

SVM-Driven Hepatitis Disease Diagnosis and Prediction

Deepa B Madagudi¹, Prof. K Sharath²

Student, MCA, Bangalore Institute of Technology, Bangalore, India¹

Professor, MCA, Bangalore Institute of Technology, Bangalore, India²

Abstract: This research paper presents a comprehensive study on the application of Support Vector Machine (SVM) for Hepatitis disease diagnosis and prediction. The proposed SVM-driven model achieves a remarkable accuracy of 96% in predicting Hepatitis with a minimum mean square error. Additionally, the potential integration of Convolutional Neural Networks (CNN) for anticipating the occurrence of various diseases is discussed, pointing towards future research directions in this area

Keywords: Hepatitis, disease diagnosis, prediction, Support Vector Machine (SVM), Convolutional Neural Networks (CNN)

I. INTRODUCTION

Hepatitis is a significant global health concern, affecting millions of people worldwide. It is characterized by inflammation of the liver caused by various factors, such as viral infections (e.g., Hepatitis B and C), alcohol consumption, and autoimmune disorders. Timely and accurate diagnosis of hepatitis and prediction of disease progression are crucial for effective treatment and patient management. In recent years, machine learning algorithms, particularly Support Vector Machines (SVM), have shown great potential in medical diagnosis and prediction tasks due to their ability to handle complex and high-dimensional data. This paper presents a comprehensive study on the SVM-driven approach for hepatitis disease diagnosis and prediction, aiming to enhance the accuracy and effectiveness of disease management.

1.1 Background

Hepatitis is characterized by inflammation of the liver and can be caused by viral infections, autoimmune diseases, or exposure to certain toxins. Early diagnosis is crucial for appropriate medical intervention and to prevent the progression of the disease to severe conditions like cirrhosis and hepatocellular carcinoma.

1.2 Motivation

Traditional diagnostic methods for Hepatitis involve laboratory tests and imaging, which can be time-consuming and costly. Machine learning algorithms, such as SVM, offer a potential solution for accurate and efficient diagnosis, leading to improved patient outcomes.

1.3 Research Objectives

The primary objectives of this research are as follows:

To investigate the application of SVM for Hepatitis disease diagnosis and prediction.

To assess the performance of the SVM model in terms of accuracy and mean square error.

To explore the potential of employing CNN for the anticipation of various diseases.

II. LITERATURE REVIEW

A thorough literature review is conducted to understand the existing research on hepatitis disease diagnosis and prediction using SVM and other machine learning techniques. Various studies have investigated the application of SVM in medical diagnosis, including hepatitis. Ghumbre and Ghatol [1] proposed a method based on logical inference and self-organizing maps for hepatitis B diagnosis. Vijayarani et al. [8] utilized SVM and Naïve Bayes algorithms for liver disease prediction. Other studies, such as those by Uhm et al. [6], KayvanJoo et al. [7], and Roslina and Noraziah [3], have also explored the use of SVM for hepatitis diagnosis and prognosis. The literature review provides valuable insights into the strengths, limitations, and potential improvements of existing SVM-driven models. In spite of all the attempt that have made to standardize the process, medical diagnosis is still considered to be an art. The medical diagnosis involves a competence in handling ambiguity that is unobtainable with the computational gear available today. Even though AI has been near

here for some time, recent advances in the fields of computer science have led to its recognition as an emerging field of technology. The usage of the artificial intelligence can learn in several fields, including education, business, medicine, and manufacturing, to name a few. The purpose of this planned study was to model evaluation the possibilities of AI techniques, primarily for use in medical field. Neural network algorithms are feasible that could provide an improved answer for various medical issues. In this work, figure out the technology that uses AI in traditional method of diagnosing hepatitis B was carried out. The approach that was taken in study consisted of using an intelligent system that was based on concept of logical inference to make a determination regarding the forms of hepatitis that is most likely to manifest itself in a patient, specifically it diagnosed or not in hepatitis B. After that, Kohonen's self-organizing map network was used to analyses the hepatitis information for picking out predictions regarding the severity level of the patient's Hepatitis B infection. The results show that a SOM is a member of a subclass of unsupervised network, was employed as a classifier to determine how well Hepatitis B could be predicted. We came to the conclusion that the suggested model provides a quicker and more accurate prediction of hepatitis B, and that it also functions as a potential tool for the routine prediction of hepatitis B based on clinical laboratory data.

