ISSN 2475-5451

Research Article

International Journal of Diabetes & Metabolic Disorders

Clinical Usefulness of Machine Learning Approaches as a Non-Invasive Technology in Reducing Hepatitis Disease Mortality

Mohsen Ghorbian*

Department of Computer Engineering, Islamic Azad University, Iran.

***Corresponding Author**
Mohsen Ghorbian, Department of Computer Engineering, Islamic Azad University, Iran.

Submitted: 2023, May 30; **Accepted:** 2023, Jun 30; **Published:** 2023, July 03

Citation: Ghorbian, M. (2023). Clinical Usefulness of Machine Learning Approaches as a Non-Invasive Technology in Reducing Hepatitis Disease Mortality. *Int J Diabetes Metab Disord,* 8(3), 390-399.

Abstract

Hepatitis is a viral infection that causes inflammation of the liver. However, other factors can cause the disease, including secondary effects from drugs, toxins, alcohol, and autoimmune hepatitis. The autoimmune form of the disease occurs when the body produces antibodies against the liver tissue, and many people worldwide are affected by it. Various clinical factors and parameters are examined in diagnosing hepatitis disease, which is analyzed by performing various tests of these factors and parameters. As a result of the vastness of the parameters under examination, it is challenging and complicated for the experts in this field to perform the analysis process on these parameters on a large scale. Healthcare experts can now identify the factors influencing the death rate of patients with high speed and accuracy thanks to emerging technologies such as machine learning, which is a subset of artificial intelligence. In this study, KNN and SVM machine learning techniques were used to analyze the positive effect of clinical parameters such as LIVER BIG, LIVER FIRM, SPLEEN PALPABLE, and ANOREXIA on patients' survival or death rates. This study investigates and analyzes the results of the implementation in two parts. The first part deals with determining the positive impact of these clinical parameters on the death and survival rate of patients, and the second part examines the performance of machine learning techniques based on the evaluation criteria of accuracy (ACC), error rate (ERR), specificity (SPE), and negative prediction value (NPV).Based on the implementation finding of machine learning techniques on data related to hepatitis patients, it has been determined that patients with positive LIVER BIG, LIVER FIRM, SPLEEN PALPABLE, and ANOREXIA clinical parameters can have a high chance of survival. On the other side, The SVM technique outperformed the KNN technique by ACC 94.05%, ERR 16.02%, SPE 93.07%, and NPV 85.7% in an analysis of the performance of machine learning techniques.

Keywords: Hematite Disease, Machine Learning (Ml), Support Vector Machine (Svm), K-Nearest Neighbor (Knn), Inflammation Liver

1. Introduction

Using computers simultaneously with technological advances in various disciplines has established the foundation for creating digital information that can be stored as valuable data in data warehouses under suitable structures [1]. Data storage in the data warehouse can be considered the first phase of using the generated data; therefore, this stored data should be utilized. This enormous volume of information is beyond the capacity of human beings to process and use, as it contains hidden relationships and patterns that can only be analyzed and used by using technologies developed explicitly for this purpose. Using these analyses can be considered as planning and decision for the future, which can lead to significant results [2]. Hence, data analysis technology is one of the most critical and practical technologies to facilitate knowledge discovery. For this purpose, Machine Learning (ML) technology is used in information analysis [3]. The field of application of ML cannot be limited to one or two fields because this technology can be used wherever there is data, and it can be used for data analysis. Still, the data related to the healthcare field that is very sensitive and critical also can be analyzed and explored [4]. In the analysis of healthcare domain data, it should be noted that since these data pertain to human health, the results should be accurate and reliable if ML techniques are applied [5]. A ML technique applied to patient data can be used to find hidden patterns and information within them [6]. Finding accurate results can reduce detection duration, reduce cost imposed, improve the treatment process, and increase patients' survival chances [7]. Considering the wide variety of existing ML techniques, it is necessary to choose techniques that can provide results with appropriate precision and accuracy; therefore, it is essential to notice a variety of factors for the choice of a technique that can provide information with a high degree of accuracy and sensitivity, along with robust analysis [8]. Hence, to apply ML techniques to the health field, it is necessary to use techniques that display the accuracy and correctness of the results at the highest level possible so that the results can be verified [9]. In light of the importance of the results obtained from applying ML techniques to healthcare data, it is essential to select reliable techniques when presenting the results [10]. Comparing these techniques is necessary to select a technique that will provide reliable and proper performance. The comparison of techniques is made through measurement criteria. A technique that can show acceptable results using these criteria and provides a more appropriate performance than other techniques can be selected as the superior technique. In the decision-making process, the results of that technique can be trusted [11]. This study examines factors and symptoms such as ANOREXIA, BIG LIVER, FIRM LIVER, and PALPABLE SPLEEN under ML techniques to determine their effect on diagnosing hepatitis. On the other hand, the second aim of this research is to compare the performance of two K-nearest neighbor (KNN) ML techniques and Support vector machine (SVM) based on four criteria: accuracy (ACC), error rate (ERR), specificity (SPE), and negative prediction value (NPV), which, finally, a technique with a better performance also be selecting.

