

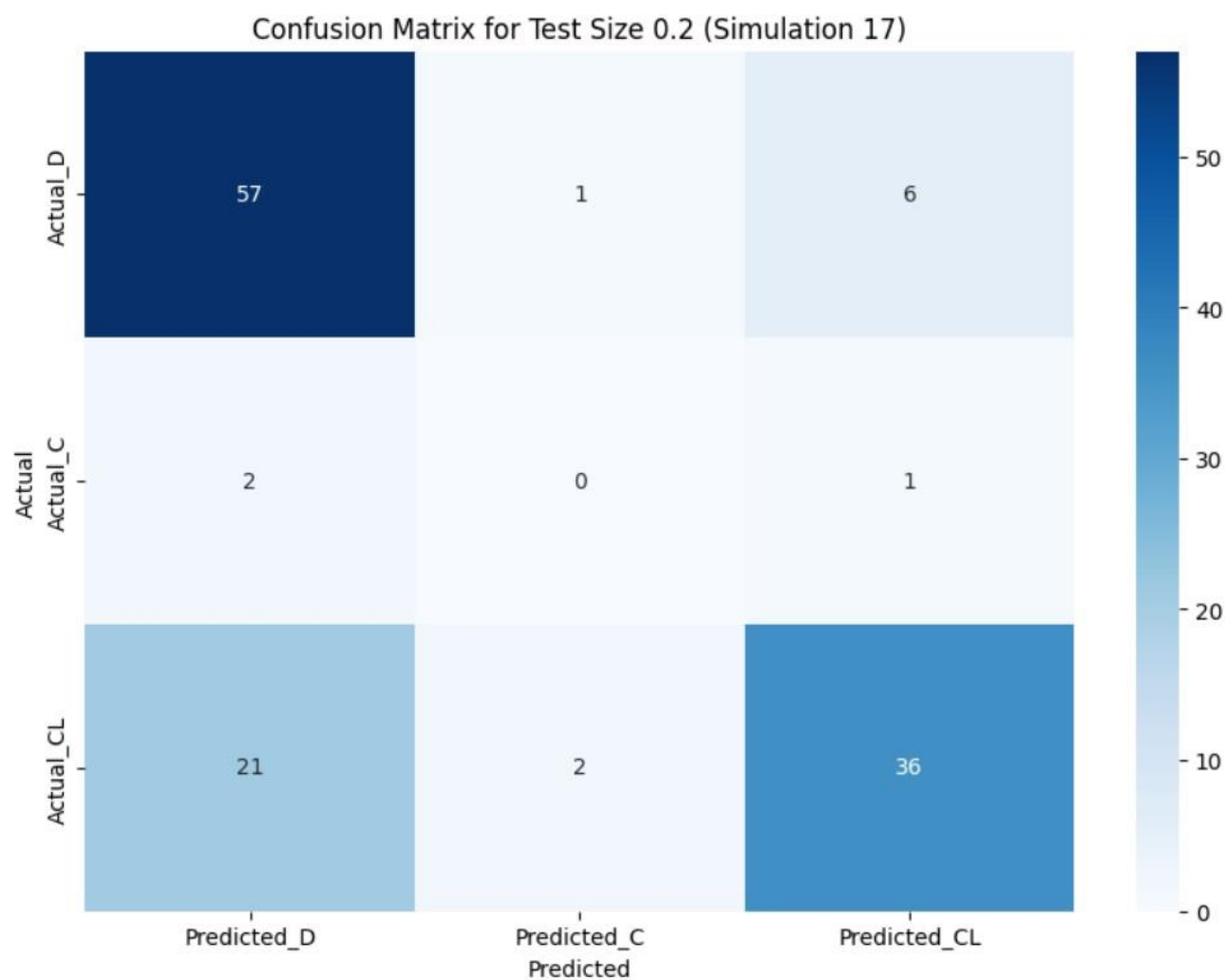


Fig 4.14 CONFUSION MATRIX FOR HIGHEST ACCURACY FOR TEST SIZE 0.2 FOR DT

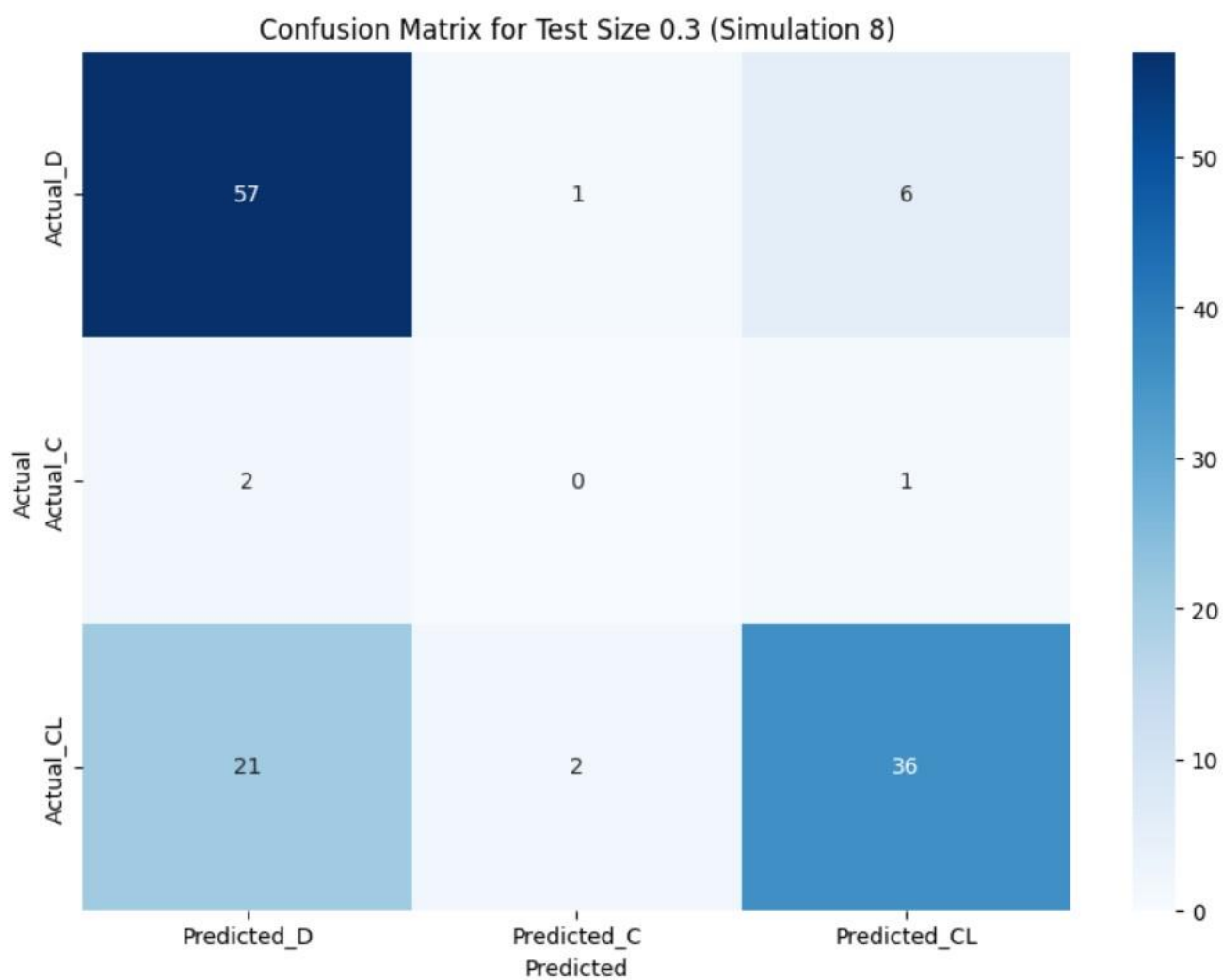


Fig 4.15 CONFUSION MATRIX FOR HIGHEST ACCURACY FOR TEST SIZE 0.3 FOR DT

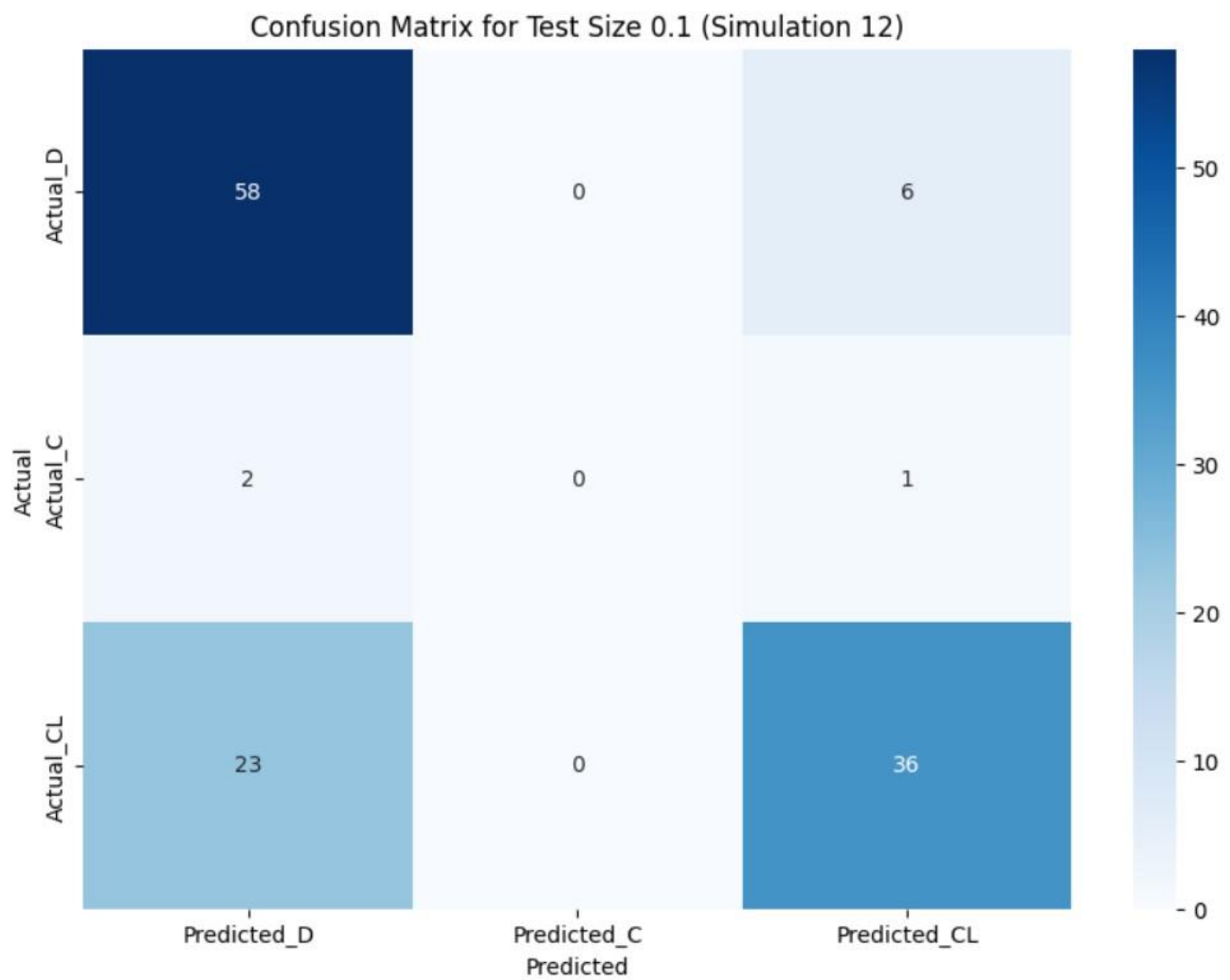


Fig 4.16 CONFUSION MATRIX FOR HIGHEST ACCURACY FOR TEST SIZE 0.1 FOR SVM

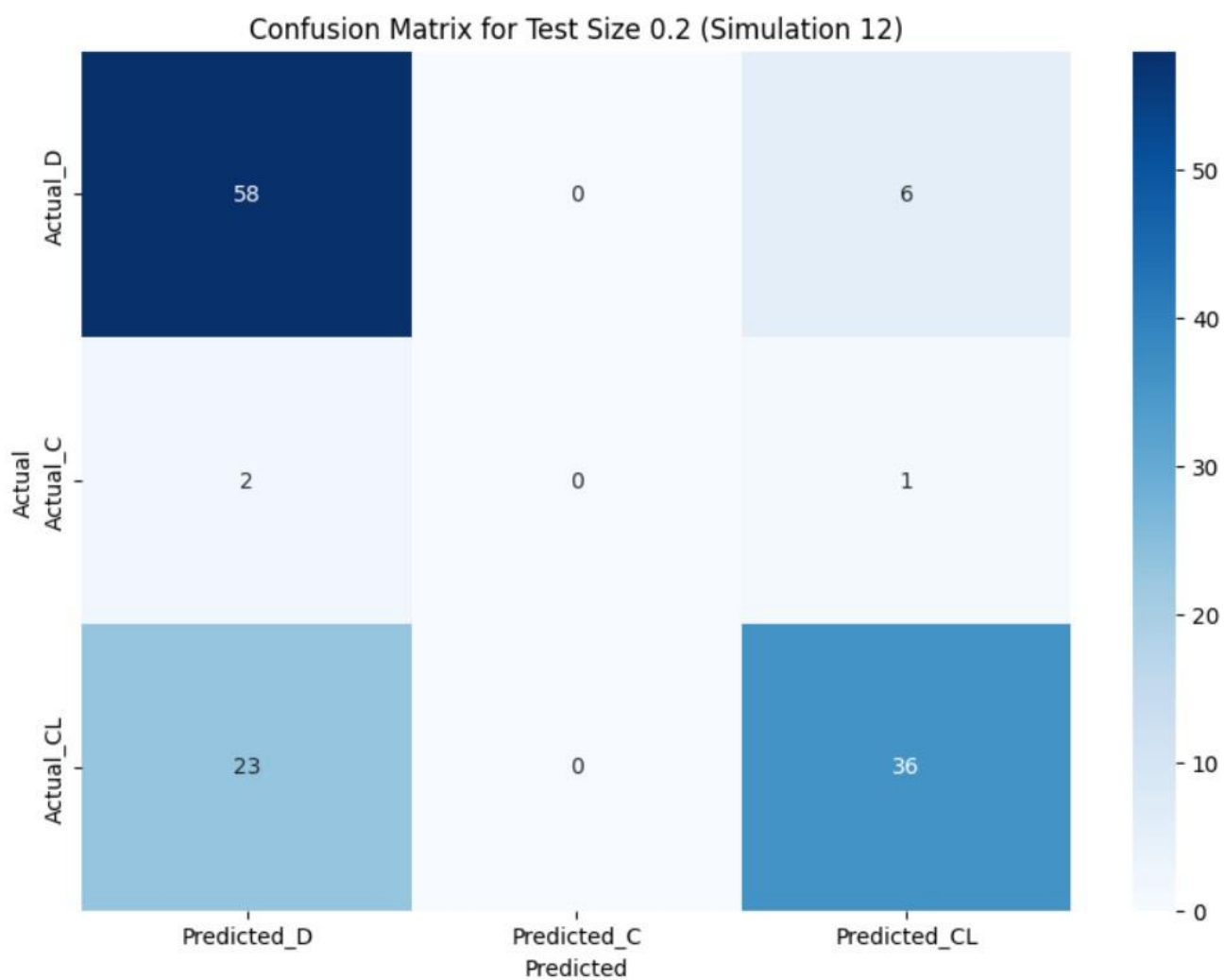


Fig 4.17 CONFUSION MATRIX FOR HIGHEST ACCURACY FOR TEST SIZE 0.2 FOR SVM

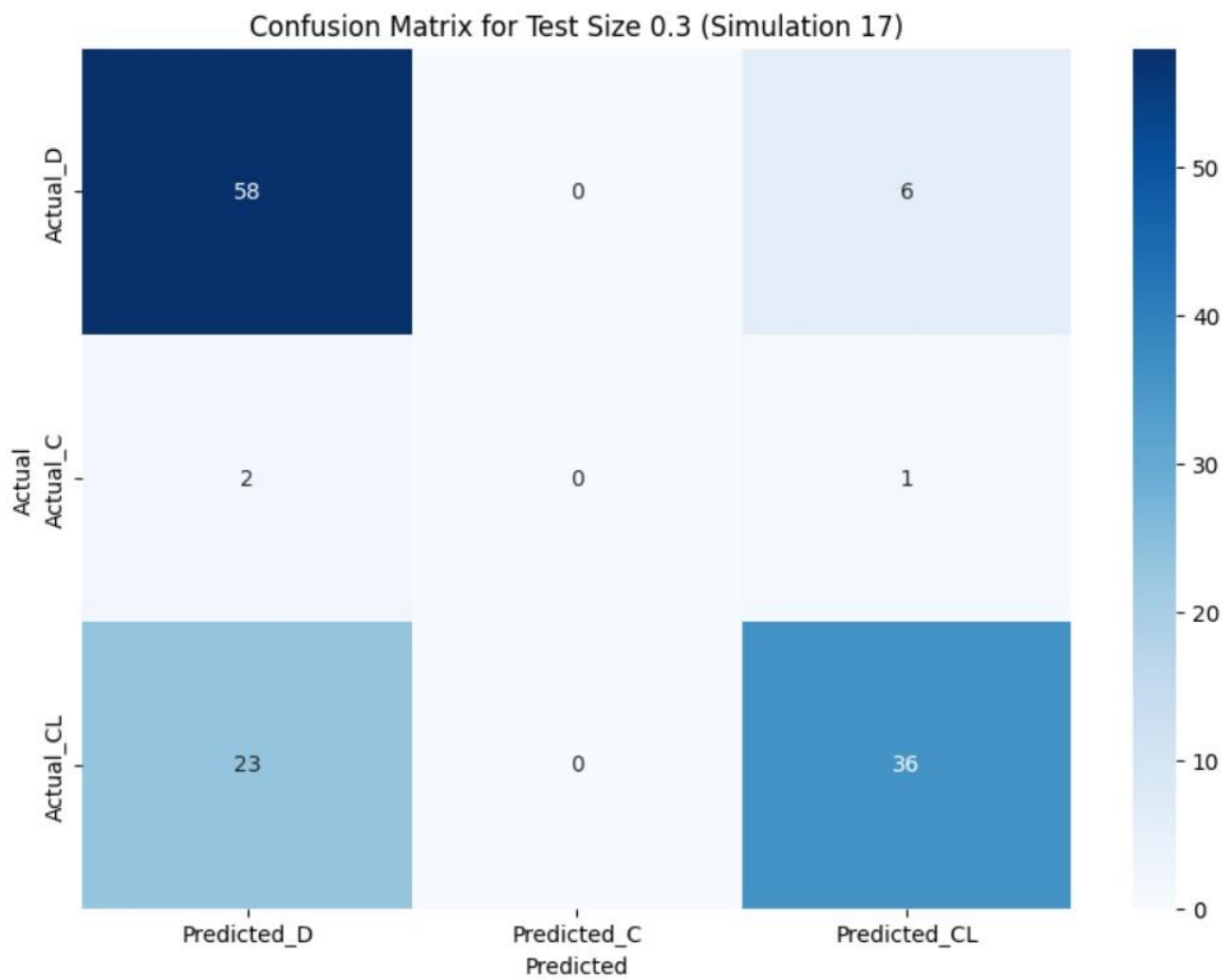


Fig 4.18 CONFUSION MATRIX FOR HIGHEST ACCURACY FOR TEST SIZE 0.3 FOR SVM

CHAPTER 5 : CONCLUSION

