

Deniz İlhan Topcu* and Nilüfer Bayraktar

Searching for the urine osmolality surrogate: an automated machine learning approach

<https://doi.org/10.1515/cclm-2022-0415>

Received April 26, 2022; accepted June 22, 2022;

published online July 4, 2022

Abstract

Objectives: Automated machine learning (AutoML) tools can help clinical laboratory professionals to develop machine learning models. The objective of this study was to develop a novel formula for the estimation of urine osmolality using an AutoML tool and to determine the efficiency of AutoML tools in a clinical laboratory setting.

Methods: Three hundred routine urinalysis samples were used for reference osmolality and urine clinical chemistry analysis. The H2O AutoML engine completed the machine learning development steps with minimum human intervention. Four feature groups were created, which include different urinalysis measurements according to the Boruta feature selection algorithm. Method comparison statistics including Spearman's correlation, Passing–Bablok regression analysis were performed, and Bland Altman plots were created to compare model predictions with the reference method. The minimum allowable bias (24.17%) from biological variation data was used as the limit of agreement.

Results: The AutoML engine developed a total of 183 ML models. Conductivity and specific gravity had the highest variable importance. Models that include conductivity, specific gravity, and other urinalysis parameters had the highest R^2 (0.70–0.83), and 70–84% of results were within the limit of agreement.

Conclusions: Combining urinary conductivity with other urinalysis parameters using validated machine learning models can yield a promising surrogate. Additionally, AutoML tools facilitate the machine learning development cycle and should be considered for developing ML models in clinical laboratories.

Keywords: automated machine learning; AutoML; conductivity; machine learning; urine osmolality.

Introduction

Urinary osmolality measures the concentration of osmotically active particles and is considered the gold standard in evaluating the renal urinary concentration capacity and hydration status. Sodium, chloride, potassium, and urea determine the urine osmolality [1].

It is commonly used in clinical practice in scenarios, such as assessing acute kidney injuries, chronic kidney disease polyuria, and hyponatremia [1]. In addition, there are some novel clinical applications, such as dysmorphic erythrocyte, obesity/insulin resistance, and diabetic nephropathy evaluation. Furthermore, there is a growing body of literature that recognizes the importance of population-based urine osmolality screening for several diseases [2–5].

The measurement of osmolality with a freezing point depression using an osmometer is the reference method in the clinical laboratory practice. However, determining osmolality using this method is not available at most institutions. Additionally, the method involves manual processing which makes this parameter less suited for high throughput determinations [1, 6]. Due to these drawbacks, various alternative methodologies have been proposed and developed for osmolality estimation, such as specific gravity (SG) or urine chemistry-based calculations, estimations from urine color, or conductivity-based calculations [5, 7, 8].

More recently, there has been a renewed interest in conductivity-based osmolality estimations, with the integration of the conductivity meters into automatic urine analyzers. This development provides the opportunity to measure osmolality rapidly and in a non-invasive manner. However, data from several studies suggests that using conductivity provides a limited accuracy for osmolality estimation. This is mainly because conductivity measurements are not sensitive to uncharged particles, such as glucose, urea, and contrast media [6, 9].

In the clinical laboratory field, the use of machine learning (ML) tools is becoming widespread for both evaluating patient results and effective laboratory management [10]. However, the development and utilization of

*Corresponding author: Deniz İlhan Topcu, MD and PhD, Department of Medical Biochemistry, Başkent University Faculty of Medicine, Ankara, Turkey, Phone: +90(532)4467779, E-mail: ditopcu@gmail.com. <https://orcid.org/0000-0002-1219-6368>

Nilüfer Bayraktar, Department of Medical Biochemistry, Başkent University Faculty of Medicine, Ankara, Turkey. <https://orcid.org/0000-0002-7886-3688>

these in healthcare settings include several challenges, such as accessing “clean” data and the need for data science knowledge and experience [10, 11]. To overcome the second problem, automated machine learning (AutoML) tools have emerged to automate certain steps of ML model development. AutoML tools offer benefits to users whose main domain is not data science, including automating the feature engineering, model building, and hyperparameter optimization steps, which all require experienced users [10, 12]. Despite the advantages of these tools and increased interest in ML studies, few studies have utilized AutoML tools in the clinical laboratory context.

The specific objective of this study was to develop a novel formula to estimate urine osmolality using cutting-edge AutoML tools to improve upon existing estimation methods. The second aim of the study is to demonstrate the utility of AutoML tools in a clinical laboratory setting. To achieve our objectives, we developed multiple ML models that combine urine conductivity measurements and different sets of traditional and extended urinalysis parameters. We compared the performance of the developed formulas and currently available estimation methods with the reference method.

Materials and methods

Study population

The study was approved by the Institutional Review Board of Baskent University (Project No: KA19/305) and conducted from September 2019 to February 2020 at the Baskent University Ankara Hospital. During this period, 300 spot urine samples were selected from patients who had already ordered urinalysis testing. The samples were chosen for their varying levels of osmolality, conductivity, glucose, SG, protein, and pH levels. To cover a wide range of osmolality results for ML model development patients were selected regardless of the underlying disease or clinical department. Left over urine samples were used for additional analyses. All measurements were completed within 1 h after the urine samples had been collected. A schematic view of the study is shown in Figure 1.

Reference osmolality measurement

Urine osmolality was measured by the freezing-point depression method using a Micro Osmometer 3,300 automatic osmometer (Advanced Instruments, MA, USA) as a reference method.

Urine osmolality calculation and clinical chemistry analyses

Urine osmolality calculation was performed using following equation:

$$\text{Urine Osmolality} = 2 \times (\text{Na}^+ + \text{K}^+) + \text{BUN}/2.8 + \text{Glucose}/18$$

Urine osmolality, mOsm/kg; Na, mmol/L; K, mmol/L; BUN, mg/dL; glucose, mg/dL.

