Nevşehir Bilim ve Teknoloji Dergisi

Research Article (Araştırma Makalesi)

Article Doi: **10.17100/nevbiltek.1256873**

Date Submitted:27/02/2023 Date Accepted: 29/12/2023

Karaciğer Fonksiyon Bozukluğu Teşhisinde Tahmin Algoritmalarının Değerlendirilmesi

Saadet Aytaç ARPACI [1], , Songül VARLI [2]

[1] *Yıldız Teknik Üniversitesi, Elektrik-Elektronik Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul*
ORCID ID: 0000-0001-6226-4210

[2] *Yıldız Teknik Üniversitesi, Elektrik-Elektronik Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul*
ORCID ID: 0000-0002-1786-6869

Öz

Akut karaciğer yetmezliği, karaciğerdeki fonksiyon bozukluğuna bağlı olarak gelişir. Kısa sürede gelişen ve vücutta ciddi hasarlara sebep olan akut karaciğer yetmezliği için erken tanı büyük önem taşır. Hekimin daha erken tanı koyabilmesi açısından makine öğrenimi yöntemlerine dayalı tahmin işlemleri, hekime karar verme sürecinde yardım sağlayabilir. Bu çalışma, akut karaciğer yetmezliği varlığının tahmini için hekime yardımcı olabilecek, yakın zamanda sunulan ve yüksek tahmin kabiliyetlerine sahip üç algoritmanın değerlendirilmesini amaçlar. Çalışmada kamuya açık veri kümeleri üzerinde XGBoost, LightGBM ve NGBoost yöntemlerinin tahmin başarımları incelenir. Bu araştırmada iki veri kümesi kullanılır; ilk veri kümesi 2008–2009 ve 2014–2015 dönemlerinde "JPAC Health Diagnostic and Control" Merkezinde toplandı. Veri kümesinde, toplam 8785 hastanın bilgisi bulunur ve hastaların çoğunda "akut karaciğer yetmezliği" geliştiğine dair bilgi yer almıyor. Ayrıca ikinci değerlendirme için Iesu vd. tarafından toplanan ve "akut karaciğer fonksiyon bozukluğu" gelişen veya gelişmeyen hastalar hakkında bilgi içeren bir veri kümesi kullanılır. Veri kümesinden edinilen bilgiye göre, 208 hastada "akut karaciğer fonksiyon bozukluğu" gelişirken, 166 hastada bu durum gelişmemiştir. Eğitim ve test aşamalarında her üç algoritmanın yüksek tahmin sonuçları verdiği ve LightGBM yönteminin daha kısa sürede sonuca ulaştığı, NGBoost yönteminin ise diğer algoritmalara göre daha uzun sürede sonuç verdiği değerlendirmeler kapsamında gözlenmiştir.

Anahtar Kelimeler: XGBoost; LightGBM; NGBoost; akut karaciğer yetmezliği; sınıflandırma.

Evaluation of the Prediction Algorithms for the Diagnosis of Hepatic Dysfunction

Abstract

Acute liver failure develops due to liver dysfunction. Early diagnosis is crucial for acute liver failure, which develops in a short time and causes serious damage to the body. Prediction processes based on machine learning methods can provide assistance to the physician in the decision-making process in order for the physician to make a diagnosis earlier. This study aims to evaluate three recently presented algorithms with high predictive capabilities that can assist the doctor in determining the existence of acute liver failure. In this study, the prediction performances of the XGBoost, LightGBM, and NGBoost methods are examined on publicly available data sets.

Sorumlu yazar e-mail: saadeta99@gmail.com

In this research, two datasets are used; the first dataset was gathered in the "JPAC Health Diagnostic and Control Center" during the periods 2008–2009 and
2014–2015. The dataset includes a total of 8785 patients' information, and it mostly does not contain patients' information that "acute liver failure" was developing. Furthermore, a dataset collected by Iesu et al., containing information on patients who developed or did not develop "acute liver dysfunction," is used for the second evaluation. According to the information obtained from the data set, "acute liver dysfunction" developed in 208 patients, while this situation did not develop in 166 patients. It is observed within the scope of the evaluations that all three algorithms give high estimation results during the training and testing stages, and moreover, the LightGBM method achieves results in a shorter time while the NGBoost method provides results in a longer time compared to other algorithms.

Keywords: XGBoost; LightGBM; NGBoost; acute liver failure; classification.

Sorumlu yazar e-mail: saadeta99@gmail.com

1. Introduction

Acute liver failure (ALF) is a rare functional disorder that occurs in the liver due to the influence of different causes. Although paracetamol (acetaminophen) toxicity is mostly the main factor that causes ALF, it also has other different aetiologies, such as viral hepatitis, pregnancy-associated liver failure, autoimmune hepatitis, liver damage due to drug toxicity, and Wilson's disease. The disease can appear itself in its acute and subacute states, which develop more slowly over several weeks, or in its hyper acute form, which develops more rapidly. The progression of the illness may eventually lead to the development of extrahepatic organ failure and the condition of cerebral oedema, which is almost characteristic of ALF. The most effective treatment method for ALF, which causes serious damage to the body and even results in death, is liver transplantation. An attempt is made to increase the survival rate with liver transplantation [1 - 3].

