How to Cite this article: Joey Mercier, Thomas Ferguson, and Navdeep Tangri, A machine learning model to predict diuretic resistance, *Kidney360*, Publish Ahead of Print, 2022, 10.34067/KID.0005562022

**Article Type:** Original Investigation

**A machine learning model to predict diuretic resistance**

**DOI:** 10.34067/KID.0005562022

Joey Mercier, Thomas Ferguson, and Navdeep Tangri

**Key Points:**
- Our machine learning algorithm was able to quickly predict those at risk for diuretic resistance using common variables
- Our machine learning algorithm could easily be implemented within electrical medical software (or as an online calculator)
- Our study provides a proof of concept/framework for the development of future machine learning models

**Abstract:**

Background: Volume overload is a common complication encountered in hospitalized patients, and the mainstay of therapy is diuresis. Unfortunately, the diuretic response in some individuals is inadequate despite a typical dose of loop diuretics, a phenomenon called diuretic resistance. An accurate prediction model that predicts diuretic resistance using pre-dosing variables could inform the right diuretic dose for a prospective patient. Methods: Two large, deidentified, publicly available and independent ICU databases from the United States were used - The Medical Information Mart for Intensive Care III (MIMIC) and the Philips eICU databases. Loop diuretic resistance was defined as less than 1400cc of urine per 40mg of diuretic dose in 24hrs. Using 24-hour windows throughout admission, commonly accessible were obtained and incorporated into the model. Data imputation was performed using a highly accurate Machine Learning method. Using XGBoost, several models were created using train and test datasets from the eICU database. These were then combined into an ensemble model optimized for increased specificity and then externally validated on the MIMIC database. Results: The final ensemble model was composed of four separate models, each using 21 commonly available variables. The ensemble model outperformed individual models during validation. Higher serum creatinine, lower systolic blood pressure, lower serum chloride, higher age, and female sex were the most important predictors of diuretic resistance (in that order). The specificity of the model on external validation was 92% yielding a positive likelihood ratio of 3.46 while maintaining overall discrimination (C-statistic 0.69). Conclusions: A diuretic resistance prediction model was created using machine learning and was externally validated in ICU populations. The model is easy to use, would provide actionable information at the bedside, and would be ready for implementation in existing electronic medical records. This study also provides a framework for the development of future Machine Learning models.

**Disclosures:** T. Ferguson reports the following: Consultancy: Quanta Dialysis Technologies; Clinpredic Ltd.; Strategic Health Resources (Tricida Inc, Protagonist Therapeutics); Ownership Interest: Klinrisk Inc.; Palpate Health Ltd.; and Honoraria: Baxter Canada. N. Tangri reports the following: Consultancy: Tricida Inc., PulseData Inc, Mesentech Inc., Renibus, Marizyme; Ownership Interest: Tricida Inc., PulseData Inc, Mesentech Inc., Clinpredict Ltd, Renibus, Marizyme, Klinrisk, Quanta; Research Funding: Astra Zeneca Inc., Tricida Inc, Janssen, Otsuka, BI-Lilly, Bayer; Honoraria: Otsuka Pharmaceuticals, Astra Zeneca Inc., BI-Lilly, Janssen, Pfizer, Bayer; Patents or Royalties: Marizyme, Klinrisk; Advisory or Leadership Role: Tricida Inc., Clinpredict, Klinrisk; and Other Interests or Relationships: National Kidney Foundation; Founder - Klinrisk, Clinpredict. J. Mercier has nothing to disclose.

**Funding:** None.

**Author Contributions:** Joey Mercier: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Software; Validation; Visualization; Writing - original draft. Thomas Ferguson: Methodology; Supervision; Writing - review and editing. Navdeep Tangri: Conceptualization; Methodology; Supervision; Validation; Writing - review and editing.
Data Sharing Statement: Partial restrictions to the data and/or materials apply (please include a detailed explanation): The data comes from two MIT PhysioNet databases, MIMIC-III (https://physionet.org/content/mimiciii/1.4/) and eICU (https://physionet.org/content/eicu-crd/2.0/) which are both publicly available databases provided the following three conditions be met: 1) be a credentialed user 2) completed required training: CITI Data or Specimens Only Research 3) sign the data use agreement for the project With regards to (3), this explicitly prohibits sharing of the data with anyone that does not have access. Thankfully, access is quite easy and straightforward to obtain, and all the raw data is available on Google Bigquery (“the cloud”) or for download (to be assembled with SQL).

Clinical Trials Registration:

Registration Number:

Registration Date:

The information on this cover page is based on the most recent submission data from the authors. It may vary from the final published article. Any fields remaining blank are not applicable for this manuscript.
Machine learning to predict diuretic resistance

**A machine learning model to predict diuretic resistance**

Joey A Mercier

Thomas W Ferguson

Navdeep Tangri

1 Department of Internal Medicine, University of Manitoba, Winnipeg, MB, Canada

2 Seven Oaks Hospital Chronic Disease Innovation Centre, Seven Oaks General Hospital, Winnipeg, MB, Canada

Corresponding author:

Joey Mercier MD

Max Rady College of Medicine, University of Manitoba

750 Bannatyne Ave, Winnipeg, Manitoba, R3E 0W2

T: 1-613-807-0599

E: jomercier@nosm.ca; mercier1@myumanitoba.ca
Key Points

- Our machine learning algorithm was able to quickly predict those at risk for diuretic resistance using common variables.
- Our machine learning algorithm could easily be implemented within electrical medical software (or as an online calculator).
- Our study provides a proof of concept/framework for the development of future machine learning models.