In this comprehensive study, the potential of Machine Learning as a transformative tool in the early diagnosis of Cirrhosis, a chronic liver disease with profound implications for patient prognosis and overall health, was explored. Leveraging a rich and extensive dataset spanning a decade of patient observations, the power of a diverse array of Machine Learning techniques, including KNearest Neighbors, Decision Trees, and, Support Vector Machine has been harnessed. This exploration underscores the immense promise that Machine Learning holds in the domain of Cirrhosis diagnosis, offering a promising avenue for the early identification and intervention of this debilitating condition.

The focal point of the findings rests on the exceptional performance of the Support Vector Machine (SVM) model. SVM emerged as the star of the study, exhibiting remarkable statistics with an accuracy of mean accuracy of 75% and highest accuracy of 85.71%.

LIMITATIONS:

The limitations underscore the need for ongoing research and improvements in the application of Machine Learning to Cirrhosis diagnosis. Recognizing these constraints informs the path toward more robust and clinically viable predictive models. Following are the two such notable limitations:

- 1) Time Dependency : One noteworthy limitation is the temporal aspect of Cirrhosis progression. Our dataset, derived from the Mayo Clinic study spanning from 1974 to 1984, may not fully encapsulate the dynamic nature of Cirrhosis over time. Future studies should consider more recent and time-resolved data to enhance the temporal accuracy of predictions.
- 2) Real-time Predictions and Integration:
While this study has established the predictive capabilities of the different models, the logical next step is to translate this capability into real-time clinical settings. Developing a user-friendly interface or application that can take input parameters from medical practitioners and instantly provide a risk assessment for Cirrhosis would immensely aid in early detection and intervention.
