# Liver Cancer Identification Grid Search RFC Model using Machine Learning

**Ms. Sunita Kumari**
*M.Tech Student*
*Sobhasaria group of institutions*
*Sikar, India*
*mandiwallsunita@gmail.com*

**Mr.Gourav Mitawa**
*Assistant Professor*
*Sobhasaria group of institutions*
*Sikar, India*
*gouravmitawa028@gmail.com*

***Article History***

***Article Submission***
*20 December 2022*
***Revised Submission***
*15 January 2023*
***Article Accepted***
*25 February 2023*
***Article Published***
*30 March 2023*

***Abstract***

*Liver is essential to the body's digestion of sugar and fats, absorption, and immunological system. This substance is present in almost everything a man takes in, breathes, or absorbs through his skin. Liver disorders are a significant health burden. It is increasing daily and is difficult to detect in its early stages since the liver may function normally even when partially damaged. Doctors have widely employed machine learning algorithms to diagnose liver illness in order to increase the efficiency of medical diagnosis. The study's primary aim is to evaluate how machine learning algorithms may be used to prevent postponing medical care, accurately diagnose liver illness, and minimize the number of erroneous diagnoses provided to sick patients. The main objective is to ensure that liver patients receive an accurate diagnosis as soon as possible.*

***Keywords****- Classifier, Random Forest, Logistic Regression, Support vector, RFC*

## I. INTRODUCTION

The liver is the most important organ in the human body. The liver of an adult weighs roughly 3 pounds. The rib cage shields the liver, which is situated on the right side beneath the right lung. It resembles a chemical plant. Digestion and the transformation of food into protein and bile depend on the liver's healthy operation. This controls the elimination and disrupts a number of intestinal nutrients. With the aid of vitamins, carbohydrates, and minerals stored in the liver, it breaks down several nutrients from the intestine and regulates the excretion of cholesterol to produce rapid energy when required. The basic unit of the body's tissues is the cell. Growing and differentiating into new cells is a property of all cells. whenever the old cell or the injured new cell. The approach frequently fails. The body doesn't make any new cells, but tissues made of ageing and damaged cells might create nodules or tumors. Males are diagnosed with liver cancer at a rate of 5 percent, while females are diagnosed at a rate of 9 percent. In 2018, there were about 840,000 new cases. Mongolia, Egypt, and Mongolia had the highest rates of liver cancer in 2018. .[1], [2]

The liver is a triangular-shaped organ that covers the entire stomach opening and is located just below the diaphragm. The right half of the body is largely occupied by the liver, which moves relentlessly in the direction of the right kidney. The liver is made up of tiny, connective tissue-coated, pinkish-cocoa tissues. The peritoneum of the stomach cavity maintains and reinforces this casing by securing and supporting the liver inside the mid-region.

Additionally, the liver performs the function of a filter by removing potentially harmful compounds from the bloodstream before they can reach the brain.

## II. RELATED WORK

A comprehensive assessment of 37 research articles, their processing, and enhancement approaches, published between 2009 and 2021, was conducted.**[Kahramanli et al. 2009]** [4] Utilize the provided neural network to develop an adaptive activation function for rule extraction. **[Lin et al. 2010]** [5] The intelligent liver diagnosis models are used to make early diagnoses of liver illnesses in this paper. ANN, the Analytic hierarchy process (AHP), and CBR are all used in the intelligent liver diagnosis models.

**[Vahini et al. 2015]** [6]involves the use of an information retrieval technique to diagnose liver cancer. We compared the performance, accuracy, and cost of C4.5, Nave Bayes, Decision trees, Support Vector Machines, Back Propagation neural networks, and classification and regression trees. All other algorithms were outperformed by C4.5.

**[Ambesange et al. 2020]** [7] In this work, the K-Nearest Neighbor model was suggested for diagnosing and forecasting liver illness. The data is transformed to improve the model's performance, and additional dimensionality reduction is performed to reduce the number of features.

