DIAGNOSIS OF LIVER DISEASE USING MACHINE LEARNING

Mr. C. Ganesh[1], Ms. M. Gayathri[2]

[1]Assistant Professor, Department of MCA, (Vivekanandha Institute of Information and Management Studies).

[2]Student, Department of MCA, (Vivekanandha Institute of Information and Management Studies).

ABSTRACT :

Liver diseases contribute significantly to global mortality rates, with early diagnosis playing a crucial role in effective treatment. However, detecting liver diseases in their early stages is challenging due to subtle symptoms. This project leverages Machine Learning (ML) techniques to develop a predictive model for liver disease diagnosis, improving early detection and aiding medical professionals. The system utilizes Artificial Neural Networks (ANN), Support Vector Machines (SVM), and ensemble learning methods to analyze patient parameters such as age, bilirubin levels, enzyme concentrations, and protein ratios. By employing advanced ML models trained on liver patient datasets, the project enhances diagnostic accuracy and provides a data-driven approach to liver disease detection.

Keywords: Liver disease, early diagnosis, subtle symptoms, machine learning, predictive model, artificial neural networks, support vector machines, patient parameters, datasets, diagnostic accuracy.

I.INTRODUCTION

Liver disease is a tricky disease to diagnose given the subtlety of the symptoms while in the early stages. Problems with liver diseases are not discovered until it is often too late as the liver continues to function even when partially damaged. Early diagnosis can potentially be life-saving. Although not discoverable to even the experienced medical practioner, the early symptoms of these diseases can be detected. Early diagnoses of patients can increase his/her life span substantially. Thus the results of this study are important both from the point of view of the computer scientist and the medical professional. This paper aims to compare 2 methods of computer aided medical diagnoses. The first of these methods is a symptomatic approach to diagnosis. This method involves the training of an Artificial Neural Network to respond to several patient parameters such as age, Bilirubin, Alkaline Phosphotase, Alamine Aminotransferase, and Aspartate Aminotransferase among others. The Neural Network classifies the patients according to whether the patient does indeed suffer from a chronic Liver Disease or not that is healthy or not. The second method studied in this paper involves a genetic approach to the diagnosis. The proposed approach is the application of Artificial Neural Networks and Multi-Layer Perceptrons to Micro-Array Analysis.

Objectives

The primary objective of this project is to develop an effective machine learning-based system for the early diagnosis of liver disease, addressing the challenges posed by the subtle symptoms in initial stages. By leveraging algorithms such as Artificial Neural Networks (ANN) with backpropagation and Support Vector Machines (SVM), the system aims to analyze various patient parameters including age, bilirubin levels, enzyme concentrations, and protein ratios to accurately predict the presence of chronic liver disease. This approach seeks to improve diagnostic accuracy, thereby assisting medical professionals in timely intervention and potentially increasing patient survival rates. Additionally, the project compares the performance of ANN and SVM techniques on a liver patient dataset, with the goal of identifying the most reliable method for liver disease prediction. The system is designed to be user-friendly, efficient, and adaptable for future enhancements, ensuring minimal maintenance and ease of use for healthcare providers. Overall, the study contributes to advancing computer-aided diagnosis through machine learning to support better health outcomes in liver disease management.

II. LITERATURE SURVEY

[1]. INTELLIGENT MODEL FOR LIVER DISEASE DIAGNOSIS

Authors: Rong-Ho Lin

Rong-Ho Lin proposed a model that integrates rule-based reasoning and case-based reasoning techniques for liver disease diagnosis. The model effectively combines historical patient cases and clinical rules to enhance decision-making in medical diagnostics. His approach demonstrated a high level of accuracy and interpretability, which made it suitable for medical practitioners. Lin emphasized the value of hybrid systems that blend traditional clinical knowledge with intelligent machine learning-based predictions to improve healthcare delivery.