Learning from data and using the predictions obtained from the models as an aid is provided by a process based on machine learning (ML) [4, 5]. Some of the most well-known basic ML prediction methods are nearest neighbour, naive bayes, and decision trees [4]. However, new proposed algorithms based on or different from these methods are still being developed. The diagnosis of many diseases can be facilitated in clinical environments with ML methods [6]. ML methods are also used in the field of hepatology for the diagnosis, progression, and prediction of mortality of liver diseases [7].

Researches that examine methods such as XGBoost [8], LightGBM [9], and NGBoost [10] developed on the basis of decision trees, which are an ML application, are also encountered in the field of health as well as in other fields. The XGBoost or LightGBM methods have been evaluated in various studies, including the estimation of Parkinson's disease [11], medical diagnosis with exhaled breath analysis [12], the estimation of chronic kidney disease [13], the estimation of type 2 diabetes risk [14], the prediction of the survival time of patients with hepatitis B [15], the estimation of the risk of gestational diabetes mellitus [16], and the prediction of blood glucose levels [17]. Furthermore, studies comparing the LightGBM and XGBoost approaches for determining the risk of liver disease [18, 19] or diabetes mellitus [20] are also available. In the research done by Noh et al. [21], the XGBoost, LightGBM, and NGBoost algorithms were examined for the estimation of survival time in hepatocellular carcinoma patients. Studies that make ALF estimation with the two models we used in this research, such as XGBoost and LightGBM, are found in the literature [22, 23]. According to the XGBoost and LightGBM methods, the NGBoost method is a relatively newer algorithm; therefore, it has been evaluated in a smaller amount of research than other, older algorithms. According to our knowledge, there have been no studies in the literature on the evaluation of the NGBoost method for ALF estimation (especially for the datasets we examined in this research [24, 25]).

Timely and early medical treatments or clinical follow-ups can significantly improve survival for ALF and also eliminate the need for liver transplantation. Given the importance of early diagnosis for ALF that develops in a short time, thanks to the hospitals' use of effective (powerful) ML methods for predicting the presence of ALF, it can also be ensured that more patients can be evaluated in a shorter time and that a health service that helps the clinician's decision-making process when working at a busy pace can be provided. The main contribution of this study is the evaluation of the three ML methods (XGBoost, LightGBM, and NGBoost) for ALF estimation, which have high performance rates and may provide this benefit to the clinician. Furthermore, the evaluation of the estimation approaches used together with the applied pre-processing methods on the data sets in this study constitutes additional contributions to the research.

The remaining article's arrangement is as follows: Section 2 provides information on the material and method. In Section 3, the results from the experiments are presented. The discussion and conclusions are presented in Sections 4 and 5, respectively.

2. Materials and Methods

The general application diagram of the study is shown in Figure 1, the same procedures were applied for both datasets. Information about each stage is presented in the next sections.

Figure 1. The general application diagram

2.1 Dataset

In this research, we used two datasets: the Kaggle acute liver failure dataset [24] and a dataset collected by Iesu et al. [25]. The first dataset includes information on 8,785 adults who are 20 years of age or older that was gathered in the “JPAC Health Diagnostic and Control Center” during the periods 2008–2009 and 2014–2015. The presence of ALF in the dataset is labelled with (label = 1). The dataset created according to 30 attributes can be accessed from the relevant web address [24]. The second data set includes a total of 374 patients' information with 208 patients whose ALF (label = 1) condition was detected by the research carried out in the “Intensive Care Unit of Erasme Hospital” in Brussels (Belgium). Extensive information about the second publicly available dataset can be obtained from the research done by Iesu et al. [25]. The obtained data sets were subjected to the processes specified in the pre-processing section and the feature elimination process before being used in the evaluation of the models.

2.2 Pre-processing

Data sets obtained from related studies have outliers or missing values. These factors, which detrimental direction impacts the prediction accuracy of the models, are corrected within the data sets during the pre-processing stage. In this process, rows with a lot of missing data or rows without labels are removed. As a result, the first dataset has 6000 samples and the second dataset has 374 samples. Performed operation for columns with a smaller amount of missing values: The median value of each column is found, and all the missing values of the column are replaced with the found median values. The outliers for each attribute are detected by means of the interquartile range (IQR). IQR represents the difference between the third quartile (A3) and the first quartile (A1), and if n>=0, values outside the range of [A1 - n(IQR), A3 + n(IQR)] are determined as outliers in the process. The prediction models that we used in this research work on numerical data. For this reason, some necessary categorical samples in both data sets containing numerical and categorical data are converted to numerical values. The "Min-Max normalization" operation is applied to convert the values to the [0, 1] scale range.