- 3) Feature Selection : Exploring advanced feature engineering techniques and rigorous feature selection methods can potentially improve the predictive performance of the model.
- 4) Multi-class Prediction and Disease Staging: Extend the scope of the research to include multi-class prediction, where the model can classify Cirrhosis into different stages or severity levels.

REFERENCES

- 1) Andreas C. M & Sarah G, “Introduction to Machine Learning with Python” O’Reilly Media, Inc., 2017.
- 2) Giuseppe Bonaccorso, “Machine Learning Algorithms”, Second Edition, Packt Publishing, 2018.
- 3) Y. Sun, K. Haapanen, and M. E. Gershwin et al, "Clin Rev Allergy Immunol", doi : 10.1007/s12016-014-8449-4
- 4) P. Aryan, et al, " A study of Liver Cirrhosis to predict its Stage " International Journal of Advance Research, Ideas And Innovations In Technology, ISSN 2454-132X Volume 7, Issue - 5 pp. 1-2, 2021
- 5) Proceedings of the International Conference on Advanced Computing & Communication Technologies (ICACCTech), pp. 1-2, 2023
- 6) J. Han, J. Pei, and, H. Tong, “Data Mining Concepts And Techniques”, Elsevier Inc., 2023.
- 7) <https://archive.ics.uci.edu/dataset/878/cirrhosis+patient+survival+prediction+dataset-1> as of 27.05.2024