Characterizing rat clinical chemistry tests for liver disease phenotypes: a large-scale, cross-study analysis using standardized electronic submission data.

***Authors:***

Daniel P. Russo1,2, C. M. Sabbir Ahmed1,2, Frederic Moulin1, Kevin P. Cross3, Kevin Snyder2

1Center for Drug Evaluation and Research, Silver Spring, MD;

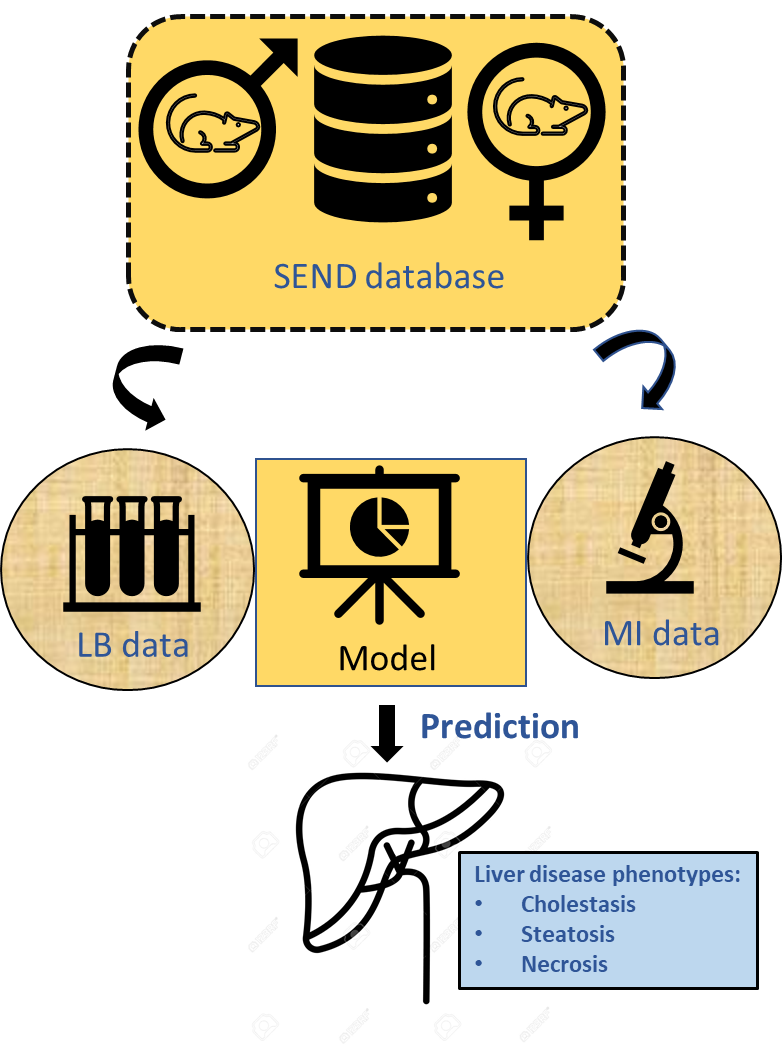
2Oak Ridge Institute for Science and Education, Oak Ridge, TN;

3Leadscope, Columbus, OH;

***Abstract***

The U.S. FDA Center for Drug Evaluation and Research requires the submission of nonclinical pharmacology and toxicology studies to support the safety of human subjects participating in clinical trials. While these data provide a valuable resource to inform the dosing, design, and safety monitoring of clinical studies, attrition during drug development due to unexpected clinical toxicity still occurs, indicating the need for development of novel methods to improve the interpretation of these nonclinical data. The FDA has begun to require the submission of Standard for Exchange of Nonclinical Data (SEND) datasets along with study reports for all repeat-dose toxicology studies initiated after December 17, 2017, providing a constant source of structured *in vivo* toxicology study data in a standardized, electronic format. SEND data submitted to the FDA provide a large, unique data landscape for informatics and machine learning-based approaches to help establish and understand the relationships between biological processes and toxicity to better identify nonclinical safety signals. In this work, we leveraged the wealth of animal toxicity data contained within > 4,300 SEND datasets containing over 397,000 animals to explore the relationship between 47 toxicity study endpoints including clinical chemistry tests, body weights, and sex to different rat liver disease phenotypes (i.e., hepatocellular necrosis, cholestasis, and steatosis). Specifically, 39,969 rats were classified as having different liver disease phenotypes based on histopathological evaluation of their liver tissue. Distinct patterns of biological perturbations and injury were observed in the clinical chemistry test results of rats with different liver disease phenotypes, suggesting that these clinical chemistry profiles can be used to inform the evaluation of liver pathology in nonclinical toxicology studies. Furthermore, these variables were also the basis for the development of machine learning classifiers to predict the probability of rat liver disease. The resulting classifiers showed acceptable overall accuracy (balanced accuracy, 57-66%) through a ten-fold cross validation process for each liver disease phenotype. In the future, this workflow could be amended and improved with the submission and incorporation of new SEND data, as well as be extend to other complex toxicity endpoints (e.g., nephrotoxicity, cardiotoxicity) to inform on potential novel biological processes and perturbations.