The glucose, urea nitrogen, potassium, and sodium levels were measured in the collected urine samples using an Abbott Architect c analyzer (Abbott Diagnostics, IL, USA). The following methods were used for the clinical chemistry tests: enzymatic hexokinase method for glucose, urease method for urea, and ion-selective electrode method for potassium and sodium.

Urinalysis

All semi-quantitative urinalysis measurements were performed using a Sysmex UC-3500 analyzer (Sysmex Corporation, Kobe, Japan). The intensity of the test pad’s response color was measured using a CMOS sensor, which determines the amount of reflected light from the pad’s surface. MEDITAPE UC-11A urine strips which provide additional albumin and creatinine measurements were used in this study to develop various ML models. SG measurements were conducted using the refractometry measurement method with a Sysmex UC-3500 analyzer.

Urine cell analysis

The Sysmex UF-4000 recognizes, counts, and classifies cells by analyzing the forward-scatter light, side-scatter light, side-fluorescent light, and depolarized side-scattered light of stained particles. The principle is based on a 488 nm blue laser flow cytometry [13].

Urine conductivity and osmolality calculation

The cell analysis and conductivity measurements were conducted using Sysmex UF-4000 equipment (Sysmex Corporation, Kobe, Japan) which measures urinary conductivity using an integrated temperature-controlled microprocessor conductivity meter.

The manufacturer osmolality calculation was performed using the following formula:

$$\text{Urine Osmolality} = 34.294 \times \text{Conductivity}$$

Urine osmolality, mOsm/kg; conductivity, mS/cm

Analytical performance

The third-party internal quality control (IQC) materials (Technopath Clinical Diagnostics, Ballina, Ireland) were evaluated every 12 h at two levels for clinical chemistry analytes. Two levels of manufacturer IQC materials were used daily for the urinalysis. For the reference osmolality study, the two levels of IQC samples (target values: 290 mOsm/kg and 850 mOsm/kg) that were provided by the manufacturer (Advanced Instruments, MA, USA) were analyzed daily before the patient samples were measured. The bias of all analytes was assessed by Randox Quality Control’s (Country Antrim, United Kingdom) external quality (EQ) program. Performance characteristics, including precision and accuracy for the measurements during the study, are shown in Table 1.

Machine learning framework

Pre-process and data cleaning: There was no missing value for any sample. As some of the categorical measurements had less than five samples in one category, the following groups were merged to obtain balanced train-test sets and cross-validation partitions: 0/10 mg/dL

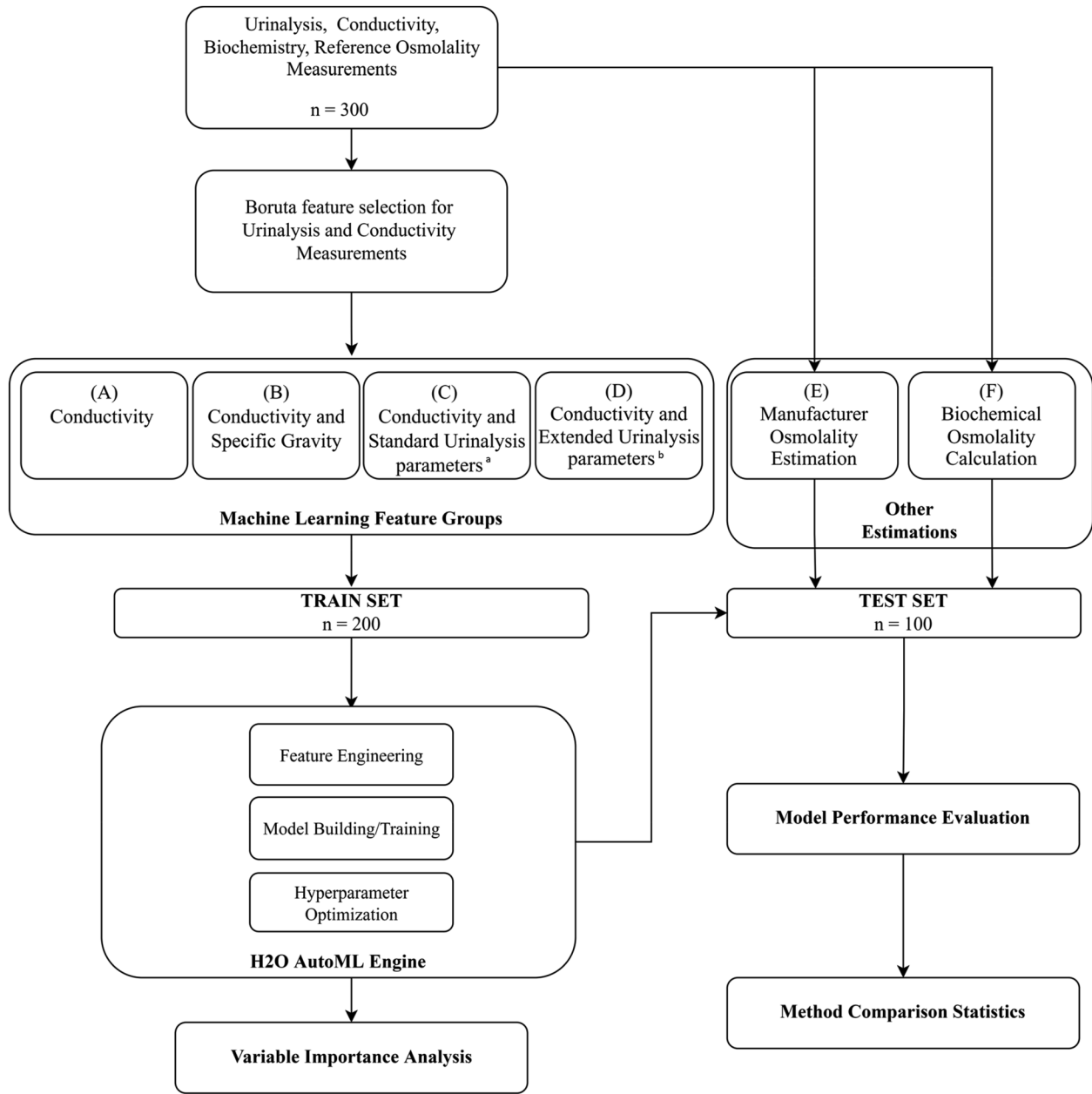


Figure 1: Study design.

