WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Prediction of Liver Disease in Patients Using Logistic Regression of Machine Learning

Harsha Vardhan [a], Dr K Ramesh babu[b], G. Lokeswar Raju[c], Sk. Saddam Hussain[d], P. Manikanta[e], P. Kavya Sri[f]

[a] *Asst. Professor, CSE, JNTUK, GVR&S CET, Guntur,522017, Andhra Pradesh, India,*
[b] *Professor, CSE, JNTUK, GVR&S CET, Guntur, 522017, Andhra Pradesh, India*
[C,d,e, f] *Research Scholars, JNTUK,GVR&S CET, Guntur, 522017, Andhra Pradesh, India*

DOI: https://doi.org/10.55248/gengpi.5.0524.1237

ABSTRACT

The existence of people living without liver cancers is one of the basic considerations of human jobs. Subsequently, for better consideration, the discovery of liver sickness at a crude stage is important. For clinical specialists, foreseeing the disease in the beginning phases because of unobtrusive signs is an undeniably challenging undertaking. Many, when it is past the point of no return, the signs become obvious. The ebb and flow work plans to expand the apparent idea of liver infection utilizing AI strategies to settle this pandemic. The vital reason for the current work zeroed in on calculations for the characterization of sound individuals from liver datasets. Jogging on their prosperity factors, this exploration likewise means to contrast the grouping calculations and with give forecast exactness results. Choice Tree, Arbitrary Backwoods, Strategic Relapse, and Backing Vector Machine calculations are utilized to Foresee the sickness,by comparing accuracy rates, the goal of the paper is to predict liver disease and select the best machine-learning algorithms.finally found Logistic regression is best than support Vector Machine in terms of accurary 72%.

Keywords: *Liver disease prediction, Logistic Regression, Support Vector Machine etc,.*

1. Introduction

The liver is a significant organ of the human body, and it is situated underneath the rib confined in the right upper mid-region. It eliminates poisons from the body and keeps up with solid blood sugar level in the body. However, body organs have self-recuperating limits, over-utilization of liquor, and openness to polluted air and water influence the liver which prompts a higher pace of Liver disappointment. Liver transplantation is the arrangement yet with greater expense and lower pace of progress. Distinguishing the liver harm at the earliest can lessen the opportunity for liver disappointment. The AI model is equipped for anticipating illnesses, because of an informational index, which is an implicit blend of key wellbeing boundaries of an individual with and without sicknesses. A successful informational index is required for building models, with the appropriate portrayal of sickness groupings.

In this work, a patient dataset is gathered from Kaggle. A few characterization AI calculations can group liver sicknesses. Rather than choosing the calculation, which gives better execution, the paper approaches how to tune the ML module for Strategic Relapse Modell in bit-by-bit ways. The Irregular Woods calculation is to construct a model since it is prepared on a few examples of information obtained from dividing information so the model isn't tuned for unmistakable information. The paper's main focus is on an in-depth analysis of how an imbalanced data set can be used to fine-tune models beyond a single saturation point. Different adjusting methods talked about and their effects on execution are organized in later segments. Segment 2 covers a writing overview, in this work, a few models worked in iterative ways, after each key step performed on the informational collection, and its effect on execution improvement is noted. The outcome and model advancement segment talks about this exhaustively with figures and results. In the last segment, ends are incorporated.

2. Literature review

In [1], authors find and observed that some machine learning algorithms such as Decision Tree, J48, and ANN provide better accuracy in the detection and prediction of liver disease.In [2], liver disease prediction has been studied and analyzed in this paper. The data is cleaned by performing various techniques such as imputation of missing values with median, label encoding to convert categorical into numerical data for easy analysis, duplicate value elimination, and outliers are eliminated using Isolation Forest to improve the performance. In [3] proposed a paper using Machine Learning algorithms to compare a Support Vector Machine and Logistic regression algorithms. In [4] proposed a paper using Logistic Regression, Decision Tree, and Random Forest Tree those Machine learning algorithms to predict a Liver Disease. In [5] proposed a paper using Logistic Regression and Multi-Layer Perception performed efficiently compared to the other basic machine learning models. In [6] proposed a paper using different classification algorithms namely Logistic Regression, Support Vector Machine, and K-Nearest Neighbour have been used for liver disease prediction. In [7] proposed a paper to comparison between different machine learning approaches on advanced liver fibrosis in Chronic Hepatitis C patients. In [8] This paper makes use of the lab test reports of the patients who have undergone Liver Function Tests. MATLAB2016 is used and developed a model by applying classification algorithms SVM, Logistic Regression, and Decision tree. In [9], ML models are built using various preprocessing techniques to balance the unbalanced data and predicted using the Random Forest algorithm. In [10] In this work, we have developed the K-Nearest Neighbour model to diagnose and predict liver disease.

