ROUND 1

Dear Authors,

I reviewed the manuscript entitled “Hybrid XGBoost model with Hyperparameter Tuning for prediction of Liver disease with better accuracy” and unfortunately decided to reject this study.

Here are some comments for this manuscript:

Specific comments

1. The authors are uncareful about the scientific article writing. Tenses are not compatible, in several places, authors used “We” for the literature and literature review is weak.

Ans:

Thank you for your time reviewing our paper and also for your observations. Language proofreading has been performed and “We” has been removed from literature review.

2. The title is almost irrelevant with the method, there is no “hybrid” XGBoost model, and it is a simple XGBoost model that is proposed.

Ans:

The learning phase of an algorithm is controlled by a set of values or weights known as hyperparameters. XGBoost has a wide variety of hyperparameters, as indicated before. XGBoost's hyperparameters may be fine-tuned to reach the highest level of accuracy possible. There are numerous learnable parameters that are tuned by the algorithm to automatically detect patterns and regularities in the data. The decision variables at each node of a tree-based model, such as XGBoost, are the learnable parameters. Large hyperparameters make XGBoost a very strong algorithm with a lot of design choices. When using a tree-based model, hyperparameters include things like how deep the tree can go, how many trees should be grown, how many variables should be taken into account when building each tree, how many samples should be taken from a leaf, and how many observations should be used to build a single tree.

The main challenge face in this stage is the optimized selection of parameters among multiple hyperparameters. It may be managed by efficient hyperparameter tuning. In this paper, the Bayesian optimization is being applied in following 4 steps as given below:

Step 1. Initialize domain space for range of values

Initially, the domain space is being finalized the input values over which is being searched. The input variables are max_depth, gamma, reg_alpha, reg_lambda, colsample_bytree, min_child_weight and n_estimators.

Step 2. Define objective function

The goal function, which can be any function that returns a real number, is then defined as the next stage in the process. XGBoost model validation error should be minimised with regard to hyperparameters in this situation. Accuracy is also an important consideration, and it should be optimised. The function should then return a negative value for the measure in question.

Step 3. Apply Optimization algorithm

Surrogate objective function construction and value selection are accomplished using this technique. This article utilises the Bayesian Optimization idea in the tuning step at this point in time. The Bayesian Optimization method is based on the Bayes Theorem and provides an efficient and effective method for solving a global optimization issue. Searches for candidate samples using an acquisition function are made more efficient by a probabilistic model of the actual objective function, known as the surrogate model.

Step 4. Evaluate hyperparameters' values

The XGBoost model is built using the scores or value pairs generated by the algorithm.

3. You added "confusion matrix" to the keywords but there is no confusion matrix inside the text.

Ans: Keyword "confusion matrix" has been replaced in "Hyperparameter Tuning"

4. Abstract does not focus on the importance of the proposed methods and their results.

Ans: The CHAID & CART model have achieved the accuracy level 71.36% & 73.24% respectively. With the help of the proposed model, the accuracy level has been achieved level up to 93.65%.

5. Introduction chapter is imprecise, and you use bullets for the objectives of the study. Design is poor.

Ans: This paper has been written with aiming of fulfilling various objectives. First objective of this paper is identifying the symptoms of the liver disease and its impact on the patient's body. Then this paper studies various machine learning approaches for predicting liver disease and evaluate the performance of decision-tree algorithms in prediction of liver disease. Next objective of this paper is concerned to propose a modified XGBoost model with hyperparameter tuning mechanism. Finally it has validated the performance of the proposed model with the existing model.