2.3 Feature Elimination

The most useful feature sets for improving the performance of classification models are found by feature selection methods; in this study also, a feature elimination process was adopted with the "Backward Feature Elimination" and "Recursive Feature Elimination with Cross-Validation" methods. In general, in the process of backward feature elimination [26]: Initially, all the features in a dataset with “N” attributes are trained. In the continuation, subsets with fewer attributes than the number of features used in the first stage are utilized to measure classification success. Attributes that show the least performance are removed. The process continues until the subset that achieves the maximum prediction success is found. In the recursive feature elimination with cross-validation method, all of the attributes in the dataset are first subjected to a performance evaluation, and the features are sorted by weighting according to the evaluation result. The least significant attribute is removed, and the process is iteratively reapplied in the framework of the k-fold cross validation layout. The best-performing features are retained [27, 28]. In our study, the evaluation of the models was carried out according to the intersection set of the attribute results obtained from both methods.

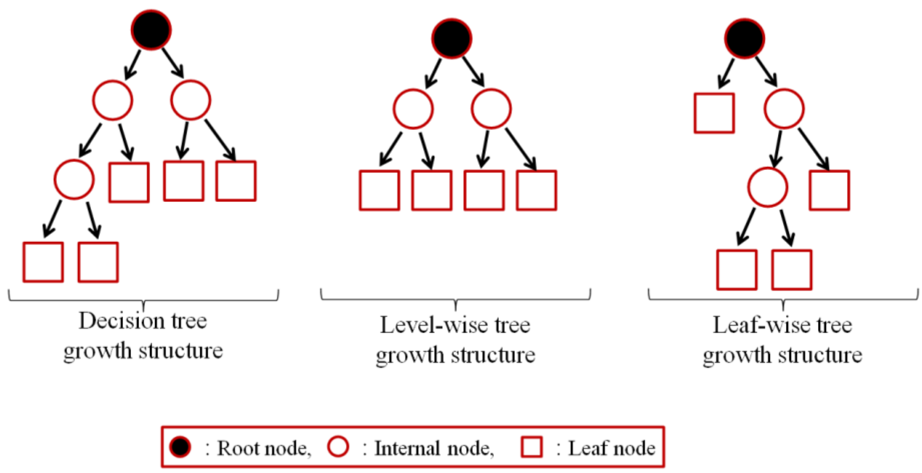

Figure 2. The decision trees growth samples [20, 36]

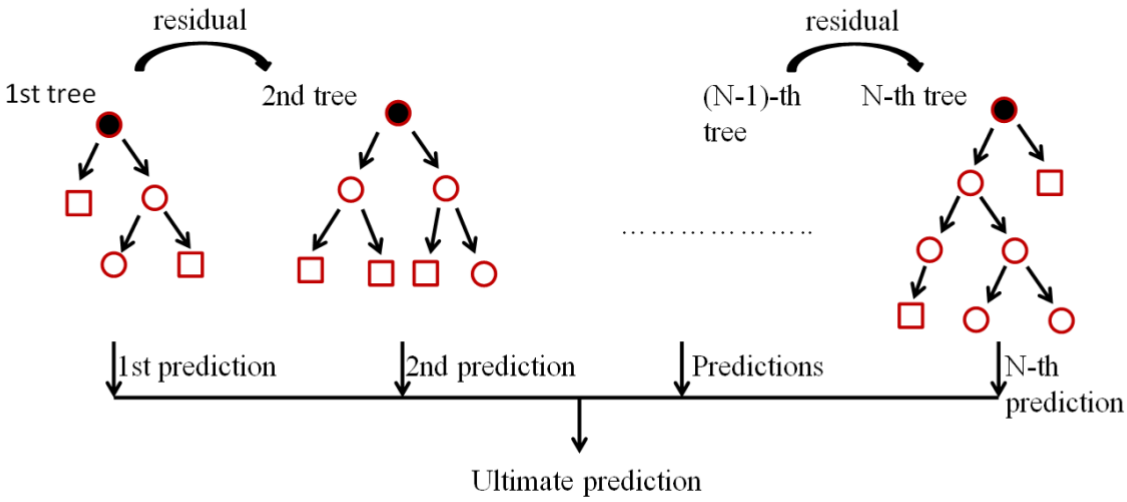

Figure 3. The boosting process

2.4 Background of Prediction Models

The classification process is performed by reaching the leaf nodes that give the classification result from a node assigned as a root in the decision trees (Figure 2). Internal nodes, on the other hand, represent the task of branching to the next internal node until they reach a leaf node by testing the attribute they receive in this process. Although the characteristics of decision trees, such as the ability to handle large amounts of data, process data in numerical or textual type, and provide high prediction performance results, are among the advantages of the method [29], the method also has some disadvantages. For example, the high-dimensional data given to the decision tree increases the search space exponentially, and thus the generalisation of the model becomes difficult. Furthermore, when there are minor changes in the training data, different splitting paths occur due to the hierarchical structure of the decision tree, and the whole tree result changes from top to bottom. Another disadvantage is that when there is a class imbalance problem in the dataset, like other standard machine learning solutions, it develops a tendency toward the majority class and becomes insufficient for all the data. Such deficiencies have been tried to be corrected by ensemble methods [30, 31]. In the learning process with the ensemble method, several classifiers are brought together to create a better performance than the prediction performance of each classifier. With the combined use of models, the risk of overfitting is reduced and performance is improved [31]. In the boosting process, which is an ensemble method, it is ensured that the weak model is weighted iteratively according to the result it gives for classification; thus, with the information produced, the model continues to learn until the final stage. In the final stage, the previous results are combined to give the most accurate estimate [32]. The principle of "training of each model depends on the pre-trained models", which forms the basis of the boosting method, is also valid for the gradient boosting method (GBM) [33]. In this method, each successive tree performs its own prediction by taking into account the error obtained with the differentiable loss function of the previous models (Figure 3). In a way that minimises the loss function, the final prediction is obtained by adding the estimation of all trees [31].

