Predicting Clinical Outcomes of Alpha-1 Antitrypsin Deficiency-Associated Liver Disease Using a Stacking Ensemble Machine Learning Model Based on UK Biobank Data

Linxi Meng
Florida State University
Will Treem
Takeda Development Center Americas, Inc
Graham Heap
Takeda Development Center Americas, Inc
Jingjing Chen (Jingjing.Chen@Takeda.com)
Takeda Development Center Americas, Inc

Article

Keywords:

Posted Date: March 21st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1445596/v1

Abstract

With many unknowns in alpha-1-antitrypsin deficiency-associated liver disease (AATD-LD), we aimed to develop a tailored stacking ensemble supervised machine learning (ML) model to predict the disease progression and clinical outcomes of AATD-LD to enable the data-driven decision-making for the clinical outcome endpoints selection as well as clinical development strategy. This analysis was carried out through a stacking ensemble learning model via meta-learning by combining five different supervised ML algorithms with 58 potential predictor variables using a nested 5-fold cross-validation with repetitions based on the UK Biobank data. Performance of the model was assessed through prediction accuracy, area under the receiver operating characteristic (AUROC), and area under the precision-recall curve (AUPRC). The importance of predictor contributions was evaluated through a feature importance permutation method. For example, the AUROC of the prediction model in patients with AATD-LD was 68.1%, 75.9%, 91.2%, and 67.7% for all-cause mortality, liver-related death, liver transplant, and all-cause mortality or liver transplant, respectively. The generalizable predictive patterns support the use of ML to address the unanswered clinical questions with clinically meaningful accuracy using real-world data. This method can be easily applied to other clinical outcomes and/or diseases.

Introduction

Alpha1antitrypsin deficiency (AATD) is an autosomal codominant genetic disorder with a prevalence range of 1 per 2,500 to 1 per 5,000 individuals in Europe and North America that causes early pulmonary disease in adults and liver disease in children and adults,[1] and which often goes under-diagnosed. Alpha-1-antitrypsin (AAT), also known as SERPINA1 (serine protease inhibitor, group A, member 1), is a 52 kDa circulating glycoprotein protease inhibitor of the serpin family. Its primary function is to inhibit neutrophil elastase and other proteases to prevent excessive protease-induced tissue damage.[2,3] AAT is normally synthesized primarily in hepatocytes and secreted in monomeric form. If the AAT proteins are malformed or deficient, it may lead to predisposition for obstructive pulmonary disease and/or liver disease.[1] The PiZZ genotype is known as the most common deficiency genotype and tends to result in the worst clinical presentation.[4] Data from real-world clinical practice has shown that over 90% of AATD is due to the PiZZ genotype.[5] The milder genotypes such as PiSZ and PiMZ are also linked to the development of lung and liver disease, mainly when unhealthy behaviors such as smoking or alcohol use are present.[4] Clinical research shows approximately 40% of adult AATD patients dying of all causes had cirrhosis at the time of death,[6] and approximately 15% of adult patients with AATD-associated liver disease (AATD-LD) required liver transplantation.[7] There is no approved therapy for AATD-LD.

The signs of AATD-LD include elevated transaminases or bilirubin, hepatitis, hepatic fibrosis or cirrhosis.[8] It is known that the liver damage may progress slowly for decades before clinical presentation. Disease progression can be accelerated significantly by other factors, including nonalcoholic fatty liver disease (NAFLD), alcoholic liver disease (ALD), hepatitis, alcohol consumption, smoking, etc. These factors can also cloud accurate diagnosis of AATD-LD.[9–12] Thus, predicting the clinical outcomes of AATD-LD and

defining patients who are more likely to progress to advanced liver disease are crucial for better understanding the disease progression of AATD-LD patients and for promoting timely medical intervention. They will also enable the biopharmaceutical companies and clinical researchers to make data-driven decision to inform clinical outcome endpoints selection as well as clinical development strategy when designing clinical trials for an AATD-LD therapy.

Recent advances in machine learning (ML) algorithms and statistical computing power have enabled ML to be applied in the medical field. Two areas that may benefit most from the application of ML are disease diagnosis and treatment outcome prediction.[13] In this work, we aimed to: (a) establish a predictive ML model of disease progression and clinical outcomes of AATD-LD based on generally available clinical information collected in daily practice; (b) improve the ML model prediction by applying a supervised stacking ensemble learning technique by combining multiple ML algorithms including random forest (RF), elastic net regularized regression (ENRR), gradient boosting (GB), and artificial neutral network multilayer perceptron (ANN-MLP) via meta-learning; and (c) improve the interpretability of predictive ML model by mapping the importance of predictor contributions through a feature importance permutation method.