Abstract

Background

Volume overload is a common complication encountered in hospitalized patients, and the mainstay of therapy is diuresis. Unfortunately, the diuretic response in some individuals is inadequate despite a typical dose of loop diuretics, a phenomenon called diuretic resistance. An accurate prediction model that predicts diuretic resistance using pre-dosing variables could inform the right diuretic dose for a prospective patient.

Methods

Two large, deidentified, publicly available and independent ICU databases from the United States were used - The Medical Information Mart for Intensive Case III (MIMIC) and the Philips eICU databases. Loop diuretic resistance was defined as less than 1400cc of urine per 40mg of diuretic dose in 24hrs. Using 24-hour windows throughout admission, commonly accessible were obtained and incorporated into the model. Data imputation was performed using a highly accurate Machine Learning method. Using XGBoost, several models were created using train and test datasets from the eICU database. These were then combined into an
Machine learning to predict diuretic resistance

ensemble model optimized for increased specificity and then externally validated on the MIMIC database.

Results

The final ensemble model was composed of four separate models, each using 21 commonly available variables. The ensemble model outperformed individual models during validation. Higher serum creatinine, lower systolic blood pressure, lower serum chloride, higher age, and female sex were the most important predictors of diuretic resistance (in that order). The specificity of the model on external validation was 92% yielding a positive likelihood ratio of 3.46 while maintaining overall discrimination (C-statistic 0.69).

Conclusions

A diuretic resistance prediction model was created using machine learning and was externally validated in ICU populations. The model is easy to use, would provide actionable information at the bedside, and would be ready for implementation in existing electronic medical records. This study also provides a framework for the development of future Machine Learning models.

Introduction

Volume overload is a complication of cardiovascular, kidney, and liver disease and is frequently encountered in hospitalized patients, notably in the intensive care unit (ICU) setting. When untreated or undertreated, volume overload can lead to increased end-organ dysfunction (kidney failure and heart failure), prolonged ventilator and ICU stay as well as increased
Machine learning to predict diuretic resistance

mortality (1, 2). Loop diuretics, including furosemide, torsemide, and bumetanide are the mainstay of treatment in hospitalized patients.

Unfortunately, some patients have a poor response to loop diuretics, a phenomenon termed diuretic resistance (3, 4). Even though a formal quantitative definition for diuretic resistance is lacking, owing to the complexities and variations in diuretic regimens, it is typically described as an inadequate amount of decongestion/diuresis despite an adequate diuretic regimen (3). Diuretic resistance is also associated with increased mortality, and proposed algorithms for its management typically involve post-hoc decision-making after a trial of therapy, potentially delaying the onset of effective therapies (3, 5–7).

Early identification of patients at high risk of developing diuretic resistance creates a unique opportunity for intervention with potential positive prognostic implications. Patients at high risk for diuretic resistance can receive higher doses upfront, and therefore potentially prevent a delay in decongestion which may lead to earlier extubation or prevent the need for dialysis therapy for fluid overload.

A prediction model that could provide actionable information on patients identified as high risk of being diuretic resistant, with appropriate early escalation of therapy as necessary, can improve clinical decision making in this setting (3). As such, we sought to develop and externally validate a model to predict diuretic resistance in patients hospitalized in ICU settings.

Materials and Methods

Data Source and Extraction

The Medical Information Mart for Intensive Case III (MIMIC-III) and the Philips eICU databases are two large, deidentified, publicly available and independent ICU databases in the
Machine learning to predict diuretic resistance

United States (8, 9). MIMIC-III contains approximately 50,000 hospital admissions to critical care units between 2001 and 2012 at the Beth Israel Deaconess Medical Center in Boston, Massachusetts, whereas eICU contains approximately 200,000 admissions monitored by Philips Healthcare across the United States between 2014 and 2015. Commonly available data, including patient demographics, routine vitals and labs, certain drugs (including diuretic doses), and weight/urine output were extracted from these databases. Comorbidities within the databases were coded using International Classification of Diseases (ICD-9). These were translated into the Charlon Comorbidity Index (CCI) given its prognostic implications (10, 11).

Patients less than 16 years of age and with less than 24 hours of ICU data were excluded. Variables with >20% missing values were excluded and data imputation using a highly accurate 3-tier novel technique which incorporates supervised machine learning was employed on remaining variables for the eICU dataset (12). The data was partitioned into 24-hour windows from time of admission to the ICU, with each 24-hour period representing a separate datapoint. Laboratory values and vitals from the previous 24 hours were averaged while both diuretic doses and urine output from the current 24 hours were summed. Patients with less than 40mg IV furosemide (or equivalent) or no IV loop diuretic doses within the current 24-hour window were excluded from the analyses.

Given that data collection from both databases was passive and that all data was deidentified in compliance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, which constitutes nonhuman subject research, no institutional or ethical approvals were required for this study (8, 9).

Definitions
Machine learning to predict diuretic resistance

Loop diuretic IV equivalents was defined using a 40:20:1 ratio (in mg) for furosemide : torsemide : bumetanide, respectively, with a 1:2 IV to oral conversion for furosemide (bumetanide and torsemide kept a 1:1 ratio) (3). These equivalent doses are in accordance with the pharmacological profiles of the loop diuretics and consistent with previous studies (3). Diuretic resistance was quantitively defined using urine output per diuretic dose, the latter being defined as 40mg IV furosemide equivalent. The cut-off for diuretic resistance was set at ≤ 1400mL/dose in 24 hours. This cut-off was selected based on a sub-analysis of the ASCEND-HF trial by Ter Maaten et al. given its associated increase in adverse outcomes at 30 days (all-cause mortality or heart failure rehospitalization) (13). Another study with a similar cut off (1,700mL/dose) also showed increased mortality at 180 days (14).