^aGlucose, pH, protein, specific gravity. ^bAlbumin, creatinine, glucose, pH, protein, specific gravity.

and 80/100 mg/dL for albumin; 5/10/20 mg/dL, 40/50 mg/dL, and 200/300 mg/dL for creatinine; and 7.5/8.0/8.5/9 for pH.

Feature selection

The Boruta feature selection algorithm was used to identify important measurements for the ML model development. The Boruta feature selection algorithm created a corresponding shadow for each attribute, whose values were obtained by shuffling the original attributes’ values across properties. Finally, the importance was classified into the following three classes to identify the important features discard (red), speculative (blue), and keep (green) [14].

According to the results of the Boruta feature selection algorithm, four different feature groups that comprised various measurements were created, and the ML model development was performed for each group. The FGs were as follows: conductivity only (FG-A), conductivity and SG (FG-B), conductivity and standard urinalysis parameters (glucose, pH, protein, SG) (FG-C) and conductivity and extended parameters (FG-D), including albumin and creatinine tests in addition to standard urinalysis parameters (Figure 1). FG-A was created to determine the relationship between conductivity and measured osmolality and enable comparisons with the manufacturers’ calculations.

Generation of train and test sets: The data was split into train (n=200) and test sets (n=100). To obtain balanced sets, splitting was performed with stratification according to the reference osmolality results. The

Table 1: Analytical performance of the quantitative measurements for ML model development.

Test	Method	Analytical performance				
		IQC level 1		IQC level 2		
		Mean	CV ^a %	Mean	CV ^a %	Bias ^b %
BUN, mg/dL	Urine chemistry	365.9	2.03	968.4	2.09	2,88
Glucose, mg/dL		32.1	3.00	366.6	1.78	1,37
Potassium, mmol/L		15.03	1.15	59.47	0.78	2,50
Sodium, mmol/L		79.45	1.09	153.4	0.75	2.78
Conductivity, mS/cm	Urinalysis	8.90	6.90	36.30	1.10	n/a ^c
Specific gravity		1.009	0.10	1.023	0.10	n/a ^c
Osmolality, mosm/kg	Freezing-point depression	300	1.68	850	1.51	2,12

CV, variation of coefficient; IQC, internal quality control; IQR, interquartile range; n/a, not applicable. ^aVariation of coefficient was calculated as standard deviation/mean \times 100 and the calculations cover the study period. ^bCalculated as average absolute percentage deviation from six external quality samples during the study period. ^cQuantitative external quality assessment was not available. BUN, mmol/L = BUN, mg/dL \times 0.357; Glucose, mmol/L = Glucose, mg/dL \times 0.0555.

same test set was used for all models and manufacturer and biochemical calculation performance metric calculations (Figure 1).

A k-fold cross-validation (k=10) was used to develop the ML models and evaluate the train set performance. All partitions were created using stratification according to the reference osmolality levels, and the same partitions were used for all FGs. The seed function was used with a constant value to provide repeatability for splitting and model development.

Utilization of the AutoML tool: In this study, the H2O automated machine learning tool was used for ML model development [15]. H2O is an open source, distributed machine learning platform that supports multiple platforms, including the R programming language. The H2O AutoML function automatically performs the pipeline of ML models for a given dataset, including data pre-processing, feature engineering, model building hyperparameter optimization, and

model performance evaluation and explainability. The H2O AutoML provides various ML models that include H2O gradient boosting machine (GBM), default random forests (DRF) and extremely randomized tree (XRT), deep neural networks, and generalized linear model (GLM) supervised algorithms [15].

In our study, the H2O AutoML was used for feature engineering, model building, and hyperparameter optimization (Figure 1). An R script was developed to facilitate the recurring use of the AutoML function for different feature groups. The provided script comprises all ML development steps and follows the same numbering with the article.

The feature engineering steps that were performed by the AutoML function included the following: (1) all numerical values were standardized before the training phase for all algorithms, (2) enum encoding was used for GBM and DRF algorithms, (3) one hot internal encoding was used for GLM and deep-learning algorithms for categorical variables [15].

Table 2: Demographic features of patients and descriptive statistics of the quantitative measurements for ML model development.

	n (%)	Minimum	Maximum	Median (IQR)
Age, years				
	Female	18	72	58.5 (28)
	Male	18	73	57 (25)
Sex	Female	139 (46.3)		
	Male	161 (53.7)		
Patient type	Outpatient	240 (80)		
	Inpatient	60 (20)		
Emergency	No	280 (93.3)		
	Yes	20 (6.67)		
BUN, mg/dL		46	1960	478 (326)
Glucose, mg/dL		1	7,520	9 (79.5)
Potassium, mmol/L		5.1	143	34.5 (35.5)
Sodium, mmol/L		2	235	72 (55.5)
Conductivity, mS/cm		0.8	35.4	11.2 (8.8)
Specific gravity		1.002	1.030	1.010 (10.2)
Osmolality, mosm/kg		96	1,150	442 (296)

IQR, interquartile range. BUN, mmol/L=BUN, mg/dL \times 0.357; Glucose, mmol/L=Glucose, mg/dL \times 0.0555.

The AutoML function was limited by 2 h of run time for the model development. During the model-building phase, the AutoML functions developed and evaluated GLM, deep-learning, DRF, GBM, and XRT supervised algorithms for each feature group (FG A-D). During model training, the AutoML function used the predefined cross-validation partitions that were described above to obtain comparable model metrics. Hyperparameter optimization was also performed within the AutoML function.