**Graphical Abstract**



# INTRODUCTION

Prior to the initiation of clinical trials, the U.S. Food and Drug Administration (FDA) Center for Drug Evaluation and Research (CDER) requires the submission of nonclinical animal toxicity data for every Investigative New Drug (IND) application [REF]. For each application, Good Laboratory Practice (GLP)-compliant single-dose or repeat-dose general toxicity studies are conducted in accordance with the guidelines set forth by the Organisation for Economic Co-operation and Development (OECD) and International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH). These studies typically evaluate the toxicological profile of the active pharmaceutical ingredient (API) in at least two mammalian animal species, i.e., one rodent and one non-rodent, across multiple dosing durations. The results of these studies are reviewed by both industry and regulatory toxicologists to inform various aspects of clinical trial design, e.g., dose selection, dosing regimen, and safety monitoring, in the interest of ensuring human safety in clinical trials. On average, CDER receives over 1,500 IND applications a year, providing a wealth of toxicity information for retrospective analysis1. Historically, however, the results of these animal experiments have been provided in an unstructured, human-readable format, e.g., a Portable Document Format (PDF) summary containing the relevant toxicity data in tables or figures making cross-study computational analysis difficult. However, beginning December 17, 2017, the FDA began requiring all repeat dose nonclinical toxicity studies submitted in support of IND applications be submitted via a data standard, called the Standard for Exchange of Nonclinical Data (SEND). The SEND data standard, developed by the Clinical Data Interchange Standards Consortium (CDISC), allows for the full representation of a nonclinical toxicity study in a standardized, machine-readable format. Specifically, SEND organizes various components of a nonclinical toxicity study into several “domains” in a relational manner, such that when taken together, they can be used to describe a study from start to finish, including most importantly, toxicity data such as histopathology findings or clinical chemistry test results. As of August 2021, over 4,300 SEND studies have been submitted to the FDA through IND applications, providing a rich amount of toxicity data for informatics.

The liver is the largest internal organ in the body, comprising approximately 2% of the total body weight and plays important functional roles including in filtering blood and metabolizing xenobiotics, secreting bile to the intestines to aid in food digestion, storing glycogen for energy retrieval, and synthesizing proteins for blood plasma2–7. Specifically, the liver’s unique role in filtering blood and nutrients ingested into the digestive tract leave it more susceptible to the exposure of toxic agents than other organs. Indeed, drug-induced hepatotoxicity accounts for greater than 50% of acute liver failure cases in the United States and has been the cause for post-market attrition of several pharmaceuticals6,8–11. Because of this, detection of potential hepatotoxicity safety signals in nonclinical studies, prior to administration to human subjects, is a critical priority. Relevant liver biomolecules, including bilirubin or liver enzymes such as alanine aminotransferase (ALT), aspartate aminotransferase (AST), and alkaline phosphatase (ALP), are routinely tested in clinical settings to monitor potential underlying drug-induced hepatotoxicity12,13. These biomarkers provide clinicians with a powerful diagnostic tool to assess underlying liver disease in patients and have even shown potential to discern between different liver disease phenotypes based on distinct biomolecule patterns14. While these biomarkers are also routinely tested as nonclinical studies as potential markers for hepatotoxicity, their relationship between liver diseases, including different liver disease phenotypes, remains incomplete.

In this work, we use the wealth of animal toxicity data available within the U.S. FDA’s SEND database to characterize drug-induced hepatotoxicity in rats. Herein, we present a workflow to leverage data across SEND studies to establish relationships between clinical chemistry tests (e.g., ALT, AST, ALP) and animal meta data (e.g., body weight and sex) with different organ disease phenotypes (e.g., acute liver necrosis, steatosis, cholestasis) identified from histopathology findings. Additionally, we explore the ability of the clinical chemistry tests and animal metadata to predict the probability of an underlying liver disease phenotypes in rats using machine learning. This workflow can be extended to other animals outside of rats, as well as other complex toxicities (e.g., nephrotoxicity) and to our knowledge is the first such large-scale study using SEND data.

# METHODS

## *SQLite Database Generation*

SEND datasets received by the FDA through August 10, 2021 were parsed and converted from their native format (Statistical Analysis System (SAS) version 5 export files, XPT files) and stored in a relational database management system (SQLite, version 3.30.1). The FDA’s Electronic Document Room (EDR) contains electronic regulatory review submissions from INDs and NDAs, Drug Safety Reports, including SEND datasets submitted in support of these applications. To identify folders containing SEND datasets, an exhaustive search of the eCTD Module 4 folders containing a ‘define.xml’ file was performed. To ensure SEND dataset validity prior to loading into the SQLite database, a series of quality checks were performed. These included that the submitted SEND dataset 1) contained a TS domain, 2) contained a DM domain, 3) contained an EX domain, 4) had readable .xpt files for all submitted domains, 5) contained domain variables allowed in either SEND Implementation Guide version 3.0 or 3.1, and 6) had one unique STUDYID across all domains submitted. After all valid datasets were inserted into the SQLite database, duplicated rows (i.e., those with exact values across all variables in a domain) were removed.

## *Training Set Preparation*

### Animal Curation

SEND studies each have different target goals, different species as animal subjects, different dosing regimens, etc. In order to further standardize and created a robust training set for analysis, we selected animals were selected based off a few different criteria. First, we restricted the modeling set to rats of any strain, labeled as RAT in CDISC Controlled terminology. Then, only rats sacrificed immediately after recovery were used. These rats were identified in a SEND study by determining the epoch during which the rat was sacrificed. Then, because there is no controlled terminology for the epoch’s a set of regular expressions were used to identify the epoch as either “screening”, “treatment”, or “recovery, as described in a previous publication15. Only rats sacrificed in the treatment phase were retained to develop a training set. In addition, because all clinical chemistry tests were normalized as fold increase above control, we limited the set of training rats to those in studies with an easily identifiable negative control group, i.e., those with a TCNTRL containing any of the following phrases: "placebo", "untreated", "sham", "negative", "saline", "peg", "vehicle", "citrate", "dextrose", "water", or "air".

### Clinical Chemistry Tests and Body Weights

The laboratory domain (LB) contains clinical chemistry tests taken during and after the treatment phase of the study. All available clinical chemistry tests for the rats in the training set were collated from this domain. There is no standard or preferred unit or tests for many common tests. For example, total bilirubin can be submitted in mg/dL or μmol/L requiring conversion factor. In order to avoid conversions, the peak clinical chemistry test responses for each rat were grouped by study and normalized as a fold change above the study negative control with respect to sex (i.e., male rats were normalized to male rat controls) so long as the study reported unique units for each test. In addition to clinical tests, rat body weight was captured by fitting a regression line to the rats’ body weights throughout the study. Along with the difference in body weight from the start and end of the study, the slope and y-intercept of the body weight regression line were all used as variables in our analysis. These values were also normalized to the study negative control with respect to sex and were named BWDIFF\_NORM, BWSLOPE\_NORM and BW\_INTERCEPT\_NORM, respectively.

