

A Peer Revieved Open Access International Journal

www.ijiemr.org

COPY RIGHT





2021IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be

obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 12th June 2022.

Link: https://ijiemr.org/downloads/Volume-11/Issue-06

Title: DETECTION OF LIVER DISEASE USING HYPER TUNED GRIDSEARCH

volume 11, Issue 06, Pages: 1532-1541

Paper Authors: Mr. Anand Thota, Ms. N Susmitha, Mr. D S S Srinivas, Ms. L Monika, Mr. T J Sai Sravan





USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER



A Peer Revieved Open Access International Journal

www.ijiemr.org

DETECTION OF LIVER DISEASE USING HYPER TUNED GRIDSEARCH

Mr. Anand Thota¹, Ms. N Susmitha², Mr. D S S Srinivas³, Ms. L Monika⁴, Mr. T J Sai Sravan⁵ Associate Professor, CSE PSCMRCET, Vijayawada, Andhra Pradesh.

2,3,4,5 Student, CSE, PSCMRCET, Vijayawada, Andhra Pradesh.

Abstract

Liver disease is considered a major health concern because of the increased risk of illness and mortality. As liver infections are slow and chronic, they are more difficult to diagnose. This is the same reason why many patients are unable to make a diagnosis until the final stage. In disease detection, machine learning approaches play an essential role. The main objective of this project is to assess to detect the liver disease with high accuracy, less computational cost, and less run time complexity. To determine liver disease, this model is developed using classification techniques that are employed as meta-classifiers. To complete the task, the technique is divided into various parts. In the first module, the traditional data is preprocessed. Data Cleaning, Data Transformations and standard scalar are the techniques that are used to compute preprocessing. Base classifiers like KNN, SVM, Decision Trees, and Naive Bayes are used, and Logistic Regression classifiers are used as meta-classifiers to process the information. Then we apply hyper tuning by using the grid search technique. These algorithms are hypertuned by a grid search to attain more accurate results. Therefore, the significance of liver disease detection is known.

Keywords:Liver disease, hyper tuning, grid search, Logistic Regression, meta classifiers, SVM, KNN, Decision Trees, Naive Bayes, and Machine learning.

ISSN: 2456-5083

1. INTRODUCTION

Liver Disease is becoming a growing serious health problem. This ends in death or partially damaged because it is difficult to predict Liver Disease in its early stages. As an outcome, in this view, this approach improves the detection of liver disease in patients, allowing people to take actions that lower the death rate. The primary goal of this project is to use the Logistic Regression algorithm as a meta classifier and the other algorithms are hyper tuned by the grid search method.to detect liver disease. To perform the assignment, we will divide the procedure into discrete modules. The traditional preprocessed in the first module. To compute, the next module includes a Logistic Regression algorithm.

Today we see that liver patients are increasing rapidly, and it is very important to predict the disease early to control the widespread of the disease. Here, we need to collect sample data from a liver-infected patient to detect liver disease. Machine learning plays a very important role in detecting liver disease. The main objective of the study is to detect liver disease based on relevant information stored in a large database. It is very important to detect liver disease early in order to save the patient's life and help us live a normal life as before. Machine learning is growing very important as it helps in analyzing and summarizing data from different perspectives.

The proposed system began with a strong desire to forecast liver disease using an effective model and to achieve a high level



A Peer Revieved Open Access International Journal

www.ijiemr.org

of accuracy when compared to previous Machine learning algorithms systems. outperform other models by a significant margin. The proposed system is used to forecast liver disease in its early stages by employing meta-classifiers. This proposed system uses a hyper tuning technique grid search approach. This proposed system helps to detect liver disease in people which helps to save the lives of people. So, in further studies, the researchers can work on a model that can detect liver disease using fewer attributes, as well as provide the required precautions that infected patients must take.

2.RELATED WORK

[1]Amrin et.al, analyzed that liver disease needs early detection and diagnosis to ensure longevity. Learning the learning algorithm of each machine, and comparing accuracy with performance parameters is the goal of this study. Liver disease can be treated early to take preventive measures. Here they used the learning models of the eight monitored machines using python programming language to detect the disease. Comparative analysis was performed by taking into account various performance metrics, eight machine learning models were created. Of these eight machine learning models, the C4.5 algorithm showed a high accuracy of 70%. They have seen that machine learning algorithms used to create robust computerassisted diagnostic systems help treat patients effectively.