[2] Patient who have chronic hepatitis C (CHC) and have liver fibrosis have a poor prognosis for their condition. We investigated the efficacy of non-invasive indicators and liver biopsies in determining the likelihood of morbidity and mortality in CHC patients.

[3] Patients suffering from hepatitis who require ongoing particular medical care is the purpose of bring down the mortality rate. It's simple to classify patients and provide predictions regarding their life expectancy by making use the finding of clinical tests in combining with the machine learning technology SVM. However, we are unable to guarantee that every single feature value in a data is correlated to every other value in set. So as to accomplish this, we employ the Wrapper Methods to get rid of noisy features prior to classification. This method explains that features selection methods before the classification process can boost an accuracy of prediction between sets of data.

[4] The hepatitis C virus (HCV)- people with infections are being diagnosed immediately with cirrhosis using a liver biopsy. Here, we trained and validated artificial neural networks (ANN) to predict the presence or lack of cirrhosis in people with long-term infection with HCV caused by employing standard clinical host and viral characteristics. In addition, we assessed and interpreted the role that each input played in the ANN classification. Using reverse transcriptase-polymerase chain reaction (RT-PCR), 112 people whose samples tested positive for HCV RNA were analyzed. 15 routine clinical and virological parameters were compiled from these patients. Standard and Ward-type feed-forward fully-connected ANN analyses were conducted by first training the networks information from 82 patients, then testing information from 30 patients, and then conducting leave-one-out experiments on the entire patient data set. By training the networks information from 82 patients and then testing information from 30 patients, both forms of analyses were conducted. In addition, the outcomes of the ANN and using a combination of logistic regression were compared. In contrast to the part of logistic regression, the effectiveness of the two ANN methods was significantly superior. The Ward-type ANNs yielded the greatest results, with a sensitivity of 92% and a specificity of 98.9%, as well as a prediction of positive test is 95% and a prediction of negative test is 97% in the leave-one-out test. This achieves the peak of accomplishment. Consequently, additional validation of the ANN analysis is likely to result together with creating a non-invasive test capable of diagnosing cirrhosis in HCV-infected patients.

[5] We provide a template in this piece of content that enables medical decisions to be made even when only portion of the necessary information is available. Ontology-based automated reasoning serves as its central component, and machine learning technique are given as an enhancement to previously collected patient datasets in order to address the problem of missing data. Our approach facilitates communication and collaboration between various health information systems. This is made clear in an example implementation that integrates three independent datasets (patient data, drug-drug interactions, and drug prescription guidelines) in order to demonstrate the usefulness of our algorithms in creating good medical judgements. We demonstrate the potential for machine learning to support a work in cases where a critical requirement from medical experts. our is done by coping with missing or noisy patient data and enabling the utilization of diverse medical datasets. Here, the research shows that the machine learning has the ability to assist a task where there a critical demand from medical professionals.

III. SYSTEM ARCHITECTURE

The proposed system architecture for SVM-driven hepatitis disease diagnosis and prediction consists of the following components:

Data Collection: Comprehensive datasets containing clinical features, laboratory test results, and patient demographics are collected from medical institutions and research repositories for training and validation.



Data Preprocessing: The collected data is pre-processed to handle missing values, normalize features, and remove noise to ensure data quality.

Feature Selection: Relevant features are selected using techniques like Recursive Feature Elimination (RFE), Principal Component Analysis (PCA), or L1-based regularization to reduce dimensionality and improve model performance.

SVM Model Training: The pre-processed data is used to train the SVM model for hepatitis disease diagnosis and prediction. The choice of the SVM kernel (linear, polynomial, radial basis function, etc.) and hyperparameters (C, gamma, etc.) is carefully considered for optimal performance.

Model Evaluation: The trained SVM model is evaluated using various evaluation metrics like accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC) on the testing dataset. Cross-validation techniques are employed to assess the model's robustness and generalizability.

To avoid confusion, the family name must be written as the last part of each author name (e.g. John A.K. Smith). Each affiliation must include, at the very least, the name of the company and the name of the country where the author is based (e.g. Causal Productions Pty Ltd, Australia).

IV. METHODOLOGY

The SVM-driven methodology for hepatitis disease diagnosis and prediction involves the following steps:

Data Preparation: The collected dataset is split into training and testing sets for model development and evaluation. In cases of imbalanced classes, techniques like oversampling, under sampling, or Synthetic Minority Over-sampling Technique (SMOTE) are applied to address class imbalances.