### Hepatotoxicity Mechanism Classification

To classify phenotypes of particular liver disease, we created regular expressions to identify terminology in pathology findings that may be associated with certain liver disease phenotypes. For example, cholestasis is characterized as a disruption of the flow of bile from the liver into the gallbladder. Pathology results describing “bile” or “biliary” could potentially indicate liver disease via cholestasis. Similarly, steatosis is characterized by a build up of lipid-containing vacuoles. Therefore, pathology results mentioning “fat”, “lipids”, “vacuolation”, “accumulation”, and “congestion” may be indicative of underlying steatosis. Lastly, more severe findings like “necrosis”, “fibrosis”, “degeneration”, “atrophy”, “apoptosis”, and “depletion” may suggest general hepatocellular necrosis. The regular expressions and the terms they were associated to capture are listed in Table 1. For each animal, these regular expressions were used to classify their liver necropsy findings as having steatosis, cholestasis, necrosis, or no observable liver disease. Since many animals, typically those in the low and medium dose groups in a study, do not receive a necropsy unless findings in the high dose groups are found, animals without any reported findings were assumed to be absence of liver disease.

## *Analytical Approaches*

### Z-score Calculation

To identify potential biomarkers for the various liver disease phenotypes the z-scores [ref: Kevin C paper] were calculate for all clinical chemistry tests and animal body weights (i.e., all predictor variables except animal sex). Z-scores were calculated for every predictor variable with respective to each liver phenotype by the following equation:

where, *x1* is a predictor variable for liver disease phenotype positive rats and *x0* is for the whole group. The values *n1* and *n0* represent the number of liver disease phenotype positive rats and the whole, respectively; *s0* is the standard deviation of all the rats.

### Machine Learning

Rat clinical chemistry responses along with rat body weights variables and sex were used to train random forest models. Prior to modeling, rats missing clinical chemistry tests had their responses imputed by taken the mean value for that test. Then, the tests were shifted into the positive range, log-scaled, then mean cantered and scaled to unit variance. The random forest models were created using the software scikit-learn. Model training was processed for each of the liver disease phenotype separately balancing the number of disease positive. Because the number of disease positive rats were much greater than the number of disease negative rats, the number of disease negative rats was randomly down sampled to match the appropriate number of disease positive rats for each phenotype. Each model was then validated using a standard 10-fold cross validation procedure. After model training, the importance of each predictor variable to the disease phenotypes was ranked using the Gini Importance.

Four metrics were used to assess model performance as standard the classification of binary variables. These included the precision, recall, specificity, and balanced accuracy and are defined as follows:

where, *TP*, *FP*, *TN*, and *FN* represent the respective counts for the true positive, false positive, true negative, and false negative cases.

All code can be found in the GitHub repository: [www.github.com/fingerscrossed/wecanpost.git](http://www.github.com/fingerscrossed/wecanpost.git).

# RESULTS AND DISCUSSION

## Animal Database Overview and Training Set

A total of 4,376 SEND datasets from 1,790 drug applications (i.e., Investigative New Drug Applications, Abbreviated New Drug Approvals, and New Drug Applications) submitted between December 2017 and August 2021 were identified from the FDA’s Electronic Document Resource and loaded into a SQLite database for a total of approximately 11 gigabytes of data. This database consists of toxicity information on 397,060 animal subjects, including rats, mice, monkeys, dog, among others, with percentage of animal species shown in Figure 1. Rats were the most populated species, consisting of 56.8% of the animals or a total of 226,547 rats. Because they were the most animal with the most data points, these 226,547 were used to develop a training set of animals to explore the relationship between clinical chemistry tests and various liver disease phenotypes. As further described in the Methods, this training set was created by curating rats to only include those from studies with an easily identifiable negative control group and those that are terminal rats (i.e., received no recovery phase in the study design) resulting in final set of 39,969 rats in the final training set. Notably, of these 39,969 rats 10,175 were identified as negative control animals, meaning the remaining 29,794 rats were included in a study as either one of the treatment groups (low, medium, or high dose) or as a positive control. Additionally, while the majority of the rats received treatment orally as the route of administration (~70%) other routes such as intravenous and respiratory and others were included.

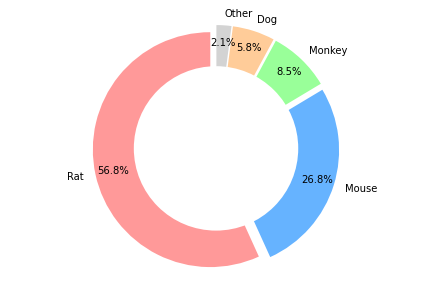


Figure 1. Frequency of species in SEND across 397,060 animals.

## Liver Disease Prevalence

OECD guidelines recommends that all animal subjects within a study undergo a full, detailed gross necropsy of all cavities and organs, including the liver, at the studies’ conclusion. Within the SEND data standard, these histopathology results are submitted via the microscopic findings (MI) domain. Liver histopathology findings are the most frequent record in the MI domain with 261,688 of 8,486,604describing liver histopathology findings. For the 39,969 rats in the training set, there were 284 unique findings recorded by pathologists with the most common findings listed in Table 1. Unsurprisingly, the most common findings are normal or benign. However, certain findings could potentially indicate higher-order liver disease phenotypes. For example, cholestasis is a liver disease characterized by a disruption of the flow of bile from the liver to the gallbladder. Therefore, findings that mention “bile” or “biliary” (e.g., bile duct, hyperplasia) could be an indication of this disruption. Similarly, steatosis, also known as fatty liver disease, results in the formation of lipid-containing vacuoles creating cellular congestion and reduced cellular function. Indeed, many liver findings mention fatty change or vacuolation and may indicate steatosis. Lastly, while less frequent, terms like necrosis and fibrosis were present and may suggest more severe, general liver necrosis. In order to automatically classify these findings that we designed a set of regular expressions relating to each of these liver disease phenotypes were developed and listed in Table 2.

