Journal of Medical Systems (2021) 45: 61
https://doi.org/10.1007/s10916-021-01736-5

SYSTEMS-LEVEL QUALITY IMPROVEMENT

An Explainable Artificial Intelligence Framework for the Deterioration Risk Prediction of Hepatitis Patients

Junfeng Peng[1] · Kaiqiang Zou[1] · Mi Zhou[2] · Yi Teng[1] · Xiongyong Zhu[1] · Feifei Zhang[1] · Jun Xu[1]

Received: 23 February 2021 / Accepted: 25 March 2021 / Published online: 13 April 2021

Abstract

In recent years, artificial intelligence-based computer aided diagnosis (CAD) system for the hepatitis has made great progress. Especially, the complex models such as deep learning achieve better performance than the simple ones due to the nonlinear hypotheses of the real world clinical data. However,complex model as a black box, which ignores why it make a certain decision, causes the model distrust from clinicians. To solve these issues, an explainable artificial intelligence (XAI) framework is proposed in this paper to give the global and local interpretation of auxiliary diagnosis of hepatitis while retaining the good prediction performance. First, a public hepatitis classification benchmark from UCI is used to test the feasibility of the framework. Then, the transparent and black-box machine learning models are both employed to forecast the hepatitis deterioration. The transparent models such as logistic regression (LR), decision tree (DT)and k-nearest neighbor (KNN) are picked. While the black-box model such as the eXtreme Gradient Boosting (XGBoost), support vector machine (SVM), random forests (RF) are selected. Finally, the SHapley Additive exPlanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME) and Partial Dependence Plots (PDP) are utilized to improve the model interpretation of liver disease. The experimental results show that the complex models outperform the simple ones. The developed RF achieves the highest accuracy (91.9%) among all the models. The proposed framework combining the global and local interpretable methods improves the transparency of complex models, and gets insight into the judgments from the complex models, thereby guiding the treatment strategy and improving the prognosis of hepatitis patients. In addition, the proposed framework could also assist the clinical data scientists to design a more appropriate structure of CAD.

Keywords Hepatitis · Model interpretation · SHapley Additive exPlanations · Local Interpretable Model-agnostic Explanations · Partial Dependence Plots

Introduction

Liver plays an important role in many essential body functions [1]. Thus, any lesion of the liver adversely affects the important physiological functions such as excretory, secretory and detoxification, which eventually leads to the poor health of patient [2]. Recent research demonstrates that hepatitis such as Hepatitis B or Hepatitis C cause the liver failure, cirrhosis, or cancer [3].

Hepatitis is defined by the World Health Organization as an inflammation of the liver and is caused by a variety of pathogenic factors such as viruses, bacteria, parasites, chemical poisons, drugs, alcohol and autoimmune [4]. Hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), hepatitis D virus (HDV) and hepatitis E virus (HEV) are the five major pathogenic viruses cause the viral hepatitis [5]. Hepatitis like HBV gradually develops into chronic hepatitis, cirrhosis and hepatocellular carcinoma, which eventually leads to a large number of deaths each year [6]. Especially, 80% of patients with HBV develop into liver cancer as the lack of timely medical intervention[7, 8].

However, early intervention on these patients with hepatitis can avoid further damage, and finally reduce

Junfeng Peng, Kaiqiang Zou and Mi Zhou are jointly of the first authorship of the paper.

This article is part of the Topical Collection on *Systems-Level Quality Improvement*

✉ Junfeng Peng
pengjunf@mail2.sysu.edu.cn

[1] Department of Computer Science, Guangdong University of Education, Guangzhou 510303, China

[2] The Third Affiliated Hospital, Sun Yat-sen University, Guangzhou 510630, China

morbidity and mortality. The Model for End-Stage Liver Disease (MELD) is widely used in liver disease diagnosis and treatment because of its simplicity and objectivity [9]. Nevertheless, it remains challenging to identify the onset of liver failure caused by hepatitis due to the complex interaction between liver and other organs [10].

Over the last few years, researchers have utilized machine learning to identify the onset of liver failure caused by hepatitis. Alexandra et al. employed Chi-squared Automatic Interaction Detector (CHAID) to forest the patients who should be screened for chronic hepatitis B or C. The results showed that the probability of HBV infection was higher in patients with ALT $\geq 0.56\ \mu\text{kat/l}$ [11]. Chen et al. utilized Support Vector Machine (SVM), Naive Bayesian Model (NBM), Random Forest (RF) and K-Nearest Neighbor (KNN) to predict the stage diagnosis of the hepatitis. The experimental results showed that RF classifier achieved the best performance among the four machine learning models. The result indicated that the complex models could be a potentially useful tool to predict the stage of hepatic fibrosis [12]. Hashem et al. took advantage of multilinear regression (MR), decision tree (DT), particle swarm optimization (PSO) and genetic algorithm (GA) to forest the advanced fibrosis risk by combining the serum bio-markers and clinical information. The study found that machine learning could be used as the alternative methods to prognose the risk of advanced liver fibrosis caused by chronic hepatitis C [13]. Tian et al. compared the eXtreme Gradient Boosting (XGBoost), RF, DT, and logistic regression (LR) to identify the optimal model to predict the HBsAg seroclearance. The study discovered that the XGBoost reached the best predictive performance for predicting HBsAg seroclearance with clinical data [14]. Singh et al. proposed a hybrid approach to evaluate the stage of hepatitis disease. Simulation results indicated that the improved ensemble learning method performed better than the other existing individual methods on the diagnosis of hepatitis [15].