3. Summary of Literature Survey

by studying all the above papers, the gap found, limitations in existing techniques, and also the accuracy is less compared to one other. The results that they achieved are not very accurate. Moreover, some literature faces regional variability's challenges as different Algorithms to predict disease. The accuracy can change based on the dataset. In this paper used by Support Vector Machine and Logistic Regression. Both algorithms accuracy quite difference.

3.1 Delimitation

The delimitations define the boundaries within which the study operates. For instance, in a study on liver disease prediction using logistic regression, the delimitations might include specific demographics of the population under study.

3.2 Limitation

Limited sample size can impact the generalizability of findings. For instance, if the study only includes a small number of patients with liver disease, the logistic regression model's accuracy might be limited.

4. Proposed architecture

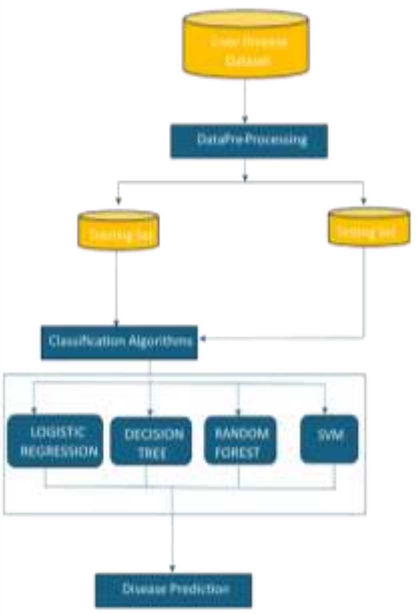


Fig: 4 Architecture of liver disease prediction using logistic regression of machine learning

Fig: 4 shows Architecture of liver disease prediction using logistic regression which will help in processing the work.the work flow shows collect the live decease data sets primary,then split in to two parts training,data sets and training & test apply to algorithm which is choose among various.finally find the decease by using machine learning.

4.1 designing phases

Liver Disease Prediction can be estimated using Machine Learning techniques by dataset parameters. Obviously characterize the issue explanation. For this situation, it is anticipate the probability of liver sickness in light of specific highlights or chance variables.

4.2 Data Collection

Characterize the issue explanation. When you have a sufficient number of patients with both positive and negative outcomes for liver disease, you can preprocess the data, select features if necessary, and then train a logistic regression model to predict the likelihood of liver disease based on the input features or risk factors. In this instance, the goal would be to predict the likelihood of liver disease based on specific features.

4.3 Algorithms

In machine learning mainly there are two types of algorithm models they are: Classification Algorithms, Regression Algorithms.

Classification Algorithms:Classification algorithms are used to predict categorical labels or classes for new data points based on past observations. The goal is to learn a mapping from input features to predefined categories.

Logistic Regression Fig 4.1 shows below explain Logistic regression is a statistical method and a type of predictive analysis used in machine learning tasks. Logistic regression is a statistical method used for binary classification problems, where the outcome variable y is categorical and has only two possible outcomes (0 and 1, or "yes" and "no"). It models the probability that a given input belongs to a certain category. For example, in medical

i) **diagnosis**, logistic regression can be used to predict whether a patient has a particular disease or not based on various features such as age, gender, blood pressure, etc.

ii) **Decision tree** Fig 4.2 shows below explain A decision tree is a machine learning algorithm used for both classification and regression tasks. It models decisions and their possible consequences by creating a tree-like structure of decisions, making it an intuitive and visually interpretable method. Decision trees split the data into branches to make predictions, using the structure of a tree consisting of nodes and leaves.