Methods

Baseline characteristics for both the eICU and MIMIC-III databases were provided with descriptive statistics, and the Chi-square and Student’s t-test were used to determine whether any significant differences between the databases existed for dichotomous and continuous variables, respectively (all continuous variables had a normal distribution). Owing to its size, the eICU database was used to create train and test datasets, set at 80% and 20% of the database, respectively. Supervised machine learning using extreme gradient boosted trees (XGBoost), a scalable end-to-end tree boosting system, was performed on the train dataset using 5-fold cross-validation (15). This degree of cross-validation was selected to prevent overfitting on the training dataset. A correlation matrix was created, and final variables were chosen by feature importance score. Hyperparameter tuning for the XGBoost model was performed using Bayesian Optimization, a global optimization scheme with gaussian processes, to maximize a certain metric (e.g.: precision, recall, area under the receiver operated curve, F1 score, etc.) (16). The
Machine learning to predict diuretic resistance

final model was then applied on the eICU test dataset (internal testing) and model parameters such as Area Under the Receiver Operated Curve (AUROC), accuracy, Brier score, F1 score, sensitivity (Sn), specificity (Sp), positive/negative predictive values (PPV/NPV), and positive/negative likelihood ratios (LR+, LR-) were calculated. This process was repeated four times, each time optimizing a different model metric using Bayesian Optimization, yielding four separate models with different hyperparameters and algorithms. The models were interrogated using Shapley additive explanation (SHAP summary plot) and partial dependence plots, the former having been shown to be superior as a means of explaining Machine Learning models (further elaboration can also be found in the Supplementary Appendix) (17, 18). Finally, the models were combined into a final ensemble model using the geometric mean of their generated probabilities, given that ensemble models were shown to outperform individual models in other applications of Machine Learning (19). This ensemble model was then optimized for specificity (in order to decrease “alert fatigue”), and then applied to the MIMIC-III dataset as a means of external validation/external testing. The overall scheme is depicted in Figure 1. To determine whether the ensemble model performed best, the individual models as well as an ensemble model with the two best models were also applied to the MIMIC-III dataset and compared. Lastly, a sensitivity analysis was performed using the cut off of 1,700mL/dose.

Data mining was performed on Google Big Query (Google Cloud Computing, Google corporation) and all other analyses (including imputation) were performed with Python 3.7 using Google Colaboratory (Google corporation).

Results

Study Population
Machine learning to predict diuretic resistance

Our selection criteria and patient allocation, as well as final sample sizes per database is summarized in Figure 2, and patient baseline characteristics are summarized in Table 1. There were 30,868 and 12,671 unique days/windows that met our inclusion criteria in the eICU and MIMIC databases, respectively. Within the eICU cohort, 52% of those days were identified as diuretic resistant, whereas this was slightly less in the MIMIC cohort at 50%. There were 11,652 and 4,480 unique patients encompassing 13,001 and 5,059 separate admissions in the eICU and MIMIC databases, respectively. On average, patients had a median of two admissions (with the most being 46 admissions) and their length of stay ranged from 24 hours to over 130 days, with a median of 4 and 5 days for eICU and MIMIC databases, respectively. Except for sex, all baseline differences were statistically significant between both databases. Apart from comorbidities, the differences in demographics were felt to be largely clinically irrelevant and instead mostly represented small differences of large samples. Most of the population were male and Caucasian. The MIMIC-III population was more comorbid, and mortality was higher, which included in-hospital mortality, as well as mortality at 30 and at 90 days.

Machine Models

The initial model derivation included 54 variables (Supplemental Appendix). These were initially trimmed down using a correlation matrix and feature importance scores, and then further trimmed down with model experimentation. An example of feature importance scores (weight and gain), a SHAP summary plot (with detailed explanations for interpretation), and results for the derivation of the individual models are included in the Supplementary Appendix. With the exception of hemoglobin used in one of the models instead of hematocrit, each model consisted of the same 21 variables. These included demographics (age, female sex, and white race), comorbidities (congestive heart failure, diabetes mellitus, chronic kidney disease, and liver
Machine learning to predict diuretic resistance

disease), laboratory (chloride, potassium, calcium, bicarbonate, creatinine, glucose, hematocrit, platelet, white blood cell count), and vitals (heart rate, systolic blood pressure, respiratory rate, temperature, and oxygen tension as measured by pulse oximetry). It is important to note that all comorbidities included in the model were higher in the MIMIC-II cohort, however, other than a history of congestive heart failure, these variables all had the lowest feature importance scores and thus did not contribute much to the model. Interrogation of the model using a SHAP summary plot (Figure 3) revealed the most important variables for predicting diuretic resistance to be higher serum creatinine, lower systolic blood pressure, lower serum chloride, higher age, and female sex (in that order). The other variables were much less predictive. Using a marginal effect size of around 0.5 as a cut off, partial dependence plots of the four most important variables (Figure 4) revealed that a serum creatinine greater than 100, systolic blood pressure less than 130, serum chloride less than 107, and age greater than 65 all had predictive ability for diuretic resistance.

Final ensemble model characteristics are summarized in Figure 5 for the AUROC and in Table 2 for model performance metrics. The specificity of the model on external validation was 92% yielding a positive likelihood ratio of 3.46 while maintaining overall discrimination (C-statistic 0.69). Classification accuracy was 70% in train/test datasets, decreasing to 60% in the external validation model. This was mostly driven by false negatives (given the optimized specificity). The ensemble model outperformed individual models, as well as an ensemble of the two highest performing individual models (more information in the Supplementary Appendix).