|  |  |
| --- | --- |
| **Table 1.** Frequency of the top 25 rat liver histopathology findings for 39,969 rats. | |
| **Finding** | **Frequency** |
| NORMAL | 7518 |
| UNREMARKABLE | 6234 |
| INFILTRATE | 3824 |
| INFILTRATE, MONONUCLEAR CELL | 3725 |
| HYPERTROPHY | 1368 |
| VACUOLATION | 1040 |
| INFLAMMATORY CELL FOCI | 994 |
| VACUOLATION, HEPATOCYTE | 825 |
| INFILTRATE, INFLAMMATORY CELL | 692 |
| HYPERTROPHY, HEPATOCYTE, CENTRILOBULAR | 604 |
| EXTRAMEDULLARY HEMATOPOIESIS | 549 |
| FOCUS OF CELLULAR ALTERATION | 489 |
| NECROSIS | 472 |
| PIGMENT | 425 |
| HYPERPLASIA, BILE DUCT | 352 |
| VACUOLATION, HEPATOCYTE, DIFFUSE | 345 |
| HEMATOPOIESIS, EXTRAMEDULLARY | 335 |
| FATTY CHANGE | 325 |
| MICROGRANULOMA | 313 |
| FOCUS OF CELLULAR ALTERATION, BASOPHILIC | 303 |
| INFILTRATION, MONONUCLEAR CELL | 280 |
| INFILTRATE MONONUCLEAR | 271 |
| VACUOLATION, HEPATOCYTE, PERIPORTAL | 266 |
| INFILTRATE, LYMPHOCYTES/MACROPHAGES | 219 |
| CYTOPLASMIC ALTERATION | 214 |

|  |  |
| --- | --- |
| **Table 2.** Regular expressions used to capture SEND findings indicative of liver disease. | |
| **Disease** | **Regular Expression** |
| Cholestasis | chol(e|o|a)|bil(i|e) |
| Steatosis | fat|lipid|vacuol|acc|steat|congest and NOT decreas|lower |
| Necrosis | necros|fibros|degen|atroph|apop|deplet |

Throughout the 39,969 rats steatosis was the most prevalent liver disease phenotype with 3,753 rats. This was followed by necrosis with 1,159 and finally cholestasis with 215. Necrosis frequently co-occurred with steatosis and cholestasis. Only 127 rats had each disease simultaneously.

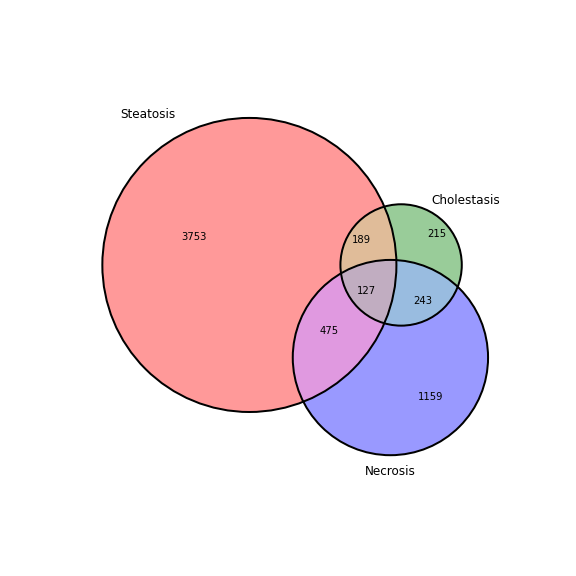


Figure 2. Venn diagram showing incidence of liver disease phenotypes across 39,969 rats provided within SEND.

## Clinical Chemistry Overview

Prior to animal sacrifice, OECD guidelines recommend blood samples to be taken and tested for concentrations of biomolecules spanning two broad groups: hematology (containing hemoglobin, erythrocytes, reticulocytes leucocytes, platelets, etc.) and general biochemistry tests (sodium potassium, glucose, cholesterols, urea, creatinine and liver enzymes such as alanine aminotransferase, bile salts etc.). Despite these guidelines, testing protocols vary greatly, in terms of number of blood samples taken, test units reported, the breadth of testing and number of subjects tested and are generally at the discretion of the study sponsor and scope of the study. Nonetheless, the current submitted clinical chemistry tests were seemingly plentiful enough to explore their relationship to various rat liver disease phenotypes.

There were 784 unique clinical chemistry lab tests submitted within the study data for the 39,969 rats in the training set and were sampled from a variety of different bodily fluids, including blood plasma or serum, urine, whole blood, etc. However, the tests were not uniformly submitted across animals with some tests being submitted more frequently than others. Of these 784 tests, there were only 42 tests for which the more than 40% of the 39,969 rats had responses and were included for analysis to liver disease phenotypes. In addition to these 42 tests, other animal meta data was included such as the animal body weight throughout the course of the study and the sex, resulting a total of 46 rat predictor variables for liver disease phenotypes. These variables are listed in Table 3.