6. There was no (very few) information about the inputs of the data, no tables were presented.

S. No.	Attribute	Information	Type
1	age	Age of the patient	Number
2	gender	Gender of the patient	String (Male/Female)
3	tot_bilirubin	Total Bilirubin	Number
4	direct_bilirubin	Direct Bilirubin	Number
5	alkphos	Alkaline Phosphotase	Number
6	sgpt	Alamine Aminotransferase	Number
7	sgot	Aspartate Aminotransferase	Number
8	tot_proteins	Total Protiens	Number
9	albumin	Albumin	Number
10	ag_ratio	Albumin and Globulin Ratio	Number
11	is_patient	Selector field used to split the data into two sets (labeled by the experts)	Categorical

7. Lots of generic information were written but specific implementations for the study is almost none.

In the study presented here, the compound 'direct bilirubin' is adjudged as most significant factor by the CHAID test, as shown in figure 1, splitting the tree into three sub trees depending upon the quantity of

bilirubin found. It also shows the adjusted p-values and chi-square value calculated at each level marking the significant difference between the corresponding sub categories. Further, on the next level, the tree splits on the basis of next important factor, 'alkaline phosphate', for category represented by node 1, i.e., people with value less than equal to 0.9 for compound direct bilirubin and, 'age' for node 2 which represents the direct bilirubin range of 0.9 to 4.1 in a body. The maximum height of the tree allowed is up to the level 5 as the further splits do not significantly affect the result of the model. The significance value kept for splitting the records is .05, using Pearson likelihood ratio for chi-square and Bonferroni method for auto adjusting the significance value and actual p-values. The threshold value for stopping the growth of tree is minimum 2% records in parent branch and 1% record in child branch. The model produced the Gini coefficient value of 0.49 exhibiting good discriminatory behaviour and showed 71.36% accuracy in correctly classifying that whether a patient has liver disease or not based on the input data.

C&RT, which is another decision tree based method but produces the binary tree classification for continuous variables, exhibited different sequence by adjudging compound total bilirubin as the least impure predictor in this work. The split-up value for total bilirubin was calculated to be 1.650 by the C&RT model at the first level, using Gini impurity index method, with a gain of 0.047. Subsequently, the other important predictors are 'aspartate_aminotransferase', 'direct bilirubin' and 'age' of the subjects, according to this model. The compound 'total bilirubin' was used at multiple levels for the split which means that this compound produced maximum gain in gini index at multiple levels. The minimum value for recording change in impurity and making a split is set to '0.0001', after a series of run to maximize the efficiency of the model. The maximum number of levels allowed here is '5' in the C&RT tree with the same criteria for stopping the growth as used in CHAID model.
As in figure 7, the result for C&RT model shows better accuracy at predicting the liver disease as compared to CHAID model at 73.24%. Out of the 583 observations, the predictions can be categorized as 389 to be true positive, 38 as true negative and 156 as false predictions in this model. The gini coefficient for the model is calculated as 0.44 which shows that the ability of the model to categorize is good enough.

For XGBoost model, the objective function optimization is performed using the logistic model, as the target is a categorical variable. The tree building model used is 'auto' with ten iterations for boosting at each level. The tree depth, with which maximum efficiency is obtained, is found to be up to level 6, wherein, minimum child weight is set as default '1' and maximum delta step is set to no constraint with value '0'. As the number of observations available is limited, therefore, the sub sample space was set to '1', to consider all the data points during each iteration. Similarly, the parameter for column and level sampling are also set to value '1'. The 'Eta' parameter value is set to default '0.3', to keep the weights stable after each iteration, whereas, gamma value to specify the loss reduction required for split up is '0'. The value for least square error and least absolute deviation represented by parameter lambda and alpha, respectively, is set to value '1' and '0' which is also used for regularizing the weights, make the model more stable and control the overfitting. The model was also executed with hyperparameter tuning setup using Bayesian optimization model and showed even better performance than the non-hyperparameter tuning setup. As shown in the figure, the model produced accuracy of 93.65 % in predicting the disease, outperforming the other models significantly. It also recorded the gini index of 0.97, categorizing it as a highly efficient model in making the distinction between a diseased and healthy body in the given context.