This article is organized as follows. We present the proposed predictive ML model for AATD-LD based on the real-world data from the UK Biobank as well as the model performance evaluation and model interpretation in the Results section. A brief discussion on the impact of this work is provided in the Discussion section. The basic concepts of the supervised stacking ensemble learning technique, a brief overview of data and analysis pipeline, and the machine learning model training and testing workflow are described in the Method section. Of note, we trained the ML model for AATD-LD and liver disease in general, for comparative purpose. Our work was focused on the prediction of disease progression and clinical outcomes of AATD-LD, but it can be applied to other clinical outcomes and/or diseases. In summary, the generalizable predictive patterns revealed in this work support the potential of ML model as a new tool to address the unanswered clinical questions with clinically meaningful accuracy using the real-world data.

Results

Participant selection.

Data were extracted from the UK Biobank for 11,583 participants with any liver disease, including 455 participants with a diagnosis of AATD-LD. Given the data limitation in the reported genotypes, only 20 AATD-LD patients were identified with the PiZZ genotype through SNP rs28929474. The demographic and disease characteristics of the participants of interest are shown in Table 1.

Variables	Category	Any liver disease (N = 11,583)	AATD-LD (N = 455)
Sex, n (%)	Female	6,097 (52.6%)	226 (49.7%)
	Male	5,486 (47.4%)	229 (50.3%)
	Missing	0	0
Race, n (%)	White	10,840 (94.1%)	426 (94.2%)
	Non-white	674 (5.9%)	26 (5.8%)
	Missing	69	3
Obesity, n (%)	Non-obese	7,180 (62.7%)	321 (72.0%)
	Obese	4,278 (37.3%)	125 (28.0%)
	Missing	125	9
Diabetes, n (%)	Non-diabetic	9,862 (85.9%)	401 (88.9%)
	Diabetic	1,625 (14.1%)	50 (11.1%)
	Missing	96	4
Smoking status, n (%)	Never smoking	5,090 (44.3%	182 (40.3%)
	Past smoker	4,493 (39.1%	194 (42.9%)
	Current smoker	1,911 (16.6%)	76 (16.8%)
	Missing	89	3
Age (years)	Mean	58.5	60.3
	Min, max	40, 70	41, 70
	Missing	0	0
BMI (kg/m^2)	Mean	29.0	27.5
	Min, max	15.0, 69.0	16.9, 52.5

Variables	Category	Any liver disease (N = 11,583)	AATD-LD (N = 455)
	Missing	125	9
Weight (kg)	Mean	82.0	78.0
	Min, max	35.8, 190.0	41.7, 151.4
	Missing	122	7
Waist (cm)	Mean	95.4	92.9
	Min, max	57, 171	62, 153
	Missing	84	4

Table 1. Summary of demographic and disease characteristics in patients with any liver disease and AATD-LD.

Clinical outcomes. The clinical outcomes of interest to assess the disease progress in this work included all-cause mortality, liver-related death, liver transplant, and all-cause mortality or liver transplant. The frequency of these outcomes recorded among study participants is shown in Table 2.

Clinical Outcomes, n (%)	Any Liver Disease (N = 11,583)	AATD-LD (N = 455)
All-cause mortality	3,524 (30%)	245 (54%)
Liver-related death	1,230 (10%)	41 (9%)
Liver transplant	124 (1%)	5 (1%)
All-cause mortality or liver transplant	3,619 (31%)	246 (54%)

Table 2. Summary of clinical outcomes in patients with any liver disease and AATD-LD.

Predictors. Fifty-eight predictor variables collected in the UK Biobank were identified as potential predictors for the clinical outcomes in this work. These potential predictor variables were categorized into 4 predictor blocks, as shown in Table 3.

Category	Description
Predictor Block 1: Demographics	• Age • Age of diagnosis • Gender • Ethnicity • BMI • Weight • Waist circumference

Predictor Block 2:
Baseline disease characteristics
	• Other underlying conditions
	  ◦ Non-alcoholic steatohepatitis (NASH)
	  ◦ Lung disease
	  ◦ Diabetes
	  ◦ Obesity

Predictor Block 3: Lifestyle and others	• Alcohol intake/status • Smoking status • Medical procedure • Major operation

Predictor Block 4
Baseline laboratory parameters
	• Blood assays
	  ◦ Albumin,
	  ◦ Alanine aminotransferase (ALT)
	  ◦ Aspartate aminotransferase (AST)
	  ◦ Alkaline phosphatase (ALP)
	  ◦ Gamma-glutamyl transferase (GGT)
	  ◦ Total bilirubin
	  ◦ Direct bilirubin
	  ◦ International Normalised Ratio (INR)
	  ◦ Haemoglobin (Hb)A1c
	• Spirometry
	  ◦ Forced vital capacity (FVC)
	  ◦ Forced expiratory volume in 1-second (FEV1)
	• Peak expiratory flow (PEF)

Table 3. Description of potential predictor variables (58 predictor variables in total). Variables in predictor blocks 2 and 3 were obtained from patient-reported questionnaires. There may be more than one predictor variable in each predictor category.