Model performance evaluation and variable importance: The previously described R script was used to assess multiple model performances simultaneously. The model performances were evaluated using the test dataset with the mean absolute error (MAE), R^2 and root mean square error (RMSE) performance metrics. The best ML models were selected from each FG according to the lowest MAE and RMSE and highest R^2 . The same metrics were calculated with the identical test set for the manufacturer (MC) and biochemical calculations (BC) as shown in Figure 1.

A variable importance (VIP) analysis was performed for the four best ML models. The percentages were derived from the relative importance of the features and used to compare the VIP.

Statistical evaluation

The data pre-processing, development of ML models, and statistical analyses were conducted using R statistical software 4.1 [16]. The R scripts of the present study's steps are available in our GitHub account (<https://github.com/ditopcu/osmoAutoML>). We also provided a web-application for validation studies of the models <https://denizt.shinyapps.io/osmoAutoML/>.

All of the test results had non-normal distribution according to the Shapiro–Wilk normality test. Therefore, non-parametric statistics were used. Spearman's correlation coefficients were calculated for each training and test set. Passing-Bablok linear regression equations were used for the comparison analysis between the reference method and osmolality estimations. The mean percentage differences were calculated from the Bland–Altman plots and were used to assess the agreement between the methods. The minimum allowable bias was used as the limit of agreement and calculated with $0.125 \times (CVI^2 + CVG^2)^{1/2}$ formula. The limit of agreement was found to be 24.17% using CVI and CVG as 28.3 and 57.9%, respectively, as stated by Chevront et al. [17].

Results

The demographic features of patients and descriptive statistics of the quantitative measurements that were used in the ML model development were given in Table 2. Table 3 provides the semi-quantitative measurement frequencies. The descriptive statistics of the excluded parameters according to the Boruta feature selection algorithm results are provided in the Supplementary Material Tables 2 and 3.

According to the results of Boruta feature selection algorithm, the urine SG, conductivity, creatinine, protein,

Table 3: Frequency of semi-quantitative measurements for ML model development.

Test	Result	n (%)	Test	Result	n (%)	Test	Result	n (%)	Test	Result	n (%)	
Albumin, mg/L	0	107 (35.7)	Creatinine mg/dL	10	76 (25.3)	Protein mg/dL	0	109 (36.3)	pH	5	101 (33.7)	
	30	110 (36.7)		50	150 (50.0)		15	30 (10.0)		6	41 (13.7)	
	80	24 (8.0)		100	63 (21.0)		30	44 (14.7)		7	25 (8.3)	
	150	25 (8.3)		200	11 (3.7)		100	82 (27.3)		5.5	86 (28.7)	
	500	34 (11.3)				300	26 (8.7)	6.5	25 (8.3)		500	13 (4.3)
						1,000	9 (3.0)	7.5	22 (7.3)		2,000	34 (11.3)

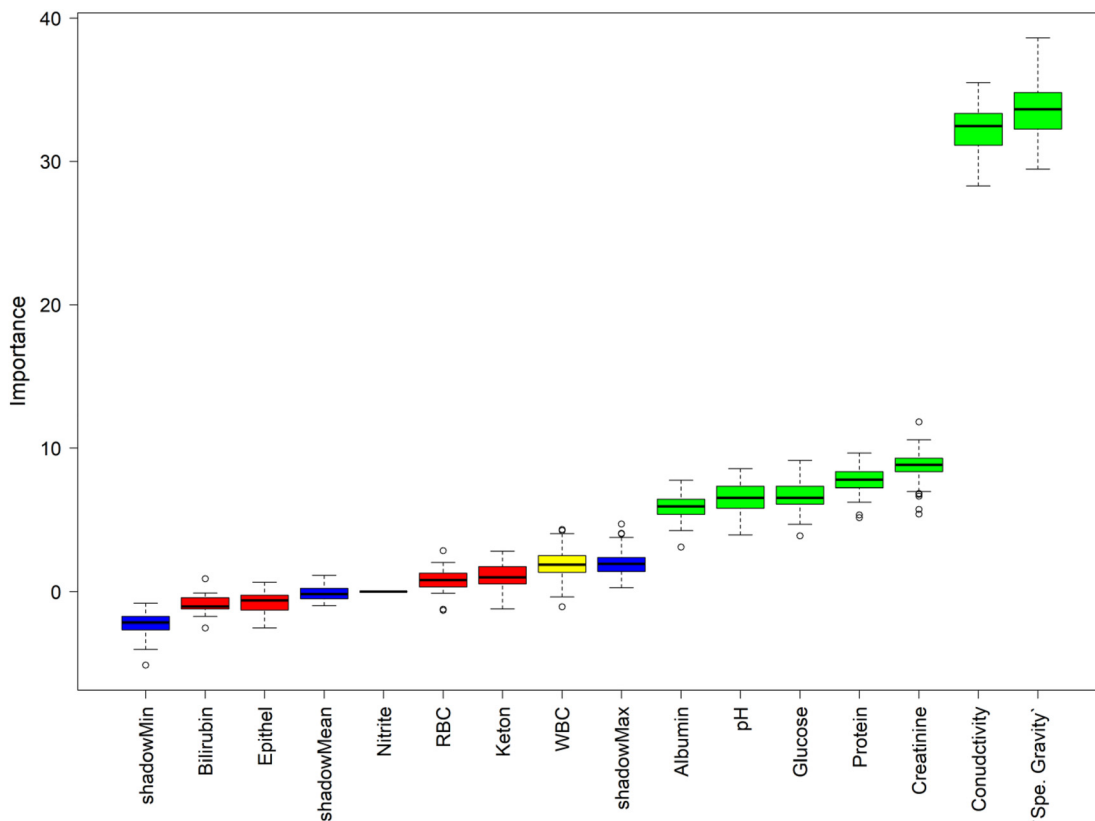


Figure 2: Boruta feature selection plot.