8. There is no conclusion. Irrelevant sentences are written for the conclusion.

Ans: In this paper, multiple machine learning algorithms have been implemented on the above said dataset of Indian liver patients. This work consists of the execution of CART and CHAID algorithm to predict the liver patients. If bilrubin, protein, alkaline phosphatase, and albumin are present in the human body, and tests like SGOT and SGPT indicate that a person needs to be diagnosed, the results are stated through the various machine learning algorithms. The decision tree algorithm i.e. CART & CHAID are found not such more accurate. This factor motivate the authors to design the proposed machine learning model (Hybrid XGBoost model) which gain the accuracy level of 93.65%.

9. There is no future study and limitations information about the study.

A proposed model faced the over-fitting issue during its execution phase. In current execution, the textual dataset of 563 patients has been used. In future, the image dataset of lung disease should be taken for avoiding other tests like SGPOT & SGPT. Such approaches are very useful for minimizing the workload of the doctors.

General Comments

1. First, what are the original findings of this manuscript?

The original findings of this manuscript are more accurate prediction of lung disease through proposed method. The accuracy level have been achieved to 93.65%.

2. What are the new hypotheses that this study proposed?

The new paper uses Bayesian optimization, which is widely considered a more efficient approach of hyparameter tuning of XGBoost model than other techniques

3. What are the new phenomena that were found through experiments in this study?

This paper has applied hybrid XGBoost model with hyperparameter tuning performed by Bayesian optimization.

4. What are the hypotheses that were confirmed through experiments in this study?

This proposed approach has gain the higher accuracy level than other decision trees.

5. Original findings claim that with XGBoost algorithm disease prediction is higher, however, technical detailed information is inadequate.

The steps and technical description of proposed algorithm has been given in section ..

6. Second, what are the quality and importance of this manuscript?

The CHAID & CART model have achieved the accuracy level 71.36% & 73.24% respectively. With the help of the proposed model, the accuracy level has been achieved up to 93.65%.

7. What are the new findings of this study?

With the help of the proposed model, the higher accuracy level has been achieved in comparison to other decision tree models.

8. What are the new concepts that this study proposes?

The hyperparameter tuning is being applied.

9. What are the new methods that this study proposed?

New tuning approach based Bayesian optimization has been applied.

10. Do the conclusions appropriately summarize the data that this study provided?

The tabular representation of dataset is being given in section 3

11. What are the unique insights that this study presented?

It has gained higher accuracy level in comparison of Ensemble method and tradition decision tree.

12. What are the keys?

Bayesian optimization is fast and efficient approach for hyperparameter tunning.

13. The manuscript compares three sttaistical tools, CHAID, CART and XGBoost. These methods are just applied with minor settings but these settings are not presented in the manuscript although the title includes hyperparameter tuning.

The steps and technical description of proposed algorithm has been given in section ..

14. Third, what are the limitations of the study and its findings?

The proposed algorithm face the problem of over-fitting issue during its execution phase

15. What are the future directions of the topic described in this manuscript?

The image dataset should be used in future.

16. What are the questions/issues that remain to be solved?

The problem of over-fitting issue should be addressed carefully.

17. What are the questions that this study prompts for the authors to do next?

The use of medical images and over-fitting issue are next to do for the authors

18. How might this publication impact basic science and/or clinical practice?

This work helps the doctors for predicting the liver disease accurately which helps in reducing their workload.

19. The manuscript proposed the XGBoost model but no details are presented. The question that "which features are important for XGBoost model?" remains unanswered. This publication has minor adding for practice because no detailed information is presented.

The technical description and steps of proposed algorithm has been given in section 3.3.

REVIEWER 2

I am really grateful for reviewing this manuscript. In my opinion, this manuscript can be published once some revision is done successfully. This study used an XGBoost model, one of the most advanced machine learning model at this point. I would like to point out that this is a great achievement. But I would like to suggest the authors to draw SHAP summary and dependence plots for identifying the direction of association between a particular feature and the dependent variable.