In general, the complex machine-learning models such as RF and XGBoost perform better than the simple models such as LR in the prediction of hepatitis [16]. However, the complex machine-learning model as a black box does not reveal its internal mechanisms. Thus, the hepatobiliary physicians can not understand the models by looking at their parameters (e.g. a XGBoost). Due to the lack of interpretability, the application of the complex machine-learning approaches in the actual clinical setting is limited.

For the sake of a broader applicability of artificial intelligence (AI) to the hepatitis diagnosis, it is imperative to provide the hepatobiliary physicians with the explanations why a certain prediction is made, that is, the internal mechanisms that lead to the prediction [17]. Thus, an explainable AI (XAI) framework combing the SHapley Additive exPlanations (SHAP) [18], Partial Dependence Plots (PDP) [19] and Local Interpretable Model-agnostic Explanations (LIME) [20] methods is proposed to provide the explanations for the complex models. Figure 1 represents the flow chart of XAI-based diagnosis process of hepatitis. The clinical data collected is sent to the XAI framework after the hepatitis patient is examined by different inspections. Then, the proposed XAI approach generates the computer aided diagnosis (CAD) results to the doctors. Finally, the doctors perform the diagnosis and the treatment to the hepatitis patient with the support of CAD.

The main contributions of this paper are summarized as follows: (1) To obtain a higher exacerbation risk identification accuracy for hepatitis, multiple complex models are explored. A public benchmark data set from UCI is applied to assess the performance of the complex models. (2) To achieve a broader applicability of the complex models to the hepatitis diagnosis, an interpretable framework is proposed to provide the global and local explanations to improve the clinical understanding of the hepatitis exacerbation risk prediction. The rest of paper is organized as follows. In “Methodology”, the research methodology that we apply is explained. In “Results”, we present the details of how the interpretability framework works. “Discussion” discusses the proposed interpretability framework. Finally, “Conclusion” concludes our paper with the future developments.

Methodology

Figure 2 illustrates the framework of the proposed XAI for the CAD system. The proposed framework provides the global and local explanations to improve the clinical understanding of the hepatitis exacerbation risk prediction. Patient record is obtained by data collecting and preprocessing. Then, the model is loaded to predict the outcome of the exacerbation risk. Next, the model explanation method is applied to achieve the global and local explanations. Finally, the prediction and the

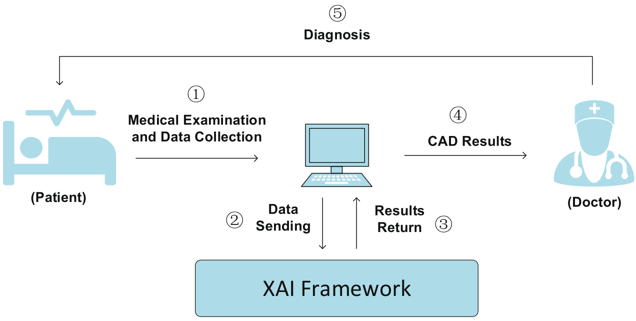

Fig. 1 Flow chart of XAI-based diagnosis process of hepatitis

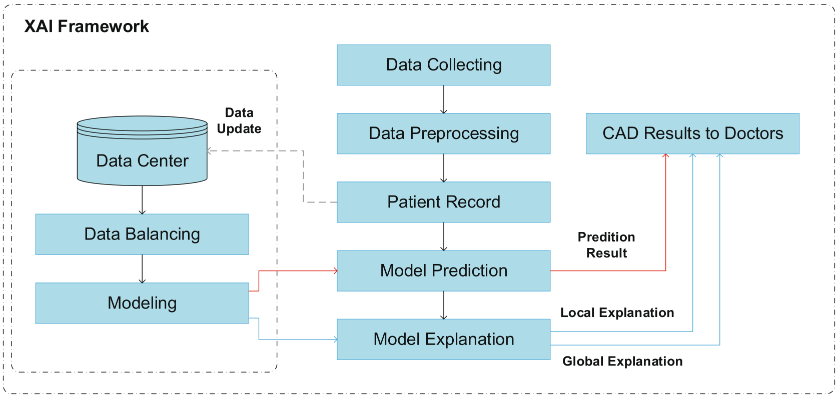

Fig. 2 The framework of XAI

explanation results are transmitted to the doctors for examining and further validating.