Feature importance values are color coded: Red: Discard, blue: speculative, green: keep. Spe. Gravity: Specific gravity.

pH, glucose, and albumin measurements were found to be important features, and these measurements were used to develop the ML model. The feature importance plot is shown in Figure 2, and the mean importance levels are provided in the Supplementary Material, Table 1.

The model counts that were developed by the H2O AutoML engine for FG-A, FG-B, FG-C, and FG-D were 58, 42, 41, and 42, respectively. The VIP analysis revealed that the conductivity measurement was the most important feature for FG-B (53%) and FG-D (47%), whereas SG and conductivity were the most important features for FG-C (37.8% and 32.8 respectively). All detailed metrics, including the hyperparameters, train set k-fold evaluation metrics, and VIP analysis, are provided in the Supplementary Material.

For the train and test sets, the performance metrics and method comparison results, including the number of results within the limit of agreement, of the developed ML models are shown in Table 4. When performance metrics evaluated for the test set, surprisingly, FG-B and FG-D have the same R^2 score (0.83) although FG-D utilizes additional creatinine and albumin measurements. Furthermore, these two groups had the smallest and most similar MAE scores.

Additionally, both the current conductivity model and manufacturers' calculation had the lowest R^2 and highest MAE scores. It is interesting that FG-C had a lower R^2 score compared to FG-B, although more features were included in FG-C.

An analysis of the method comparison results reveals that FG-B and FG-D, as demonstrated above, had the highest correlation coefficient, lowest mean differences in the Bland-Altman plots and both groups have the most results within the limit of agreement. Additionally, these two models met the Passing-Bablok method agreement, which is 95% confidence interval (CI) for the intercept, including zero, and CI for the slope, covering one. The other models did not meet these requirements. The Passing-Bablok regression analysis and Bland-Altman plots are shown in Figures 3 and 4, respectively.

Discussion

The initial objective of the study was to develop a novel urine osmolality surrogate using an AutoML tool and

Table 4: Performance metrics and method comparison results for developed models.

Feature group	Method	Model metrics			Correlation r (95% CI) ^d	Passing-Bablok regression analysis		Bland-Altman plot Mean absolute difference (mean % difference)	Number of results within the limit of agreement ^c , %
		Dataset	R ²	MAE		RMSE	Slope (95% CI)		
A. Conductivity	AutoML GLM	Train	0.60	102	140	0.80 (0.75-0.85)	0.80 (0.71-0.89)	108.44 (81.84-141.73)	68
		Test	0.66	88	124	0.82 (0.74-0.87)	0.93 (0.83-1.03)	56.60 (10.13-96.31)	70
B. Conductivity, SG	AutoML GBM	Train	0.90	43	71	0.95 (0.94-0.96)	0.94 (0.91-0.97)	28.17 (14.09-43.00)	95
		Test	0.83	56	87	0.90 (0.86-0.93)	1.01 (0.95-1.08)	-8.97 (-33.63-19.26)	88
C. Conductivity Standard urinalysis ^a	AutoML GLM	Train	0.81	53	97	0.93 (0.91-0.95)	1.00 (0.96-1.03)	11.19 (-4.63-27.04)	90
		Test	0.79	57	99	0.88 (0.83-0.92)	1.08 (1.02-1.12)	-24.73 (-45.79-2.56)	84
D. Conductivity Extended urinalysis ^b	AutoML GBM	Train	1.00	10	14	0.99 (0.98-1.00)	0.97 (0.97-0.98)	11.67 (8.48-15.02)	100
		Test	0.83	54	88	0.90 (0.85-0.93)	0.96 (0.91-1.02)	13.42 (-12.33-29.80)	90
BC biochemical calculation	Formula	Train	0.88	51	83	0.93 (0.91-0.95)	0.94 (0.91-0.96)	-2.02 (-9.94-6.62)	93
		Test	0.70	65	120	0.83 (0.76-0.88)	0.94 (0.91-0.97)	-2.01 (-13.34-8.11)	89
MC conductivity	Manufacturer formula	Train	0.60	109	157	0.80 (0.75-0.85)	0.97 (0.87-1.07)	-29.90 (-63.83-6.14)	58
		Test	0.67	102	140	0.82 (0.74-0.87)	1.12 (1.00-1.25)	-93.24 (-148.16-46.02)	59

CI, confidence interval; GBM, gradient boosting machine; GLM, generalized linear model; MAE, mean absolute error; r, Spearman's correlation coefficient. ^aSpecific gravity, glucose, pH, protein. ^bAlbumin, creatinine, specific gravity, glucose, pH, protein. ^cLimit of agreement: 24.17% (Minimum allowable bias). ^dp<0.05. Topcu and Bayraktar: AutoML for urine osmolality estimation 1917 Additionally, FG-D variable importance values revealed

compare its performance to the current alternatives and reference method. To our knowledge this is the first study which uses an AutoML tool in a clinical laboratory setting. Our results show that developed models, which include conductivity and other urinalysis parameters, appear to successfully estimate urine osmolality. The second aim of the study was to assess the advantages of novel AutoML tools in a clinical laboratory setting. We demonstrated that the utilization of AutoML tools facilitates the model development cycle.

Our first finding that developed ML models, using both urine conductivity and other urinalysis measurements, can estimate urine osmolality is supported by evaluation of mean percentage differences and number of results within the agreement limits. FG-B, FG-C and FG-D had mean percentage differences between 3.2%-4.7 and 84%-90% of estimated results which were within the limit of agreement (Table 4 and Figure 4). Passing-Bablok regression analysis also supports that FG-B and FG-D had acceptable comparison results with reference osmolality measurement (Table 4 and Figure 3).