**[Kalsoom et al.2021]** [8] suggested an approach combining unsupervised machine learning with a supervised mechanism for correctly segmenting liver tumors. The features like LBP and HOG have been extracted, and classification is done using KNN. The overall accuracy of 97% is achieved. As compared to SVM and Ensemble, which have shown the accuracy of 85% and 49%, respectively, KNN has performed better.

Most of the researchers used predefined selected attributes to determine the accuracy level, but sometimes they didn't find better results so that more attributes can be added for achieving better accuracy.

Most of the researchers have not used feature selection and extraction techniques that are very important for getting the good performance of the machine learning algorithm. Very little work has been done for data preprocessing for both numerical data and image data together that impact the accuracy of the ML model.

## III. PROPOSED METHOD

We proposed a method to evaluate the diagnostic performance of several machine learning algorithms on liver disease datasets. We will combine dimensionality reduction and optimization approaches with the chosen classification algorithms for the purpose of assessing result variation and to conduct a performance study on classification algorithms and to choose the most appropriate classification models for liver disorders.

We proposed a method to detect the live cancer using Support Vector Machine, Logistic Regression, Random Forest, multilayer perceptron, K Nearest Neighbor, Gradient boosting. We have taken two data sets The ILPD (Indian Liver Patient Dataset) dataset was retrieved from UC Irvine's data repository (UCI). Including 416 individuals with the liver illness and 167 healthy people without liver disease, it includes 583 total data points. The second data set is UCI Machine Learning Repository's BUPA Liver Disorders dataset.

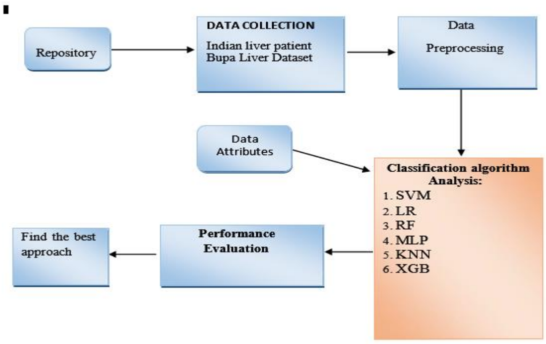


*Fig.1. Process flow diagram*

*A. Experimental Result and Analysis :*

Each algorithm's performance was compared to a benchmark set of data. We observed that both the data set i.e ILPD and BUPA liver disorder in the RFC and XGB has performed better. So we have chosen both for our model and compared them. So when they were used along with hyperparameter tuning using Gid Search CV both performs better. So for Data set 1, normal RFC has detected the disease with 82.4% accuracy, whereas grid search RFC shows 84.4% accuracy that is an overall 2% improvement and for XGB, normal XGB predicted the disease with 74.8% accuracy whereas grid search XGB predicted the disease with 81.2 % accuracy that is overall 6.4% improvement.

For data set 2, normal RFC has predicted the disease with 79.16% accuracy whereas grid search RFC has shown improvement of 3.34% and predicted the disease with 82.5% accuracy. Normal XGB has predicted with an accuracy of 77.5% accuracy and grid search XGB with an accuracy of 80% is an overall improvement of 2.5%. Overall for both the data i.e data set1 and dataset 2it is observed that Grid Search Random Forest has performed better with an overall accuracy of 84.4% and 82.5% accuracy.