**Table 3.** Clinical chemistry tests and rat meta data used as predictor variables of liver disease.

|  |  |  |
| --- | --- | --- |
| **Test** | **Tissue** | **Description** |
| ALB | SERUM | *Albumin* |
| ALBGLOB | SERUM | *Albumin/Globulin* |
| ALP | SERUM | *Alkaline Phosphatase* |
| ALT | SERUM | *Alanine Aminotransferase* |
| APTT | PLASMA | *Activated Partial Thromboplastin Time* |
| AST | SERUM | *Aspartate Aminotransferase* |
| BASO | WHOLE BLOOD | *Basophils* |
| BILI | SERUM | *Bilirubin* |
| CA | SERUM | *Calcium* |
| CHOL | SERUM | *Cholesterol* |
| CK | SERUM | *Creatine Kinase* |
| CL | SERUM | *Chloride* |
| CREAT | SERUM | *Creatinine* |
| EOS | WHOLE BLOOD | *Eosinophils* |
| FIBRINO | PLASMA | *Fibrinogen* |
| GGT | SERUM | *Gamma Glutamyl Transferase* |
| GLOBUL | SERUM | *Globulin* |
| GLUC | SERUM | *Glucose* |
| HCT | WHOLE BLOOD | *Hematocrit* |
| HGB | WHOLE BLOOD | *Hemoglobin* |
| K | SERUM | *Potassium* |
| LGUNSCE | WHOLE BLOOD | *Large Unstained Cells* |
| LYM | WHOLE BLOOD | *Lymphocytes* |
| MCH | WHOLE BLOOD | *Ery. Mean Corpuscular Hemoglobin* |
| MCHC | WHOLE BLOOD | *Ery. Mean Corpuscular HGB Concentration* |
| MCV | WHOLE BLOOD | *Ery. Mean Corpuscular Volume* |
| MONO | WHOLE BLOOD | *Monocytes* |
| NEUT | WHOLE BLOOD | *Neutrophils* |
| PH | URINE | *pH* |
| PHOS | SERUM | *Phosphate* |
| PLAT | WHOLE BLOOD | *Platelets* |
| PROT | SERUM | *Protein* |
| PROT | URINE | *Protein* |
| PT | PLASMA | *Prothrombin Time* |
| RBC | WHOLE BLOOD | *Erythrocytes* |
| RDW | WHOLE BLOOD | *Erythrocytes Distribution Width* |
| RETI | WHOLE BLOOD | *Reticulocytes* |
| SODIUM | SERUM | *Sodium* |
| SPGRAV | URINE | *Specific Gravity* |
| TRIG | SERUM | *Triglycerides* |
| UREAN | SERUM | *Urea Nitrogen* |
| VOLUME | URINE | *Volume* |
| WBC | WHOLE BLOOD | *Leukocytes* |
| BWDIFF\_NORM | NA | *Body Weight Diff. from Beg. to End of Study* |
| BWSLOPE\_NORM | NA | *Body Weight Regression Slope* |
| BWINTCEPT\_NORM | NA | *Body Weight Regression y-intercept* |
| SEX | NA | *Male/Female* |

## Biomarker Profiles for Liver Disease Phenotypes

Calculating z-scores can help delineate trends for subsets of data compared to the total population16. To explore the relationship between the predictor variables and liver disease phenotypes, z-score profiles were created by calculating z-scores for each predictor variable with respect to each liver disease phenotype (Figure 3). The profiles for different liver disease phenotypes were distinct but with notable commonalities. For example, liver enzymes such as alkaline phosphatase (ALP), alanine aminotransferase (ALT), and aspartate aminotransferase (AST) were significantly raised across each liver disease phenotype, consisted with their status as biomarkers where elevated levels are signs of underlying liver disease in humans12,13. Similarly, albumin (ALB), a protein produced primarily in the liver, as well overall protein blood levels (PROT) were reduced. This is consistent with impaired liver function as the liver is a site of protein synthesis for blood proteins17,18. Rats with liver necrosis had the most pronounced changes compared to normal as well as hematological changes indicative of anemia such as decreased red blood cells (RBC), hemoglobin (HGB) and hematocrit levels (HCT). Rats with liver cholestasis showed similar changes to liver necrosis, yet less drastic. One notable difference between necrosis and cholestasis rats were increases in urea nitrogen (UREA), creatinine (CREAT), and specific gravity of the urine (SPGRAV), all of which are well-established biomarkers of kidney function19,20. These observed increased could likely be secondary effects due to the buildup of bile and bile acids in blood and increased kidney workload induced by cholestasis. Indeed, cholestasis-induced renal failure has been observed in the clinical21. The z-score profile for steatosis was mostly characterized by much slighter changes hematological abnormalities consistent with a immunological response to inflammation, for example, elevated white blood cells (WBC), lymphocytes (LYM), and basophiles (BASO). This could be due to early stage of disease that was caught by our classification or that currently the submitted tests do not describe this liver disease phenotype particularly well.