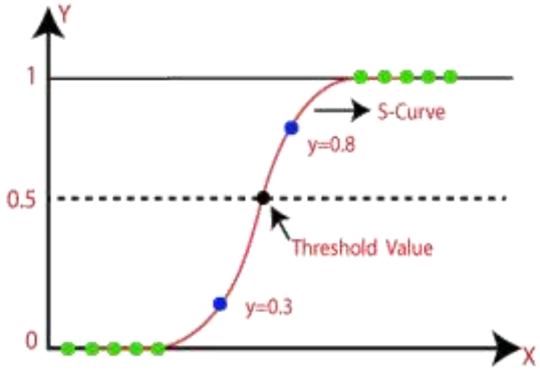


Fig: 4.1 Logistic regression

A Decision Tree is a versatile and intuitive machine-learning algorithm used for both classification and regression tasks. It's a tree-like structure where an internal node represents a feature (or attribute), the branch represents a decision rule, and each leaf node represents the outcome (or class label). The below diagram explains the general structure of a decision tree:

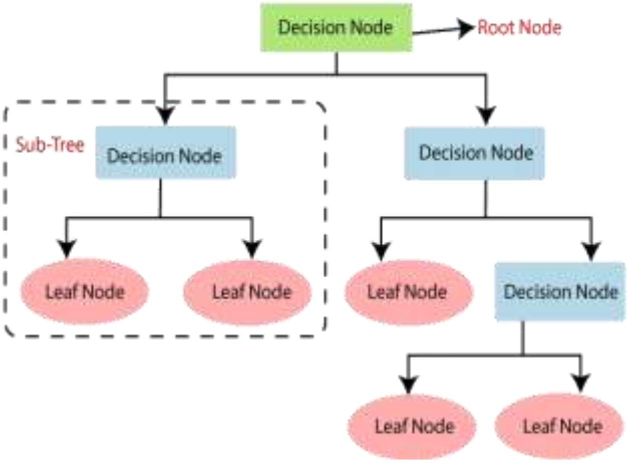
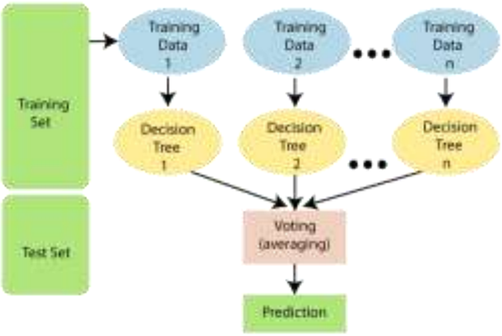


Fig: 4.2 Decision tree algorithms

iii)Random Forest Fig 4.3 shows below explain A Random Forest is an ensemble learning method used in machine learning that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.Random Forest is an ensemble learning method used for both classification and regression tasks. It operates by constructing a multitude of decision trees during training and outputs the class that is the mode of the classes (classification) or the mean prediction (regression) of the individual trees. Here is an explanation of the Random Forest algorithm: The below diagram explains the working of the Random Forest.

Fig: 4.3 Random Forest algorithms



iv) Support Vector Machine Fig 4.4 shows below explain support Vector Machine (SVM) is a powerful, supervised machine learning algorithm used for both classification and regression, though it is more commonly used for classification tasks. The basic idea of SVM is to find a hyper plane in an N-dimensional space (N the number of features) that distinctly classifies the data points. Support Vector Machines (SVM) is a powerful supervised learning algorithm used for classification and regression tasks. In this explanation, we'll focus on the

SVM algorithm for classification, particularly for linearly separable Support Vector Machines or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