[2] Geetha et.al, presented the prediction of liver disease. It has a huge financial burden on the healthcare industry due to the increase in the number of patients and the progression of diseases. Predicting

sickness in its early stages based on tiny signals is a challenging assignment for medical specialists. The main objective of the study was to develop a machine learning model using data collected from the Indian liver Patient Database used to predict liver disease. Considered classification algorithms are Support Vector Machine and Logistic Regression. The Support Vector Machine (SVM) model was found to outperform the other method and attained 75.04%. Accuracy.

[3] M.ArdraMeghana et.al, explained the importance of predicting the outcome of liver disease in order to provide appropriate treatment. Machine learning algorithms are introduced in order to detect Liver disease at its early stages. To predict liver disease various classification algorithms like SVM, KNN, Random Forest Logistic Regression and Decision Trees are applied to the Indian Liver Patient Dataset(ILPD). Logistic Regression performed the best in predicting liver disease compared to other classifiers and obtained 81% accuracy.

[4] Ambesange et.al identified that many people are suffering from liver disease. To predict liver disease in early-stage with a dataset fetched from the northeast of Andhra Pradesh which uses a feature selection technique Correlation Matrix along with K-Nearest Neighbor(KNN). The model is hyperparameter tuned by the grid search approach. When the abovementioned techniques are applied to the dataset the combination of K-Nearest Neighbor (KNN) with the Correlation Matrix along with hyper tuning technique

using grid search approach resulted in an accuracy of 91%.



A Peer Revieved Open Access International Journal

www.ijiemr.org

[5] Hartaik et.al, studied and focused on the symptoms of liver disease. To predict the disease various classification algorithms likeK-Nearest Neighbor, and Naive Bayes are applied to the liver patient dataset. The main goal of this work is to build a data mining model and compare its performance with that of other learning models of modern technology. The main functions found in the Liver are alkaline, Age, Gender, Tb, db, Alkphos, Sgpt, Sgot, Tp, Alb, A_G.Naive Bayes andKNN areused to diagnose and train data sets. Naive Bayes and KNN divide data sets into training and testing sets. To stand out, these features are embedded in machine learning models. Naive Bayes provides precise prediction results for liver disease with an accuracy of 72.5%.

3. PROPOSED SYSTEM:

This paper focuses on a Machine Learning system that uses parameters like Problem-Solving skills, Technical Activities, and Non-Technical Activities to detect liver disease. The Indian Liver Patient Dataset is used as a sample dataset taken from Kaggle, and preprocessing is done to remove noise and unwanted data. We replace the missing values by mean and mode and standard scalar for preprocessing. Base classification and meta classification are two approaches followed in the model. The classification uses the K-Nearest Neighbor, Support Vector Machine, Decision Trees, and Naive Bayes techniques while meta classification uses Logistic Regression. The base classifications make predictions on the training data and these are sent to the stack to make final predictions using ensemble technique by meta classifier,

Logistic Regression. The performance of these classifiers is improved by using the hyper tune grid search method.

To develop a Liver Disease Detection, we will perform the below steps.

Data pre-processing

Classification

Final prediction

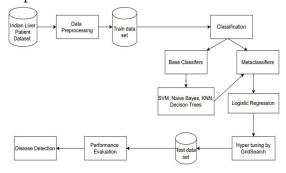


Figure 1: Model Architecture

Fig2: represents the system architecture of the proposed system. The architecture first performs preprocessing technique. preprocessing contains the data cleaning and the data transformation. In data cleaning, it identifies and fills null values by mean and encodes the given labels into a numerical representation. Whereas the data transformation used a standard scalar technique. After successfully completing preprocessing the data is divided into training and testing data, train data set goes to classification where the base classifiers and the meta classifiers are defined for the system Evolution. The classifiers are hyper-tuned by the grid search approach. Going on to the Testing data it's predicted and directly goes to the System Evolution.