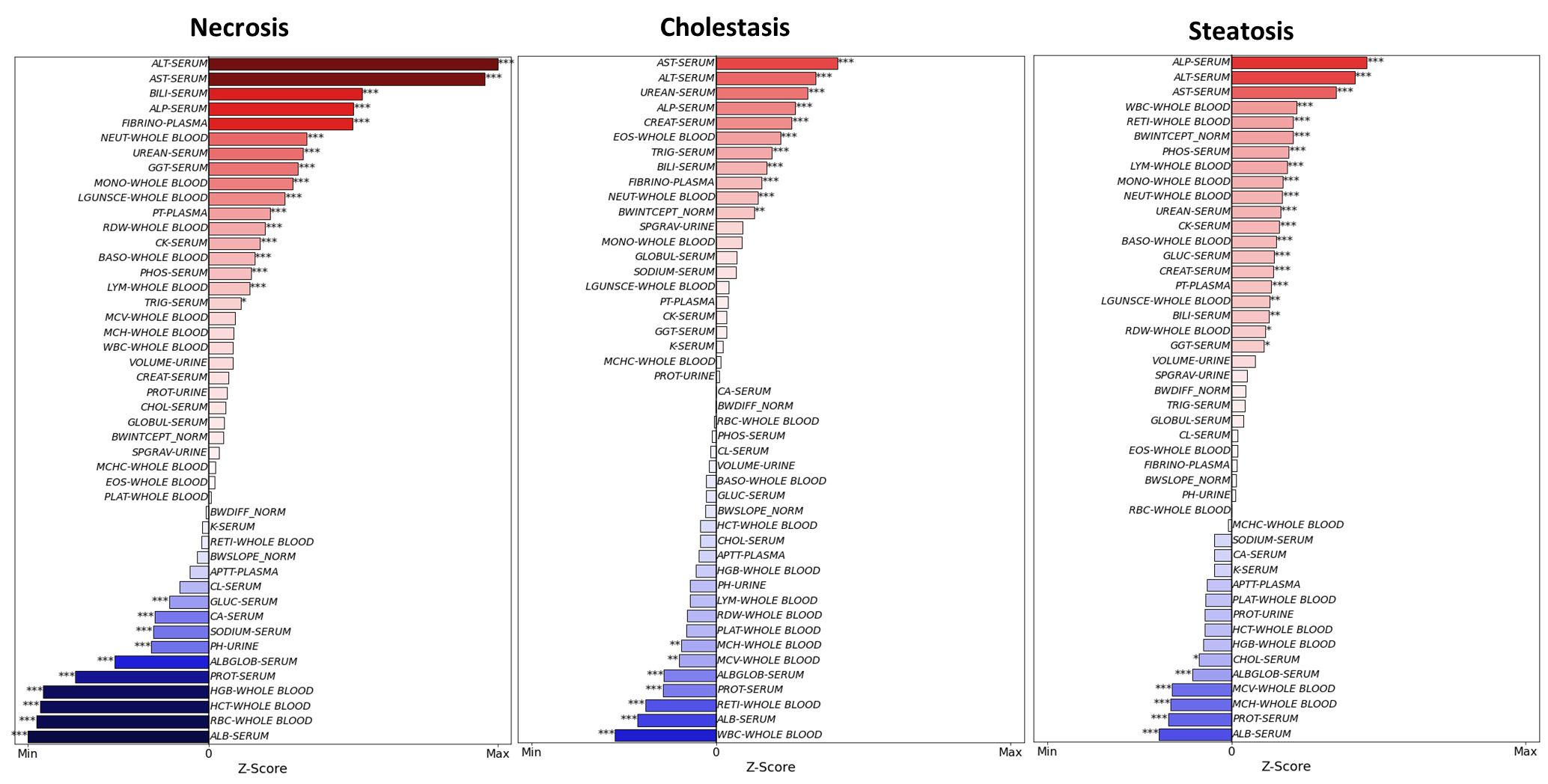


Figure 3. Z-score profiles of predictor variables for each liver disease phenotype.

## Machine Learning

The clinical chemistry tests along with animal meta data were used to train machine learning models to predict the probability of liver disease phenotypes in rats. The results from these random forest models are shown in Figure 4A. The models showed balanced accuracy from 58 to 66% with a universally lower recall (37-41%) than precision (63 – 82%) as seen in Figure 4A. The models could likely be improved to as the SEND database grows and includes more clinical chemistry tests with appreciable amounts of tests as well as the expansion into other relevant SEND domains. For example, animal food consumption is recorded in the Food and Water Consumption (FW) domain. Changes in eating behavior is noted in liver disease and can be a sign of underlying liver disease22. Additionally, rat behavior available in the Clinical Observation (CL) domain could help identify the eating habits, fatigue and/or other clinical behaviors associated with liver disease. Lastly, the incorporation of chemical structure information (e.g., molecular descriptors or chemical fingerprints) could improve model performance while also providing a link to the chemical features underpinning biological processes associated with liver disease phenotypes.

In addition to random forest model classifications, data was assessed by using the trained models to predict the probability of liver disease phenotypes in the training set, shown as histograms of probability predictions of liver disease phenotype positive and negative rats shown in Figure 4B. Cholestasis models appear to show the best separation, followed by necrosis, then finally steatosis (Figure 4B). The large peaks of probability scores, particularly those of necrosis and steatosis, are further evidence that more features could potentially increase predictions and improve discrimination.