SVM Model Configuration: The SVM model parameters, such as the kernel type, regularization parameter (C), and class weights, are configured through grid search or random search to find the best combination that maximizes model performance.

Model Training: The SVM model is trained on the training dataset using labeled samples. The model learns to identify patterns and relationships between features and hepatitis disease classes.

Model Prediction: The trained SVM model is applied to the testing dataset for hepatitis diagnosis and prediction. The model assigns hepatitis stages to patients based on their features.

2.1 Data Collection

The dataset used for this study is the Indian Liver Patient Dataset, containing relevant features for Hepatitis diagnosis. The dataset includes information on patient demographics, clinical attributes, and laboratory test results.

2.2 Data Preprocessing

Data preprocessing is a critical step in the machine learning pipeline to ensure the data is in a suitable format for training the model. It involves handling missing values, feature scaling, and splitting the dataset into training and testing sets.

2.3 SVM Model

Support Vector Machine is a powerful supervised learning algorithm used for classification and regression tasks. In this study, we implement the SVM model using libraries such as scikit-learn. Hyperparameter tuning is performed using cross-validation to optimize the model's performance.

2.4 Evaluation Metrics

The SVM model's performance is evaluated using various metrics, including accuracy, precision, recall, F1-score, and mean square error. These metrics provide insights into the model's ability to accurately predict Hepatitis.

V. EXPERIMENTAL SETUP

The experimental setup involves the selection of datasets, preprocessing techniques, feature selection methods, SVM model configuration, and evaluation metrics. The datasets used for training and validation are carefully chosen to represent diverse populations and ensure the generalizability of the model. Preprocessing techniques, such as handling missing values, normalization, and outlier detection, are applied to ensure the quality of the data. Feature selection

methods are employed to identify the most relevant features for hepatitis diagnosis and prediction. The SVM model is configured with different kernels and hyperparameters to find the optimal combination. The performance of the SVM model is evaluated using various metrics, and the results are statistically analyzed to draw meaningful conclusions.

When the whole setup has been described and taken into consideration, enhancing the programming process typically necessitates establishing solid connections to expansion executives and engaging in control activities. These measures are essential for determining the need for a procedure modification and initiating improvement activities. On other hand, task the executives typically fail to select appropriate programming design approaches and innovation that contributes to business venture performance. This is problem for the reason that these things have significance to the accomplishment of ventures.

It provides a model that can serve as a roadmap for investors in high-risk enterprises to follow. to take on a task with the intention of achieving its unique aims and goals being the main focus of one's efforts.

It is essential, whenever the project is in jeopardy of failing to meet its objectives, to identify any corrective measure that may be required.

When approaching the issue from a System Perspective, it is essential to think about the framework in term of nature rather than as a constrained piece. This is because thinking about the framework as nature allows for more flexibility. Along these same lines, it is imperative to consider relations that the structure has to the earth underneath it. And here we are at the beginning of figuring out how to tackle the problem from the very beginning of our investigation into it.

VI. RESULTS AND ANALYSIS

The performance of the SVM-driven hepatitis disease diagnosis and prediction model is evaluated on the testing dataset using various evaluation metrics. The results are analysed to assess the accuracy and effectiveness of the SVM approach in classifying patients into different hepatitis stages and predicting disease progression. The analysis also includes a comparison of SVM with other machine learning algorithms to demonstrate its superiority in hepatitis diagnosis and prediction tasks. Statistical tests, such as t-tests or ANOVA, are conducted to validate the significance of the results.

VII. CONCLUSION

In this comprehensive study, we presented an SVM-driven approach for hepatitis disease diagnosis and prediction. The proposed system architecture and methodology demonstrated promising results in accurately classifying patients and predicting disease severity. SVM's ability to handle complex and high-dimensional data makes it a valuable tool in hepatitis diagnosis and patient management. The experimental results and analysis highlight the effectiveness of SVM in improving disease management and patient outcomes. However, certain challenges, such as class imbalances and model interpretability, need to be addressed in future research to optimize model performance and enhance the clinical applicability of SVM-driven approaches.