[2]. MULTI-CLASS MOLECULAR CANCER CLASSIFICATION USING SVM

Authors: Ryan Rifkin et al

Though originally applied to cancer classification, the methodology by Rifkin et al. has significantly influenced the application of SVM in biomedical diagnosis. The study used SVMs to classify tumors based on gene expression data with approximately 80% accuracy. This approach laid the groundwork for applying SVMs to liver disease datasets, especially in tasks involving micro-array data classification. Their analytical method proved useful for handling high-dimensional biomedical data efficiently

[3]. COMPARATIVE STUDY OF CLASSIFICATION ALGORITHMS

Authors: Bendi Venkata Ramana et al

This comprehensive study critically evaluated several classification algorithms (including Decision Trees, Naïve Bayes, ANN, and SVM) for diagnosing liver disease. The researchers used the Indian Liver Patient Dataset (ILPD) and demonstrated that combining ensemble methods like bagging and boosting with basic classifiers improved prediction performance. The study highlighted the importance of evaluating not just accuracy but also precision, sensitivity, and specificity, and concluded that ANN provided the most reliable results for this dataset.

III.EXISTING SYSTEM

A. Micro Array Analysis

Among the most influential work in Micro-Array Analysis can be attributed to Rifkin et al [2]. Their work is attributed to a Support Vector Machine to accurately (80%) predict the origin of tumors collected from samples obtained at Massachusetts General and other medical institutions.

Kun-Hong Liu and De-Shuang Huang also solved the problems of cancer origin identification using Micro Array analysis. Several other technologies for Micro-Array analysis have been developed over the last decade. The most common ones are spotted cDNA and oligonucleotide microarrays which are discussed in this paper. Pioneers in the field include researchers from Brown and Stanford (Duggan et al Chipping Forecast 1999) where cDNA samples were hybridized to glass slides onto which the corresponding genes of interest were robotically deposited.

B. SVM and Neural Networks

Akin Ozcift and ArifGulten constructed a rotation forest ensemble classifier that was tested with success on Parkinson's, heart disease and diabetes. Some of the most useful work was done by BendiVenkataRamana et al who successfully compared various machine learning algorithms on the basis of Accuracy, Precision, Sensitivity, and Specificity when classifying this very liver patient data set. They proposed the use of Bayesian classification combined with Bagging and Boosting for improved accuracy. Bayesian classification is a simple yet powerful algorithm and works on the assumption that all variables are independent of one another. They also proposed ANOVA and MANOVA (Analysis of Variance and Multivariate Analysis of Variance) for a population comparison between the ILPD and UCI dataset.

DISADVANTAGES:

- Liver diagnosis often involves imaging modalities like ultrasound or CT, which are not incorporated into the existing system.
- The micro-array and patient datasets used are often limited in size and diversity, making it difficult to generalize the model across various demographics and clinical settings.
- Micro-array analysis methods require large-scale gene expression data, which is expensive, complex to interpret, and not always available in routine clinical practice.

IV.PROPOSED SYSTEM

The proposed system aims to provide a reliable and efficient diagnostic tool for detecting liver disease by leveraging advanced machine learning algorithms. The system utilizes Artificial Neural Networks (ANN) with a backpropagation learning strategy and Support Vector Machines (SVM) for classification tasks. The primary dataset used is the Indian Liver Patient Dataset (ILPD), which contains multiple clinical attributes such as age, gender, total and direct bilirubin levels, enzyme concentrations, and protein ratios. In this system, ANN is trained using a supervised learning approach, where the network is optimized over five folds of cross-validation to ensure generalizability and robustness. The forward and backward propagation steps in ANN enable the model to minimize prediction error and adjust internal weights for improved accuracy. In parallel, the SVM model is employed to create an optimal hyperplane that separates the input data into classes based on liver disease presence or absence. The dataset is preprocessed to ensure all attributes are in numerical form, and model training is conducted with a clear focus on performance evaluation metrics such as accuracy, precision, and error rate. The proposed approach offers a comparative analysis of ANN and SVM methods, ultimately allowing medical professionals to choose the most suitable model for real-time liver disease prediction. By combining data-driven decision-making with scalable machine learning techniques, the proposed system significantly improves the reliability and speed of liver disease diagnostics.

ADVANTAGES:

- Utilizes machine learning algorithms to detect liver disease at early stages, improving the chances of timely medical intervention.
- Employs ANN with backpropagation and SVM to increase accuracy, sensitivity, and specificity in predictions.
- Provides a performance comparison between ANN and SVM, enabling selection of the most effective model for real-time application.
- Incorporates clinically relevant attributes like bilirubin levels, enzyme activity, and protein ratios for informed decision-making.