Data collecting and preprocessing

To evaluate the feasibility of the proposed framework, a public classification benchmark on hepatitis from UCI machine learning repository is used in the empirical study [21]. The benchmark contains a mixture of integer and real value attributes about patients affected by the hepatitis. The task of our proposed framework is to predict the disease deterioration risk. Distribution of survival (low risk) and death groups (high risk) in patients with hepatitis is shown in Table 1. The benchmark contains 155 patients with hepatitis and 19 features. The attribute description of hepatitis and its abbreviation are shown in Table 1.

As indicated in Table 1, patients those die are labeled as class 1, while those survive are labeled as class 2. The data set contains 75 instances with missing values. To deal with the missing values of the hepatitis data set, the nominal or binary features are set to the majority value while the continuous attributes are set to the average value. The proportion of hepatitis patients with death outcome is 20.6%, while the proportion of survival outcome is 79.4%. To overcome the model bias caused by data imbalance, Synthetic Minority Oversampling Technique (SMOTE) is applied to balance the data set [22].

Model selection and prediction

The easiest way to obtain the model explanations is to apply the interpretable models (simple models) to the clinical data. Linear/logistic regression, decision tree, naive bayesian and k nearest neighbor are the most commonly used explanatory models. However, the simple models such as logical regression can only represent linear relationships between the input and output, which often oversimplify the complex relationships in reality and usually reach the unsatisfactory predictive performance. In the low-risk scene (e.g. a music recommender system), it may be good enough that the simple model performs well on a test dataset. But in the high-risk medical scene, the prediction performance provides the reliability for the model. While the explanations give the clinicians the deeper understanding about the problem, the data and the reason why a model might fail. Thus, the prediction performance and the explanation are both important to the clinicians when designing the CAD system [23].

To achieve a high prediction performance, the complex models SVM, Xgboost and RF are employed to build the model. SVM is a convex optimization problem that achieves data partitioning by searching the hyperplane with maximum intervals [24]. Xgboost is an optimized distributed gradient promotion model, which is designed to be efficient, flexible and portable [25]. RF is generated by the ensemble of decision trees. It is widely used in the analysis and modeling of medical scenarios due to its rapidity, high accuracy, and robustness [26]. In the field of data science, SVM, XGBoost and RF are the most popular models. In particular, SVM, XGBoost and RF are also the most commonly used in hepatitis-assisted decision-making systems. However, it is not enough just to know what is predicted in the high-risk medical scene by the black box models.

Model-agnostic explanations

To obtain the explanation of the complex models, the model-agnostic interpretation methods, the recent advances in machine learning, are applied to achieve the explanations of the complex models while retaining a good prediction performance. Compared with model-specific explanation method, the model-agnostic interpretation is more flexible by separating the model from explanations [27]. Model-agnostic interpretation methods can be divided into two categories: local explanation and global explanation [28].

Table 1 Distribution of survival and death groups in patients with hepatitis. Values are expressed as mean ± standard deviation

		Survival Group(low risk)	Death Group(high risk)
Number of cases		123(79.4%)	32(20.6%)
Sex	Male	107(87.0%)	32(100.0%)
	Female	16(13.0%)	0(0.0%)
Age(year)		39.8 ± 12.8	46.6 ± 9.8
Steroid(STR)	Without	56(45.5%)	20(62.5%)
	With	67(54.5%)	12(37.5%)
Antivirals(ATV)	Without	22(17.9%)	2(6.2%)
	With	101(82.1%)	30(93.8%)
Fatigue(FTG)	Without	71(57.7%)	30(93.8%)
	With	52(42.3%)	2(6.2%)
Malaise(MLS)	Without	38(30.9%)	23(71.9%)
	With	85(69.1%)	9(28.1%)
Anorexia(ANR)	Without	22(17.9%)	10(31.3%)
	With	101(82.1%)	22(68.8%)
LiverBig(LB)	Without	22(17.9%)	3(9.4%)
	With	101(82.1%)	29(90.6%)
LiverFirm(LF)	Without	47(38.2%)	13(40.6%)
	With	76(61.8%)	19(59.4%)
SpleenPalpable(SPP)	Without	18(14.6%)	12(37.5%)
	With	105(85.4%)	20(62.5%)
Spiders(SPD)	Without	29(23.6%)	22(68.8%)
	With	94(76.4%)	10(31.2%)
Ascites(ASC)	Without	6(4.9%)	14(43.8%)
	With	117(95.1%)	18(56.2%)
Varices(VRC)	Without	7(5.7%)	11(34.4%)
	With	116(94.3%)	21(65.6%)
Bilirubin(BLRB)		1.2 ± 0.7	2.5 ± 1.9
AlkPhosphate(APSP)		102.0 ± 45.6	118.1 ± 45.7
Sgot(SG)		82.5 ± 85.5	99.0 ± 96.9
AlbuMint(ABM)		4.0 ± 0.5	3.3 ± 0.6
ProTime(PRT)		64.5 ± 16.6	51.5 ± 15.2
Histology(HTL)	Without	78(63.4%)	7(21.9%)
	With	45(36.6%)	25(78.1%)

LIME is the most commonly used local explanation method. While PDP and SHAP are the most popular global interpretable approaches.