2.5 XGBoost Method

The authors added various optimizations and improvements to GBM to increase scalability in XGBoost [8, 31]. For example, (a) the missing data is placed in the nodes in a way that will give the best result (minimizing loss); (b) a more efficient weighting method is applied over all possible splits to optimize the splitting threshold. The most significant improvement was made to the XGBoost model with the addition of a regulation component to the loss function presented in GBM (Equation 1).

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad \text{(Equation 1)}$$

where l is a convex and differentiable loss function that measures how close the prediction $\hat{y}_i$ is to the target y_i. Ω controls the complexity of the model and helps prevent overfitting. It is defined more broadly as follows (Equation 2):

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda\|w\|^2 \quad \text{(Equation 2)}$$

where T represents the number of leaves on the tree and w represents the output scores of the leaves. The value of γ controls the minimum loss reduction gain required to split an internal node [34]. A scoring function (Equation 3) is applied to the system according to the differentiable values of the loss function to perform the best split.

$$gain = \frac{1}{2}[G_l + G_r - G_p] - \gamma \quad \text{(Equation 3)}$$

Gl and Gr are the optimal values of the left and right nodes, respectively, and Gp represents the optimal value obtained before division; (c) the nodes of the algorithm are advanced level-wise (Figure 2) and from top to bottom in order to generate the best tree according to all possible outcomes. If the value of the regulation parameter gamma (γ) is higher, the pruning process is started and the tree size is reduced, thus preventing the overfitting situation.

2.6 LightGBM Method

It has been observed that the XGBoost [8] method shows better prediction accuracy compared to other machine learning methods [35]. After the introduction of the XGBoost method, the LightGBM [9] method was proposed in 2017. Leaf-wise tree growth (Figure 2) applied LightGBM is another method developed on the basis of GBM, like XGBoost. According to the experiments carried out, it is reported that LightGBM mostly requires less memory, is faster, and gives more accurate results compared to XGBoost [36]. The first of the two improvement techniques that come to the fore in the LightGBM method is “Gradient-based One-Side Sampling”. In the estimation of the division points, the GBM method scans all the data samples for each feature, which causes the process to take longer to execute. Guolin Ke et al. [9] took into account that the training error of the samples that obtained small gradients was small and therefore gave a good training result, and they stated that it was unnecessary to process these samples. In the LightGBM method, instead of continuously processing samples that have a small gradient, they processed randomly selected samples that have a small gradient and all samples with a large gradient. Thus, the data size was reduced in the LightGBM method. In addition, with the second improvement method, the "Exclusive Feature Bundling" algorithm, which prevents features with zero values from entering the calculation by optimising a histogram-based algorithm, was applied to the model.

2.7 NGBoost Method

The NGBoost [10] method proposed by Duan et al. in 2020 was developed based on GBM, like the XGBoost and LightGBM methods. Most traditional estimation approaches are based on giving a single best prediction result for the problem. The NGBoost method evaluates the data it receives over a conditional probability distribution and gives a prediction scoring result by comparing the prediction distribution with the observation labels. It performs this process within the framework of the "base learner, parametric probability distribution, and scoring rule" components.

3. Results

In this study, the performances of the XGBoost, LightGBM, and NGBoost algorithms were measured on the datasets we used at the training stage and at the prediction stage of the test samples. The evaluation of the models was performed using the Python programming language. In both datasets, 20% of the samples were randomly selected and used for testing. The training results were obtained by applying 5-fold cross validation.

Table 1. The average training success expressed as a percentage and the average training time expressed in seconds

	Accuracy (%)	Training Time (s)
XGBoost	**89.86**	0.3033
LightGBM	88.47	**0.1054**
NGBoost	89.22	9.8129

Table 1 shows the average training success (in percentage) and average training time (in seconds) of the algorithms for both datasets. According to the average values in Table 1, although LightGBM performed the training phase in the shortest time, the XGBoost algorithm achieved the highest training accuracy (89.86%).

$$Accuracy = \frac{(TN+TP)}{(TP+TN+FP+FN)} \quad \text{(Equation 4)}$$

$$Precision = \frac{TP}{(TP+FP)} \quad \text{(Equation 5)}$$

$$Sensitivity = \frac{TP}{(TP+FN)} \quad \text{(Equation 6)}$$

$$Specificity = \frac{TN}{(TN+FP)} \quad \text{(Equation 7)}$$

$$AUC = \frac{(\frac{TP}{(TP+FN)}+\frac{TN}{(TN+FP)})}{2} \quad \text{(Equation 8)}$$

The test prediction performances of the XGBoost, LightGBM, and NGBoost algorithms were measured with accuracy (Equation 4) [14, 37], precision (Equation 5) [14, 37], sensitivity (Equation 6) [14, 37], specificity (Equation 7) [14], and AUC (Equation 8) [38] metrics. In the equations, TP defines "true positive," TN defines "true negative," FP defines "false positive," and FN defines "false negative." In Table 2, the average metric values obtained from the test samples are shown as percentages. Table 2 also presents the average prediction times (seconds) of the algorithms.