2. Related Work

In light of the advancement of science and the emergence of advanced technologies, it can be said that technology has become a priority in all fields due to simplifying and making problem solving more straightforward and accessible. Technology has become an integral part of every field, and this characteristic is not restricted to one particular field. Healthcare is one of the many fields in which humans have always been involved, and the experts in this field have always sought to improve the quality of life of their patients. To improve the quality of human life, these experts have always aimed to prevent, increase the speed of disease diagnosis, reduce the cost of diseases, and slow down the process and growth rate of diseases. As technology has advanced in this field, it has also led to various benefits, such as the developing of advanced surgical instruments and various diagnostic tests. Hence, this has contributed to improving the quality of life of patients. ML technology is one of a variety of technologies that, through exploring patient-related information and discovering the relationships between the information, create patterns that provide experts in this field with a strategic perspective for achieving patterns based on the relationships discovered between the information. Because of these patterns, they can diagnose or predict diseases more quickly, reducing costs for the patient and improving hope that they will recover more quickly. Hence, it is natural that experts would benefit from the ability to gain information with high accuracy and precision. Therefore, the strategy for making models and introducing the best performance must be reliable. Therefore, ML experts have recently attempted to analyze patient data in collaboration with medical experts. Rajeswari and Reena [2010] proposed a new approach and implemented it on data on individuals with liver problems. They achieved a new model with the desired level of prediction accuracy. By implementing their proposed method on 345 samples of data pertaining to patients with liver problems, they could accurately predict the causes of liver problems [12]. Nguyen et al. [2007] proposed a novel method by which they could significantly enhance the accuracy of the predicted model compared to other methods. They applied their proposed method to the dataset of patients with hepatitis disease. Based on the constructed prediction models, they could finally identify the factors influencing the incidence of hepatitis accurately. The findings of their study will significantly aid physicians in reducing the time required to diagnose diseases [13]. A study conducted by Kim et al. [2007] applied support vector machine (SVM) and decision tree ML techniques to liver disease patient data. Their understanding of the liver's sensitivity enabled them to anticipate the likelihood that this sensitivity may evolve into chronic hepatitis. In light of the results of this research, physicians will be able to detect hepatitis in susceptible patients with liver problems more quickly, resulting in a significantly higher chance of patient survival [14]. Radwan et al. [2013] utilized a wide variety of techniques for ML based on decision trees to analyze the data of patients with hepatitis. By creating models, they could predict the outcomes of using antiviral drugs for hepatitis and the complications that occurred as a result. This research can lead to a reduction in the costs imposed on the patient and an increase in the recovery process of the disease [15]. Yokoi et al. [2005], implementing ML techniques on the information obtained from the urine tests of patients with hepatitis and analyzing the results, were able to achieve a model that can assist experts in the field of hepatitis treatment and the treatment process of this disease accelerate [16]. Sato et al. [2002] accomplished patterns by implementing the decision tree technique on the information of hepatitis patients and analyzing the obtained results. Then, by combining the obtained patterns, these researchers were able to achieve a mechanism that could make the work of experts very easy because to obtain such results through testing, experts needed to perform very hard and long-term tests, so the process of diagnosing hepatitis can be done more quickly [17].

3. Implementation Process

One of the fundamental challenges in ML technology is choosing a technique appropriate for the field of activity. Specifically, when the data is related to healthcare, the significance of the results is doubled as experts make decisions about how to treat or diagnose diseases based on this information, which can directly impact an individual's health. So it is essential to select techniques that can provide the highest level of confidence and correctness in the obtained results to combine the correct decision of the situation with the best analysis of the situation and to obtain the best possible result. During this research, we have attempted to select two k-nearest neighbor techniques and SVM for selecting these two techniques. The reason for this was the widespread use and the high quality of the analyses provided by these two techniques, which can be considered the reason for the choice. An SVM is a ML algorithm that can be applied to regression and classification problems and has a supervised approach. Due to its robustness, it is commonly used to solve classification problems. This algorithm first represents the data points in an n-dimensional space. Based on statistical approaches, the algorithm determines the best line that distinguishes between the various classes in the data. [18]. The K-nearest neighbor technique divides the data into groups based on their characteristics. When a new sample is being examined, this sample is placed in the group most similar to the sample currently under examination. This technique provides an example of a simple data classification technique. Hence, using the K-nearest neighbor technique, the accuracy of the analysis results can be enhanced by placing a new sample in the classification close to that of the previous data and using the classification information from the previous data

[19]. 'Hepatitis' in the dictionary refers to liver swelling, which can be caused by various factors, such as smoking, drinking alcohol, or using chemical substances. It is essential to point out that hepatitis can have different types, among which there are the following:

- Hepatitis type A
- Hepatitis type B
- Hepatitis type C

It is possible to distinguish different types of Hepatitis based on the type of virus responsible for inflaming and swelling the liver. People can protect themselves from contracting Hepatitis type A and Hepatitis type B diseases by vaccinating themselves against them. However, people cannot be protected against contracting Hepatitis type C disease through vaccination as no vaccine is available. As a result, a person could contract two or more types of Hepatitis simultaneously. The transmission of hepatitis type C is through blood directly or through tattoos, drug use, or traditional medicine. Hepatitis type C is considered one of the most dangerous types of Hepatitis [20]. The data used in this research are related to 150 patients with hepatitis. A prestigious university in the United States, the University of California at Irvine, has made it available on the Kaggle website. This information can be classified into different categories based on its content, and this dataset provided information about gender and the mortality rate for those individuals who have undergone these tests. A preliminary analysis of some important and influential factors has been conducted in this study to examine their impact on the survival and death of individuals with hepatitis disease. In this study, novel technologies, such as machine learning, are used to predict the probability of survival or death for patients with hepatitis while assessing the effectiveness of the proposed techniques based on the defined measurement criteria. Information obtained from the patients includes various agents and factors examined during the examination. Due to the diversity of existing factors and various examinations, it has been attempted to focus on four factors out of the various factors available in the patients' information as the factors that will be analyzed in this study. A patient's survival or death depends on the factors selected, which can also be called variables. Essentially, these factors are examined concerning the mortality or survival of individuals with hepatitis. As a result of the selection process, the following factors were considered:

- Anorexia
- Liver Big
- Liver Firm (fatty liver)
- Palpable Spleen

A diagram illustrating the implementation process is shown in Figure 1.