3. DATA PRE-PROCESSING:

Data preprocessing is the method of changing raw data into an understandable format for machine learning. Preprocessing has been shown to improve the effectiveness of machine learning systems in previous research. In the process of data cleaning, the proposed system uses various datasets. Because we can't work with raw



A Peer Revieved Open Access International Journal

www.ijiemr.org

data, this is a crucial phase in the data mining process. Data cleaning has an impact on the efficiency of machine learning systems. During the data cleaning process, the suggested system discovered certain missing values, which has a negative influence on the system's accuracy. So, we used two ways to preprocess the data: label encoding and standard scalar.

3.1.1 Preprocessing using Label Encoding and replacing missing values with mean

The process of replacing null values is done by mean. In supervised learning, it is a pre-processing step for crucial structured dataset. More than all of the attributes in the dataset have few unique values in their respective columns, and numerical values have a larger role in calculation time than categorical data. As a result, the system used the label encoding technique to convert categorical data into numerical data. Each unique value for the attribute is assigned a number by the label encoding.

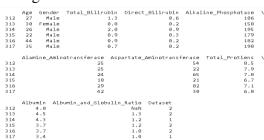


Figure 2: Before Label Encoding and replacing missing values

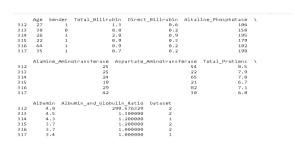


Figure 3: After Label Encoding and replacing missing values

3.1.2 Preprocessing using Standard scalar:

Data transformation is also a key preprocessing step, which is applied as standard feature scaling in the proposed system, which is calculated as the "difference between the attribute and the whole mean is divided by the standard deviation". This standard scalar mechanism makes the mean value concentrated around 0 and standard deviation values around 1.

3.2 CLASSIFICATION:

ISSN: 2456-5083

Classification is a predictive modeling technique that helps systems recognize objects and classify data based on training data. Classification is a method of distributing data into a certain number of classes. The main goal of a classification problem is to determine which categories/classes the new data will belong to in the classifier. An algorithm is used to assign the input data to one of the category. In machine learning, classification is a supervised learning approach in which the computer program is trained to learn from the data



A Peer Revieved Open Access International Journal

www.ijiemr.org

input given to it and then uses applied learning to classify further observations. Among the various classification algorithms, here the system uses classification techniques such as:

- Base classifiers
- Meta classifiers.

3.3 BASE CLASSIFIERS:

3.3.1 KNN:

The K-Nearest Neighbor algorithm dependent on machine learning. Based on classification and regression, KNN is used to predict difficulties. A supervised machine learning algorithm is also known as KNN. The non-parametric learning algorithm works because it makes no assumptions regarding The KNN algorithm separates the dataset and groups new data together. The KNN algorithm is also known as the lazy learning algorithm. When training data is provided, the lazy learning agenda is that it does not undertake any training. KNN is also well-suited to data mining. If k is set to one, the nearest neighbor determines the class of the given data point.

3.3.2 SVM:

A support vector machine (SVM) is related to machine learning that can be used for both classification and regression. The supervised learning machines in intelligence and artificial machine learning provide both inputs and desired data that are labeled classification. This algorithm is also wellsuited for classification. SVM main aim creates a hyper-plane to classify the classes. The SVM algorithm separates the dataset and groups new data together An SVM, like other supervised learning machines, must be trained with labeled data. For classification, items are labeled into groups.

3.3.3 Naive Bayes:

For predictive modeling and classification problems, Naive Bayes is a simple but surprisingly strong probabilistic machine learning method. Spam filtering, prediction, document sentiment classification, and other applications of Naive Bayes are common. It is a popular algorithm because it is simple to write in code and predictions can be made quickly, which boosts the solution's scalability. The Naive Bayes algorithm has long been the algorithm of choice for practical-based applications, especially when responses to user requests are required. The purpose. The target of Bayesian learning is to construct a design of class label distribution with a precise definition of the objective attribute. Then theorem of Bayes is applied with the Naive Bayes.