Lastly, the sensitivity analysis using a 1,700mL/40mg dose showed consistent results, with results summarized in Figure 9 and Table 1 of the Supplementary Appendix.
Discussion

Machine learning algorithms are increasingly being used to help predict clinical outcomes, including sepsis and AKI, with implementation of sepsis algorithms associated with decreased length in hospital stay (20, 21, 22). Here we developed and externally validated a new machine learning algorithm to predict diuretic resistance in patients admitted to the ICU. Our final model can be optimized to be highly specific (>90 %), while preserving overall discrimination (C statistic 0.69) and can potentially lead to better decision making for diuretic doses in patients with volume overload.

Diuretic resistance is increasingly recognized as an independent predictor of poor outcomes, and several recent studies have examined risk factors associated with diuretic resistance. (6, 23–26). In a study performed as a post hoc analysis of the PROTECT randomized trial, investigators defined diuretic response as a change in baseline weight at day four per 40mg furosemide equivalent dose based on the total dose of furosemide from days 1 through 3 (6). A poor diuretic response independently predicted mortality and rehospitalization, which was subsequently confirmed in an analysis of the RELAX-AHF trial and in another prospective study in the ICU setting (5, 14, 23–25). More recently, investigators used a natriuretic response prediction equation based on a spot 2h urine sodium post diuretic administration and found for a strong association total 6h natriuresis (26). Unfortunately, none of these previous studies integrated these factors in a risk prediction model that can be used prospectively by clinicians at the bedside. Interestingly, most of these previous studies found older age, female sex, lower blood pressure (mostly systolic), higher baseline creatinine, higher urea, and occasionally lower potassium/chloride as factors commonly associated with diuretic resistance, which was similar to our findings.
With respects to the four most important variables identified in our model, there are well-established mechanistic insights that help explain their relevance physiologically. For creatinine, it is well known that chronic kidney disease predicts poor diuretic response, which is largely attributable to a decrease in glomerular filtration, accumulation of uremic anions, and proteinuria (3, 4). Hypochloremia has also been implicated in both diuretic resistance and increased mortality in heart failure (27–29). Lower systolic pressure is associated with lower perfusion pressures to the kidney and typically a marker for more severe disease. Lastly, age is a well-established predictor of poor outcomes given that older individuals tend to be more comorbid, frailer, and typically have less physiologic reserve. There are also the effects of aging on drug pharmacokinetics (30).

It is important to note that we chose to model diuretic resistance as defined by a urine output of ≤ 1,400 mL per 40mg IV furosemide dose (or equivalent) for two reasons. The first reason is its association with increased mortality and increased rehospitalizations in two separate independent trials (13, 14). The second reason stems from the inaccuracies and inconsistencies between day to day the weights from both MIMIC-III and eICU databases, yielding any definition with weight unreliable for our purposes. While we do acknowledge that daily weights are typically better at capturing actual fluid balance, a previous analysis showed no significant differences between using weights and urine output in terms of defining diuretic resistance and its associated outcomes (13). Furthermore, the reliability of urine output charting in the ICU is quite high given the high prevalence of catheterized patients as well as the low nursing to patient ratios. Lastly, the quantitative definition of diuretic resistance was extracted from a study of patients with heart failure, and there are well-described differences that exist in decompensated heart failure, such as shifting of the diuretic response curve (3). To assess to relevance of this, a
subset analysis including only patients with a history of heart failure was performed and it did not improve model performance and discriminability.

Clinically, as an example, physicians can identify patients at high risk of diuretic resistance and consider a starting dose of 80mg IV furosemide instead of 40mg, decide on early sequential blockade, and perhaps even consider earlier nephrology consult and possible early CRRT if resistance is confirmed. The idea is that earlier identification and treatment would lessen the time in an overloaded state and its associated complications, such as local ischemia, and could hopefully result in a reduced length of stay within the ICU. This could have significant implications from resource utilization and economic perspectives. Additionally, combining our prediction algorithm with other tools such as the natriuretic response prediction equation from a 2h spot urine sodium could potentially greatly improve in-hospital diuresis. From a research perspective, a randomized control trial on the impacts of an algorithm guided strategy on effectiveness of diuresis and length of ICU stay should be performed.

The machine learning ensemble model described in this paper had reasonable performance for test and validate datasets. Interestingly, sensitivity was highest in the validation cohort, at the cost of specificity. Given the interest in deploying this model to electronic medical charting software, high specificity was a prespecified requirement of the model to decrease “alert fatigue”. As such, we determined that an alert threshold corresponding to at least 90% specificity would be reasonable for future implementation studies. We believe that this threshold would balance the correct identification of diuretic resistant individuals against the potential risks of escalating therapy too quickly.

Strengths
Machine learning to predict diuretic resistance

This study has other notable strengths. The databases included provide a multicentre perspective that spans nearly two decades, and although the original sample size may have been truncated due to our inclusion criteria, the number of actual samples generated from 24-hour windows provided a large database for analysis. This captured the real-world heterogeneity of diuretic resistance that exists clinically. The model also uses variables that are easily accessible and has a built-in algorithm to handle missing values. Lastly, although the model generates probabilities, it ultimately produces a binary output to simplify its use at the bedside, either through integration on electronic medical software, or manually via an online calculator, as an example. From a software perspective, the use of Python ensures portability, compatibility, and easy integration with any existing programming platforms. This paper could also serve as a proof of concept/framework for the development of other prediction tools, or simply for the use of future Machine Learning projects/prediction models.