Thank you for your time reviewing our paper and also for your observations. As per reviewer's suggestion, the SHAP values (impact on model output) has been demonstrated below:

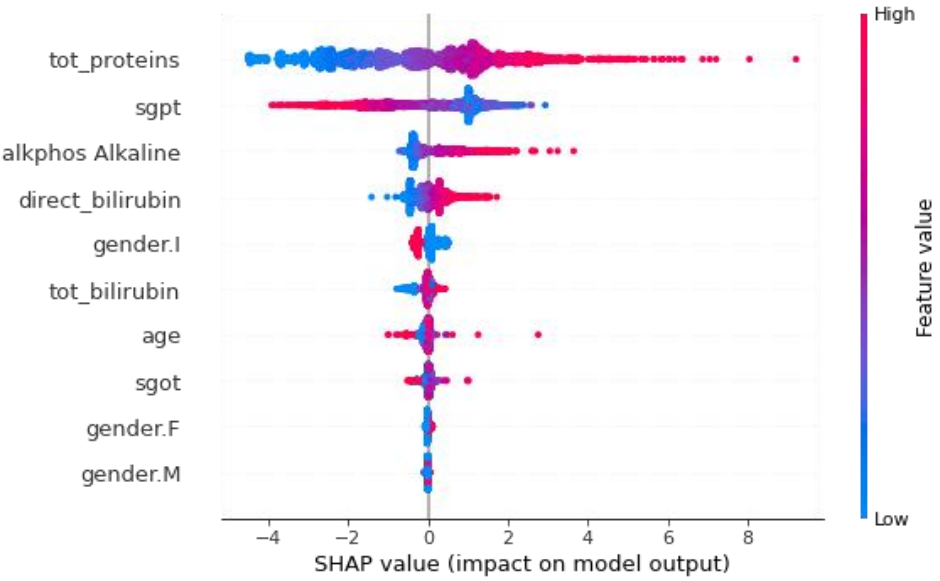

Figure 5 SHAP value (Impact on model output)

Here, the SHAP value of a particular feature for a particular observation measures a difference between what the model (e.g., XGBoost) predicts for the probability of the dependent variable for the observation with and without the predictor.

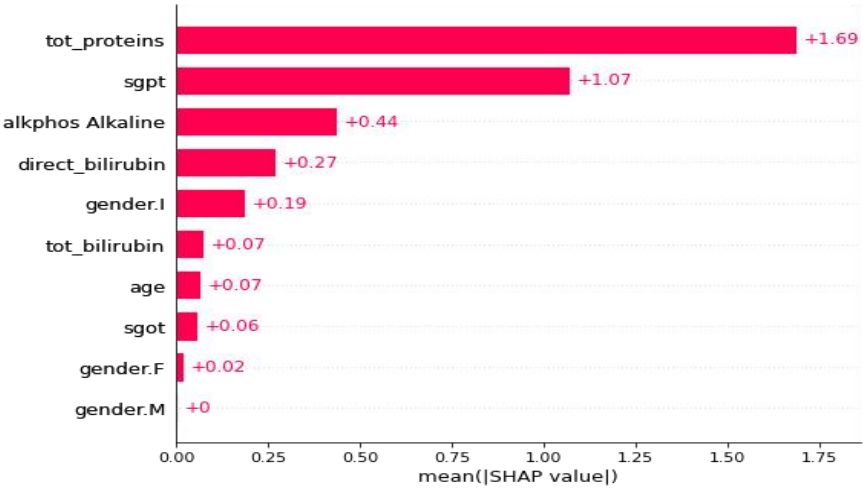

Figure 6 mean (SHAP value)

Indeed, the SHAP dependence plot reveals an interaction between two features regarding their effects on the probability prediction of the dependent variable.