3.3.4 Decision Trees:

ISSN: 2456-5083

decision tree comes under of machine supervised learning that classifies or predicts outcomes based on the output to given questions. The model is supervised learning, which means it is trained and tested on a set of data with desired consideration. A clear response or decision may not always be provided by the decision tree. The construction of a decision tree entails selecting the attributes and criteria that will result in the tree. The tree is then trimmed to remove any branches that may interfere with accuracy. Pruning entails identifying outliers, or data points that deviate from the norm and may cause



A Peer Revieved Open Access International Journal

www.ijiemr.org

computations to be thrown off by giving too much weight to infrequent events in the data. Decision trees with few nodes and branches show data well.

3.3.5 META CLASSIFIER using Logistic Regression:

Logistic regression comes under the supervised machine learning technique that is used to predict the many situations. In binary classification situations, the probability of an outcome, occurrence, or observation is calculated. The model produces a binary or dichotomous result with only two options: yes/no, 0/1, or true/false. The link between one or more independent variables is investigated using logical regression, which divides data into discrete classes. It's often used in predictive modelling, where the model calculates the mathematical likelihood of whether or not a given event corresponds given category. In classification issues, where the outcome variable exposes one of two groups, logistic regression is typically utilized (0 and 1).

4. Hyper Tuning using GridSearch:

Tuning is used for increasing the performance of a machine without leading to overfitting or increasing variance. This is processed by picking appropriate "hyperparameters." A machine learning model's hyperparameters can be thought of as "dials" or "knobs." The better way is to go for alternative ways of hyperparameters and select the subset that results in model performance. This is referred to as hyperparameter optimization or tweaking by grid search. GridSearchCV is an approach for determining the

optimized parameter values from a given set of parameters in a grid. Grid search is ideal for double-checking combinations that have previously been performed successfully. GriSearchCV is generally used to combine an estimator with a grid search preamble with hyperparameters.

5. EXPERIMENTAL RESULTS:

The dataset used for this model consists of 584 records and 11 columns in which 10 are features and 1 is a class label.

1	Age of the	Gender o	Total Bilin	Direct Bili	Alkphos A	Sgpt Alan	Sgot Aspa	Total Prot	ALB Albu	A/G Ratio	Result
2	65	Female	0.7	0.1	187	16	18	6.8	3.3	0.9	1
3	62	Male	10.9	5.5	699	64	100	7.5	3.2	0.74	1
1	62	Male	7.3	4.1	490	60	68	7	3.3	0.89	1
5	58	Male	1	0.4	182	14	20	6.8	3.4	1	1
6	72	Male	3.9	2	195	27	59	7.3	2.4	0.4	1
7	46	Male	1.8	0.7	208	19	14	7.6	4.4	1.3	1
3	26	Female	0.9	0.2	154		12	7	3.5	1	1
9	29	Female	0.9	0.3	202	14	11	6.7	3.6	1.1	1
0	17	Male	0.9	0.3	202	22	19	7.4	4.1	1.2	2
1	55	Male	0.7	0.2	290	53	58	6.8	3.4	1	1
2	57	Male	0.6	0.1	210	51	59	5.9	2.7	0.8	1
3	72	Male	2.7	1.3	260	31	56	7.4	3	0.6	1
4	64	Male	0.9	0.3	310	61	58	7	3.4	0.9	2
5	74	Female	1.1	0.4	214	22	30	8.1	4.1	1	1
6	61	Male	0.7	0.2	145	53	41	5.8	2.7	0.87	1
7	25	Male	0.6	0.1	183	91	53	5.5	2.3	0.7	2
8	38	Male	1.8	0.8	342	168	441	7.6	4.4	1.3	
9	33	Male	1.6	0.5	165	15	23	7.3	3.5	0.92	2
0	40	Female	0.9	0.3	293	232	245	6.8	3.1	0.8	1
1	40	Female	0.9	0.3	293	232	245	6.8	3.1	0.8	1
2	51	Male	2.2	1	610	17	28	7.3	2.6	0.55	1
3	51	Male	2.9	1.3	482	22		7	2.4	0.5	1
4	62	Male		3	542	116	66	6.4	3.1	0.9	1
25	40		1.9	1	231	16	55	4.3	1.6	0.6	1

Figure 4: Dataset

In the proposed system, the KNN, SVM, Naive Bayes, and Decision Trees algorithms alone were obtained with 72%, 70.29%, 61%, and 70.29% accuracies To get better accuracy we use the meta classifier and the Logistic Regression algorithm obtained with 93.54%% accuracy. The output from KNN, SVM,

Naive Bayes, and Decision Trees are passed as input to the Logistic Regression algorithm. It was obtained with an accuracy of 93.54%.