Limitations

This study has several limitations. To begin, the retrospective nature brings about all its inherent limitations, and, although externally validated, the lack of prospective validation does limit potential applications. It is also unclear to what effect predicting loop diuretic resistance early will actually affect patient outcomes (e.g.: decreased mortality, decreased length of stay, etc.). The derivation and validation were performed on an ICU population, so generalizability to medicine wards or the ED remains unknown. There was a large amount of missing/incomplete data that would have likely been relevant and may have improved model performance owing to the proposed pathophysiologic mechanisms underlying loop diuretic resistance (3). These include, but are not limited to serum albumin, urine studies (electrolytes, creatinine, albumin, protein), BNP, troponin, lactate, blood pH and so on. There was also incomplete data on dialysis
Machine learning to predict diuretic resistance

and so a complete analysis for this was impossible, yet likely relevant. Lastly, each 24-hour window was treated as a separate independent datapoint, whereas the median length of stay per patient was four days in the eICU cohort and six days in the MIMIC-III cohort (not including repeated admissions).
Machine learning to predict diuretic resistance

**Disclosures**

T. Ferguson reports the following: Consultancy: Quanta Dialysis Technologies; Clinpredic Ltd.t; Strategic Health Resources (Tricida Inc, Protagonist Therapeutics); Ownership Interest: Klinrisk Inc.; Palpate Health Ltd.; and Honoraria: Baxter Canada. N. Tangri reports the following: Consultancy: Tricida Inc., PulseData Inc, Mesentech Inc., Renibus, Marizyme; Ownership Interest: Tricida Inc., PulseData Inc, Mesentech Inc., Clinpredict Ltd, Renibus, Marizyme, Klinrisk, Quanta; Research Funding: Astra Zeneca Inc., Tricida Inc, Janssen, Otsuka, BI-Lilly, Bayer; Honoraria: Otsuka Pharmaceuticals, Astra Zeneca Inc., BI-Lilly, Janssen, Pfizer, Bayer; Patents or Royalties: Marizyme, Klinrisk; Advisory or Leadership Role: Tricida Inc., Clinpredict, Klinrisk; and Other Interests or Relationships: National Kidney Foundation; Founder - Klinrisk, Clinpredict. J. Mercier has nothing to disclose.

**Funding**

None.

**Author Contributions**

Joey Mercier: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Software; Validation; Visualization; Writing - original draft.
Navdeep Tangri: Conceptualization; Methodology; Supervision; Validation; Writing - review and editing.
Thomas Ferguson: Methodology; Supervision; Writing - review and editing.

**Data Sharing Statement**

Partial restrictions to the data and/or materials apply: The data comes from two MIT PhysioNet databases, MIMIC-III (https://physionet.org/content/mimiciii/1.4/) and eICU (https://physionet.org/content/eicu-crd/2.0/) which are both publicly available databases provided the following three conditions be met: 1) be a credentialed user 2) completed required training: CITI Data or Specimens Only Research 3) sign the data use agreement for the project. With regards to (3), this explicitly prohibits sharing of the data with anyone that does not have access. Thankfully, access is quite easy and straightforward to obtain, and all the raw data is available on Google Bigquery ("the cloud") or for download (to be assembled with SQL).
Supplementary Materials

Initial Variables

Shapley additive explanation (SHAP) values

Initial Variable Selection

Supplemental Figure 1: Feature Importance Plot – Weight
Supplemental Figure 2: Feature Importance Plot – Gain
Supplemental Figure 3: Feature Importance Plot – SHAP summary plot

Individual Model Characteristics (eICU)

Model 1 Train and Test Metrics
Supplemental Figure 4: Model 1 ROC
Model 2 Train and Test Metrics
Supplemental Figure 5: Model 2 ROC
Model 3 Train and Test Metrics
Supplemental Figure 6: Model 3 ROC
Model 4 Train and Test Metrics
Supplemental Figure 7: Model 4 ROC

Validation/External Testing

Model 1 Validation (MIMIC)
Model 4 Validation (MIMIC)
Model 1 + 4 Ensemble Validation (MIMIC)
Supplemental Figure 8: Models 1 + 4 Ensemble Model Validation ROC (eICU and MIMIC)

Sensitivity Analysis

Supplemental Figure 9: AUC (1,700cc/40mg model)
Supplemental Table 1: Comparison of Ensemble Model Performance Metrics (1,700cc/40mg model)

Supplemental References
Machine learning to predict diuretic resistance

References


Machine learning to predict diuretic resistance


Machine learning to predict diuretic resistance


Machine learning to predict diuretic resistance


Machine learning to predict diuretic resistance

Tables

Table 1: Baseline Characteristics of Both eICU and MIMIC-III Databases

<table>
<thead>
<tr>
<th></th>
<th>eICU</th>
<th>MIMIC</th>
<th>differences&lt;sup&gt;a&lt;/sup&gt;, p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Patients, n</td>
<td>11,652</td>
<td>4,480</td>
<td></td>
</tr>
<tr>
<td>Mean age (SD), y</td>
<td>67.4 (14.3)</td>
<td>69.0 (14.4)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (%), n</td>
<td>6,548 (56.2)</td>
<td>2,463 (55.0)</td>
<td>0.16</td>
</tr>
<tr>
<td>Female (%), n</td>
<td>5,104 (43.8)</td>
<td>2,017 (45.0)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian (%), n</td>
<td>9,206 (79.0)</td>
<td>3,366 (75.1)</td>
<td>&lt; 0.001&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Black (%), n</td>
<td>1,234 (10.6)</td>
<td>325 (7.3)</td>
<td></td>
</tr>
<tr>
<td>Asian (%), n</td>
<td>147 (1.3)</td>
<td>79 (1.8)</td>
<td></td>
</tr>
<tr>
<td>Hispanic (%), n</td>
<td>296 (2.5)</td>
<td>137 (3.1)</td>
<td></td>
</tr>
<tr>
<td>Indigenous (%), n</td>
<td>92 (0.8)</td>
<td>3 (0.07)</td>
<td></td>
</tr>
<tr>
<td>Other/Unknown (%), n</td>
<td>677 (5.8)</td>
<td>570 (12.7)</td>
<td></td>
</tr>
<tr>
<td>Comorbidities&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congestive heart failure (%), n</td>
<td>3,197 (27.4)</td>
<td>2,305 (51.5)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Renal failure (%), n</td>
<td>1,695 (14.5)</td>
<td>1,149 (25.6)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Liver disease (%), n</td>
<td>363 (3.1)</td>
<td>634 (14.2)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Diabetes mellitus (%), n</td>
<td>1,998 (17.1)</td>
<td>1,282 (28.6)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Cancer&lt;sup&gt;d&lt;/sup&gt; (%), n</td>
<td>1,758 (15.1)</td>
<td>355 (7.9)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Peripheral vascular disease (%), n</td>
<td>3,540 (30.4)</td>
<td>656 (14.6)</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>
Machine learning to predict diuretic resistance