The advantages of using urine conductivity for osmolality measurement estimation were also evaluated by Oyaert et al. [6]. They conducted a multi-step study on 102 samples to evaluate the urine conductivity and reference osmolality methods. First, they evaluated the relationship between the urine conductivity and direct measurement of osmolality using a regression analysis. The analysis resulted in value of R²=0.539. In our study, FG-A included only conductivity as a predictor and a model was developed using the GLM algorithm. In this model, the R² value was calculated as 0.66. In the final step Oyaert et al. created a multiple linear mixed model that uses urinary conductivity, SG, and urine creatinine quantitative reflectance values as predictors. In this model, the conductivity, SG, and creatinine levels were determined to be the strongest predictors. This model was evaluated using 36 patients and the value of R²=0.89 was found. In our study, the conductivity, SG, and creatinine levels were determined to be the strongest predictors in the Boruta feature selection algorithm which is similar with findings of Oyaert et al. Additionally, FG-D variable importance values revealed that the four major parameters were as follows: conductivity (44.8%), SG (40.4%), protein (4.0%), and creatinine (2.9%). An evaluation of the model's performances demonstrated that both R² values were similar. However, in the evaluation of the Bland-Altman graph in the current study, a difference of 10.71 mOsm/Kg was observed with this model, while Oyaert et al. obtained a lower (1.3 mOsm/Kg) difference. Interestingly, our study's results demonstrated that the creatinine-free model had a similar

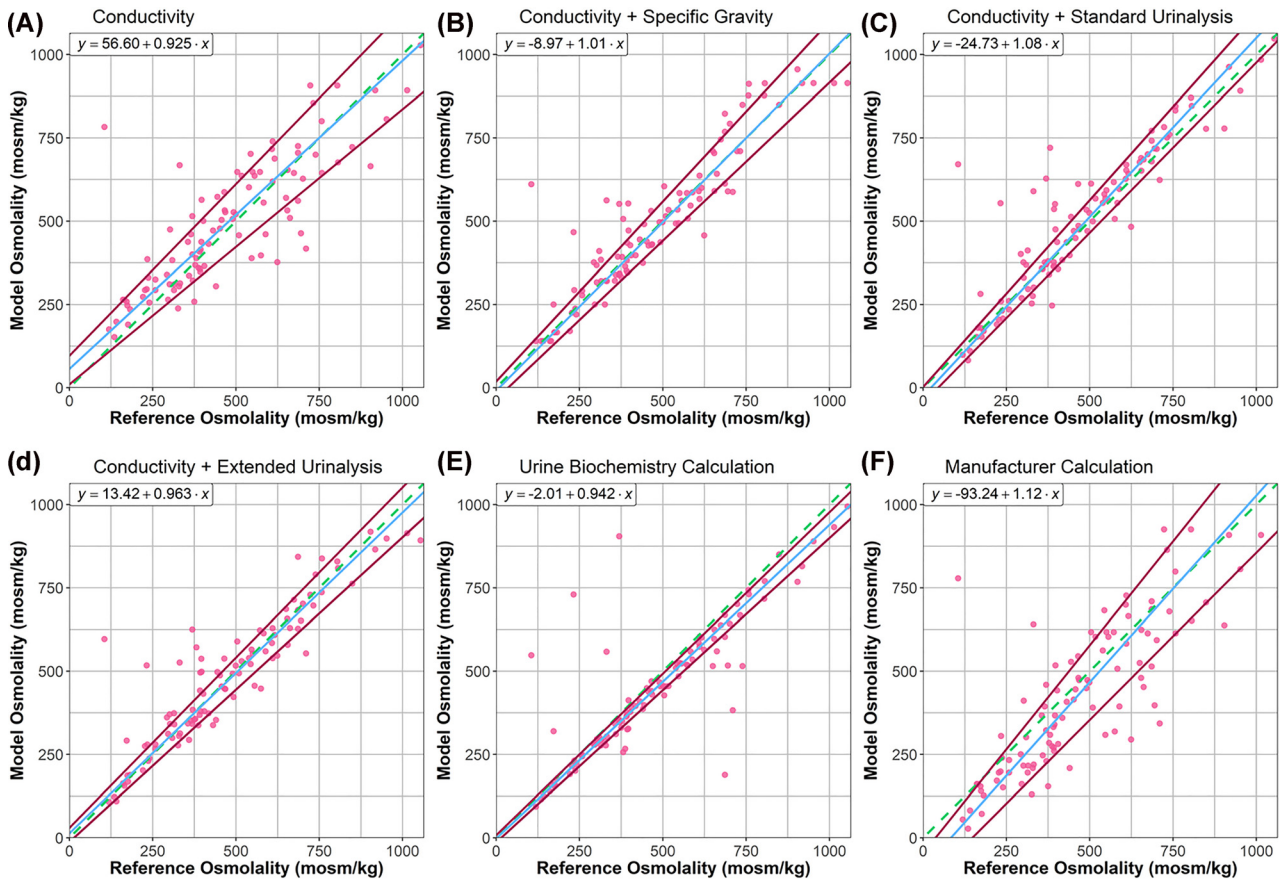


Figure 3: Passing-Bablok regression plots for the test dataset.

Identity lines ($y=x$) are dashed green, confidence intervals are claret and regression lines are blue. Plots for the train dataset are provided in the supplementary material (Supplementary Figure 1).

performance to the creatinine model. These two different outcomes can be explained by the creatinine levels that were used semi-quantitatively in our study, whereas reflectance values were used by Oyaert et al. Another study that compared the manufacturer's osmolality estimation with the reference osmolality measurement was performed by Yoo et al. using 270 urine samples [18]. In this study, a regression analysis was performed, and $R^2=0.667$ was found. This result is similar to our calculation, which was obtained using both the FG-A and manufacturer calculation ($R^2=0.66$ and 0.67 respectively). These two studies along with our study, show that the sole use of conductivity can be misleading in the estimation of osmolality.