***Predictive model performance and prediction accuracy*.** The full dataset was split into the *Training Set* and *Test Set* prior to the model building through a nested 5-fold cross-validation method. Table 4 displays the model performance measures using the stacking ensemble learning algorithm in the *Training Set* and *Test Set* for the four clinical outcomes of interest in patients with AATD-LD, while Table 5 displays the

model performance measures in patients with any liver disease for comparative purpose. The prediction accuracy and the *area under the receiver operating characteristic (*AUROC) were reported as a performance measure to indicate the capability of a classification model to distinguish between classes. A prediction accuracy or AUROC score close to 1 indicates good model separability. The area under the precision-recall curve (AUPRC) was reported as a performance metric for imbalance data. An AUPRC score better than the baseline fraction of positive cases indicates good performance.

For the purpose of comparison, the model performance measures were also reported for each individual ML algorithm that was used to train the stacking ensemble learning algorithm including random forest (RF), gradient boosting (GB), elastic net regularized regression (ENRR), and artificial neural networks multilayer perceptron (ANN-MLP) (Appendix A and B). For illustration purposes, Figure 1 and Figure 2 display the receiver operating characteristic (ROC) curves of the final best model based on the stacking ensemble learning algorithm in the *Test Set* from one Training-Test split in patients with AATD-LD and patients with any liver disease, respectively.

The results showed that the stacking ensemble learning algorithm and the 5 individual ML models all worked generally well with complex data and massive scope of predictors and showed similar prediction performance. In particular, the stacking ensemble learning algorithm performed generally better than each individual ML algorithm in both *Training Set* and *Test Set* (i.e., associated with a higher prediction accuracy and AUROC).

In summary, the model prediction performance is generally acceptable with clinically meaningful accuracy.

- For patients with AATD-LD, the mean prediction accuracy was 0.632 for all-cause mortality, 0.914 for liver-related death, and 0.989 for liver transplant and 0.633 for all-cause mortality or liver transplant respectively. The mean AUROC was 0.681 for all-cause mortality, 0.759 for liver-related death, 0.912 for liver transplant and 0.677 for all-cause mortality or liver transplant in the Test Set, respectively.
- For patients with any liver disease, the mean prediction accuracy was 0.756 for all-cause mortality, 0.913 for liver-related death, and 0.989 for liver transplant and 0.755 for all-cause mortality or liver transplant; and the mean AUROC was 0.770 for all-cause mortality, 0.835 for liver-related death, 0.859 for liver transplant and 0.777 for all-cause mortality or liver transplant in the Test Set, respectively.

Clinical Outcomes	Accuracy		AUROC		AUPRC	
	Training	Test	Training	Test	Training	Test
All-cause mortality	0.828 ± 0.091	0.632 ± 0.038	0.899 ± 0.080	0.681 ± 0.035	0.911 ± 0.073	0.709 ± 0.032
Liver-related death	0.991 ± 0.011	0.914 ± 0.009	0.997 ± 0.007	0.759 ± 0.108	0.979 ± 0.043	0.411 ± 0.170
Liver transplant	1.000 ± 0.001	0.989 ± 0.000	1.000 ± 0.000	0.912 ± 0.133	1.000 ± 0.000	0.414 ± 0.416
All-cause mortality or liver transplant	0.837 ± 0.087	0.633 ± 0.029	0.903 ± 0.076	0.677 ± 0.025	0.917 ± 0.067	0.703 ± 0.040

Table 4. Mean (± standard deviation) model performance measures for stacking ensemble learning in *Training Set* and *Test Set* across the nested 5-fold cross-validation with repetitions in patients with AATD-LD (N=455), respectively.

Clinical Outcomes	Accuracy		AUROC		AUPRC	
	Training	Test	Training	Test	Training	Test
All-cause mortality	0.806 ± 0.025	0.756 ± 0.008	0.852 ± 0.034	0.770 ± 0.009	0.737 ± 0.049	0.629 ± 0.016
Liver-related death	0.991 ± 0.011	0.913 ± 0.004	0.999 ± 0.002	0.835 ± 0.009	0.998 ± 0.003	0.517 ± 0.023
Liver transplant	0.999 ± 0.001	0.989 ± 0.003	1.000 ± 0.000	0.859 ± 0.045	1.000 ± 0.000	0.142 ± 0.048
All-cause mortality or liver transplant	0.815 ± 0.039	0.755 ± 0.006	0.863 ± 0.046	0.777 ± 0.010	0.764 ± 0.067	0.636 ± 0.010

Table 5. Mean (± standard deviation) model performance measures in *Training Set* and *Test Set* across the nested five-fold cross-validation with repetitions in patients with any liver diseases (N=11,583), respectively.