- The system can be retrained or updated with new data, making it adaptable to evolving medical knowledge and larger datasets.

V. METHODOLOGY

OVERVIEW OF THE PROJECT

The project titled "Diagnosis of Liver Disease Using Machine Learning" aims to address one of the most critical challenges in healthcare — the early and accurate detection of liver diseases. Liver diseases often go unnoticed during their initial stages due to subtle or non-specific symptoms. This silent progression can lead to severe complications and even fatality if left undiagnosed. Therefore, the need for an intelligent, automated, and data-driven diagnostic tool has become increasingly essential in modern medicine. This project utilizes cutting-edge Machine Learning (ML) techniques to build a predictive system capable of diagnosing liver diseases based on key biochemical parameters and patient demographics. Two primary ML algorithms are implemented and evaluated: Artificial Neural Networks (ANN) and Support Vector Machines (SVM). The objective is to compare the predictive performance of these models and identify the most effective approach for real-time application in clinical settings. It includes patient attributes and direct bilirubin, levels of liver enzymes (Alkaline Phosphotase, Alamine Aminotransferase, Aspartate Aminotransferase), total proteins, albumin, and albumin-to-globulin ratio. These parameters are critical indicators of liver function and are commonly used in medical diagnostics. The ANN model employs a backpropagation algorithm within a multi-layer perceptron architecture. This model learns through supervised training, adjusting weights to minimize the error between predicted and actual outputs. The architecture includes an input layer, one or more hidden layers, and an output layer that classifies the presence or absence of liver disease. The ANN is trained using 5-fold cross-validation to ensure robustness and generalization of the model. The SVM model, on the other hand, constructs hyperplanes in a high-dimensional space to separate the dataset into distinct classes. It is particularly effective for binary classification problems and is known for its strong theoretical foundations in statistical learning. However, in this project, SVM was observed to have relatively lower performance in comparison to ANN due to the complexity and non-linearity of the data. The system is designed to be modular and scalable. It reads patient data from CSV files, preprocesses the data to ensure numerical consistency, trains both ANN and SVM models, and provides diagnostic predictions. The output is presented in a user-friendly format, clearly indicating whether the patient is likely to have liver disease and encouraging consultation with medical professionals for further evaluation. Testing and validation play a critical role in this project. A comprehensive testing strategy—including unit testing, integration testing, validation testing, recovery testing, security testing, and performance testing—ensures that the software is reliable, robust, and user-acceptable. The application is designed to be lightweight, platform-independent, and easily maintainable, using Python 3.6 and open-source libraries, making it a cost-effective tool for deployment in hospitals, clinics, or research institutions. Experimental results show that the ANN model achieved an accuracy of approximately 75%, outperforming the SVM model, which achieved an accuracy of around 30%. These neural network models in handling complex medical data and making intelligent predictions.

MODULES

a) Initialize Network
b) Forward Propagate
c) Back Propagate Error
d) Train Network and Predict values

MODULE DESCRIPTION

a) Initialize Network
Each neuron has a set of weights that need to be maintained. The input layer is a row from our training dataset. The first real layer is the hidden layer. This is followed by the output layer that has one neuron for each class value. Initialize the network weights to small random numbers in the range of 0 to 1. initialize_network() that creates a new neural network with three input parameters, the number of inputs, the number of neurons to have in the hidden layer and the number of outputs.

b) Forward Propagate
The output of a neural network is computed by passing the input data sequentially through each layer until the final layer produces a result. This process is essential for generating predictions during both the training phase—where errors are identified and corrected—and for making predictions on new, unseen data after training is complete. Forward propagation typically involves three key steps: activating neurons, applying transfer functions, and passing the data forward through the network.

c) Back Propagate Error
Backpropagation derives its name from the method used to adjust the weights within a neural network. It works by first calculating the difference between the predicted output and the actual target value. This error is then propagated in reverse—from the output layer back to the hidden layers—allowing the network to identify which weights contributed to the error. The process involves two main steps: calculating the derivative of the transfer function and distributing the error backward to update the weights accordingly.

d) Train Network and Predict values
The training process involves updating the network with newly adjusted weights. A function called predict() is used to perform predictions by identifying the index of the output neuron with the highest probability. This function operates under the assumption that class labels have been encoded as integers beginning from zero.