Table 2. Test results

	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	AUC (%)	Prediction Time (s)
XGBoost	87.96	96.24	83.97	69.15	76.55	**0.0056**
LightGBM	**89.59**	95.52	**87.24**	66.12	76.68	0.0058
NGBoost	88.09	**98.40**	83.03	**73.49**	**78.26**	0.089

According to Table 2, the most efficient algorithm in the prediction phase in general is LightGBM, with an accuracy value of 89.59% and a short prediction time of 0.0058 seconds. The LightGBM method most accurately detected the absence of ALF (label = 0) with a sensitivity value of 87.24%. The presence of ALF (label = 1) was best predicted by the NGBoost algorithm with a specificity rate of 73.49%. The classification ability of the NGBoost method is better due to its higher precision and AUC value. However, the NGBoost algorithm took longer than other methods to complete its operations during the training and prediction stages. As can be seen in Tables 1 and 2, the results given by the models are quite close to each other. The success of the models was also examined statistically by McNemar's test [37, 39]. As a result of the statistical evaluation, it was observed that there was no significant difference ($p>0.05$) between the prediction abilities of the algorithms.

The size of the datasets evaluated by the algorithms is reduced by feature selection methods; thus, noise and confusion (complexity) caused by too much data, and hence the error that may occur in the result, are prevented. However, each of the features in the reduced dataset may have different levels of importance for the classification performance of the models. For example, Figure 4 show the importance rankings of the features in the second dataset for XGBoost, LightGBM, and NGBoost, respectively. According to our findings, the important common features in the first dataset for all models are “age, hypertension, good cholesterol, and minimum blood pressure,” while the important common features in the second dataset are highest INR (INR = international normalized ratio), PT (prothrombin time), highest BIL (BIL=bilirubin), and min PLT (PLT=platelet).