LIME as a local explanation method trains the local surrogate models to provide the interpretability for the complex models. First, LIME creates a new dataset by data perturbation. Then, LIME trains an interpretable model such as decision tree on the new dataset. Finally, the corresponding prediction performance of the black box model is compared with that of the interpretable model. LIME is defined as follows:

$$\gamma(x) = \underset{g\in G}{\arg\min} L(f, g, \pi_x) + \Omega(g) \tag{1}$$

where the loss function L is used to measure how close the interpretable model g is to the prediction of the original complex model f. f is the original complex model. g denotes the interpretable model for the instance x (e.g., logistic regression). G indicates the family of the interpretable models. π_x represents proximity of the sampled instances to the instance x. $\Omega(g)$ is the complexity of model g.

PDP demonstrates the marginal effect of the single feature on the predicted outcome for the complex machine learning model. PDP represents the relationship (linear, monotonous or more complex) between the outcome and input. The partial dependence function $\hat{f}_{x_s}$ defined as:

$$\hat{f}_{x_s}(x_s) = \frac{1}{n}\sum_{i=1}^{n}\hat{f}_{x_s}(x_s, x_c^i) \tag{2}$$

where $\hat{f}_{x_s}(x_s)$ is the partial function which displays the global relationship of a input feature with the predicted

Table 2 Comparison results on the hepatitis dataset using K-fold validation

Models	Interpretable models				Complex models		
	LR	CART	KNN	NBM	SVM	XGBoost	RF
K=5	85.4%	85.4%	77.2%	74.0%	**88.2%**	87.4%	**88.2%**
K=10	85.7%	87.0%	79.2%	72.7%	86.9%	89.8%	**91.0%**
K=20	87.5%	85.9%	78.9%	76.7%	88.3%	89.6%	**91.9%**

outcome. s is a feature set containing only one or two features, x_s denotes the set of features is to be plotted by $\hat{f}_{x_s}(x_s)$, x_c indicates the other features used in the machine learning model f. x_c^i expresses the actual feature values from the dataset for the features in which we are not interested, n is the number of instances of the dataset.

SHAP uses the Shapley values to measure the feature impact for the complex model. Shapley values is defined as the (weighted) average of marginal contributions [29]. It is characterized by the impact of feature value on the prediction across all possible coalitions [30]. Shapley value is defined as:

$$\phi_j(x) = \sum_{s \subseteq \{x_1, x_2, \ldots, x_m\} \backslash \{x_j\}} \frac{|s|!(m-|s|-1)!}{m!}(val(s \cup \{x_j\} - val(s))) \quad (3)$$

where $\phi_j(x)$ is the Shapley value of x_j, x_j represents a feature value, s is a feature subset of the model, m depicts the number of features, val is the prediction for feature values in set s.

Results

Prediction results

We implement the interpretable framework on the development platform of Python 3.6.4. We calculate the overall accuracy of the simple and complex models using K-fold cross-validation. Generally, K is set to 5,10 or 20. The evaluation of the simple and complex models on the hepatitis data are shown in Table 2. When K=5, the prediction accuracy of the developed LR, Classification and Regression Tree (CART), KNN, NBM, SVM, XGBoost and RF with K-fold-cross validation is 85.4%, 85.4%, 77.2%, 74.0%, 88.2%, 87.4% and 88.2%. SVM and RF perform better than the other models. When K=10, the prediction accuracy of the developed LR, CART, KNN, NBM, SVM, XGBoost and RF with K-fold-cross validation is 85.7%, 87.0%, 79.2%, 72.7%, 86.9%, 89.8% and 91.0%. RF perform better than the other models. When K=20, the prediction accuracy of the developed LR, CART, KNN, NBM, SVM, XGBoost and RF with K-fold-cross validation is 87.5%, 85.9%, 78.9%, 76.7%, 88.3%, 89.6% and 91.9%.

We can find that the developed the complex models such as SVM, XGBoost and RF achieve better performance than the simple ones. Especially, RF obtains the best predictive performance. This is mainly due to the data collected fits better with RF. However, RF is a black box model. To get the explanation of RF, the global and local interpretation methods are applied while retaining the good prediction performance.