The performance evaluation metrics are calculated as

1)Accuracy:

ISSN: 2456-5083

Accuracy is determined to measure the best machine learning model. There are



A Peer Revieved Open Access International Journal

www.ijiemr.org

various machine learning models. By using them we obtain high accuracy. Based on input and trained data, we determined the link and correlation between the variables in the dataset.

Naive Bayes, and Decision Trees are passed as input to the Logistic Regression algorithm. It was obtained with an accuracy of 93.54%.

The performance evaluation metrics are calculated are

Accuracy: Total Number of correct predictions

Total number of samples

Accuracy Score obtained for KNN: 72%

Accuracy Score obtained for SVM: 70.29%

Accuracy Score obtained for Naive Bayes: 61%

Accuracy Score obtained for Decision Trees: 71%

The accuracy obtained for Proposed System: 93.54%

2)Precision:

Precision is defined as one of the indicators to identify the machine learning model performance. By using the machine learning model, the quality of the precision is made. The quality of precision is based on positive precision and negative precision. The true positives are divided by the total number of the given positive predictions.

Precision: True positives

True positives+False positives - (2)

3)Recall:

Recall is used to identify the number of positives found during the recall session. It is also specified as the correct number of hits has been found.

Recall: True positives True positives +False Negatives - (3)

Recall Obtained for Proposed System: 87.2%.

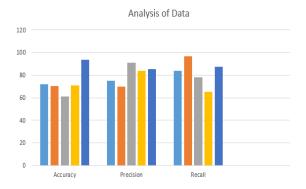


Figure 5: Analysis of Data

Table 1: Comparison of Performance metrics

S. NO	Algorithm	Accuracy	Precision	Recall
1.	KNN	72	75	84
2	SVM	70.29	70	97
3	Naive Bayes	61	91	78
4	Decision trees	71	84	65
5	Proposed System	93.54	85.3	87.2

5. CONCLUSION:



A Peer Revieved Open Access International Journal

www.ijiemr.org

Our proposed system helps in detecting patients who are suffering from liver disease. It is tough to recognize liver disease at its early stages. In this system we used SVM, KNN, Decision Tree, and Naive Bayes as Base-Classifiers and a meta-classifier Logistic Regression, we have identified the efficiency of a single model on the dataset. These algorithms are hyper tuned by the grid search approach to attain more accurate results with respect to their accuracy, precision, and recall. The future study of this system is to detect liver disease along with its stage. This work may allow for the development of more effective and reliable ways of detecting liver disease at its stage and will provide a better healthcare system by lowering the costs, time, and effort.

6. REFERENCES

- Amrin1, Omar Pahlevi2, UniversitasBinaSarana Indonesia InformatikaDOI:https://doi.org/ 10.33395/sinkron.v5i1.10589
- 2. Geetha, C., & Arunachalam, A. (2021).**Evaluation** based Approaches for Liver Disease Prediction using Machine Learning Algorithms. 2021 International Conference on Computer Communication and **Informatics** (ICCCI). doi:10.1109/iccci50826.2021.9 402463

- 3. M.ArdraMeghana, N. KalyanSaradhiP.Sahithi, R. SaiKiran. Department of Computer Science & Engineering Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh, India, Mrs A. Aparna, M.Tech, (Ph. D) Assistant Professor, **CSE** Department of GudlavalleruEngineering College, Gudlavalleru, Andhra Pradesh, India. http://www.jctjournal.com/galle ry/9-june2021.pdf
- 4. Ambesange, S., Nadagoudar, R., Uppin, R., Patil, V., Patil, S., &Patil, S. (2020). LiverDiseases Prediction using KNN with Hyper Parameter Tuning **IEEE** Techniques. 2020 Bangalore Humanitarian Technology Conference (B-HTC). doi:10.1109/bhtc50970.2020.9297949
- 5. Mallikarjuna
 Reddy A, RupaKinnera G, Ch
 andrasekhara
 Reddy T, Vishnu
 Murthy G. 2019. Generating
 cancelable fingerprint
 template using triangular
 structures. Journal of