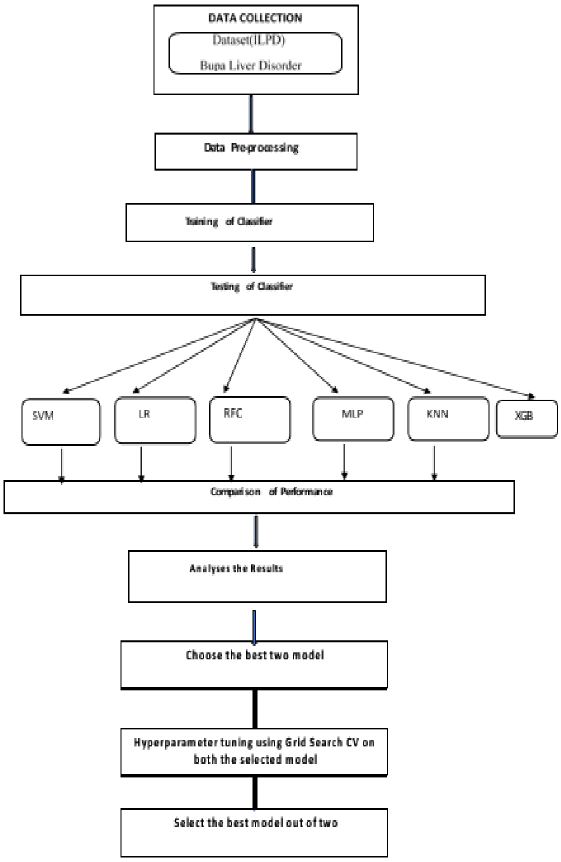


*Fig. 2 Proposed System*

## IV. RESULT ANALYSIS

We observed that both the data set i.e ILPD and BUPA liver disorder in the RFC and XGB has performed better. So we have chosen both for our model and compared them. So when they were used along with hyperparameter tuning using Gid Search CV both performs better. So for Data set 1, normal RFC has detected the disease with 82.4% accuracy, whereas grid search RFC shows 84.4% accuracy that is an overall 2% improvement and for XGB, normal XGB predicted the disease with 74.8% accuracy whereas grid search XGB predicted the disease with 81.2 % accuracy that is overall 6.4% improvement.

For data set 2, normal RFC has predicted the disease with 79.16% accuracy whereas grid search RFC has shown improvement of 3.34% and predicted the disease with 82.5% accuracy. Normal XGB has predicted with an accuracy of 77.5% accuracy and grid search XGB with an accuracy of 80% is an overall improvement of 2.5%.

ISSN: 2250-0839

Overall for both the data i.e data set1 and dataset 2it is observed that Grid Search Random Forest has performed better with an overall accuracy of 84.4% and 82.5% accuracy.

*Table 1 Comparison of RFC and Grid Search RFC for dataset 1*

| Parameters | RFC | Grid Search RFC |
|---|---|---|
| **Accuracy** | 83.2 | 84.4 |
| **Precision Score** | 89.60 | 90.0 |
| **Recall Score** | 73.10 | 75.63 |
| **F1 Score** | 80.55 | 82.19 |
| **MEA** | 16.8 | 15.6 |
| **Root Mean Square Error** | 40.98 | 39.49 |
| **ROC_AUC Score** | 82.73 | 83.99 |

*Table 2 Comparison of RFC and grid search RFC for dataset 2*

| Parameters | RFC | Grid Search RFC |
|---|---|---|
| **Accuracy** | 83.2 | 84.4 |
| **Precision Score** | 89.60 | 90.0 |
| **Recall Score** | 73.10 | 75.63 |
| **F1 Score** | 80.55 | 82.19 |
| **MEA** | 16.8 | 15.6 |
| **Root Mean Square Error** | 40.98 | 39.49 |
| **ROC_AUC Score** | 82.73 | 83.99 |