VI.SYSTEM ARCHITECTURE
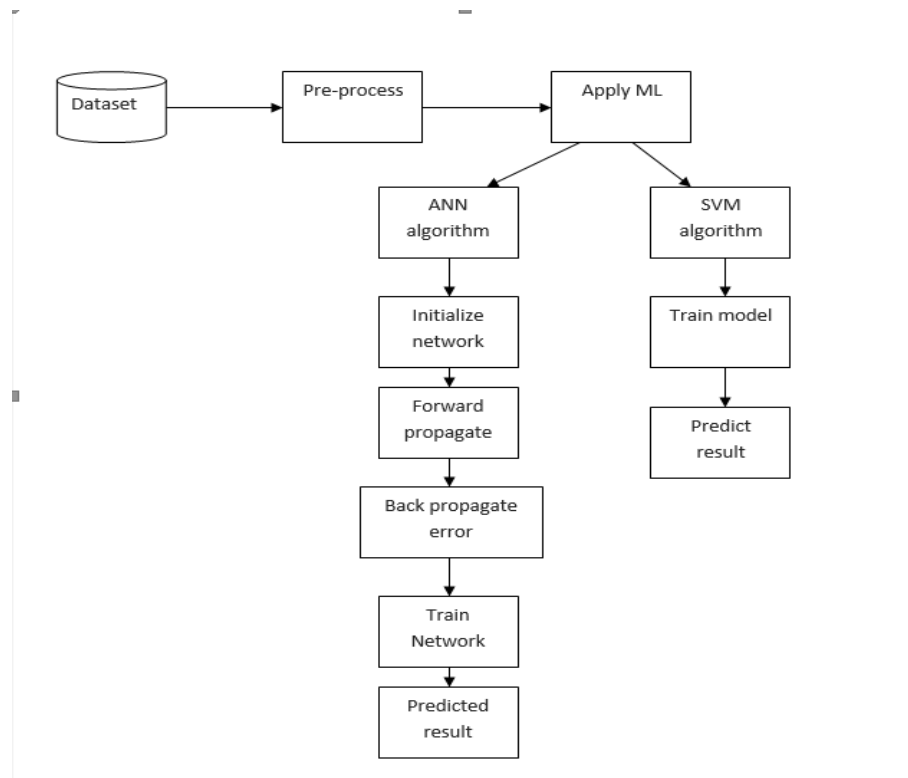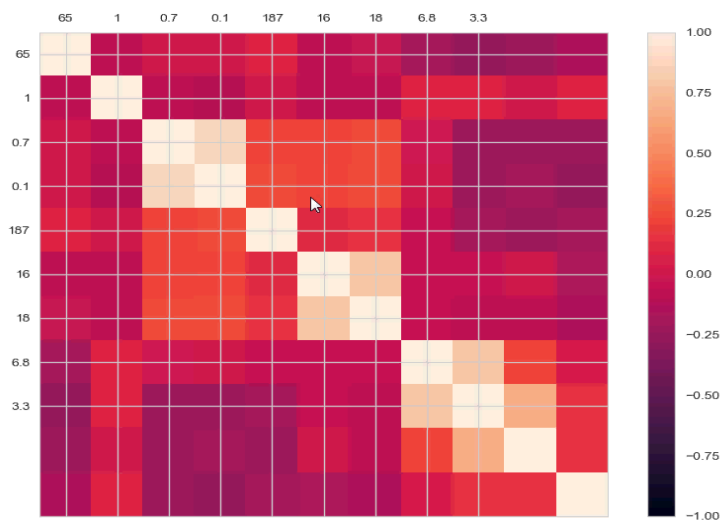