The biochemical analysis of serum urine, sodium, potassium, urea, and glucose is another widely used method of estimating urine osmolality. The study carried out by Youhanna et al. which involved 4,247 individuals from a total of four large cohorts and 146 patients with renal failure, osmolality that was calculated using a biochemical formula was compared with the reference

method, in which the correlation analysis and Bland–Altman graphs were applied separately for each cohort [5]. The study results revealed a linear relationship between the estimated and reference osmolality in all cohorts, where the correlation coefficient was calculated between 0.98 and 0.99. The evaluation of the Bland–Altman plots demonstrated that the mean difference in the cohorts was ± 24 mmol/L, while the mean difference in the renal failure group was 15 mmol/L. In our study, the difference between the biochemical measures was similar to biochemical analysis of urine samples is commonly available in clinical laboratories. Conversely, the requirement of an additional analysis, equipment, and cost for this calculation reduces the effectiveness of a biochemical osmolality calculation in large populations.

Another of aspect of our study was the utilization of AutoML tools. We used an open-source H2O tool to develop multiple highly optimized ML models. Only the basic data preparation and model performance evaluation were handled manually. All other ML steps that require

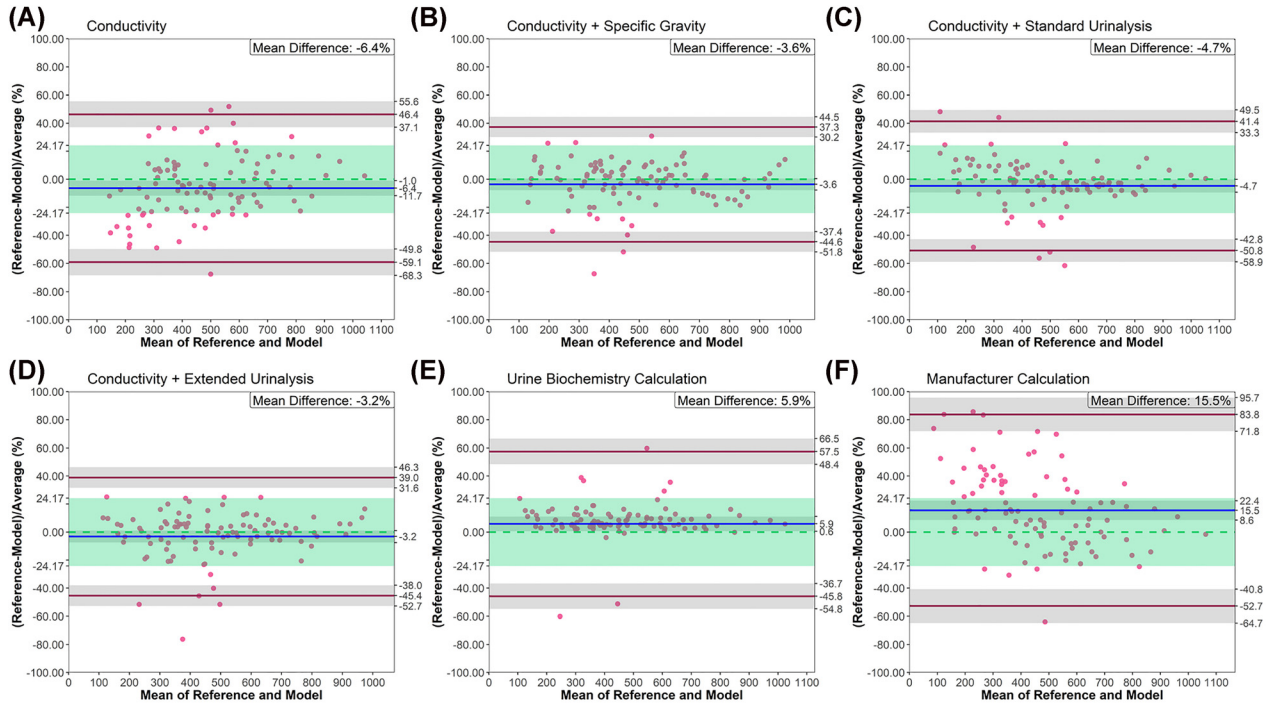


Figure 4: Bland-Altman plots for the test dataset.

$y=0$ lines are dashed green, mean differences (%) are blue, 95% limits of agreement are claret, confidence intervals are solid gray. Second y axis represents confidence intervals. Green area indicates the limit of agreement (minimum allowable bias 24.17%). Plots for the train dataset are provided in the supplementary material (Supplementary Figure 2).

experience were performed using an AutoML tool (Figure 1). Many studies have shown that ML can contribute to patient safety and effective cost management by transforming Big Data into information in the healthcare sector [10, 11, 19]. To support the human expertise need for ML, AutoML tools were developed. AutoML has the added benefit of allowing simpler solutions, faster generation of models without the need for frequent manual intervention that frequently outperform manually developed models [20]. Although, these tools are frequently utilized in other industries, their applications are limited in healthcare, and even more so in the clinical laboratory setting [10]. Our study showed that despite the need to develop and optimize multiple models for each feature group, numerous models were created automatically with the aid of the H2O engine with minimum manual intervention and human labor which is consistent with the statement of Rashidi et al. [11]. All models were fine-tuned by H2O engine using hyperparameter optimization which is normally considered as difficult and time-consuming step for ML development.

A key strength of our study was the utilization of the AutoML tool, which allowed the development of numerous ($n=183$) and different ML models. Moreover, the developed R script facilitated the evaluation of model performance and

method comparison. Furthermore, we included patients with wide-spectrum urinalysis and urine osmolality results, which is significant for successful model training.