3.1 Primary Processing

ML techniques require formal and standardized information structures to function correctly and optimally. Consequently, the data related to hepatitis patients are not exempt from this rule. They should have a structure appropriate to the techniques used in the ML process to be effective. In light of this, modifying the initial raw data is necessary.

3.1.1 Standardization of Patient Information

The ML techniques selected for this research are supervised learning techniques based on classification approaches, so it is necessary to organize and display the information on hepatitis patients in columns and rows. However, because this information contains a primary structure, it must be modified to obtain the desired result with just a small amount of modification. In other words, it turned them into a standard format.

3.1.2 Convert the Results into P and N

The patients' information relates to the tests that they underwent. After converting the data to the basic standard form, some changes need to be made to the values of some features. The results of these tests are displayed as numbers. As these numbers have values of 2 and 1, the results of these tests are either positive or negative, according to the explanations given in connection with the information provided by the patient's test. In other words, when a value of 1 is placed in front of a variable, it indicates that the patient's test result for that variable was negative; on the other hand, when placed in front of a variable of value 2, it indicates that the patient's test result was positive. Therefore, to utilize optimally and increase the performance of ML techniques and to increase the readability and more straightforward understanding of the results, we have substituted the values of P and N, which represent positive and negative values, respectively, instead of the values of 2 and 1.

3.2 Feature Selection and Data Preparation

Available information on patients includes all types of examinations. In order to prepare for the implementation of ML techniques, due to a large number of experiments and the extent of clinical parameters tested, it is necessary to select some clinical parameters in the form of required features and factors for analysis.

3.2.1 Features Selection

A total of 20 clinical parameters have been analyzed and checked in the dataset of hepatitis patients. In other words, 20 clinical parameters have been analyzed and checked for hepatitis disease. In light of the fact that all of these clinical parameters are of relatively not equal importance; therefore, it is to examine them outside this research's scope. Hence, this research has attempted to select only four experimental and critical clinical parameters, and in the following, ML techniques are used to examine and analyze their effect on the death or survival of patients, and the results will be analyzed. The reason for choosing these four parameters, which can also be called factors or features, is that they have the most significant impact on the treatment process of the affected people in terms of their importance, frequency, and effectiveness. Table 1 illustrates the selected parameters.

Properties		Values
CLASS	**LIVE**	**Death**
ANOREXIA	P (Positive)	N(Negative)
LIVER BIG	P (Positive)	N(Negative)
LIVER FIRM	P (Positive)	N(Negative)
SPLEEN PALPABLE	P (Positive)	N(Negative)

Table 1: Features Selection

3.2.2 Placement of Missing Values

Among the features selected as main variables, some fields in some experiments do not have values; therefore, the absence of these values can impact the results and analyses resulting from implementing ML techniques. It is possible to complete these empty fields using the methods provided to eliminate the effects caused by the absence of some values. Many methods have been proposed for filling the empty fields; however, the averaging method has been used in this study. The empty fields in columns are completed by averaging the values of all fields in the column related to the selected feature. Table 2 shows the missing values in each of the selected features.

Features	Missing Values
SPLEEN PALPABLE	7
LIVER FIRM	14
ANOREXIA	13
LIVER BIG	9

Table 2: Missing Values for Selected Features

3.3 Modeling Process

After the data preparation stage, data modeling can be introduced as a main part of the ML implementation process. In this stage, ML techniques will be implemented on data whose structure is appropriate and standard. Modeling by ML techniques will involve obtaining prior knowledge and using this prior knowledge to make a new model. In light of this, it is possible to say that there will always be two components involved in the implementation of ML techniques, which include a part of the data used to teach the techniques and a part of the data used to make predictions based on their learning from the previous training. Hence, the more knowledge a ML technique has of datasets, the more accurate and reliable his or their predictions will be.

3.3.1 Selection and Implementation of Ml Techniques

Due to the vast number of available ML techniques, selecting techniques appropriate for the desired target is necessary before applying them to the dataset. On another side, it is necessary to prepare the data according to the selected techniques. In this research, the selected techniques are classification-based supervised techniques. Hence, the data should be organized based on these two techniques.

3.3.2 Model Creation

After preparing the data and choosing the appropriate techniques, it is necessary to implement the selected techniques on the data for the modeling process. Hence, this is the final stage in the implementation process for machine learning. In this stage, a ML model can be obtained by implementing ML techniques on the prepared data, resulting in a result. An analysis of the obtained results will be presented in this section. The analysis of the obtained results consists of two parts. The first part predicts the impact of the selected characteristics on the mortality rate of hepatitis patients. The second section will evaluate the efficiency of the ML techniques and introduce a more efficient technique. As a result, evaluation criteria can be used to assess the ML techniques' performance. Hence, this research has tried to use criteria such as ACC, ERR, SPE, and NPV to check ML techniques' performance. These criteria will assess the performance of two KNN and SVM techniques.