VII.EXPERIMENTAL RESULTS

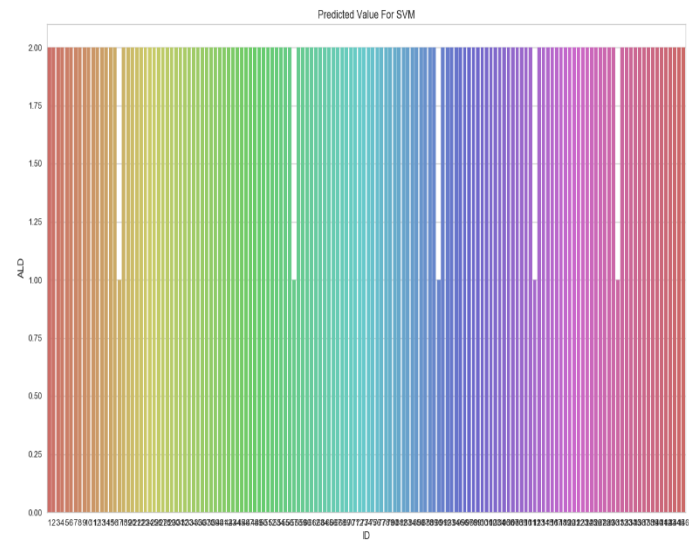**Figure 1: Scatter plot for SVM algorithm is plotted below**

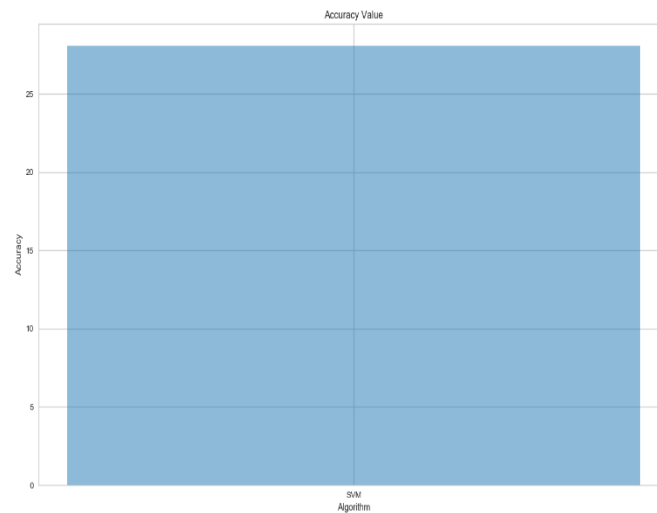**Figure 2: Liver disease is predicted through SVM algorithm and plotted below**
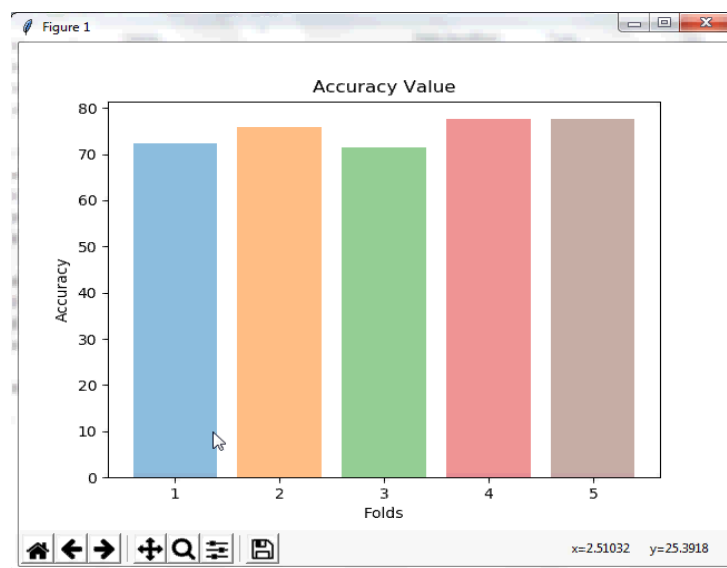