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group 1</th>
<th>Group 2</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other (%)</td>
<td>4,381 (37.6)</td>
<td>3,827 (85.4)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Total Admissions, n</td>
<td>13,001</td>
<td>5,059</td>
<td></td>
</tr>
<tr>
<td>Median admissions per patient (IQR; range), n</td>
<td>2 (2; 1 - 46)</td>
<td>2 (2; 1 - 31)</td>
<td></td>
</tr>
<tr>
<td>Median length of stay per admission (IQR; range), h</td>
<td>98 (130; 24 - 3170)</td>
<td>140 (218; 25 - 2804)</td>
<td></td>
</tr>
<tr>
<td>Days/windows with diuretic resistance</td>
<td>16,121 (52.2)</td>
<td>6,354 (50.1)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Days/windows with other classes of diuretics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thiazide (%)</td>
<td>166 (0.5)</td>
<td>531 (8.4)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Potassium Sparing (%)</td>
<td>856 (2.8)</td>
<td>407 (6.4)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Carbonic Anhydrase Inhibitor (%)</td>
<td>0</td>
<td>101 (1.6)</td>
<td>n/a</td>
</tr>
<tr>
<td>Mortality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In hospital (%)</td>
<td>1,560 (12.0)</td>
<td>834 (16.5)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>At 30 days (%)</td>
<td>1,489 (11.5)</td>
<td>901 (17.8)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>At 90 days (%)</td>
<td>1,556 (12.0)</td>
<td>1,351 (26.7)</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

SD: standard deviation; IQR: interquartile range

- a Chi-square for dichotomous variables, Student’s t-test for continuous variables
- b Grouped as white vs. non-white given small numbers of non-white patients
- c Coded as per Charlson Comorbidity Score (CCI)
- d Includes lymphoma, solid cancers, metastatic disease – i.e.: multiple CCI categories compiled into one
Machine learning to predict diuretic resistance

e Includes: myocardial infarction, cerebrovascular accident/transient ischemic attack, dementia, chronic obstructive pulmonary disease, connective tissue disease, peptic ulcer disease, hemiplegia, and AIDS

f Expressed as a percent of total days (30,868 for eICU and 6,354 for MIMIC-III); thiazide diuretics include hydrochlorothiazide, indapamide, and metolazone; potassium sparing diuretics include both renal tubular epithelial sodium channel blockers (amiloride and triamterene) and mineralocorticoid receptor antagonists (eplerenone and spironolactone); carbonic anhydrase inhibitor include acetazolamide.

g Not cumulative
Table 2: Comparison of Ensemble Model Performance Metrics

<table>
<thead>
<tr>
<th></th>
<th>eICU</th>
<th>MIMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Train</td>
<td>Test</td>
</tr>
<tr>
<td>AUROC</td>
<td>0.769</td>
<td>0.774</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.696</td>
<td>0.701</td>
</tr>
<tr>
<td>Brier Score</td>
<td>0.198</td>
<td>0.196</td>
</tr>
<tr>
<td>F1 score</td>
<td>0.714</td>
<td>0.720</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>72.9%</td>
<td>73.0%</td>
</tr>
<tr>
<td>Specificity</td>
<td>66.0%</td>
<td>66.9%</td>
</tr>
<tr>
<td>PPV</td>
<td>70.0%</td>
<td>71.0%</td>
</tr>
<tr>
<td>NPV</td>
<td>69.1%</td>
<td>69.0%</td>
</tr>
<tr>
<td>LR+</td>
<td>2.14</td>
<td>2.20</td>
</tr>
<tr>
<td>LR-</td>
<td>0.41</td>
<td>0.40</td>
</tr>
</tbody>
</table>
Machine learning to predict diuretic resistance

**Figure Legend**

Figure 1: Machine Learning Workflow Diagram

Caption: SMILES: xgboost MIsSing vaLues In timE Series; imputation method as described by Zhang et al. (12).