Global explanations

To get insights into the impact of each predictor to the output of complex model, we compute the mean SHAP values of random forest. Figure 3 demonstrates the average

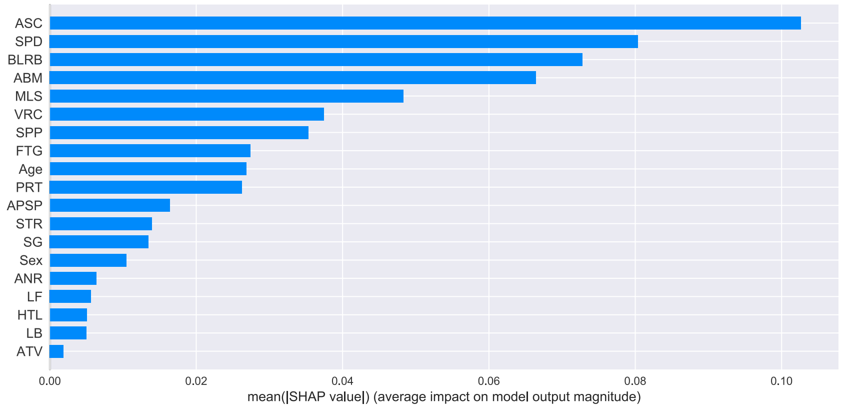

Fig. 3 Average feature impact of the developed RF classifier

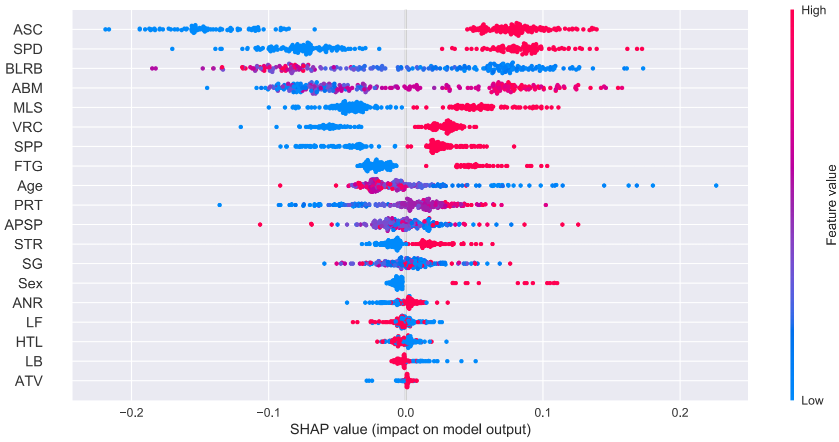

Fig. 4 Averaged feature-importance estimates of random forest

feature impact of the developed RF classifier. We can find that ascites, spiders, bilirubin, albuMin, malaise, varices, and SpleenPalpable have more impact than the others. The explanations for the feature impact are broadly in accordance with the literature and prior knowledge from hepatobiliary physicians.

Figure 4 represents the averaged feature-importance estimates extracted from random forest classifier. Horizontal axis (x-axis) represents the Shapley value which denotes the average feature value marginal contribution on the output across all possible coalitions. Shapley value with less than 0, equal to 0 and greater than 0 means the negative contribution, no contribution and positive contribution, respectively. Left longitudinal coordinate (y-axis) indicates the features which are sorted by the importance in reverse order. Right longitudinal coordinate expresses the value of the features from low to high. We can see that ascites is the most important feature on average, and the developed random forest classifier is more likely to consider the hepatitis patients as high risk when the feature value of ascites becomes larger. Compared with the traditional features importance, the interpretable framework we proposed can assist the hepatobiliary physicians to predict the deterioration risk of hepatitis.

SHAP aids the hepatobiliary physicians to probe the feature contribution of the developed model. While it is also clinically meaningful to explore how each feature affects the model decision-making. Thus, PDP is applied to achieve to visualize the linear, monotonous or more complex relationship between the output and a feature. To visualize the PDP with the continuous features, we examine the effects of the bilirubin and alkphosphate on the predicted output.

Figure 5 shows the relationship between the bilirubin and the prediction of patient outcomes. It can be seen that there exists a complex relationship between the output and the feature bilirubin. First, the impact of bilirubin on the output increases when the value changes from 0.3 to 0.9. Then, the impact falls when the value changes from 0.9 to 2.5. Finally, the impact remains the same when the value is greater than 2.5.

Similarly, Fig. 6 depicts the relationship between the alkphosphate and the prediction of patient outcomes. Similarly, there also exists a complex relationship between the output and the feature alkphosphate. First, the impact of alkphosphate on the output falls when the value changes from 26 to 106. Then, the impact increases when the value changes from 106 to 250. Finally, the impact remains the same when the value exceeds 250.