Feature importance. The important predictors with strong contribution to the outcome prediction in the final best model were identified through a feature importance permutation method. Figure 3 and Figure 4 display the top 25 important predictors ranked by feature importance score in patients with AATD-LD and patients with any liver disease, respectively. Of note, the final importance score was obtained as the sum of the feature importance scores calculated through the permutation importance method[14] from each of the candidate ML models used to train the stacking ensemble learning algorithm.

- For patients with AATD-LD, the highest-ranked common predictors of death-related clinical outcomes (i.e., all-cause mortality and liver-related death) were alcohol intake, liver function tests such as GGT

levels or total bilirubin, and smoking status. In particular, the genetic AAT deficiency (e.g., rs28929474 genotypes with Pi type of “ZZ”) appeared to be the highest-ranked predictor of liver-related death, followed by GGT out of range, other serious medication conditions/disability, alcohol intake frequency, GGT, total bilirubin, smoking status, and albumin level out of range. The 10 highest-ranked predictors of liver transplant include GGT, alcohol intake frequency, body fat percentage, other serious medication condition/disability, ethnic background, sex and smoking status.

- For patients with any liver disease, the prediction pattern was similar to that of patients with AATD-LD. The highest-ranked common predictors of death-related clinical outcomes are age at recruitment, liver cancer, liver function tests, and smoking status. In particular, liver cancer, smoking status, liver function tests (i.e., AST, ALT, and GGT levels) and heavy alcohol drinking seem to contribute the most to the prediction of liver-related death. The highest-ranked predictors of liver transplant include alcohol intake frequency, liver cancer, underlying diabetes, smoking status, liver function tests and albumin level. It is worth noting that all three measures of alcohol intake — alcohol intake frequency, heavy alcohol drinking, and alcohol usually taken with meals — seem to play an important role in predicting liver transplant.

Discussion

We aimed to develop a clinically meaningful and accurate predictive model of disease progression and clinical outcomes of AATD-LD based on generally available clinical information collected in daily practice. This work has demonstrated the feasibility of applying a stacking ensemble supervised learning technique via meta-learning by combining five different ML algorithms to a large-scale complex and massive real-world database, the UK Biobank. This methodology was applied to understand the mechanism underlying the multivariate disease progression and clinical outcome prediction of AATD-LD and liver diseases. Of note, one limitation of ML is lack of data or lack of good data. We would recommend using an iterative imputation strategy to resolve model fitting problems due to missing or incomplete information in the predictor variables. Oversampling technique may help to overcome data imbalance challenge during model fitting process and improve model prediction performance without overfitting. In addition, we would also recommend using the nested k-fold cross-validation to optimize the stability of the prediction results.

In summary, the generalized predictive patterns suggest the easily obtained demographic, baseline disease characteristics, lifestyle information and laboratory tests can predict the disease progression and clinical outcomes of AATD-LD and liver diseases with clinically meaningful accuracy. To help to interpret the ML results and disease progression, we adopted a feature importance permutation method by combining feature importance scores across multiple candidate ML models to identify and rank the predictor variables based on their contributions to the predictive model. The identified important predictor variables all appeared clinically relevant, although the predictor variables are slightly different for patients with AATD-LD and patients with any liver disease, which is not unexpected. For example, the genotype of

PiZZ (e.g., rs28929474 genotypes with Pi type of “ZZ”) appeared to be the top one contributor to liver-related death and GGT to liver transplant in AATD-LD patients, while liver cancer was the top one contributor for liver-related death and alcohol intake for liver transplant in patients with any liver disease. One interesting finding is the focus on GGT, which is usually dismissed as unimportant in predicting the course of chronic liver disease because (1) it is more a sign of intrahepatic cholestasis and injury to the canalicular membrane or biliary epithelium than to hepatocytes; or even a sign of obstructive jaundice secondary to either intrahepatic or extrahepatic obstruction of bile ducts; and (2) it can easily be mildly perturbed by moderate alcohol intake, smoking, and multiple common medications. Using the feature importance permutation method to detect important predictors with a strong contribution to outcome, it appears that in AATD-LD GGT out of range is more important than laboratory parameters more reflective of liver metabolism and transport (total bilirubin), liver synthetic function (albumin), liver injury (AST), or cholestasis (alkaline phosphatase). In a previous study of PiZZ AATD patients with lung and liver disease, it is important to note that GGT is also found in the lung and that serum GGT is related to static lung function, chronic bronchitis, sputum purulence, history of acute exacerbations, and smoking status in addition to alcohol consumption, cirrhosis and serum markers of liver disease.[15] GGT was independently correlated with airflow obstruction and was associated with chronic bronchitis and independently associated with mortality. This suggests that the importance of out-of-range GGT in AATD may originate from its dual source of origin in the two most affected and damaged organs in AATD, the liver and the lung. This work may suggest the components of a clinical composite score that will help to predict disease progression to these clinical outcomes. In addition, it is worth noting that lung manifestations of AATD (e.g., FVC) appear as important predictors of all-cause mortality, but not liver-related mortality.