Figure 2: Consort Diagram

Figure 3: Model SHAP Summary Plot

Figure 4: Model Partial Dependence Plots for Creatinine, Systolic Blood Pressure, and Chloride

Figure 5: Area Under Receiver Operated Curve
Figure 2

Unique patients in MIMIC-III
n = 46,476

exclude: age < 16yrs or < 24hrs of ICU data
n = 13,508

exclude: no IV loop diuretic dose or < 40mg IV furosemide equivalents in 24hrs
n = 28,488

Unique patients in eICU
n = 139,367

exclude: age < 16yrs or < 24hrs of ICU data
n = 6,585

exclude: no IV loop diuretic dose or < 40mg IV furosemide equivalents in 24hrs
n = 121,130

Unique patients in study
n = 4,480

Total ICU admissions in study
n = 5,059

Total days in study
n = 12,671

With diuretic resistance
n = 6,354 (50.1%)

Total ICU admissions in study
n = 13,001

Total days in study
n = 30,868

With diuretic resistance
n = 16,121 (52.2%)
Figure 5

ROC Curve

- Train, AUC = 0.7686
- Test, AUC = 0.7740
- Validate, AUC = 0.7165
- Validate, AUC = 0.6902 (specificity optimized)
Initial Variables
The following are the 54 variables that were used in the analysis:

- Demographics: age, sex, ethnicity
- Comorbidities: congestive heart failure, cardiac arrhythmias, valvular disease, pulmonary and circulatory disorders, peripheral vascular disease, hypertension, paralysis, other neurological disorders, chronic pulmonary disease, diabetes (complicated and uncomplicated disease), hypothyroidism, renal failure, liver disease, peptic ulcer disease, HIV/AIDS, lymphoma, metastatic cancer, solid tumor, coagulopathy, obesity, weight loss, fluid and electrolyte abnormalities, blood loss/anemia, deficiency anemias, alcohol abuse, drug abuse, psychoses, depression
Machine learning model to predict diuretic resistance

- Laboratory: sodium, chloride, potassium, bicarbonate, creatinine, urea, glucose, anion gap, calcium, magnesium, phosphate, hemoglobin, hematocrit, platelet count, white blood cell count
- Vitals: heart rate, systolic blood pressure, diastolic blood pressure, mean arterial pressure, respiratory rate, temperature, oxygen tension

Shapley additive explanation (SHAP) values

Traditional regression models have coefficients that give insight into how each variable (or feature) tends to impact the outcome. On the other hand, machine learning models are typically quite complex “black box” models that provide little in the way of how certain inputs can predict an outcome. XGBoost, as well as other machine learning methods, do provide feature importance scores (such as weight, gain, etc.) which is a measure into each variable’s contribution towards the final prediction (1). Unfortunately, these are plagued by inconsistencies (2). SHAP values utilizes an approach from cooperative game theory that comes with desirable properties (3). In their paper, Nohara and colleagues explain and show how SHAP values are superior to traditional feature importance scores (4). A SHAP summary plot will include feature importance (by the amount a certain variable can alter the SHAP value) and it will also show how each individual values from these variables tend to affect the SHAP value. This last attribute of the SHAP summary plot is very similar to partial dependence plots, which are usually used to help explain how certain variables contribute to the outcome of a model. The advantage of also using the partial dependence plots is that they provide a visualization of how a variable impacts a model based on that variable’s actual value (instead of a SHAP value, which may be harder to interpret).

Initial Variable Selection

The plots below provide an idea on how variables were selected with feature importance. To elaborate on the inconsistencies of feature importance identified above, feature importance scores of weight and gain generated by XGBoost are included, followed by a SHAP summary plot. Weight refers to the number of times a particular feature occurs in the trees of the model. Gain refers to the improvement in accuracy brought by adding this feature to a branch of a tree. Dichotomous variables tend to have either large or low gain and typically have a low weight since they can only be used once per tree given their dichotomous nature. On the other hand, continuous variables tend to have larger weight, and gain is likely a better reflector of actual variable importance. The issue with using only gain is that our model has both dichotomous and continuous variables. SHAP summary plots help bridge this issue given its properties described above.

Supplemental Figure 1: Feature Importance Plot – Weight

This plot shows how dichotomous variables are all at the lower end of feature importance
Supplemental Figure 2: Feature Importance Plot – Gain
This plot shows how dichotomous variables tend to have higher gain despite having low weight. By comparing weight and gain (for continuous variable), we can also assume that the most important variables are probably creatinine, chloride, age, and systolic blood pressure. It is unclear how the dichotomous variables help the prediction, however.

Supplemental Figure 3: Feature Importance Plot – SHAP summary plot
This plot provides a better estimate of feature/variable importance. Feature position along the vertical axis indicates its importance (highest on the plot is the most important, in this case, s_Cr). Each individual value for each feature/variable is plotted as a separate point, with the horizontal axis indicating how much that particular point affects the SHAP value (i.e.: how much it tends to affect the outcome). In our model, a positive SHAP value indicates that it is more likely to predict diuretic resistance, while negative values tend to predict no resistance. As points accumulate, they spread along the vertical axis of that feature. The color for each point is a gradient based on that variable’s value. As an example, for creatinine (s_Cr), which is the most important feature in our model, the red points refer to higher values of creatinine, blue points indicate lower values, and purple indicate values in the middle. By observing color, we can tell that higher creatinine is predictive of diuretic resistance, as is lower systolic blood pressure.
Machine learning model to predict diuretic resistance
Individual Model Characteristics (eICU)
The following includes the performance metrics of each individual model used in the ensemble model.