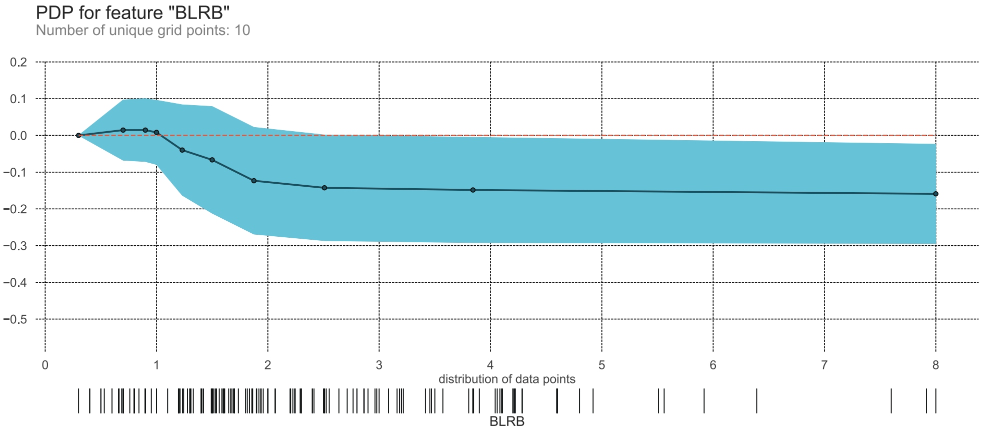

Fig. 5 Partial dependence plot of feature Bilirubin

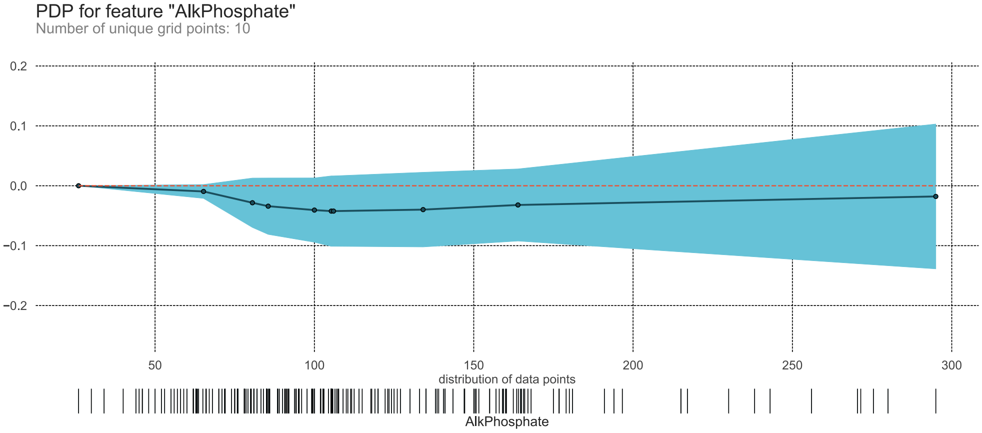

Fig. 6 Partial dependence plot of feature AlkPhosphate

Local explanations

After taking into account the global explanation for the predicted outcome of hepatitis patients, it is also essential to comprehend whether the condition of a specific hepatitis patient will get worse. To explain why individual outcome prediction for the hepatitis patients are carried out by the black box machine learning model, LIME is employed to train a local surrogate models instead of training a global surrogate model. Figure 7 shows the LIME explanations for one instance randomly selected from hepatitis dataset.

The top left diagram shows the predicted outcome of hepatitis patients with probability. Class 1 indicates the hepatitis patients with death outcome, while class 2 represents hepatitis patients with survival outcome. The developed RF classifier predicts the randomly selected hepatitis patient with 93% probability survival (7% probability death). The orange color represents the target class 1 whereas the blue color represents the target class 2. It can be seen that the weight for each feature with their predicted class is denoted by color. They represent the local positive or negative weights assigned to each feature. The greater the weight is, the longer the color bar becomes.