3.3.2.1 Specificity
An essential criterion in assessing the performance of ML techniques is specificity. As a result, to calculate the SPE criterion, the number of times the model true negative cases model correctly predicted the class is divided by the total number of false positive and true negative cases predicted. In this case, this evaluation criterion is calculated according to Equation 1 [21].

$$SPE = \frac{TN}{FP+TN} \qquad (1)$$

3.3.2.2 Accuracy
The accuracy criterion is among a model's most essential and commonly used evaluation criteria. As a result, to calculate the ACC criterion, the number of times the model correctly predicted the class is divided by the number of times the class was predicted .Hence, this evaluation criterion is calculated using Equation 2 [22].

$$ACC = \frac{TP+TN}{TP+FN+FP+TN} \qquad (2)$$

3.3.2.3 Negative Prediction Value
NPV, or negative prediction value, is a widely known evaluation criterion. NPV is obtained by calculating the ratio of predicted true negative cases to the total of predicted true negative and false negative cases. This evaluation criterion is calculated according to Equation 3 [23].

$$NPV = \frac{TN}{FN+TN} \qquad (3)$$

3.3.2.4 Error rate
The error rate evaluation criterion is the inverse of the correctness criterion's performance. Hence, to calculate the ERR criterion, the number of cases whose classes are incorrectly predicted is divided by the total number of cases predicted by the model. In other words, the ERR criterion is calculated by subtracting the accuracy criterion value from the numerical value of 1. This evaluation criterion is calculated using Equation 4 or 5 [24].

$$ERR = \frac{FN+FP}{TP+FN+FP+TN} \qquad (4)$$

$$ERR = 1 - ACC \qquad (5)$$

The constituent components of equations 1 through 4 represent predictions reached by techniques implementation on samples and created models. Those parameters are as follows:
- TP: Positive samples True identified.
- TN: Negative samples True identified.
- FP: Positive samples False identified.
- FN: Negative samples False identified.
-

3.4 Results Analysis
The purpose of implementing this research is to assess the effect of the selected characteristics on the mortality rate of people with hepatitis. Therefore, each characteristic is analyzed individually. By using evaluation criteria, it attempted to examine and analyze the performance of ML techniques and to introduce a technique with superior performance. Therefore, at this stage, it is attempted to analyze the obtained results.

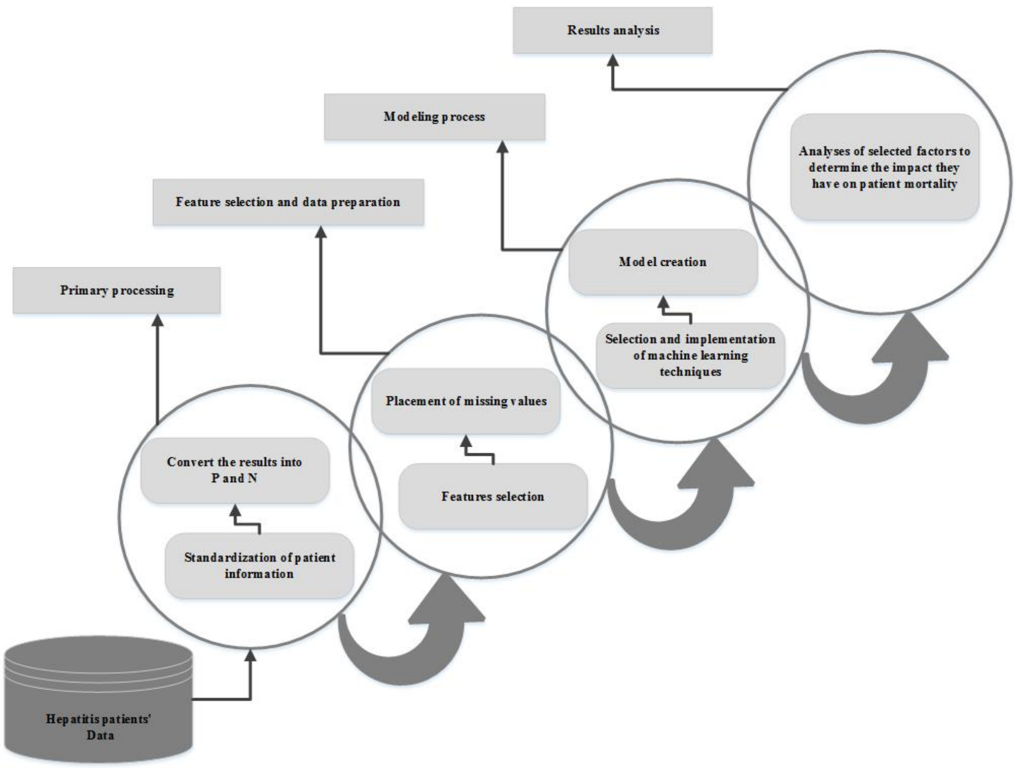

Figure 1: The Implementation Process in a Step-By-Step Manner

4. Implementation Findings

In this section, in the beginning, has attempted to evaluate the results of the implementation of the two techniques of the KNN and the SVM on the four characteristics of ANOREXIA, LIVER BIG, LIVER FIRM, and SPLEEN PALPABLE. This section includes two types of answers and will be analyzed. In the first type of answer, we examine the positive impact that each of the four characteristics ANOREXIA, LIVER BIG, LIVER FIRM, and SPLEEN PALPABLE, has on the survival of a person with hepatitis. As a type of the second answer, we will discuss the performance results of the implemented techniques, which are evaluated according to the four criteria of ACC, ERR, SPE, and NPV, and introduce a technique that has demonstrated a higher level of performance. Several conditional parameters are considered in this section to evaluate the effect of the selected characteristics on the mortality rate due to hepatitis disease. If the features meet those conditional parameters, they are known as features with a positive impact rate. In other words, the affected person can survive when the examination result is positive. The defined condition parameters for this research are the confidence interval value of 80% and the support value of 50%. The reason for choosing a value of 80% for the confidence interval can be seen because of the high sensitivity of the decision-making process in the field of healthcare, and considering the issue of human health, it is necessary to consider a high confidence interval, so that the results obtained from accuracy and have enough confidence. The support value of 50% means that when, out of the 150 available samples, the desired characteristic is detected as positive in at least 76 cases, the individual is likely to survive most of the time. As a result, the person still has a chance of survival if they obtain a positive score for this characteristic in the tests. In Table 3, the status of the selected variables in the 150 existing samples is presented.