This work may lead to greater insights in clinical practice and assist clinicians in effectively identifying high-risk patients with AATD-LD and/or liver diseases, mitigating the burden of diagnosis, and in managing the disease progression and treatment. It may also enable a data-driven strategy for the biopharmaceutical companies to select clinical outcome endpoints and target patient populations in clinical research when developing a treatment for AATD and/or liver diseases.

There are a few limitations of this work. First, given the data limitation of the UK Biobank, only the first 4 digits of International Classification of Diseases (ICD) code were available to identify patients with AATD-LD, which might affect the precision of AATD-LD patient selection. For example, E88.01 is the ICD10 code for AATD-LD, while only E88.0 was recorded in the UK Biobank. Secondly, there are very few AATD-LD patients with genotype information in the UK Biobank, which limited our ability of further exploring the predictive pattern of disease progression and clinical outcomes in a subset of AATD-LD patients with PiZZ genotype. Lastly, one of the foci for liver disease research is to understand the patient disease progressive journey, in particular that of rapid disease progression. For future research, we will further explore the potential predictors of rapid disease progression of AATD-LD.

Methods

This section provides a brief overview of data and analysis pipeline, data assembly and process prior to the modeling training, and the machine learning model building workflow in this work. We also provide the details of statistical techniques applied to improve the model performance and interpretation including feature selection, oversampling technique and feature importance. The principles that we demonstrated in this work can be readily applied to other clinical outcomes and/or disease indications. All methods were carried out in accordance with relevant guidelines and regulations.

Data and analysis pipeline. Patient data were extracted from the United Kingdom (UK) Biobank (https://biobank.ctsu.ox.ac.uk), a large-scale biomedical database and research resource of 500,000 participants aged 40 to 69 years recruited throughout the UK between 2006 and 2010. The database includes participants with a wide range of serious and life-threatening illnesses, who have undergone measures, provided blood, urine and saliva samples, and detailed information about themselves, and agreed to have their health followed.

Data were extracted (UK Biobank application #26041) for participants with a diagnosis of any liver disease according to ICD codes were selected for study. Participants with a diagnosis of AATD-LD (identified by ICD code or questionnaire) and PiZZ genotype (SNP rs28929474) were subsequently identified. The data were pre-processed before modeling (e.g., centering and scaling the predictors, imputing the missing predictor information via multiple imputation). The process flow for data assembly, processing, and analysis is shown in Fig. 5.

Feature engineering. To prevent the modeling barriers from the overfitting or multicollinearity, redundant features were eliminated through feature selection methods prior to the model training. The final set of predictor variables for the model training was selected through the joint application of seven feature selection methods including: (1) filter methods, such as Pearson correlation and Chi-squared correlation; (2) wrapper methods, such as feature elimination recursive; and (3) embedded methods such as Lasso and three tree-based models (Fig. 6). Predictor variables selected by at least 4 of the 7 selection methods were assigned to one of four predictor blocks (Fig. 5).

Oversampling technique. To address the imbalanced classification challenge where there were too few records of a minority class for the model to effectively learn and to improve the model performance on the minority class, the synthetic minority oversampling technique (SMOTE)[16] was applied to the clinical outcomes with data imbalance including liver-related death and liver transplant. The new synthetic records were generated using the existing samples of the minority class by linear interpolating for the minority class. AUPRC was used as a performance measure for data imbalance.

Machine learning model building. The stacking ensemble learning algorithm[17] was applied in this work for a better model prediction performance. The stacking ensemble is a meta-learning algorithm that combines the predictions from multiple well-performing machine learning models including classification tree and/or regression methods to make the final model perform better than any single model in the ensemble. We applied and combined the learning from RF, GB, ENRR, and ANN-MLP in this work.