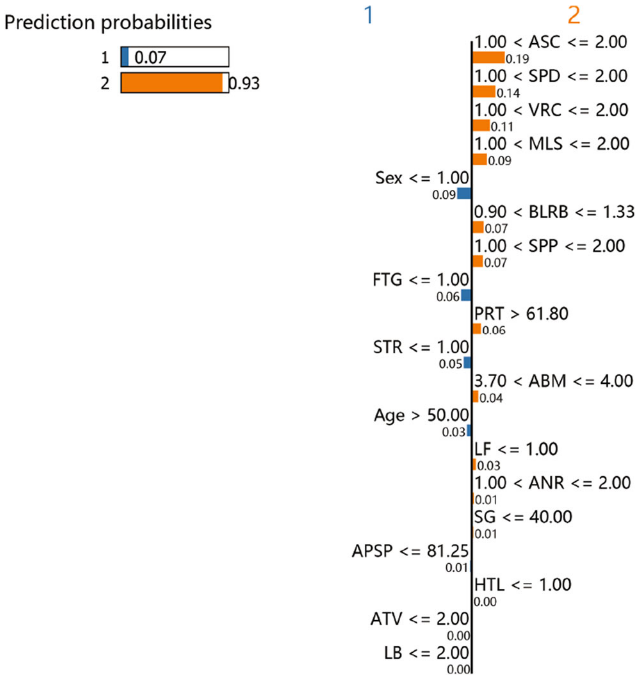

Fig. 7 LIME explanations for one instance from hepatitis dataset

Discussion

We investigate the use of XAI frameworks and the example of such application to support the healthcare of hepatitis. Our research can be summarized as follows: First, both interpretable and complex models are utilized to identify the exacerbation risk in patients with hepatitis. Especially, to improve the prediction accuracy, the complex models based on decision trees are introduced. Second, the global and local explanation methods are employed to avoid the obscurity of the complex models. Third, the predictors such as ascites, spiders, bilirubin, albumin, malaise, varices and spleenpalpable seem to display more important clinical significance than the other predictors. This can assist the hepatobiliary physicians to get insight into the predictions made by the clinical decision support system, and thereby they can make more accurate clinical diagnosis.

Lundberg et al. employed a single complex model (XGBoost) and the explanation method SHAP to predict the intraoperative hypoxaemia events based on the electronically recorded data before they occur [17]. Due to the complexity of clinical decision-making, it is often more convincing to adopt multiple models and interpretation methods. Different from the prescience system Lundberg (2018) developed, we stress the integration of multiple complex models and interpretable methods to improve the clinical understanding of the hepatitis exacerbation risk.

There are some limits in our research. First, to make the experiment objectivity and justice, the present study uses a benchmark on the hepatitis from UCI Machine Learning Repository. The number of patients is relatively small. To ensure the generalization ability of the model, K-fold cross validation is applied. However, more hepatitis data would be needed to conclusively validate the results. Especially, the real time data of hepatitis is the key to realizing the monitor of the exacerbation risk in patients with hepatitis. In the future, we will collect more hepatitis patients from the real world. Second, we employ three typical model-agnostic methods to improve the complex models explanation. However, the other interpretability methods such as counterfactual explanation, which may be conducive to improve the explanation, are ignored for now because the construction of counterfactual samples in medicine often requires rich human and material resources.

Conclusion

In the study, we propose an interpretable machine learning framework, which combines the complex models and the explanation methods that are developed recently, to reliably forecast the exacerbation risk of hepatitis. To evaluate the feasibility of the proposed framework, a benchmark on the hepatitis from UCI Machine Learning Repository is used. The results shows that random forest achieved the best overall accuracy (91.9%). The detailed evaluation of the proposed framework is shown in Table 1. The explanation results generated by the proposed framework agree with the characteristics of the hepatitis, which may improve the diagnostic accuracy of clinicians. In addition, our proposed framework could help the hepatobiliary physicians choose the right structure when they design the CAD system. Our work highlights the values of XAI frameworks in interpreting blackbox models such as RF, which supports the use of AI in healthcare. Further research can focus on the collection of the real time hepatitis data and the exploration of the novel model-agnostic methods.

Author Contributions JFP conceived and designed the study, JFP, KQZ, and MZ drafted the manuscript. JFP, KQZ, MZ, YT, XYZ, FFZ and JX critically revised the manuscript. All authors read and approved the final manuscript.

Funding This work has been supported by Science and Technology Planning Project of Guangzhou (No. 201804010280), Foundation for Young Innovative Talents in Higher Education of Guangdong, China (No. 2017KQNCX140).

Declarations

Ethics approval and consent to participate This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of Interests The authors declare that they have no conflict of interest.