Clinical Features	Available samples (150 case)			SVM				KNN				Death		LIVE	
	Missing values	P	N	NPV	ERR	SPE	ACC	NPV	ERR	SPE	ACC	P	N	P	N
	count	count	count	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	count	count	count	count
ANOREXIA	13	119	18	87.6	16.5	89.6	90.7	71.7	19.6	76.5	79.9	17	7	110	16
LIVER BIG	9	121	20	81.8	14.6	92.4	96.6	71.8	16.8	81.7	84.3	20	5	105	20
LIVER FIRM	14	91	45	83.9	15.9	94.8	91.1	72.9	21.3	75.6	82.8	12	18	69	51
SPLEEN PALPABLE	7	123	20	89.8	17.9	95.5	97.8	72.4	18.2	79.7	88.2	21	13	109	7

Table 3: The Summary of 150 Samples' Status

4.1 Liver Big Feature

The results obtained from the implementation of ML techniques on the data set of people with hepatitis indicate the fact that people who have been diagnosed with a positive LIVER BIG trait in their examinations have a high probability of survival, as shown in Figure 2.

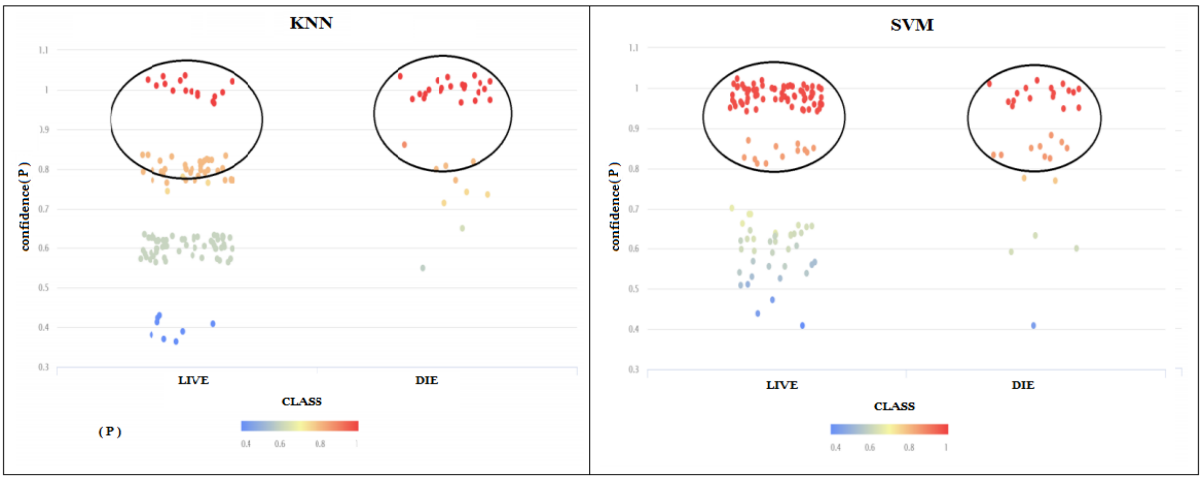

Figure 2: Predicting the Probability of Patient Survival Based on Liver Big Results

4.2 Anorexia Feature

According to the results obtained from employing ML techniques in the data set of people with hepatitis, people diagnosed with a positive ANOREXIA trait in their examinations have a high chance of surviving, as shown in Figure 3.

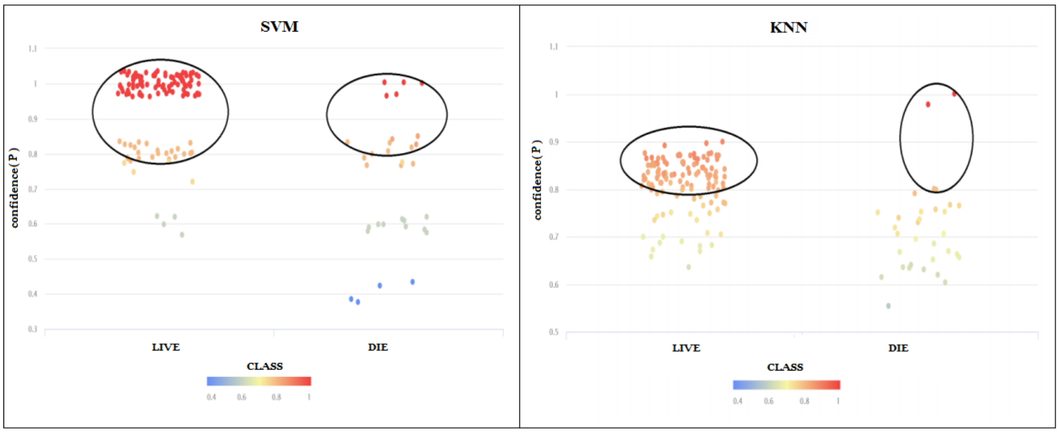

Figure 3: Predicting the Probability of Survival of a Patient Based on the Results of Anorexia

4.3 Spleen Palpable Feature

In the application of ML techniques on the data set of people with hepatitis, it has been determined that people with a positive diagnosis of SPLEEN PALAPABLE in their examinations have a high probability of surviving, as shown in Figure 4.