Figure 3: Accuracy for SVM algorithm

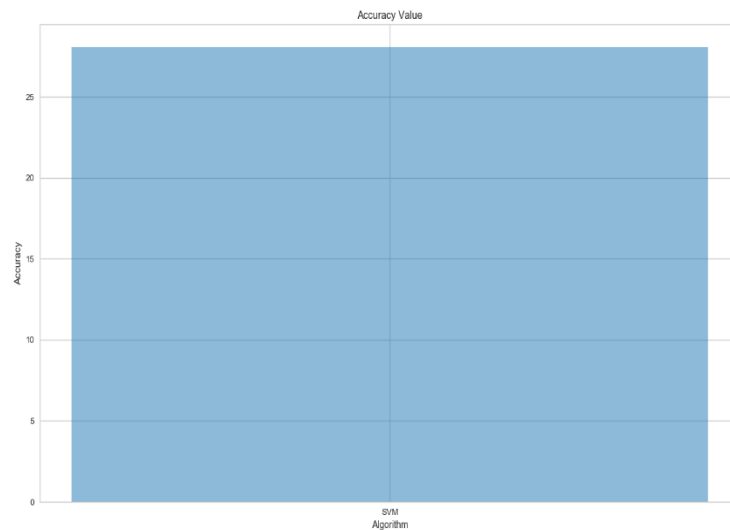**Figure 4: Accuracy for different folds of ANN algorithm**

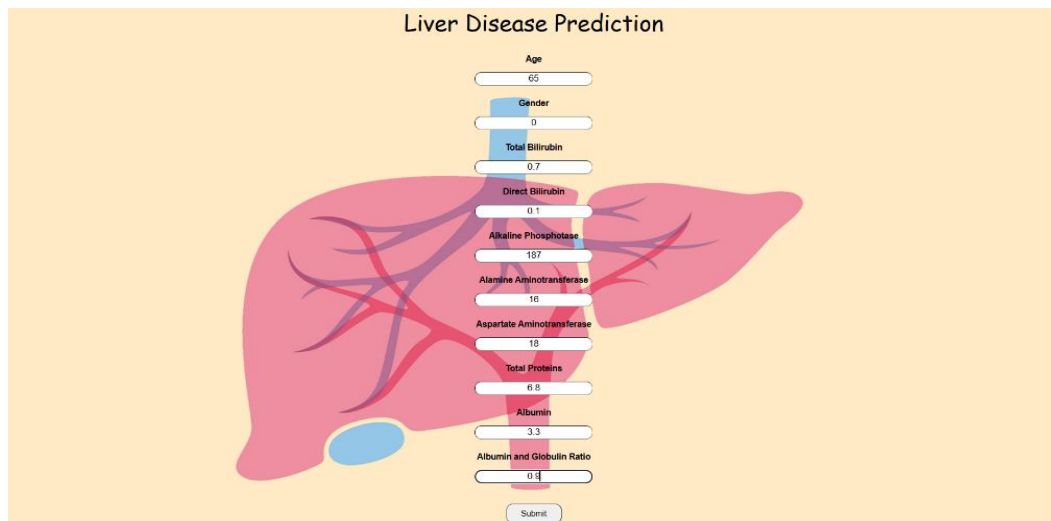**Figure 5: Accuracy for SVM Classification algorithm**
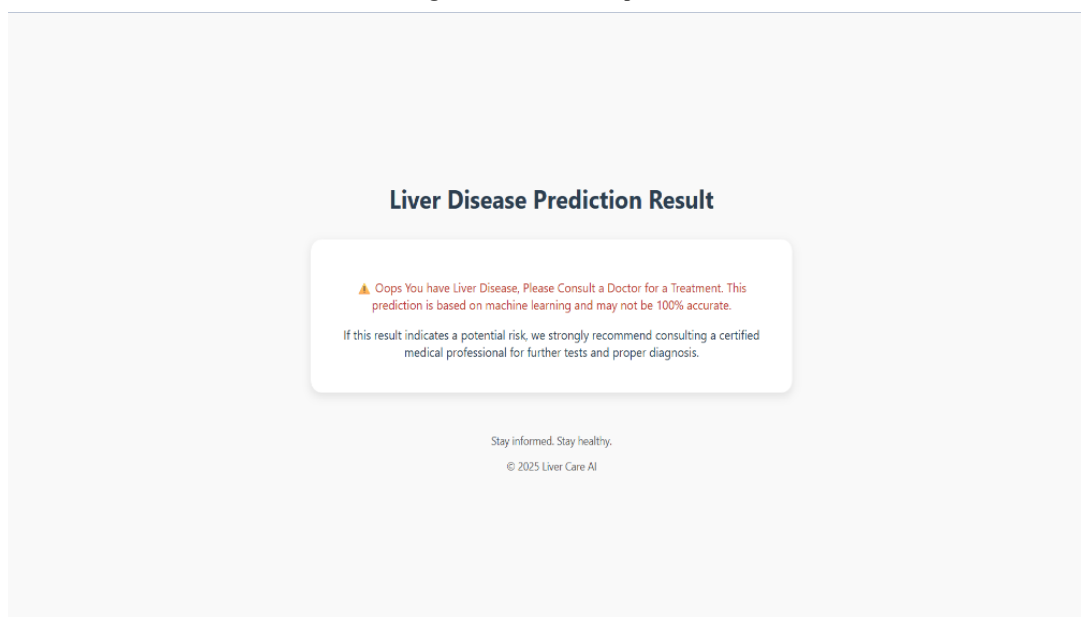