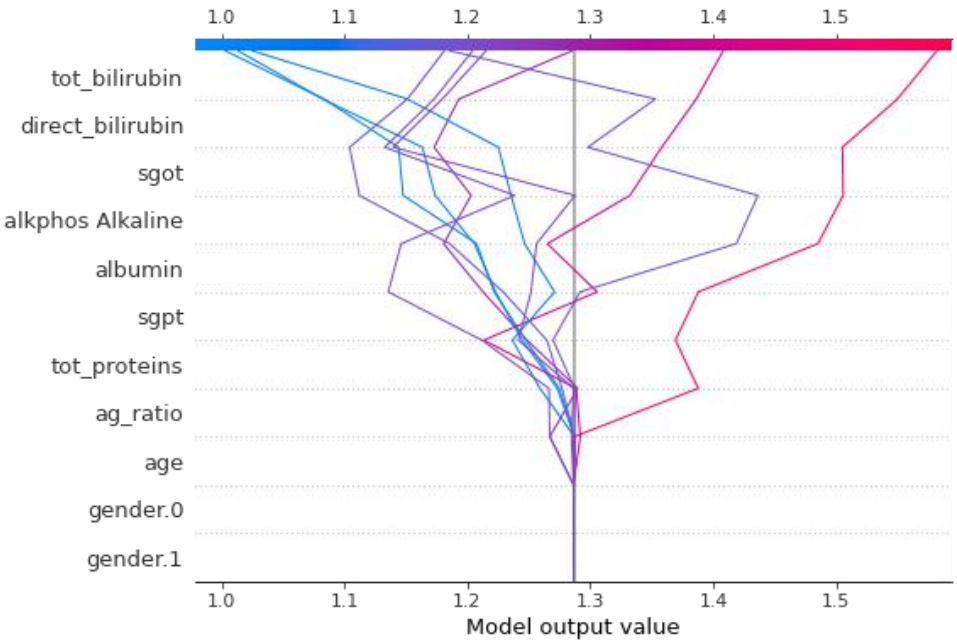

Figure 7 Model Output value

ROUND 2

There are still some vocabulary problems inside the text and even in Tables.

Thank you for your time reviewing our paper and also for your observations. Language proofreading has been performed

In handling the data, what are three ways to copping the data?

There are following three ways as given below:

1. Deleting Rows with missing values
2. Impute missing values
3. Prediction of missing values

There are still tense inconsistencies.

Language proofreading has been performed

Please simply explain GINI coefficient and its importance.

The Gini Coefficient is a machine learning metric used to assess the efficiency of binary classifier models. In the range of 0 to 1, the Gini Coefficient can be used. The higher the Gini coefficient, the better the model.

You mentioned precise and recall but you did not explain them.

There are two ways to measure precision: the number of properly categorised positive samples (True Positives) and the overall number of positively classified samples (either correctly or incorrectly). Using precision, we can see how reliable the machine learning model is when it comes to determining whether or not the model is positive

The recall is computed by dividing the total number of Positive samples by the number of Positive samples that were correctly categorised as Positive. The recall assesses the model's capability to identify positive samples in a data set. The more positive samples are found, the higher the recall.

Please interpret Figure 8.

As in figure 8, it is shown that the CHAID model gained the 71.36% accuracy level in predicting the liver disease. The AUC & Gini value of this model are 0.746 & 0.493 respectively.

Please interpret Figure 9 shortly.

As in figure 9, the result for C&RT model shows better accuracy at predicting the liver disease as compared to CHAID model at 73.24%. The AUC & Gini value of this model are 0.724 & 0.4448 respectively.

Discussion and Conclusion parts are still too short.

In this paper, multiple machine learning algorithms have been implemented on the above said dataset of Indian liver patients. This work consists of the execution of CART and CHAID algorithm to predict the liver patients. If bilrubin, protein, alkaline phosphatase, and albumin are present in the human body, and tests like SGOT and SGPT indicate that a person needs to be diagnosed, the results are stated through the various machine learning algorithms. The decision tree algorithm i.e. CART & CHAID are found not such more accurate. This factor motivate the authors to design the proposed machine learning model (Hybrid XGBoost model) which gain the accuracy level of 93.65%. This proposed model faced the over-fitting issue during its execution phase. In current execution, the textual dataset of 563 patients has been used. In future, the image dataset of lung disease should be taken for avoiding other tests like SGPOT & SGPT. Such approaches are very useful for minimizing the workload of the doctors.