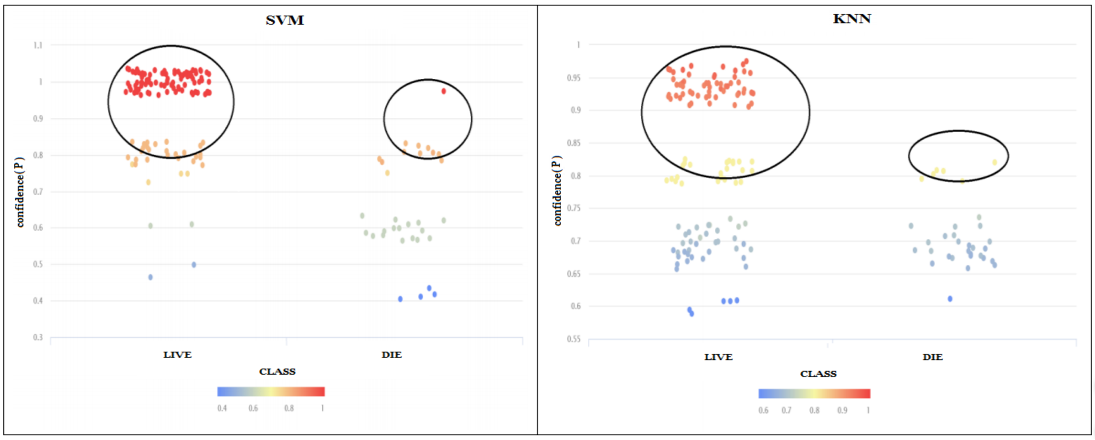

Figure 4: Predicting the Probability of Patient Survival Based on Spleen Palapable Results

4.4 Liver Firm Feature

As a result of implementing ML techniques on the data set of people with hepatitis, the results indicate a high probability of survival for people diagnosed with a positive LIVER FIRM trait in their examinations following successful treatment for hepatitis, as shown in Figure 5.

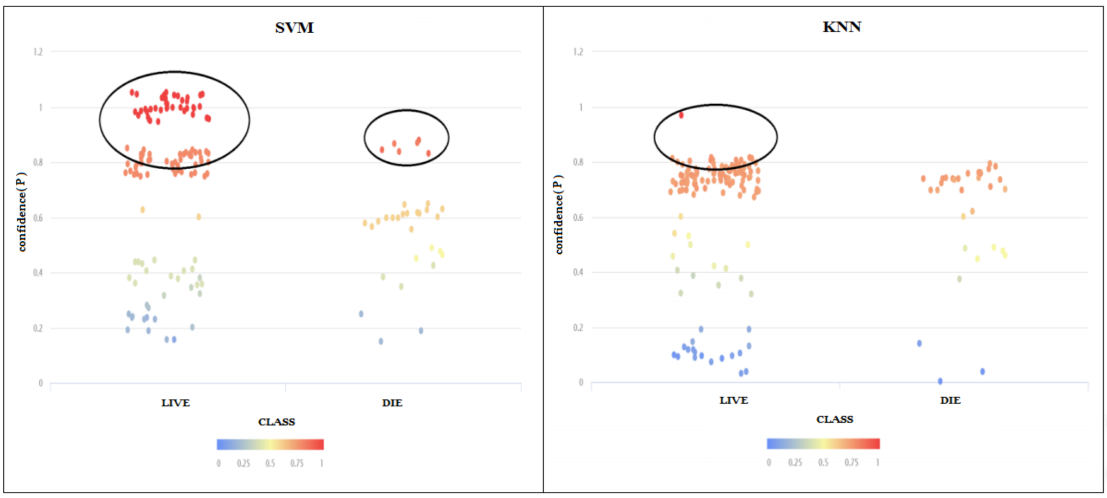

Figure 5: Predicting the Probability of Survival of a Patient Based on the Results of Liver Firm

5. Analyzing Ml Techniques

Based on the results obtained from the implementation of machine learning techniques on the clinical parameters LIVER BIG, LIVER FIRM, SPLEEN PALPABLE, and ANOREXIA, as well as the evaluations based on ACC, ERR, SPE, and NPV, it can be concluded that in all implementations, the SVM technique performed better than the KNN technique in predicting the survival of patients with hepatitis, as shown in Figure 6.