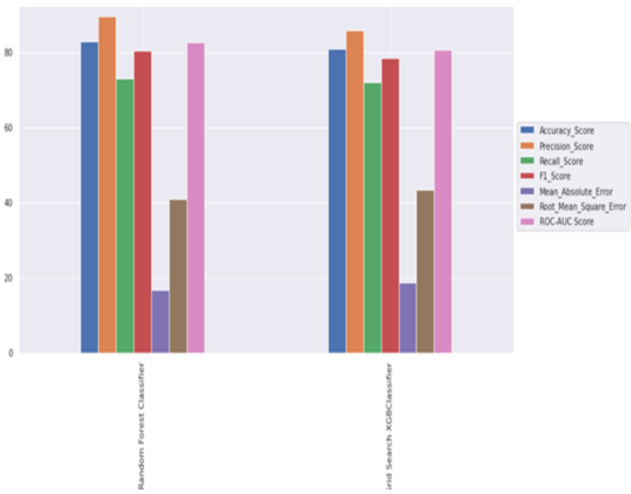


*Figure 4.1 Graph showing Performance of Grid Search RFC and XGB on BUPA data*

ISSN: 2250-0839

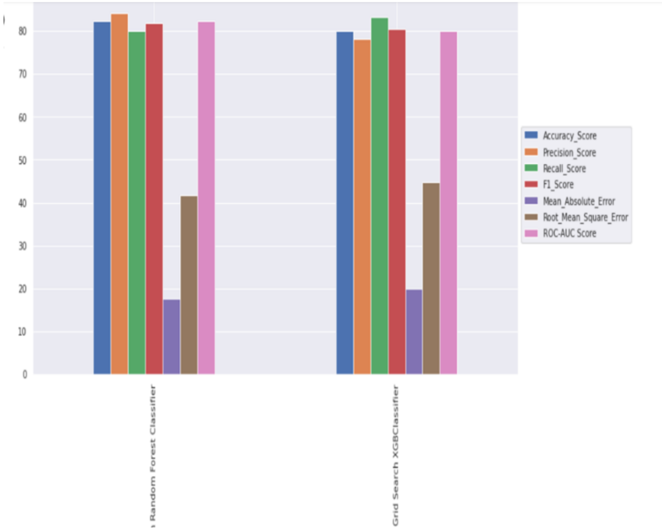


*Figure 4.2 Graph showing performance of Grid search RFC and XGB on ILPD data*

## V. CONCLUSION

The liver disease is on the rise, it's critical to anticipate the likelihood of developing liver cancer. Despite advances in medical technology, detecting liver disease is still a challenging task. Early identification of liver illness is challenging due to the liver's ability to function normally even after part of it has been damaged. Diagnostic efficiency and effectiveness were improved by developing models for classifying the presence or absence of liver disease. As a result, professionals are more equipped to provide patients with appropriate care. In this study, we developed the Grid Search RFC model to diagnose and forecast liver disease. To improve performance, the data is converted and further reduced to minimize the number of jobs in the model. Classification and prediction algorithms are evaluated using a variety of performance metrics, including accuracy, precision, recall, and F-1 scores, root means square errors, root mean square errors, and ROC AUC scores. The model hyperparameters are set via grid search. With an accuracy rate of 84.4 percent for Dataset 1 (ILPD) and 82.5 percent for Dataset 2 (BUPA), the RFC model excelled.

## REFERENCES

[1] J. Rehm, A. v. Samokhvalov, and K. D. Shield, "Global burden of alcoholic liver diseases," *Journal of Hepatology*, vol. 59, no. 1, pp. 160–168, 2013, doi: 10.1016/j.jhep.2013.03.007.

[2] B. F, F. J, S. I, S. RL, T. LA, and J. A, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 68, no. 6, pp. 394–424, Nov. 2018, doi: 10.3322/CAAC.21492.

[3] H. Kahramanli and N. Allahverdi, "Rule extraction from trained adaptive neural networks using artificial immune systems," *Expert Systems with Applications*, vol. 36, no. 2 PART 1, pp. 1513–1522, 2009, doi: 10.1016/J.ESWA.2007.11.024.

[4] R. H. Lin and C. L. Chuang, "A hybrid diagnosis model for determining the types of the liver disease," *Computers in Biology and Medicine*, vol. 40, no. 7, pp. 665–670, 2010, doi: 10.1016/j.compbiomed.2010.06.002.

[5] R. Vahini and B. Sathees Kumar, "A survey on liver disorder using classification techniques in data mining," *International Journal of Applied Engineering Research*, vol. 10, no. 55, pp. 2796–2799, 2015.

[6] S. Ambesange, R. Nadagoudar, R. Uppin, V. Patil, S. Patil, and S. Patil, "Liver Diseases Prediction using KNN with Hyper Parameter Tuning Techniques," *Proceedings of B-HTC 2020 - 1st IEEE Bangalore Humanitarian Technology Conference*, Oct. 2020, doi: 10.1109/B-HTC50970.2020.9297949.

[7] A. Kalsoom, A. Moin, M. Maqsood, I. Mehmood, and S. Rho, "An Efficient Liver Tumor Detection using Machine Learning," pp706–711,2021,doi: 10.1109/csci51800.2020.

ISSN: 2250-0839