A Peer Revieved Open Access International Journal

www.ijiemr.org

Computational and Theoretical
Nanoscience 16(5–6):1951-1955

- 6. A Mallikarjuna Reddy, Vakulabharanam Venkata Krishna, LingamguntaSumalatha and AvukuObulesh, Age Classification Using Motif Statistical and Feature Derived On Gradient Facial Images, Recent Advances in Computer Science and Communications (2020)13:965. https://doi.org/10.2174/22132 75912666190417151247.
- 7. Hartatik, H., Tamam, M. B., &Setyanto, A. (2020). Prediction for Diagnosing Liver Disease in Patients using KNN and Naïve Bayes Algorithms. 2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS). doi:10.1109/icoris50180.2020.93 20797
- 8. Kannapiran, Thirunavukkarasu& Singh, Ajay &Irfan, Md&Chowdhury, Abhishek. (2018). Prediction of Liver Disease using Classification

- Algorithms. 1-3. 10.1109/CCAA.2018.8777655.
- 9. Rakshith D B ,
 MrigankSrivastava , Ashwani
 Kumar, Gururaj S P, 2021, Liver
 Disease Prediction System using
 Machine Learning Techniques,
 INTERNATIONAL JOURNAL
 OF ENGINEERING
 RESEARCH & TECHNOLOGY
 (IJERT) Volume 10, Issue 06
 (June 2021).
- Reddy, 10. Mallikarjuna A., Venkata Krishna, V. and Sumalatha. L. Face recognition approaches: A survey. International Journal of Engineering and Technology (UAE), 4.6,6(7)(2018) 117-121. doi: 10.14419/ijet.v7i4.6.20446.
- 11. M. R. Ayaluri, K. Sudheer Reddy, S. R. Konda, and S. R. Chidirala, "Efficient steganalysis using convolutional auto encoder network to ensure original image quality," PeerJ Computer Science, vol. 7, p. e356, 2021.
- 12. A. Sivasangari, B. J. Krishna Reddy, A. Kiran and P. Ajitha,



A Peer Revieved Open Access International Journal

www.ijiemr.org

- "Diagnosis of Liver Disease using Machine Learning Models," 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2020, pp. 627-630, doi: 10.1109/I-SMAC49090.2020.9243375.
- 13. R. Zhao, X. Wen, H. Pang and Z. Ma. "Liver disease prediction using W-LR-XGB Algorithm," 2021 International Conference on Computer, Blockchain and **Financial** Development (CBFD), 2021, 245-248, doi: pp. 10.1109/CBFD52659.2021.000

55.

- 14. M. Hassoon, M. S. Kouhi, M. Zomorodi-Moghadam and M. Abdar, "Rule Optimization of Boosted C5.0 Classification Using Genetic Algorithm for Liver disease Prediction," 2017 International Conference on Computer and Applications (ICCA), 2017, pp. 299-305,doi:10.1109/COMAPP.201 7.8079783.
- 15. C Ramakrishna, G Kiran Kumar, A Mallikarjuna Reddy, and P. Ravi, "A

ISSN: 2456-5083

- Survey on various IoT
 Attacks and its
 Countermeasures," Int. J.
 Eng. Res. Comput. Sci. Eng.,
 vol. 5, no. 4, pp. 2394–2320,
 2018, [Online]. Available:
 http://ijercse.com/specissue/a
 pril-2018/27.pdf.
- 16. Swarajya Lakshmi Papineni, nigdhaYarlagadda, HaritaAkkineni,A.Mallikarju na Reddy. Big Data Analytics the Fusion Applying of Multicriteria Approach Decision Making with Deep Algorithms Learning International Journal of Engineering Trends and Technology, 69(1), 24-28, doi: 10.14445/22315381/IJETT-
- 17. A.Mallikarjuna, B.

 KarunaSree, "Security towards Flooding Attacks in Inter Domain Routing Object using Ad hoc Network"

 International Journal of Engineering and Advanced Technology (IJEAT), Volume-8 Issue-3, February 2019.

V69I1P204.