- RF is a ML algorithm for classification, which consists of a large number of individual decision trees and uses bagging and feature randomness for training to create an uncorrelated forest of trees. The final prediction from random forest model is the class selected by most trees.
- GB is a ML algorithm that uses boosting technique and grows trees in a stage-wise, gradual, additive and sequential manner. Two GB algorithms were applied in this work, including eXtreme Gradient Boosting (XGBOOST), which splits the tree level-wise; and light GBM, which has faster training speed and higher efficiency.
- ENRR is an application of regularized regression with penalties to avoid extreme parameters that could cause overfitting. ENRR combines two commonly used regularization techniques (Lasso and Ridge) into a hybrid penalized model.
- ANN is one of the deep-learning algorithms inspired by the structure and function of the human brain. MLP is a class of feedforward ANN. We applied multiple-input single-output neural network forecasting in this work.

To optimize the stability of the prediction results, a nested five-fold cross-validation with independent random partitions was conducted with 100 repetitions. The nested cross-validation has an inner loop cross-validation nested in an outer cross-validation, where the inner loop was used for model selection and hyperparameter tuning and the outer loop was used for model performance evaluation. The ML model was trained using the *Training Set* and evaluated using the *Test Set* through the nested five-fold cross-validation. Of note, the SMOTE oversampling technique was applied to the *Training Set* for liver-related death and liver transplant. To avoid the noisy estimate of model performance by a single run of nested five-fold cross-validation, we conducted different splits of *Training* and *Test* data by repeating the nested five-fold cross-validation 100 times to stabilize the performance of the ML models. The model performance was evaluated by prediction accuracy, AUROC, and AUPRC. The mean result and standard deviation across all iterations were reported. It is worth pointing out that the mean result is considered as a more accurate and stable estimate to the underlying performance of model prediction.

This analysis was carried out using Python 3.8 and Keras 2.5.0. Figure 7 presents the workflow of the stacking ensemble learning algorithm in this work.

Feature importance. Feature importance refers to a class of techniques for assigning scores to input features in a predictive model that indicates the relative importance of each feature when making a prediction, which can provide insight and better understanding into the data and a ML prediction model. We applied the permutation importance[14] to each of the five ML models to obtain the permutation importance scores and calculated the final feature importance score by summing up these importance scores (Appendix C). The important predictors were identified and ranked based on the final importance score.

Data and code availability

The data underlying this article is a part of the UK Biobank dataset (application #26041), but not publicly available. The data, data processing, feature extraction, machine learning, and analysis code will be shared by the corresponding author upon reasonable request.

Declarations

Acknowledgements

We thank Dr. Erin Smith, Dr. Michael Williams, and Rajesh Mikkilineni for their contribution and support. We also thank Dr. Nick Rusbridge, Oxford PharmaGenesis, Oxford, UK, for editorial and submission support (funded by Takeda).

Author contributions

J.C. conceived of the presented idea and developed the theory. L.M. performed the analysis. T.W. and G.H. contributed to the data interpretation. All authors discussed the results and contributed to the final manuscript.

Additional information

Conflicts of interest:

1. JC, WT, GH – Employee of Takeda and receive stock/stock options
2. LM – Takeda intern at the time of this work

Funding:

This study was sponsored by Takeda Development Center of the Americas, Inc.. Editorial support was provided by Oxford Pharmagenesis.

References

1. Nelson, D. R., Teckman, J., Di Bisceglie, A. M. & Brenner, D. A. Diagnosis and management of patients with α1-antitrypsin (A1AT) deficiency. Clin. Gastroenterol. Hepatol. **10**, 575–580 (2012).
2. Kim, M., Cai, Q. & Oh, Y. Therapeutic potential of alpha-1 antitrypsin in human disease. Ann. Pediatr. Endocrinol. Metab. **23**,131–135 (2018).
3. Strnad, P., McElvaney, N. G. & Lomas, D. A. $Alpha_1$-antitrypsin deficiency. N. Engl. J. Med. **382**, 1443–1455 (2020).
4. Santos, G. & Turner, A. M. Alpha-1 antitrypsin deficiency: an update on clinical aspects of diagnosis and management. Faculty Reviews **9**, 1. https://doi.org/10.12703/b/9-1 (2020).
5. de Serres, F. J. & Blanco, I. Prevalence of α1-antitrypsin deficiency alleles PI*S and PI*Z worldwide and effective screening for each of the five phenotypic classes PI*MS, PI*MZ, PI*SS, PI*SZ, and

PI*ZZ: a comprehensive review. Ther. Adv. Respir. Dis. **6**, 277–295 (2012).