A limitation of this study is that the effects of iodine based radiocontrast agents were not evaluated. It is known that large molecules such as radiocontrast media could increase SG more than osmolality [6]. Another limitation of ML studies relates to the external validation of the produced models [10, 11]. Our study is limited by the lack of external validation however we provided a web-tool for further model evaluation. Open platforms that were used in this study, such as “Shiny” could facilitate the validation of developed ML models [21, 22]. An additional limitation of this study was related to the utilization of AutoML tools. Normally, AutoML tools can automate all ML model development steps. However, in our study, feature selection and data splitting steps were performed manually to provide an identical data set for different feature sets.

In conclusion, we found that urine osmolality can be estimated using GBM model which utilizes conductivity and all other urinalysis parameters, but further research is required to evaluate the performance of developed models in selected cohorts. Our findings also illustrate that AutoML tools can provide reliable models and facilitate ML model development. Taken together, these findings

suggest, clinical laboratories should take advantage of these tools while developing machine learning models.

Research funding: This study was supported by Baskent University Research Fund [KA19/305] and Sysmex Turkey.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: Authors state no conflict of interest.

Informed consent: Not applicable.

Ethical approval: The study was approved by the institutional review board of Baskent University (Project No: KA19/305).

References

- De Jesús Vidal-Mayo J, Olivas-Martínez A, Pérez-Díaz I, López-Navarro JM, Sánchez-Landa E, Carrillo-Maravilla E, et al. Calculated versus measured urine osmolality: accuracy of estimated urine density. *Rev Investig Clin* 2018;70:310–8.
- Wright AE, Wragg R, Lopes J, Robb A, McCarthy L. Prediction of need for intervention in posterior urethral valves: use of urine osmolality. *J Pediatr Surg* 2018;53:316–20.
- Kavouras SA, Suh H-G, Vallet M, Daudon M, Mauromoustakos A, Vecchio M, et al. Urine osmolality predicts calcium-oxalate crystallization risk in patients with recurrent urolithiasis. *Urolithiasis* 2021;49:399–405.
- Lee MJ, Chang TI, Lee J, Kim YH, Oh KH, Lee SW, et al. Urine osmolality and renal outcome in patients with chronic kidney disease: results from the KNOW-ckd. *Kidney Blood Press Res* 2019;44:1089–100.
- Youhanna S, Bankir L, Jungers P, Porteous D, Polasek O, Bochud M, et al. Validation of surrogates of urine osmolality in population studies. *Am J Nephrol* 2017;46:26–36.
- Oyaert M, Speeckaert MM, Delanghe JR. Estimated urinary osmolality based on combined urinalysis parameters: a critical evaluation. *Clin Chem Lab Med* 2019;57:1169–76.
- Walawender L, Patterson J, Strouse R, Ketz J, Saxena V, Alexy E, et al. Mobile technology application for improved urine concentration measurement pilot study. [Internet]. *Front Pediatr* 2018;6:160.
- Oyaert MN, Himpe J, Speeckaert MM, Stove VV, Delanghe JR. Quantitative urine test strip reading for leukocyte esterase and hemoglobin peroxidase. [Internet]. *Clin Chem Lab Med* 2018;56:1126–32.
- Picálek J, Kolafa J. Molecular dynamics study of conductivity of ionic liquids: the Kohlrausch law. *J Mol Liq* 2007;134:29–33.
- Waring J, Lindvall C, Umeton R. Automated machine learning: review of the state-of-the-art and opportunities for healthcare. [Internet]. *Artif Intell Med* 2020;104:101822.
- Rashidi HH, Tran N, Albahra S, Dang LT. Machine learning in health care and laboratory medicine: general overview of supervised learning and Auto-ML. *Int J Lab Hematol* 2021;43:15–22.
- Bagrow JP. Democratizing AI: non-expert design of prediction tasks. *PeerJ Comput Sci* 2020;6:1–23.
- Previtali G, Ravasio R, Seghezzi M, Buoro S, Alessio MG. Performance evaluation of the new fully automated urine particle analyser UF-5000 compared to the reference method of the Fuchs-Rosenthal chamber. [Internet]. *Clin Chim Acta* 2017;472:123–30.
- Kursa MB, Rudnicki WR. Feature selection with the boruta package. *J Stat Software* 2010;36:1–13.
- Ledell E, Poirier S. H2O AutoML: scalable automatic machine learning [Internet]. AutoML Org; 2020. Available from: <https://scinet.usda.gov/user/geospatial/#tools-and-software%0Ahttps://www.slideshare.net/Oxdata/intro-to-automl-handson-lab-erin-ledell-machine-learning-scientist-h2oai%0Ahttps://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html>.
- Core Team RR. A language and environment for statistical computing [Internet]. Vienna: Austria; 2021. Available from: <https://www.r-project.org/>.
- Cheuvront SN, Ely BR, Kenefick RW, Sawka MN. Biological variation and diagnostic accuracy of dehydration assessment markers. *Am J Clin Nutr* 2010;92:565–73.
- Yoo DW, Lee SM, Moon SY, Kim IS, Chang CL. Evaluation of conductivity-based osmolality measurement in urine using the Sysmex UF5000. *J Clin Lab Anal* 2021;35:1–7.
- Feretakis G, Sakagianni A, Loupelis E, Kalles D, Skarmoutsou N, Martsoukou M, et al. Machine learning for antibiotic resistance prediction: a prototype using off-the-shelf techniques and entry-level data to guide empiric antimicrobial therapy. *Healthc Inform Res.* 2021;27:214–21.
- He X, Zhao K, Chu X. AutoML: a survey of the state-of-the-art. [Internet]. *Knowl Base Syst* 2021;212:106622.
- Chang W, Cheng J, Allaire JJ, Sievert C, Schloerke B, Xie Y, et al. Shiny: web application framework for R [Internet]; 2021. Available from: <https://cran.r-project.org/package=shiny>.
- Burnett JL, Dale R, Hou CY, Palomo-Munoz G, Whitney KS, Aulenbach S, et al. Ten simple rules for creating a scientific web application. [Internet]. *PLoS Comput Biol* 2021;17:1–12.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/cclm-2022-0415>).