Model 1 Train and Test Metrics
MODEL REPORT | train | test
Accuracy:   | 0.7121 | 0.6506
Precision Score: | 0.7195 | 0.6674
Recall (sensitivity): | 0.7538 | 0.6870
F1 Score: | 0.7362 | 0.6770
AUC Score: | 0.7893 | 0.7116

Supplemental Figure 4: Model 1 ROC

Model 2 Train and Test Metrics
MODEL REPORT | train | test
Accuracy:   | 0.6765 | 0.6523
Precision Score: | 0.6869 | 0.6700
Recall (sensitivity): | 0.7224 | 0.6850
F1 Score: | 0.7042 | 0.6774
AUC Score: | 0.7442 | 0.7067

Supplemental Figure 5: Model 2 ROC
Machine learning model to predict diuretic resistance

Model 3 Train and Test Metrics

<table>
<thead>
<tr>
<th>MODEL REPORT</th>
<th>train</th>
<th>test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.6717</td>
<td>0.6398</td>
</tr>
<tr>
<td>Precision Score</td>
<td>0.6819</td>
<td>0.6579</td>
</tr>
<tr>
<td>Recall (sensitivity)</td>
<td>0.7198</td>
<td>0.6757</td>
</tr>
<tr>
<td>F1 Score</td>
<td>0.7003</td>
<td>0.6667</td>
</tr>
<tr>
<td>AUC Score</td>
<td>0.7360</td>
<td>0.6916</td>
</tr>
</tbody>
</table>

Supplemental Figure 6: Model 3 ROC

Model 4 Train and Test Metrics

<table>
<thead>
<tr>
<th>MODEL REPORT</th>
<th>train</th>
<th>test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.7225</td>
<td>0.6568</td>
</tr>
<tr>
<td>Precision Score</td>
<td>0.7282</td>
<td>0.6717</td>
</tr>
<tr>
<td>Recall (sensitivity)</td>
<td>0.7650</td>
<td>0.6965</td>
</tr>
<tr>
<td>F1 Score</td>
<td>0.7461</td>
<td>0.6839</td>
</tr>
<tr>
<td>AUC Score</td>
<td>0.7976</td>
<td>0.7113</td>
</tr>
</tbody>
</table>

Supplemental Figure 7: Model 4 ROC
Validation/External Testing

From the results above we can see that Models 1 and 4 are clearly superior to the other two. One would wonder whether using those as individual models, or whether a combination of only those two models would yield a superior prediction. This was tested and the four-model ensemble model outperformed them all when it came to external validation (AUC 0.7165) – this may be due to more “smoothing”, or less susceptibility to noise. The results of these tests are included below. This would support the notion that ensemble models typically outperform single models (with the caveat that gradient boosted trees/XGBoost are themselves a form of ensemble model). Similar results of superior performance with ensemble models compared to individual models has been reported in other types of predictive modeling (5).

Model 1 Validation (MIMIC)

MODEL REPORT
Accuracy: 0.6402
Precision: 0.6110
Recall (sensitivity): 0.8185
F1 Score: 0.6997
AUC Score: 0.7065

Model 4 Validation (MIMIC)

MODEL REPORT
Accuracy: 0.6387
Precision: 0.6092
Recall (sensitivity): 0.8214
F1 Score: 0.6996
AUC Score: 0.7057

Model 1 + 4 Ensemble Validation (MIMIC)

MODEL REPORT - Validate
Accuracy: 0.6414
Precision: 0.6067
Recall (sensitivity): 0.8096
F1 Score: 0.6936
AUC Score: 0.7114
Brier: 0.2245

Characteristics, Validate
sens: 0.8096
spec: 0.4722
ppv: 0.6067
npv: 0.7114
LR+: 1.5339
LR-: 0.4033

MODEL REPORT - Validate (Spec optimized)
Accuracy: 0.6000
Precision: 0.7676
Recall (sensitivity): 0.2901
F1 Score: 0.4210
AUC Score: 0.6866
Brier: 0.2238
Machine learning model to predict diuretic resistance

Characteristics, Validate (Spec optimized)
sens: 0.2901
spec: 0.9117
ppv: 0.7676
npv: 0.5608
LR+: 3.2836
LR-: 0.7787

Supplemental Figure 8: Models 1 + 4 Ensemble Model Validation ROC (eICU and MIMIC)

Sensitivity Analysis
The analysis was run using a cut off of 1,700cc/40mg to determine whether this would impact model performance in any way. Although there are some differences, the results below indicate that they are quite small and clinically irrelevant.
Machine learning model to predict diuretic resistance

Supplemental Figure 9: AUC (1,700cc/40mg model)

Supplemental Table 1: Comparison of Ensemble Model Performance Metrics (1,700cc/40mg model)

<table>
<thead>
<tr>
<th></th>
<th>Train</th>
<th>Test</th>
<th>Validate</th>
<th>Validate (specificity optimized)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUROC</td>
<td>0.754</td>
<td>0.761</td>
<td>0.702</td>
<td>0.679</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.684</td>
<td>0.687</td>
<td>0.668</td>
<td>0.515</td>
</tr>
<tr>
<td>Brier Score</td>
<td>0.202</td>
<td>0.200</td>
<td>0.211</td>
<td>0.227</td>
</tr>
<tr>
<td>F1 Score</td>
<td>0.728</td>
<td>0.731</td>
<td>0.740</td>
<td>0.385</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>68.2%</td>
<td>68.4%</td>
<td>77.9%</td>
<td>25.0%</td>
</tr>
<tr>
<td>Specificity</td>
<td>68.6%</td>
<td>69.3%</td>
<td>50.0%</td>
<td>92.4%</td>
</tr>
<tr>
<td>PPV</td>
<td>78.1%</td>
<td>78.5%</td>
<td>70.5%</td>
<td>83.6%</td>
</tr>
<tr>
<td>NPV</td>
<td>56.8%</td>
<td>57.2%</td>
<td>59.2%</td>
<td>44.4%</td>
</tr>
<tr>
<td>LR+</td>
<td>2.17</td>
<td>2.23</td>
<td>1.54</td>
<td>3.30</td>
</tr>
<tr>
<td>LR-</td>
<td>0.46</td>
<td>0.46</td>
<td>0.45</td>
<td>0.81</td>
</tr>
</tbody>
</table>
Machine learning model to predict diuretic resistance

References