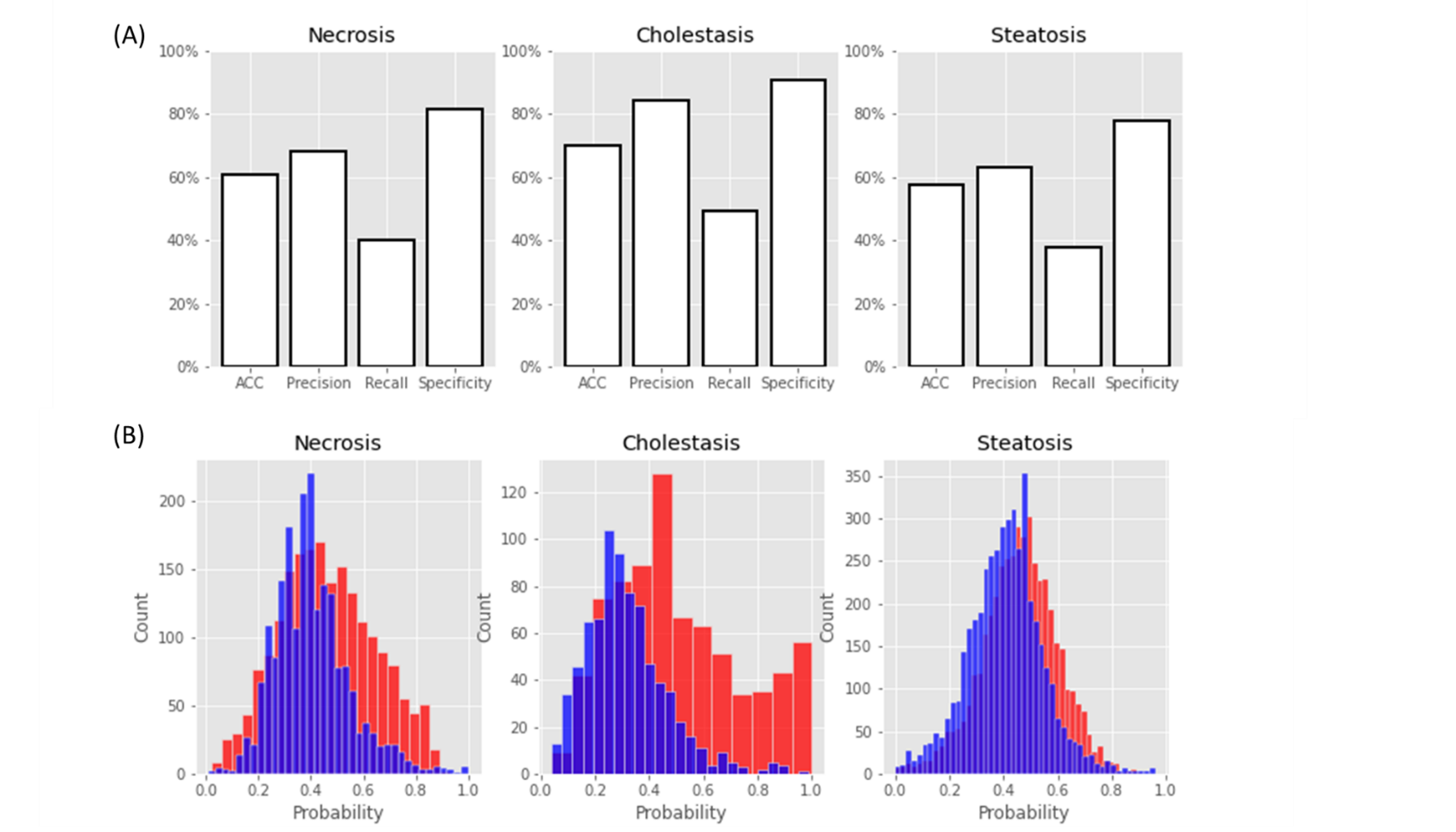


Figure 4. (A) Predictive performance of models for each liver disease phenotypes. (B) Histogram of probability of phenotype for disease negative (blue) and disease positive (red) rats.

The trained model forests were used to identify the most predictive features for each liver disease phenotype via the gini importance method. The values, averaged across all the models in the cross validation, are shown in Figure 5, with lower values (shown in red) contribute more to the model than those in blue. These profiles had similarities to the z-score profiles. For example, in necrosis and cholestasis, protein levels alanine aminotransferase (ALT) contributed to the model along with tests that suggest anaemia i.e., red blood cells (RBC) and white blood cells (WBC). The most contributory tests to the steatosis model consisted mostly of those describing changes in body weight (BWDIFF\_NORM and BWINTERCEPT\_NORM) likely due to the increase in liver size seen in fatty livers. Additionally, as seen in the z-score profiles, the specific gravity of urine was also seen.). This analysis confirms that random forest models are capable identifying relationships between clinical chemistry tests and different liver disease phenotypes better discriminating variables are still in need to improve model predictions.

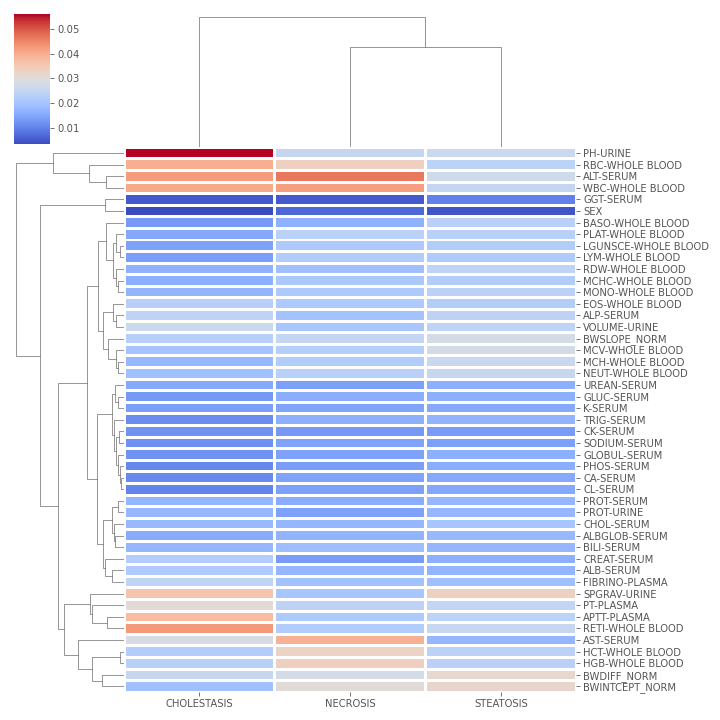


Figure 5. Gini importance score for random forest models. Left: Cholestasis, Center: Necrosis, Right: Steatosis. The color gradient is show as the maximum responses in red and the minimum in blue. The rows are sorted in descending order by the average across all models (last column on right).

# conclusions

In this work, applied informatics and machine learning methods to leverage the wealth of toxicity study data contained with SEND data submitted to the U.S. FDA. In this initial attempt of large-scale cross-study analysis applying these techniques, we characterize rat hepatotoxicity using their clinical chemistry tests and other animal metadata. By doing so, distinct patterns relating to different liver disease phenotypes can be seen, providing researchers with important information on the relationship between these variables’ hepatotoxicity. While these data are able to modestly classify liver disease in rats using machine learning, the wealth of information in SEND data to be provided in the future will likely improve these predictions. Lastly, this workflow can be extended to other animals (e.g., dog, monkey) and toxicity endpoints (e.g., nephrotoxicity, cardiotoxicity), which will help to predict complex toxicity data and understand drug-induced toxicity

1. **REFERENCES**

(1) Lapteva, L.; Pariser, A. R. Investigational New Drug Applications: A 1-Year Pilot Study on Rates and Reasons for Clinical Hold. *J. Investig. Med.* **2016**, *64* (2), 376–382. https://doi.org/10.1136/jim-2015-000010.

(2) Remmer, H. The Role of the Liver in Drug Metabolism. *Am. J. Med.* **1970**, *49* (5), 617–629. https://doi.org/10.1016/S0002-9343(70)80129-2.

(3) Hofmann, A. F. The Continuing Importance of Bile Acids in Liver and Intestinal Disease. *Arch. Intern. Med.* **1999**, *159* (22), 2647–2658. https://doi.org/10.1001/archinte.159.22.2647.