References

1. Pratt, D. S., and Kaplan, M. M., *Evaluation of liver function. Harrisons Principles of Internal Medicine*, pp. 1711–1715. New York: McGraw-Hill, 2002.
2. Acharya, U. R., Koh, J. E. W., Hagiwara, Y. K., Tan, J. H., Gertych, A., Vijayananthan, A., Yaakup, N. A., Abdullah, H. J. J., Fabell, M. K. B. M., and Yeong, C. H., Automated diagnosis of focal liver lesions using bidirectional empirical mode decomposition features. *Comput. Biol. Med. Vol.* 94:11–18, 2018. https://doi.org/10.1016/J.COMPBIOMED.2017.12.024.
3. Lok, A. S. F., Chronic hepatitis B. *New Engl. J. Med.* 346(22):1682–1683, 2002. https://doi.org/10.1056/NEJM200205303462202.
4. Organization (WHO), Hepatitis B, 2002.
5. Longo, D., Fauci, A., Kasper, D., Hauser, S., Jameson, J., and Loscalzo, J., *Harrisons manual of medicine*. New York City: McGraw Hill Professional, 2019.
6. Lee, W. M., and Hepatitis, B., Virus infection. *New Engl. J. Med.* 337(24):1733–1745, 1997.
7. Hews, S., Eikenberry, S., Nagy, J. D., and Kuang, Y., Rich dynamics of a hepatitis B viral infection model with logistic hepatocyte growth. *J. Math. Biol.* 60(4):573–590, 2010. https://doi.org/10.1007/s00285-009-0278-3.
8. Lin, R. H., and Chuang, C. L., A hybrid diagnosis model for determining the types of the liver disease. *Comput. Biol. Med.* 40(7):665–670, 2010. https://doi.org/10.1016/J.COMPBIOMED.2010.06.002.
9. Cholongitas, E., Marelli, L., Shusang, V., Senzolo, M., Rolles, K., Patch, D., and Burroughs, A. K., A systematic review of the performance of the model for end-stage liver disease (MELD) in the setting of liver transplantation. *Liver Transplant.* 12(7):1049–1061, 2006.
10. Luca, A., Angermayr, B., Bertolini, G., Koenig, F., Vizzini, G., Ploner, M., Peck Radosavljevic, M., Gridelli, B., and Bosch, J., An integrated MELD model including serum sodium and age improves the prediction of early mortality in patients with cirrhosis. *Liver Transplant.* 13(8):1174–1180, 2007.
11. Lukáová, A., Babi, B., Paraliová, Z., and Parali, J., *How to increase the effectiveness of the hepatitis diagnostics by means of appropriate machine learning methods. Information Technology in Bio- and Medical Informatics*. Berlin: Springer International, 2015.
12. Chen, Y., Luo, Y., Huang, W. et al., Machine-learning-based classification of real-time tissue elastography for hepatic fibrosis in patients with chronic hepatitis B. *Comput. Biol. Med.* 89:18–23, 2017.
13. Hashem, S., Esmat, G., Elakel, W. et al., Comparison of machine learning approaches for prediction of advanced liver fibrosis in chronic hepatitis c patients. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 15(3):861–868, 2018.
14. Tian, X., Chong, Y., Huang, Y. et al., Using machine learning algorithms to predict hepatitis b surface antigen seroclearance. *Comput. Math. Methods Med.* 2019:1–7, 2019.
15. Singh, A., Mehta, J. C., Anand, D. et al., An intelligent hybrid approach for hepatitis disease diagnosis: Combining enhanced k?means clustering and improved ensemble learninge. *Expert Syst.*, e12526, 2020.
16. Molnar, C., Interpretable machine learning. Retrieved from https://christophm.github.io/interpretable-ml-book/, 2018.

17. Lundberg, S. M., Nair, B., Vavilala, M. S. et al., Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* 2(10):749–760, 2018.
18. Lundberg, S. M., Lee, S. I., and Vavilala, M. S., A unified approach to interpreting model predictionsy. *Neural Inf. Process. Syst.* 30:4768–4777, 2017.
19. Friedman, J. H., Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 29(5):1189–1232, 2001.
20. Ribeiro, M. T., Singh, S., and Guestrin, C., Why should i trust you?: Explaining the predictions of any classifier. In: *North American Chapter of the Association for Computational Linguistics.*, pp. 97–101, 2016.
21. Blake, C. L. U. C. I., Repository of Machine Learning Databases. Dept. of Information and Computer Science. Univ. of California, Irvine. http://archive.ics.uci.edu/ml/datasets/Hepatitis, 1997.
22. Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P., SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16(1):321–357, 2001.
23. Kim, B., Rajiv, K., and Oluwasanmi, O. K., Examples are not enough, learn to criticize! criticism for interpretability. *Neural Inf. Process. Syst.* 29:2280–2288, 2015.
24. Vapnik, V., and Chervonenkis, A., The necessary and sufficient conditions for consistency in the empirical risk minimization method. *Pattern Recognit. Image Anal.* 1(3):283–305, 1991.
25. Chen, T. Q., and Guestrin, C., XGBoost: a scalable tree boosting system. *Knowl. Discov. Data Mining*,785–794, 2016.
26. Breiman, L., Random Forests. *Mach. Learn.* 45(1):785–794, 2001.
27. Ribeiro, M. T., Sameer, S., and Carlos, G., Model-agnostic interpretability of machine learning ICML. In: *Workshop on Human Interpretability in Machine Learning*, 2016.
28. Du, M., Liu, N., and Hu, X., Techniques for interpretable machine learning. *Commun. ACM* 63(1):68–77, 2016.
29. Thomson, W., and Roth, A. E., The Shapley value: essays in honor of Lloyd S. Shapley. *Economica* 58(229):123, 1991.
30. Štrumbelj, E, Kononenko, I., and Hu, X., Explaining prediction models and individual predictions with feature contributions. *Knowl. Inf. Syst.* 41(3):647–665, 2014.