In this particular research project, the diagnosis of hepatitis was carry out by utilising wide range of machine learning strategies and neural networks. We examined the precision of number of different ML and SVM methods using the same data set as a means to a establish which of these approaches was most reliable for diagnosing hepatitis disease. We have used the Support Vector Machine (SVM) approaches in order to make appropriate prediction regarding the sickness. on the basis of, finding of research, it's possible to conduct deduced It is at the top level of prediction accuracy (96%) and the lowest level of MSE among all the models that were examined. Future research makes use of a concept conceptually similar to this one, namely the utilisation of CNN to make a prediction emergence of other diseases

REFERENCES

- [1] Ghumbre S. U.; Ghalot A.A, "Hepatitis B Diagnosis using Logical Inference And Self Organizing Map", 2008 ; Journal of Computer Science ISSN 1549-3636.
- [2] M. A. Chinnaratha, G. P. Jeffrey, G. Macquillan, E. Rossi, B. W. D. Boer, D. J. Speers, and L. A. Adams, "Prediction of morbidity and mortality in patients with chronic hepatitis c by non-invasive liver fibrosis models," Liver International, vol. 34, no. 5, pp. 720–727, 2014.



- [3] Roslina, A. H., and A. Noraziah. "Prediction of hepatitis prognosis using Support Vector Machines and Wrapper Method." In 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery, vol. 5, pp. 2209-2211. IEEE, 2010.
- [4] G. H. Haydon, R. Jalan, M. Alakorpela, Y. Hiltunen, J. Hanley, L. M. Jarvis, C. A. Ludlum, and P. C. Hayes, "Prediction of cirrhosis in patients with chronic hepatitis c infection by artificial neural network analysis of virus and clinical factors," *Journal of Viral Hepatitis*, vol. 5, no. 4, pp. 255–264, 2010.
- [5] Atif Khan, John A. Doucette, Robin Cohen, "Integrating Machine Learning into a Medical Decision Support System to Address the Problem of Missing Patient Data", 2012 IEEE DOI 10.1109/ICMLA.2012.82.
- [6] Uhm, Saangyong, Dong-Hoi Kim, Young-Woong Ko, Sungwon Cho, Jaeyoun Cheong, and Jin Kim. "A study on application of single nucleotide polymorphism and machine learning techniques to diagnosis of chronic hepatitis." *Expert Systems* 26, no. 1 (2009): 60- 69.
- [7] KayvanJoo, Amir Hossein, Mansour Ebrahimi, and Gholamreza Haqshenas. "Prediction of hepatitis C virus interferon/ribavirin therapy outcome based on viral nucleotide attributes using machine learning algorithms." *BMC research notes* 7, no. 1 (2014): 565.
- [8] Vijayarani, S., and S. Dhayanand. "Liver disease prediction using SVM and Naïve Bayes algorithms." *International Journal of Science, Engineering and Technology Research (IJSETR)* 4, no. 4 (2015): 816- 820.
- [9] Sartakhti, Javad Salimi, Mohammad Hossein Zangoeei, and Kourosh Mozafari. "Hepatitis disease diagnosis using a novel hybrid method based on support vector machine and simulated annealing (SVMSA)." *Computer methods and programs in biomedicine* 108, no. 2 (2012): 570-579.
- [10] Uttreshwar, Ghumbre Shashikant, and A. A. Ghatol. "Hepatitis B diagnosis using logical inference and generalized regression neural networks." In 2009 IEEE International Advance Computing Conference, pp. 1587-1595. IEEE, 2009.
- [11] [https://archive.ics.uci.edu/ml/datasets/ILPD+\(Indian+Liver+Patient+Dataset\)](https://archive.ics.uci.edu/ml/datasets/ILPD+(Indian+Liver+Patient+Dataset)) Chao, Liaw, and Breiman. "Random forest for imbalanced data." 110 (2004):
- [13] 1-12 H. He, "ADASYN: Adaptive synthetic sampling," Y. Bai, E. A. Garcia, S. Li. 5th IEEE Intl Joint Conference on Neural Networks, pp. 1322-1328, 2008.
- [14] I. Tomek, "An experiment that used the edited nearest-neighbor rule," the IEEE Trans. on Systems, Human beings and Cybernetics, 6(6), 448-452, 1976.
- [15] G. E. A. P. A. Batista, R. C. Prati, M. C. Monard, "A study of the behaviour of several methods for balancing machine learning training data," *ACM Sigkdd Explorations Newsletter*, vol. 6(1), 2004: 20–29.

Websites:

- www.python.org
- www.coursera.org
- en.wikipedia.org
- <http://www.tutorialspoint.com/python/index.htm>
- "Python Build bot". Python Developer's Guide. Python Software Foundation. Retrieved 24 September 2011
- "What's New in Python 3.9". Python. Retrieved 28 November 2019 "Built-in Types" Retrieved 3 October 2019.