(4) Airhart, J.; Vidrich, A.; Khairallah, E. A. Compartmentation of Free Amino Acids for Protein Synthesis in Rat Liver. *Biochem. J.* **1974**, *140* (3), 539–545. https://doi.org/10.1042/bj1400539.

(5) Miller, L. L.; Bly, C. G.; Watson, M. L.; Bale, W. F. The Dominant Role of the Liver in Plasma Protein Synthesis; a Direct Study of the Isolated Perfused Rat Liver with the Aid of Lysine-Epsilon-C14. *J. Exp. Med.* **1951**, *94* (5), 431–453. https://doi.org/10.1084/jem.94.5.431.

(6) Zimmerman, H. J. *Hepatotoxicity: The Adverse Effects of Drugs and Other Chemicals on the Liver*; Lippincott Williams & Wilkins, 1999.

(7) Kanungo, S.; Wells, K.; Tribett, T.; El-Gharbawy, A. Glycogen Metabolism and Glycogen Storage Disorders. *Ann. Transl. Med.* **2018**, *6* (24). https://doi.org/10.21037/atm.2018.10.59.

(8) Onakpoya, I. J.; Heneghan, C. J.; Aronson, J. K. Post-Marketing Withdrawal of Analgesic Medications Because of Adverse Drug Reactions: A Systematic Review. *Expert Opin. Drug Saf.* **2018**, *17* (1), 63–72. https://doi.org/10.1080/14740338.2018.1398232.

(9) Onakpoya, I. J.; Heneghan, C. J.; Aronson, J. K. Post-Marketing Withdrawal of 462 Medicinal Products Because of Adverse Drug Reactions: A Systematic Review of the World Literature. *BMC Med.* **2016**, *14*. https://doi.org/10.1186/s12916-016-0553-2.

(10) Hunter, E.; Johnston, P.; Tanner, G.; Pinson, W.; Awad, J. Bromfenac(Duract)-Associated Hepatic Failure Requiring Liver Transplantation. *Am. J. Gastroenterol.* **1999**, *94* (8), 2299–2301. https://doi.org/10.1111/j.1572-0241.1999.01321.x.

(11) Kohlroser, J.; Mathai, J.; Reichheld, J.; Banner, B.; Bonkovsky, H. Hepatotoxicity Due To Troglitazone: Report of Two Cases and Review of Adverse Events Reported To The United States Food and Drug Administration. *Am. J. Gastroenterol.* **2000**, *95* (1), 272–276. https://doi.org/10.1111/j.1572-0241.2000.01707.x.

(12) McGill, M. R. The Past and Present of Serum Aminotransferases and the Future of Liver Injury Biomarkers. *EXCLI J.* **2016**, *15*, 817–828. https://doi.org/10.17179/excli2016-800.

(13) Meunier, L.; Larrey, D. Drug-Induced Liver Injury: Biomarkers, Requirements, Candidates, and Validation. *Front. Pharmacol.* **2019**, *10*. https://doi.org/10.3389/fphar.2019.01482.

(14) Hoofnagle, J. H.; Björnsson, E. S. Drug-Induced Liver Injury — Types and Phenotypes. *N. Engl. J. Med.* **2019**. https://doi.org/10.1056/NEJMra1816149.

(15) Carfagna, M. A.; Anderson, J.; Eley, C.; Fukushima, T.; Horvath, J.; Houser, W.; Larsen, B.; Page, T.; Russo, D.; Sloan, C.; Snyder, K.; Thompson, R.; Ullmann, G.; Whittaker, M. Leveraging the Value of CDISC SEND Data Sets for Cross-Study Analysis: Incidence of Microscopic Findings in Control Animals. *Chem. Res. Toxicol.* **2021**, *34* (2), 483–494. https://doi.org/10.1021/acs.chemrestox.0c00317.

(16) Cross, K. P.; Myatt, G.; Yang, C.; Fligner, M. A.; Verducci, J. S.; Blower, P. E. Finding Discriminating Structural Features by Reassembling Common Building Blocks. *J. Med. Chem.* **2003**, *46* (22), 4770–4775. https://doi.org/10.1021/jm0302703.

(17) Amacher, D. E.; Adler, R.; Herath, A.; Townsend, R. R. Use of Proteomic Methods to Identify Serum Biomarkers Associated with Rat Liver Toxicity or Hypertrophy. *Clin. Chem.* **2005**, *51* (10), 1796–1803. https://doi.org/10.1373/clinchem.2005.049908.

(18) Amacher, D. E. A Toxicologist’s Guide to Biomarkers of Hepatic Response. *Hum. Exp. Toxicol.* **2002**, *21* (5), 253–262. https://doi.org/10.1191/0960327102ht247oa.

(19) Lopez-Giacoman, S.; Madero, M. Biomarkers in Chronic Kidney Disease, from Kidney Function to Kidney Damage. *World J. Nephrol.* **2015**, *4* (1), 57–73. https://doi.org/10.5527/wjn.v4.i1.57.

(20) Zhang, Q.; Davis, K. J.; Hoffmann, D.; Vaidya, V. S.; Brown, R. P.; Goering, P. L. Urinary Biomarkers Track the Progression of Nephropathy in Hypertensive and Obese Rats. *Biomark. Med.* **2014**, *8* (1), 85–94. https://doi.org/10.2217/bmm.13.106.

(21) Holt, S.; Marley, R.; Fernando, B.; Harry, D.; Anand, R.; Goodier, D.; Moore, K. Acute Cholestasis-Induced Renal Failure: Effects of Antioxidants and Ligands for the Thromboxane A2 Receptor. *Kidney Int.* **1999**, *55* (1), 271–277. https://doi.org/10.1046/j.1523-1755.1999.00252.x.

(22) Deems, R. O.; Friedman, M. I.; Friedman, L. S.; Munoz, S. J.; Maddrey, W. C. Chemosensory Function, Food Preferences and Appetite in Human Liver Disease. *Appetite* **1993**, *20* (3), 209–216. https://doi.org/10.1006/appe.1993.1021.

(23) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45* (1), 5–32. https://doi.org/10.1023/A:1010933404324.