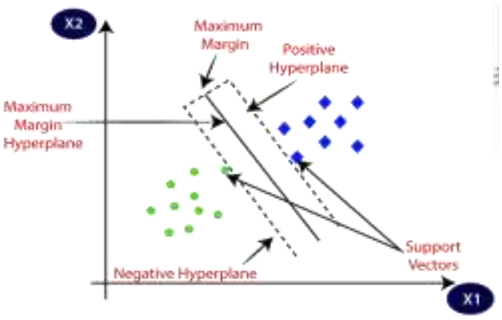


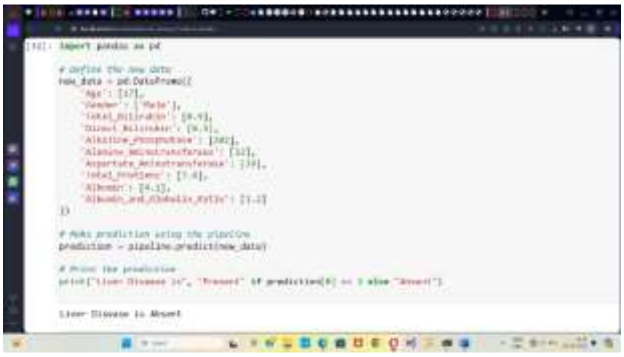
Fig: 4.4 Support Vector Machine

5. Results

 Fig: 5.1 Data Set	 Fig: 5.2 Plot Graph

Fig: 5.3 Logistic Regression

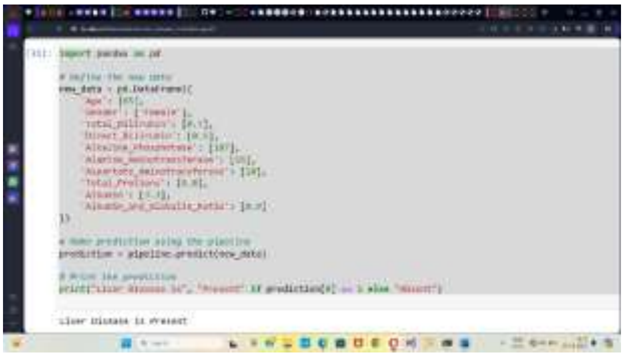
Fig: 5.4 SVM



```
import pandas as pd

# Define the new data
new_data = pd.DataFrame({
    'Age': [17],
    'Gender': ['Male'],
    'Total_Bilirubin': [0.9],
    'Direct_Bilirubin': [0.3],
    'Alkaline_Phosphotase': [202],
    'Alamine_Aminotransferase': [22],
    'Aspartate_Aminotransferase': [19],
    'Total_Protiens': [7.4],
    'Albumin': [4.1],
    'Albumin_and_Globulin_Ratio': [1.2]
})

# Make prediction using the pipeline
prediction = pipeline.predict(new_data)

# Print the prediction
print("Liver Disease is", "Present" if prediction[0] == 1 else "Absent")
```

```
Liver Disease is Absent
```

Fig: 5.5 Disease Absent



```
import pandas as pd

# Define the new data
new_data = pd.DataFrame({
    'Age': [65],
    'Gender': ['Female'],
    'Total_Bilirubin': [0.7],
    'Direct_Bilirubin': [0.1],
    'Alkaline_Phosphotase': [187],
    'Alamine_Aminotransferase': [16],
    'Aspartate_Aminotransferase': [18],
    'Total_Protiens': [6.8],
    'Albumin': [3.3],
    'Albumin_and_Globulin_Ratio': [0.9]
})

# Make prediction using the pipeline
prediction = pipeline.predict(new_data)

# Print the prediction
print("Liver Disease is", "Present" if prediction[0] == 1 else "Absent")
```

```
Liver Disease is Present
```

Fig: 5.6 Disease Present

From figs 5.1 to 5.6 shows and clearly tells about data set used,and their graphs ,simulations of Logistic regression,SVM,dicease absent,Dicease presents respectively.

Table 5.1 comparison of results

Algorithm	Accuracy
SVM	70
Linear regression	**72**

From the table 5.1shows results comparison between SVM,Linear regression methods to find the disease in terms of accuracy to find the liver decease in human body. the above table tells us clearly linear regression has better results than SVM.

6. Conclusion

It explored the effectiveness of machine learning algorithms for Liver Disease Prediction, with a focus on accuracy. The data is cleaned by performing various techniques such as imputation of missing values with median, and label encoding to convert categorical into numerical data for easy analysis. In this paper, used two machine learning algorithms are Support Vector Machine algorithms gets 7% accuracy, and the Logistic Regression Algorithm got a 72% accuracy . The conclusion of a liver disease prediction study would highlight the model's performance, its ability to identify key risk factors, potential areas for improvement, and actionable recommendations for healthcare providers, policymakers, and individuals at risk of liver disease. The goal is to improve early detection, patient outcomes, and public health strategies related to liver health.

7. Future scope

For future scholars improvement suggestion is, Gather the different live datasets and using multiple machine learning models to improve the accuracy of prediction liver disease prediction using logistic regression is characterized by advancements in data integration. by harnessing the power of predictive analytics and leveraging interdisciplinary collaborations, researchers can continue to refine and optimize logistic regression models to enhance the early detection, prevention, and management of liver diseases.