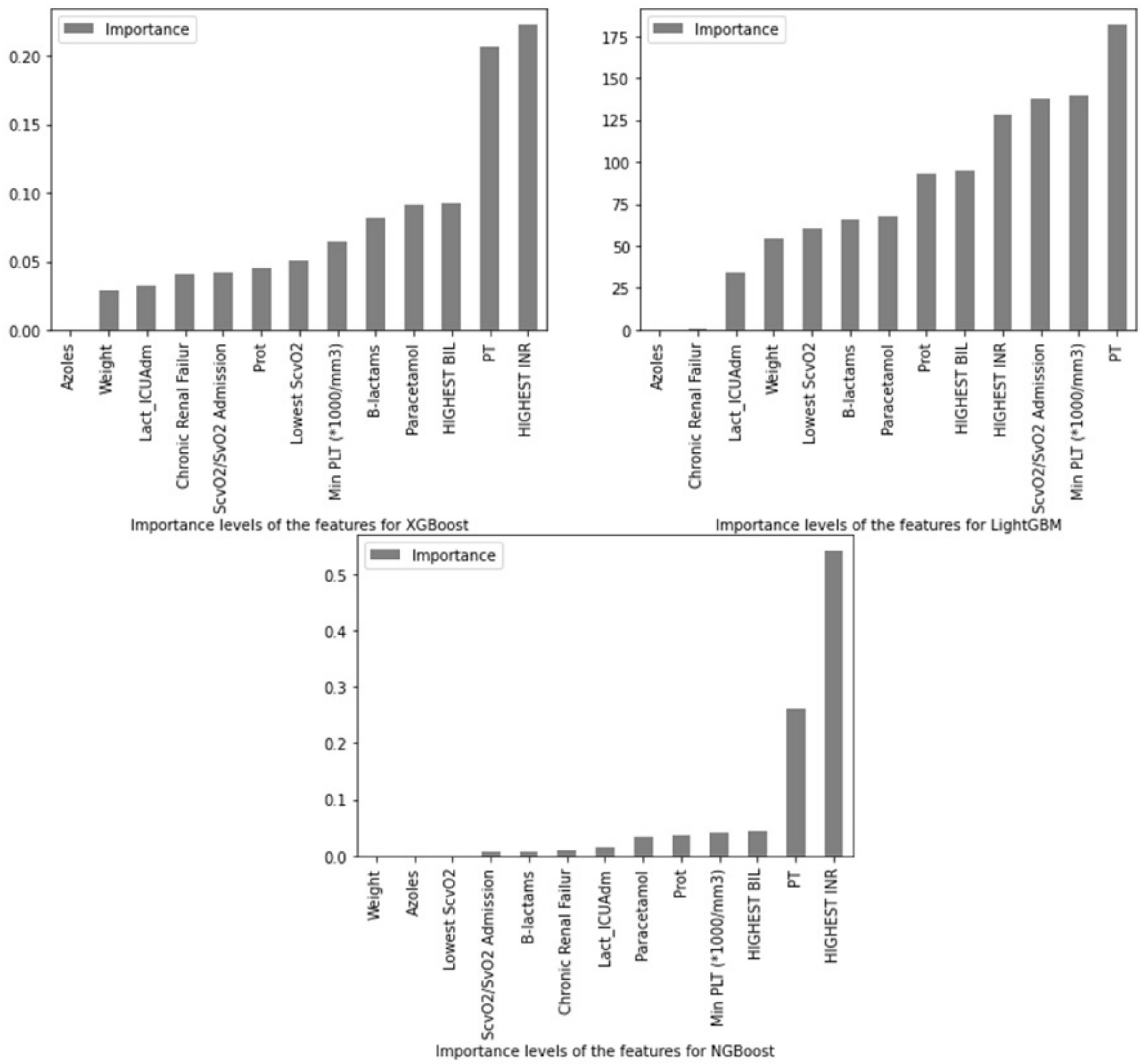

Figure 4. The importance ranking of the features in the second dataset for the XGBoost, LightGBM, NGBoost model

4. Discussion

The general opinion is that the later developed algorithms will achieve better performance than the previous algorithm, but based on the comparative analyses made on different datasets and the results we obtained from this research, we can say that, in general, the success rates of the XGBoost, LightGBM, and NGBoost algorithms are close to each other and high.

In the pre-processing stage, missing values and outliers in the datasets are processed with different algorithms based on specific mathematical methods. The success of the examined models is tried to be increased depending on these operations. The research on pre-processing methods remains current and continues to develop. Zhang and Gong [22] compared the XGBoost and LightGBM methods on the first dataset we used in this study. The pre-processing strategy (feature selection, etc.) they [22] applied during the preparation of the dataset and the amount of labelled data they [22] used are the main differences that differ from this study and affect the classification performance. Furthermore, the models we applied in our research are algorithms that can work on big data with high efficiency. The small sample size of the publicly available datasets to which these models will be applied to predict the ALF status constitutes a limitation in terms of a more effective comparison of the models.

The developers have also provided hyper parameter tuning support for the XGBoost, LightGBM, and NGBoost methods. In this study, the methods are compared according to the default parameter settings. Since hyper parameter optimization for algorithms needs to be examined in great detail, additional measurements can be made as a continuation of this research.

The algorithms we evaluated in this study perform their operations with numerical variables due to their nature. Therefore, the effect of categorical variables on algorithms cannot be understood. The CatBoost [40] method, which can also work on categorical properties, was introduced recently. The studies carried out to the present day point to the success of the method. The ALF estimation results, which we have obtained within the scope of existing algorithms, can also be compared with the CatBoost method in the future.

The superiorities of the XGBoost [12, 14, 15] and LightGBM [41] methods compared to traditional ML methods are shown in the studies. However, in the literature studies comparing the LightGBM and XGBoost algorithms, it is generally (mostly) shown that the LightGBM method gives better prediction results [20, 23]. [21, 42 - 44] investigations are some of the studies in which the NGboost method shows less performance success compared to the LightGBM and XGBoost methods. A small amount of research has been done up to now for NGBoost, which is a newer prediction algorithm according to XGBoost and LightGBM; therefore, it would not be correct to generalize the results. More research with NGBoost on different datasets is required in order to assess the efficacy of the method and clarify its superiority over other algorithms.

As seen in Tables 1 and 2 in the result section, NGBoost had the longest processing time among the examined algorithms. The long processing time for NGBoost, which does not show a difference compared to other algorithms in terms of the metric values and statistical prediction accuracy results, weakens the power of the algorithm. In this direction, when we add the time factor to the evaluation in addition to the prediction success, the observations obtained from our study highlight the importance of the LightGBM method.

5. Conclusions

In this article, three prediction algorithms are evaluated for the estimation of acute liver failure status. Pre-processing, feature selection, and evaluation are some of the various stages that our study went through. In the study, the prediction performances of the XGBoost, LightGBM, and NGBoost methods on publicly available datasets were examined. Our study presents some evaluations within the scope of the results we observed: (a) The prediction of acute liver failure development can be made by state of the art classification models such as XGBoost, LightGBM, and NGBoost with high accuracy rates. (b) Furthermore, the LightGBM method contributes to achieving an earlier result, while the NGBoost algorithm gives a result with a longer time. (c) The use of these methods in hospital areas can help the physician make an early diagnosis, and the adverse consequences caused by acute liver failure can be prevented.

In this research, we focused on three current and high-performance methods (XGBoost, LightGBM, and NGBoost) developed on the basis of GBM. In the future, the performance of different machine learning methods for predicting acute liver failure can be investigated.