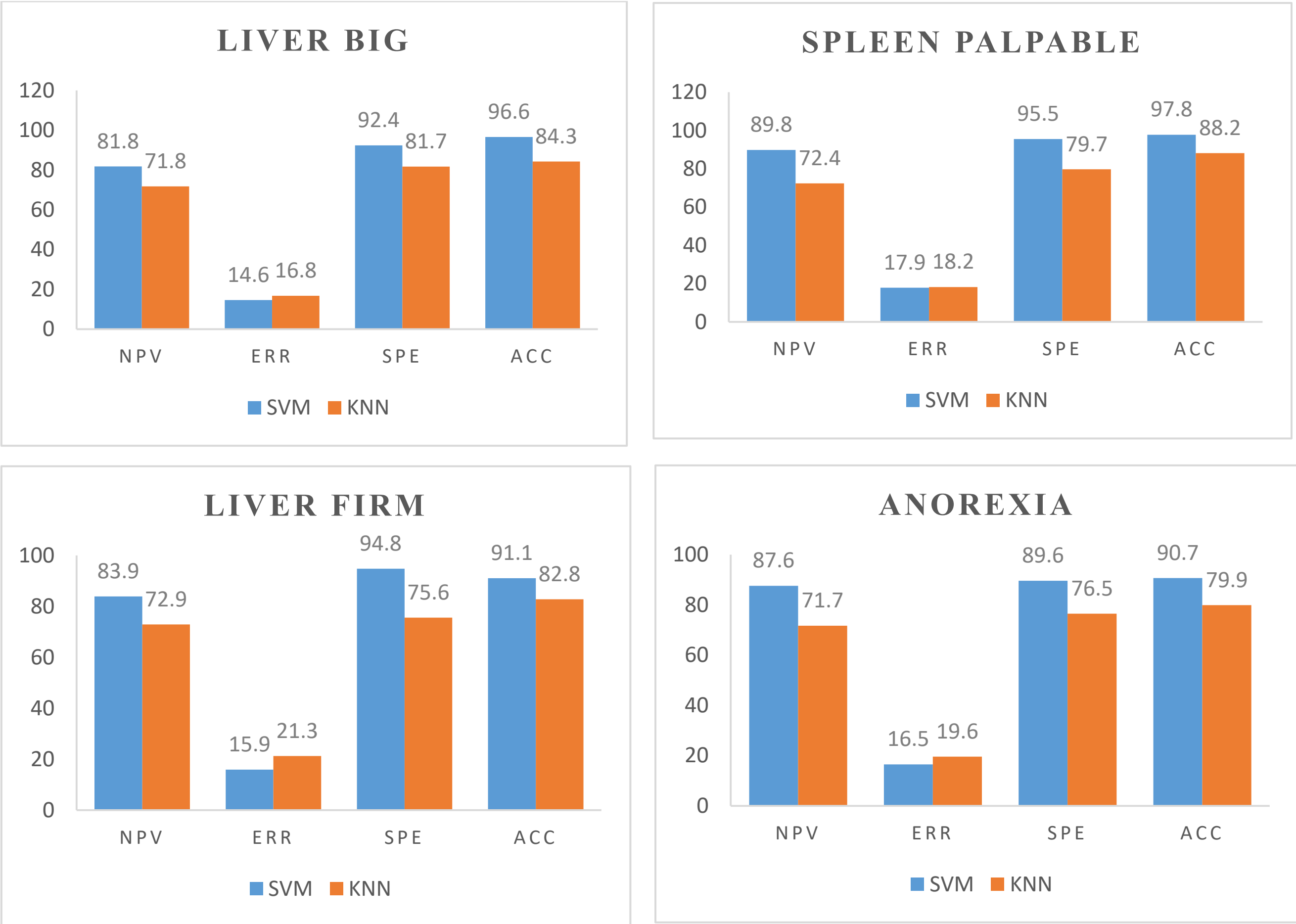

Figure 6: Evaluating the Performance Of Knn And Svm Techniques in Predicting Patient Survival in Cases Of Liver Big, Liver Firm, Spleen Palpable, and Anorexia Positivity Clinical Parameters Using Acc, Err, Spe, and Npv Evaluation Criteria.

To obtain a comprehensive analysis of the performance of implementing ML techniques, including KNN and SVM, we have attempted to average each technique's implementation results at different aggregated stages and then obtain an average. The type of performance achieved by the implemented techniques is shown in Table 4.

ML technique	ACC	SPE	ERR	NPV
SVM	94.05	93.07	16.2	85.7
KNN	83.8	78.3	18.9	72.2

Table 4: Comparison of the Performance of Svm And Knn

Based on all the evaluations conducted, it is evident that the SVM technique performs better than the KNN technique, as shown in Figure 7.