8. References

1. Williams, Roger. "Global challenges in liver disease." Hepatology 44.3 (2006): 521-526.

2. Clark, Jeanne M., Frederick L. Brancati, and Anna Mae Diehl. "Nonalcoholic fatty liver disease." Gastroenterology 122.6 (2002): 1649-1657.

3. De Zeng, Min, et al. "Guidelines for the diagnosis and treatment of nonalcoholic fatty liver diseases." Journal of digestive diseases 9.2 (2008).

4. Vijayarani, S., and S. Dhayanand. "Liver disease prediction using SVM and Naïve Bayes algorithms." International Journal of Science, Engineering and Technology Research (IJSETR) 4.4 (2015): 816-820.

5. Rahman, A. Sazzadur, et al. "A comparative study on liver disease prediction using supervised machine learning algorithms." International Journal of Scientific & Technology Research 8.11 (2019): 419-422.

6. Nahar, Nazmun, and Ferdous Ara. "Liver disease prediction by using different decision tree techniques." International Journal of Data Mining & Knowledge Management Process 8.2 (2018): 01-09.

7. Priya, M. Banu, P. Laura Juliet, and P. R. Tamilselvi. "Performance analysis of liver disease prediction using machine learning algorithms." Int. Res. J. Eng. Technol 5.1 (2018): 206-211.

8. Durai, Vasan, Suyan Ramesh, and Dinesh Kalthireddy. "Liver disease prediction using machine learning." Int. J. Adv. Res. Ideas Innov. Technol 5.2 (2019): 1584-1588.

9. Mostafa, Fahad, et al. "Statistical machine learning approaches to liver disease prediction." Livers 1.4 (2021): 294-312.

10. Veeranki, Sreenivasa Rao, and Manish Varshney. "Intelligent techniques and comparative performance analysis of liver disease prediction." International Journal of Mechanical Engineering 7.1 (2022): 489-503.

11. Kalaiselvi, R., K. Meena, and V. Vanitha. "Liver Disease Prediction Using Machine Learning Algorithms." 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA). IEEE, 2021.

12. Mutlu, Ebru Nur, et al. "Deep learning for liver disease prediction." Mediterranean Conference on Pattern Recognition and Artificial Intelligence. Cham: Springer International Publishing, 2021.

13. Gupta, Ketan, et al. "Liver disease prediction using machine learning classification techniques." 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT). IEEE, 2022.

14. Vats, Varun, et al. "A comparative analysis of unsupervised machine techniques for liver disease prediction." 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT). IEEE, 2018.

15. Nahar, Nazmun, et al. "A comparative analysis of the ensemble method for liver disease prediction." 2019 2nd International Conference on Innovation in Engineering and Technology (ICIET). IEEE, 2019.

16. Afrin, Saima, et al. "Supervised machine learning based liver disease prediction approach with LASSO feature selection." Bulletin of Electrical Engineering and Informatics 10.6 (2021): 3369-3376.

17. Khan, Md Ashikur Rahman, et al. "An effective approach for early liver disease prediction and sensitivity analysis." Iran Journal of Computer Science 6.4 (2023): 277-295.

18. Singh, Jagdeep, Sachin Bagga, and Ranjodh Kaur. "Software-based prediction of liver disease with feature selection and classification techniques." Procedia Computer Science 167 (2020): 1970-1980.

19. Pirola, Carlos J., and Silvia Sookoian. "Multiomics biomarkers for the prediction of nonalcoholic fatty liver disease severity." World journal of gastroenterology 24.15 (2018): 1601.

20. Kotronen, Anna, et al. "Prediction of non-alcoholic fatty liver disease and liver fat using metabolic and genetic factors." Gastroenterology 137.3 (2009): 865-872.

21.implementation of crop water requirement using machine learning

22.https://scholar.google.com/citations?view_op=view_citation&hl=en&user=sHqj0cIAAAAJ&cstart=20&pagesize=80&citation_for_view=sHqj0cIAAAAJ:kNdYIx-mwKoC

23.https://scholar.google.com/citations?view_op=view_citation&hl=en&user=sHqj0cIAAAAJ&cstart=20&pagesize=80&citation_for_view=sHqj0cIAAAAJ:7PzlFSSx8tAC

24.https://scholar.google.com/citations?view_op=view_citation&hl=en&user=sHqj0cIAAAAJ&cstart=20&pagesize=80&citation_for_view=sHqj0cIAAAAJ:L8Ckcad2t8MC