6. References

[1]. Arshad M. A., Murphy N., Bangash M. N.,” Acute liver failure” ***Clinical Medicine Journal***, 20 (5), 505-508, 2020 DOI: 10.7861/clinmed.2020-0612

[2]. Kayaalp C., Ersan V., Yılmaz S., “Acute liver failure in Turkey: A systematic review” ***Turkish Journal of Gastroenterology***, 25(1), 35 – 40, 2014 DOI: 10.5152/tjg.2014.4231

[3]. Sugawara K., Nakayama N., Mochida S., “Acute liver failure in Japan: definition, classification, and prediction of the outcome” ***Journal of Gastroenterology***, 47, 849–861, 2012 Available from: https://doi.org/10.1007/s00535-012-0624-x

[4]. Saberi-Karimian M., Khorasanchi Z., Ghazizadeh H., Tayefi M., Saffar S., Ferns G. A., Ghayour-Mobarhan M.,” Potential value and impact of data mining and machine learning in clinical diagnostics” ***Critical Reviews in Clinical Laboratory Sciences***, 58(4), 275-296, 2021 DOI: 10.1080/10408363.2020.1857681

[5]. Park D. J., Park M. W., Lee H., Kim Y. J., Kim Y., Park Y. H., “Development of machine learning model for diagnostic disease prediction based on laboratory tests” ***Scientific Reports***, 11, 7567, 2021 Available from: https://doi.org/10.1038/s41598-021-87171-5

[6]. Mostafa F., Hasan E., Williamson M., Khan H., “Statistical machine learning approaches to liver disease prediction” ***Livers***, 1(4), 294-312, 2021 Available from: https://doi.org/10.3390/livers1040023

[7]. Ahn J. C., Connell A., Simonetto D. A., Hughes C., Shah V. H., “Application of artificial intelligence for the diagnosis and treatment of liver diseases” ***Hepatology***, 73(6), 2546-2563, 2021 Available from: https://doi.org/10.1002/hep.31603

[8]. Chen T., Guestrin C., “XGBoost: A scalable tree boosting system” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, California, USA, 2016

[9]. Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T. Y., “LightGBM: A highly efficient gradient boosting decision tree” *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, California, USA, 2017

[10]. Duan T., Avati A., Ding D.Y., Thai K. K., Basu S., Ng A., Schuler A.,”NGBoost: Natural gradient boosting for probabilistic prediction“ *Proceedings of the 37th International Conference on Machine Learning*, Virtual Event, 2020

[11]. Abdurrahman G., Sintawati M., “Implementation of XGBoost for classification of parkinson’s disease” *3rd International Conference on Combinatorics, Graph Theory, and Network Topology*, East Java, Indonesia, 2019

[12]. Paleczek A., Grochala D., Rydosz A., “Artificial breath classification using XGBoost algorithm for diabetes detection” ***Sensors***, 21(12), 4187, 2021 Available from: https://doi.org/10.3390/s21124187

[13]. Aydin Z. E., Ozturk Z. K., “XGBoost feature selection on chronic kidney disease diagnosis” *Proceedings of the IV International Conference on Data Science and Applications*, Virtual Event, 2021

[14]. Wang L., Wang X., Chen A., Jin X., Che H., “Prediction of type 2 diabetes risk and its effect evaluation based on the XGBoost model” ***Healthcare***, 8(3), 247, 2020 Available from: https://doi.org/10.3390/healthcare8030247

[15]. Ali N., Srivastava D., Tiwari A., Pandey A., Pandey A. K., Sahu A., "Predicting life expectancy of Hepatitis B patients using machine learning" *IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics*, Ballari, India, 2022

[16]. Hou F., Cheng Z., Kang L., Zheng W., “Prediction of gestational diabetes based on LightGBM” *Proceedings of the 2020 Conference on Artificial Intelligence and Healthcare*, Taiyuan, China, 2020

[17]. Wang Y., Wang T., “Application of Improved LightGBM Model in Blood Glucose Prediction” ***Applied Sciences***, 10(9), 3227, 2020 Available from: https://doi.org/10.3390/app10093227

[18]. Shobana G., Umamaheswari K., “Prediction of liver disease using gradient boost machine learning techniques with feature scaling” *5th International Conference on Computing Methodologies and Communication*, Erode, India, 2021

[19]. Sinthuja U., Hatti V., Thavamani S., “Analysis and prediction of liver disease for the patients in India using various machine learning algorithms” *International Conference on Advances in Data Computing, Communication and Security*, Kurukshetra, India, 2021

[20]. Rufo D. D., Debelee T. G., Ibenthal A., Negera W. G., “Diagnosis of diabetes mellitus using gradient boosting Machine (LightGBM)” ***Diagnostics***, 11(9), 1714, 2021 Available from: https://doi.org/10.3390/diagnostics11091714

[21]. Noh B., Park Y. M., Kwon Y., Choi C. I., Choi B. K., Seo K., Park Y. H., Yang K., Lee S., Ha T., Hyon Y., Yoon M., “Machine learning-based survival rate prediction of Korean hepatocellular carcinoma patients using

multi-center data" ***BMC Gastroenterology***, 22, 1-9, 2022 Available from: https://doi.org/10.1186/s12876-022-02182-4

[22]. Zhang D., Gong Y., "The comparison of LightGBM and XGBoost coupling factor analysis and prediagnosis of acute liver failure" ***IEEE Access***, 8, 220990-221003, 2020 DOI: 10.1109/ACCESS.2020.3042848

[23]. Sengupta D., Mondal S., Basu S., De A. K., Nath S., Pandey A., "Classification of acute liver failure using machine learning algorithms" *IEEE International Conference on Electronics, Computing and Communication Technologies*, Bangalore, India, 2022