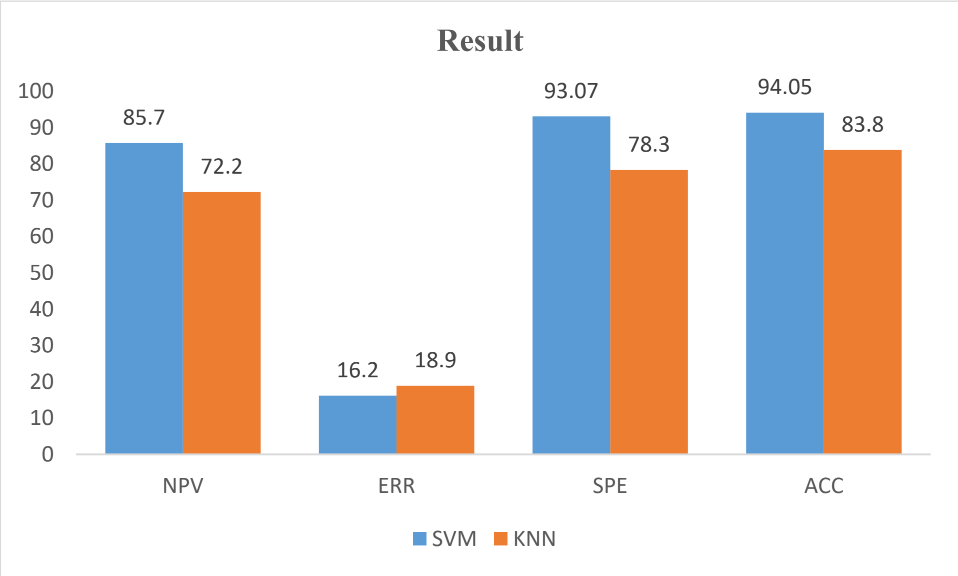

Figure 7: Analyzing the Performance of Svms And Knns

6. Conclusion

Hepatitis disease has a high mortality rate, and patients must undergo various examinations to monitor their condition. The examinations conducted have a wide range of clinical parameters. For scientists and health care professionals, examining the effects of each of these parameters on patient survival or death can provide valuable information and enable them to determine the appropriate type of treatment, treatment process, and even the appropriate level of care for the patient. Achieving such information would be an extremely difficult and time-consuming task for professionals. Machine learning technology is an emerging technology that takes its power from artificial intelligence. Hence, in sensitive areas such as healthcare, machine learning technology can be highly effective in the decision-making processes of healthcare professionals. This study analyzes the positive effect of clinical parameters like LIVER BIG, LIVER FIRM, SPLEEN PALPABLE, and ANOREXIA on patients' survival or death rate by SVM machine learning techniques and KNN. The results of the analysis of the implementations are examined in two parts. The clinical parameters are examined in the first part to determine their positive impact on patient survival and death rates. In the second part, machine learning techniques are assessed by evaluating their performance based on ACC, ERR, SPE, and NPV. According to the implementation of machine learning methods on patient data with hepatitis, it has been determined patients with the LIVER BIG, LIVER FIRM, SPLEEN PALPABLE, and ANOREXIA clinical parameters can have a high chance of survival rate. Based on an analysis of the performance of the machine learning techniques, the SVM technique performed far more effectively than KNN in terms of ACC 94.05%, ERR 16.02%, SPE 93.07%, and NPV 85.7%. A more thorough analysis of another clinical parameter and using more powerful techniques like neural networks can cause more accurate results, and the resulting models will have more excellent reliability and effectiveness for healthcare professionals.

References

1. Ng, R. T., & Pei, J. (2007). Introduction to the special issue on data mining for health informatics. ACM SIGKDD Explorations Newsletter, 9(1), 1-2.
2. Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive data mining for medical diagnosis: An overview of heart disease prediction. International Journal of Computer Applications, 17(8), 43-48.
3. Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., ... & Steinberg, D. (2008). Top 10 algorithms in data mining. Knowledge and information systems, 14, 1-37.
4. Padhy, N., Mishra, D. P., & Panigrahi, R. (2012). The survey of data mining applications and feature scope. arXiv preprint arXiv:1211.5723.
5. Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2013). Data mining with big data. IEEE transactions on knowledge and data engineering, 26(1), 97-107.
6. Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: a comparison of three data mining methods. Artificial intelligence in medicine, 34(2), 113-127.
7. Glover, S., Rivers, P. A., Asoh, D. A., Piper, C. N., & Murph, K. (2010). Data mining for health executive decision support: an imperative with a daunting future!. Health Services Management Research, 23(1), 42-46.
8. Gharehchopogh, F. S., Molany, M., & Mokri, F. D. (2013). Using artificial neural network in diagnosis of thyroid disease: a case study. International Journal on Computational Sciences & Applications (IJCSA) Vol, 3, 49-61.
9. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37-37.
10. Maroco, J., Silva, D., Rodrigues, A., Guerreiro, M., Santana, I., & de Mendonça, A. (2011). Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vec-

tor machines, classification trees and random forests. BMC research notes, 4(1), 1-14.

11. Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. In AI 2006: Advances in Artificial Intelligence: 19th Australian Joint Conference on Artificial Intelligence, Hobart, Australia, December 4-8, 2006. Proceedings 19 (pp. 1015-1021). Springer Berlin Heidelberg.
12. Rajeswari, P., & Reena, G. S. (2010). Analysis of Liver Disorder using Datamining Algorithm. Global journal of computer science and technology, 10(14).
13. Ho, T. B., Nguyen, C. H., Kawasaki, S., Le, S. Q., & Takabayashi, K. (2007). Exploiting temporal relations in mining hepatitis data. New Generation Computing, 25, 247-262.
14. Uhmn, S., Kim, D. H., Cho, S. W., Cheong, J. Y., & Kim, J. (2007, October). Chronic hepatitis classification using SNP data and data mining techniques. In 2007 Frontiers in the Convergence of Bioscience and Information Technologies (pp. 81-86). IEEE.
15. Zayed, N., Awad, A. B., El-Akel, W., Doss, W., Awad, T., Radwan, A., & Mabrouk, M. (2013). The assessment of data mining for the prediction of therapeutic outcome in 3719 Egyptian patients with chronic hepatitis C. Clinics and research in hepatology and gastroenterology, 37(3), 254-261.
16. Abe, H., Tsumoto, S., Ohsaki, M., Yokoi, H., & Yamaguchi, T. (2007, November). Evaluation of learning costs of rule evaluation models based on objective indices to predict human hypothesis construction phases. In 2007 IEEE International Conference on Granular Computing (GRC 2007) (pp. 458-458). IEEE.
17. Ohsaki, M., Sato, Y., Yokoi, H., & Yamaguchi, T. (2002, December). A rule discovery support system for sequential medical data, in the case study of a chronic hepatitis dataset. In Workshop Notes of the International Workshop on Active Mining, at IEEE International Conference on Data Mining (p. 121).
18. Yin, X., & Han, J. (2003, May). CPAR: Classification based on predictive association rules. In Proceedings of the 2003 SIAM international conference on data mining (pp. 331-335). Society for Industrial and Applied Mathematics.
19. Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN model-based approach in classification. In On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings (pp. 986-996). Springer Berlin Heidelberg.
20. Lauer, G. M., & Walker, B. D. (2001). Hepatitis C virus infection. New England journal of medicine, 345(1), 41-52.
21. Cios, K. J., & Moore, G. W. (2002). Uniqueness of medical data mining. Artificial intelligence in medicine, 26(1-2), 1-24.
22. Lucas, P. (2004). Bayesian analysis, pattern analysis, and data mining in health care. Current opinion in critical care, 10(5), 399-403.
23. Steinberg, D. M., Fine, J., & Chappell, R. (2009). Sample size for positive and negative predictive value in diagnostic research using case–control designs. Biostatistics, 10(1), 94-105.
24. Witt, O., Deubzer, H. E., Milde, T., & Oehme, I. (2009). HDAC family: What are the cancer relevant targets?. Cancer letters, 277(1), 8-21.