Figure 6: Details of the patient

Figure 7: Final Prediction of the Liver Disease

VIII.CONCLUSION

This study investigates two machine learning approaches—Artificial Neural Networks (ANN) and Support Vector Machines (SVM)—for predicting chronic liver disease. Diagnosing liver conditions is particularly challenging due to the subtle and often overlooked nature of early symptoms. In 2014 alone, chronic liver disease was responsible for approximately 38,170 deaths out of 2.6 million total deaths in the United States. The role of computational prediction in medical diagnostics is becoming increasingly vital. This work evaluates how different ML models can enhance diagnostic accuracy. While molecular biology-based methods may be influenced by factors like age, diet, and ethnicity, chemical analysis tends to offer more consistent results. Nevertheless, ongoing research in molecular biology holds significant promise for deepening our understanding of human physiology and advancing life-saving medical innovations.

XI.FUTURE WORK

Although the proposed system using Artificial Neural Networks (ANN) and Support Vector Machines (SVM) has demonstrated promising results in liver disease prediction, there are several avenues for enhancing its capabilities and applicability in real-world medical settings:

• Integration of Deep Learning Architectures
• Real-Time Prediction System with GUI
• Multi-Disease Prediction Framework

The outlined future directions aim to significantly enhance the predictive power, reliability, and practical usability of the liver disease diagnosis system. By integrating cutting-edge deep learning architectures, incorporating larger and more diverse datasets, and adopting explainable AI techniques, the system can evolve into a more accurate and transparent diagnostic tool. Furthermore, embedding the solution within real-time, user-friendly interfaces and extending it to mobile and web platforms ensures wider accessibility, especially in under-resourced regions..

REFERENCE:

1. Rong-Ho Lin, "An Intelligent Model for Liver Disease Diagnosis," Artificial Intelligence in Medicine, 2009"
2. Ryan Rifkin, Sridhar Ramaswamy, Pablo Tamayo, Sayan Mukherjee, Chen-Hsiang Yeang, Micheal Angelo, Christine Ladd, Micheal Reich, Eva Latulippe, Jill P Merisov, Tomaso Poggio, William Gerald, Massimo Loda, Eric S Lander, Todd R Golub, "An Analytical Method For Multi-Class Molecular Cancer Classification ", 2003
3. Akin Ozcivit and Arif Gulten "Classifier Ensemble Construction With Rotation Forest To Improve Medical Diagnosis Performance Of Machine Learning Algorithms",2011
4. Kun-Hong Liu and De-Shuang Huang. "Cancer classification using Rotation forest", Computers in Biology and Medicine, 2008
5. BendiVenkataRamana, Prof. M.Surendra Prasad Babu and Prof. N. B. Venkateswarlu, "A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis". International Journal of Engineering Reasearch and Development, 2012
6. V.N. Vapnik, "Statistical Learning Theory", Wiley Publications, 1998
7. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Delving Deep into Rectifiers", Microsoft Research, 2009
8. Beilharz TH, Preiss T: Translational profiling: the genome-wide measure of the nascent proteome. Brief Funct Genomic Proteomic, 2009.
9. Gros F: From the messenger RNA saga to the transcriptome era. C R Biol. 2003, 326: 893-900.
10. Shackel NA, Gorrell MD, McCaughan GW: Gene array analysis and the liver. Hepatology. 2002, 36: 1313-1325. 10.1053/jhep.2002.36950.