[24]. Kumar R., "Acute liver failure dataset", Available from: https://www.kaggle.com/datasets/rahul121/acute-liver-failure [Accessed 20 December 2022]

[25]. Iesu E., Franchi F., Cavicchi F. Z., Pozzebon S., Fontana V., Mendoza M., Nobile L., Scolletta S, Vincent J. L., Creteur J., Taccone F. S., "Acute liver dysfunction after cardiac arrest" ***PLoS ONE***, 13(11), e0206655, 2018 Available from: https://doi.org/10.1371/journal.pone.0206655

[26]. Lin J. L., Peng Z. Q., Lai R. K., "Improving pavement anomaly detection using backward feature elimination" *20th International Conference on Business Information Systems*, Poznan, Poland, 2017

[27]. Misra P., Yadav A. S., "Improving the classification accuracy using recursive feature elimination with cross-validation" ***International Journal on Emerging Technologies***, 11 (3), 659-665, 2020

[28]. Mustaqim A. Z., Adi S., Pristyanto Y., Astuti Y., "The effect of recursive feature elimination with cross-validation (RFECV) feature selection algorithm toward classifier performance on credit card fraud detection" *International Conference on Artificial Intelligence and Computer Science Technology*, Yogyakarta, Indonesia, 2021

[29]. Chang Y., Chen X., "Estimation of chronic illness severity based on machine learning methods" ***Wireless Communications and Mobile Computing***, 2021, 1-13, 2021 Available from: https://doi.org/10.1155/2021/1999284

[30]. Kern C., Klausch T., Kreuter F., "Tree-based machine learning methods for survey research" ***Surv Res Methods***, 13(1), 73-93, 2019

[31]. Sagi O., Rokach L., "Ensemble learning: A survey" ***WIREs Data Mining Knowl Discov***, 8(4), 1-18, 2018 Available from: https://doi.org/10.1002/widm.1249

[32]. Mayr A., Binder H., Gefeller O., Schmid M., "The evolution of boosting algorithms. From machine learning to statistical modelling" ***Methods Inf Med***, 53(6), 419-427, 2014

[33]. Friedman J. H., "Greedy function approximation: A gradient boosting machine" ***The Annals of Statistics***, 29(5), 1189–1232, 2001

[34]. Bentéjac C., Csörgő A., Martínez-Muñoz G., "A comparative analysis of gradient boosting algorithms" ***Artificial Intelligence Review***, 54, 1937–1967, 2021 Available from: https://doi.org/10.1007/s10462-020-09896-5

[35]. Kim C., Park T., "Predicting determinants of lifelong learning intention using gradient boosting machine (GBM) with grid search" ***Sustainability***, 14(9), 5256, 2022 Available from: https://doi.org/10.3390/su14095256

[36]. Ma X., Sha J., Wang D., Yu Y., Yang Q., Niu X.," Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning" ***Electronic Commerce Research and Applications***, 31, 24-39, 2018 Available from: https://doi.org/10.1016/j.elerap.2018.08.002

[37]. Dalianis H., "Evaluation metrics and evaluation", Clinical Text Mining, *Springer*, *Cham*, Switzerland, 2018 Available from: https://doi.org/10.1007/978-3-319-78503-5_6

[38]. Hussain S, Mustafa M. W., Al-Shqeerat K. H. A., Saeed F., Al-rimy B. A. S., "A novel feature-engineered–NGBoost machine-learning framework for fraud detection in electric power consumption data" ***Sensors***, 21(24), 8423, 2021 Available from: https://doi.org/10.3390/s21248423

[39]. McNemar Q., "Note on the sampling error of the difference between correlated proportions or percentages" ***Psychometrika***, 12(2), 153–157, 1947 Available from: https://doi.org/10.1007/BF02295996

[40]. Prokhorenkova L., Gusev G., Vorobev A., Dorogush A. V., Gulin A., "CatBoost: unbiased boosting with categorical features" *NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems*, Montréal, Canada, 2018

[41]. Han L., Yang T., Pu X., Sun L., Yu B., Xi J., "Alzheimer's disease classification using LightGBM and Euclidean distance map" *IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference*, Chongqing, China, 2021

[42]. Zheng P., Yu Z., Li L., Liu S., Lou Y., Hao X., Yu P., Lei M., Qi Q., Wang Z., Gao F., Zhang Y., Li Y., “Predicting blood concentration of tacrolimus in patients with autoimmune diseases using machine learning techniques based on real-world evidence” ***Front. Pharmacol***, 12, 727245, 2021 DOI: 10.3389/fphar.2021.727245

[43]. Muzumdar P., Basyal G. P., Vyas P., “An empirical comparison of machine learning models for student’s mental health illness assessment” ***Asian Journal of Computer and Information Systems***, 10(1), 1-10, 2022 Available from: https://www.ajouronline.com/index.php/AJCIS/article/view/6882

[44]. Kim E., Han K. S., Cheong T., Lee S. W., Eun J., Kim S. J., “Analysis on benefits and costs of machine learning-based early hospitalization prediction” ***IEEE Access***, 10, 32479-32493, 2022 DOI: 10.1109/ACCESS.2022.3160742