6. Elzouki, A. N. & Eriksson, S. Risk of hepatobiliary disease in adults with severe alpha 1-antitrypsin deficiency (PiZZ): is chronic viral hepatitis B or C an additional risk factor for cirrhosis and hepatocellular carcinoma? Eur. J. Gastroenterol. Hepatol. **8**, 989–994 (1996).
7. Townsend, S. A. *et al.* Systematic review: the natural history of alpha-1 antitrypsin deficiency, and associated liver disease. Aliment. Pharmacol. Ther. **47**, 877–885 (2018).
8. American Thoracic Society & European Respiratory Society. American Thoracic Society/European Respiratory Society statement: standards for the diagnosis and management of individuals with alpha-1 antitrypsin deficiency. Am. J. Respir. Crit. Care Med. **168**, 818–900 (2003).
9. Wiegand, J. & Berg, T. The etiology, diagnosis and prevention of liver cirrhosis. Dtsch Arztebl. Int. **110**, 85–91 (2013).
10. Mitra, S., De, A. & Chowdhury, A. Epidemiology of non-alcoholic and alcoholic fatty liver diseases. Transl. Gastroenterol. Hepatol. **5**, 16. doi: 10.21037/tgh.2019.09.08 (2020).
11. Hamesch, K. & Strnad, P. Non-invasive assessment and management of liver involvement in adults with alpha-1 antitrypsin deficiency. Chronic Obstr. Pulm. Dis. **7**, 260–271 (2020).
12. Townsend, S. A., Edgar, R. G., Ellis, P. R., Kantas, D., Newsome, P. N. & Turner, A. M. Systematic review: the natural history of alpha-1 antitrypsin deficiency, and associated liver disease. Aliment. Pharmacol. Ther. **47**, 877–885 (2018).
13. Sidey-Gibbons, J. A. M. & Sidey-Gibbons, C. J. Machine learning in medicine: a practical introduction. BMC Med. Res. Methodol. **19**, 64 (2019).
14. Fisher, A., Rudin, C. & Dominici, F. All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. J. Mach. Learn. Res. **20**, 177 (2019).
15. Holme, J., Dawkins, P. A., Stockley, E. K., Parr, D. G. & Stockley, R. A. Studies of gamma-glutamyl transferase in alpha-1-antitrypsin deficiency. COPD. **7**, 126–132 (2010).
16. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002).
17. Wolpert, D. H. Stacked generalization. Neural Netw. **5**, 241–259 (1992).

Figures

Figure 1

ROC curves for the trained classifiers in the *Test Set* from one *Training-Set* split for patients AATD-LD (N=445). Given the low liver transplant event incidence in AATD-LD patients in the *Test Set*, there was insufficient data to populate the ROC curve for liver transplant and a box plot was presented instead.

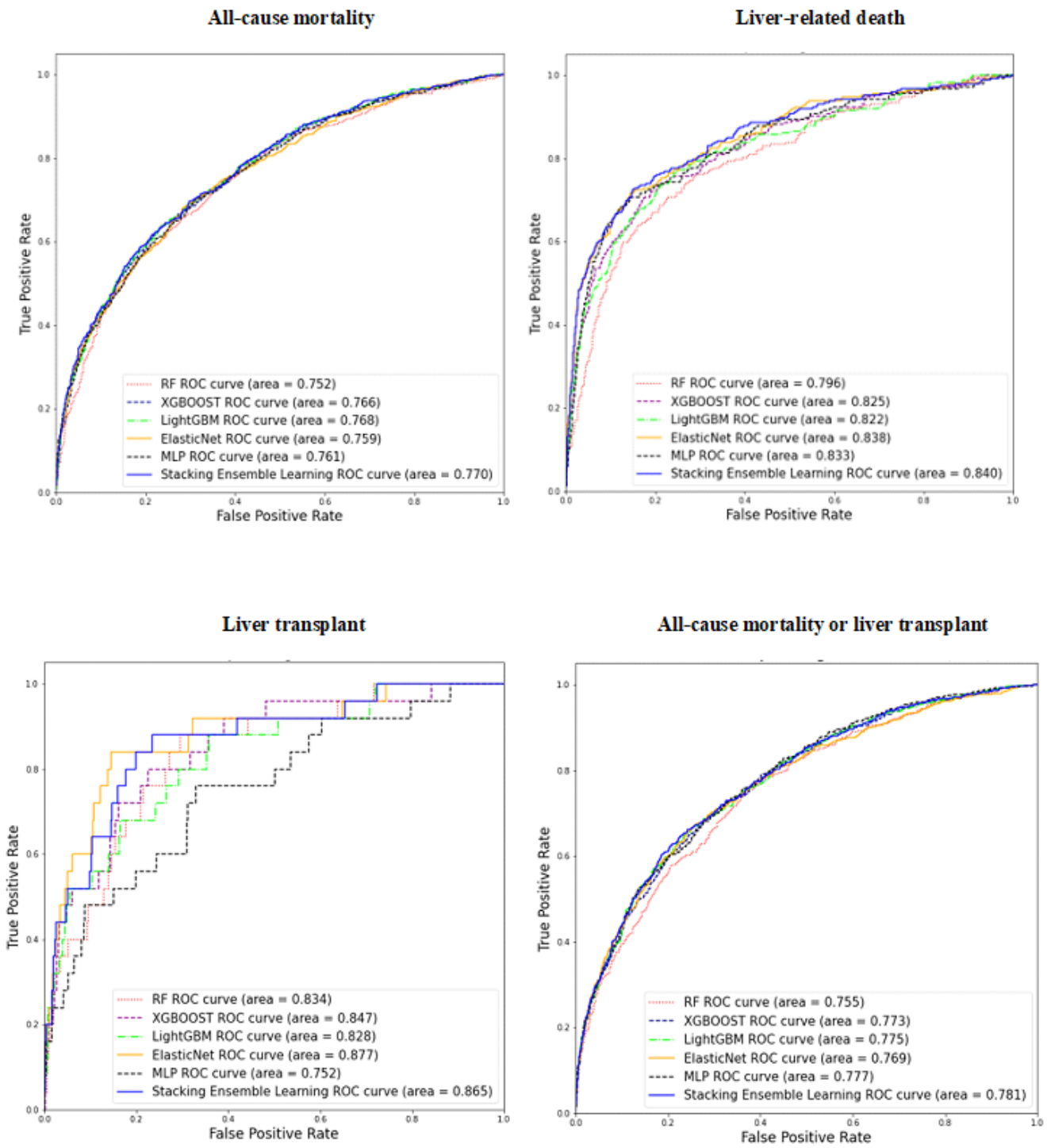
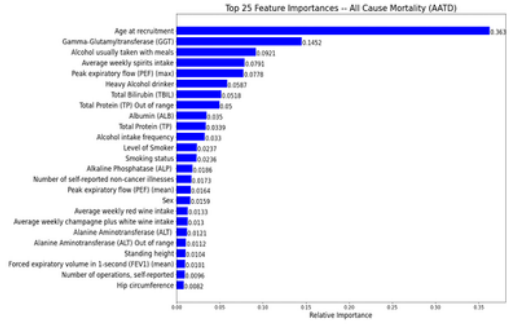


Figure 2

ROC curves for the trained classifiers in the *Test Set* from one *Training-Set* split for patients with any liver disease (N=11,583).

All-cause mortality



Liver-related death

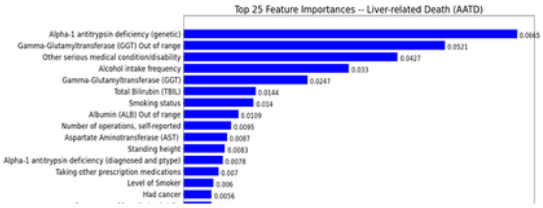


Figure 3

Feature importance in the final stacking ensemble learning model for patients AATD-LD (N=455).

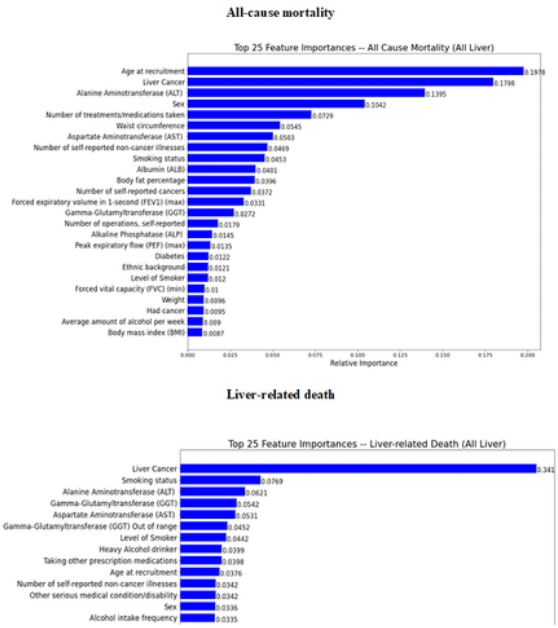


Figure 4

Feature importance in the final stacking ensemble learning model for patients with any liver disease (N=11,583).

Figure 5

Flow chart of data assembly, processing, and analysis.

Figure 6

Feature selection strategy prior to model training.

Figure 7

The workflow of stacking ensemble learning.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AppendixABC.docx