DOI: *10.4081/aiua.2025.14333*

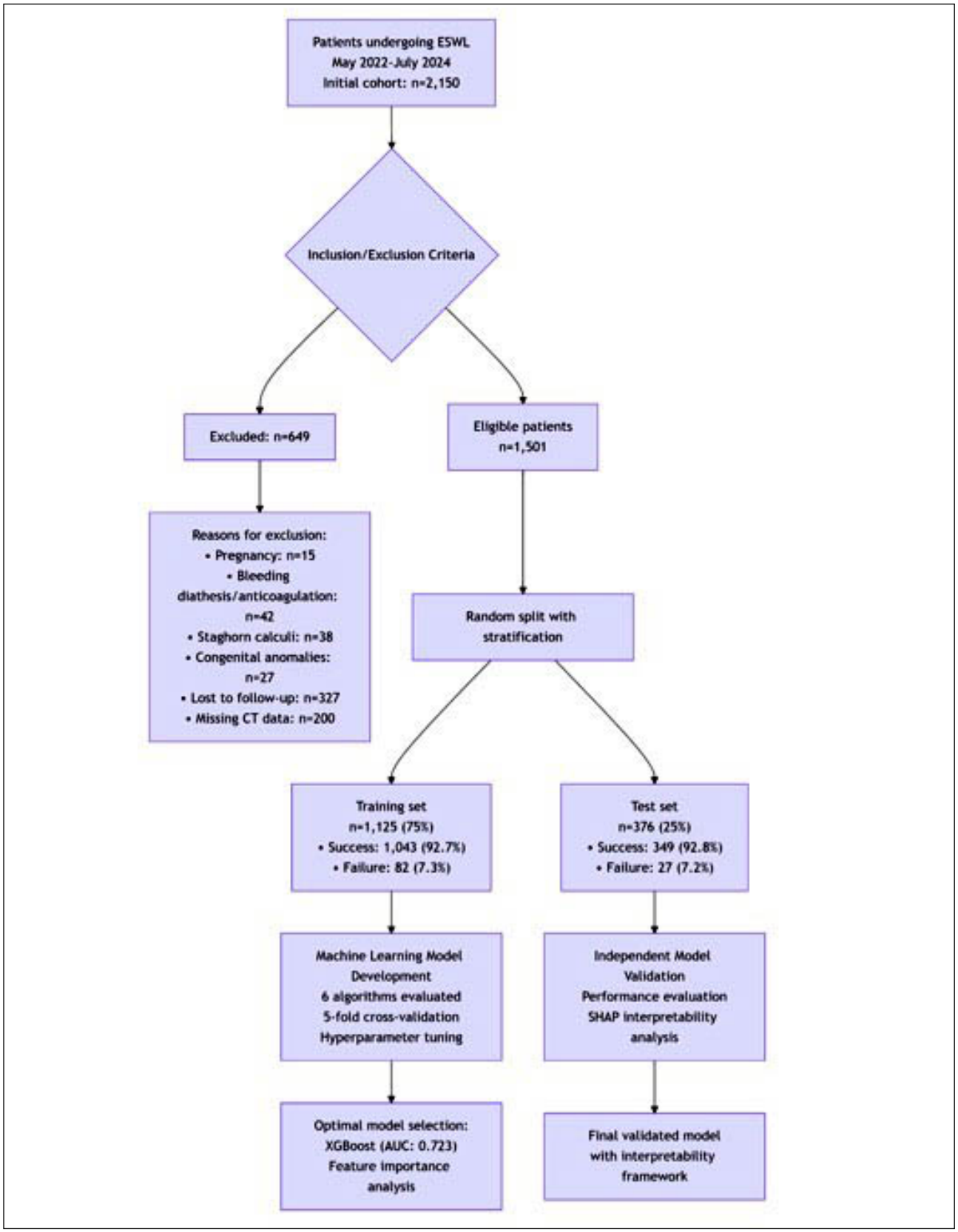
ORIGINAL PAPER - SUPPLEMENTARY MATERIAL

Interpretable machine learning prediction of Extracorporeal Shock Wave Lithotripsy outcomes for urinary stones: A retrospective cohort study

Morshed Salah [1, 2], Maged Al-Ghashmi [1], Abu Baker [1], Hatem Kamkoum [1], Salvan Alhabash [1], Hossameldin Alnawasra [1], Abdoulhafid Elmogassabi [1], Mohammed Ebrahim [1], Mohamed Abdelkareem [1], Faisal Ahmed [3]

[1] Urology Section, Hazm Mebaireek General Hospital, Hamad Medical Corporation, Doha, Qatar;
[2] College of Medicine, Qatar University, Doha, Qatar;
[3] Department of Urology, School of Medicine, Ibb University, Ibb, Yemen.

Supplementary Figure S1.
Study Flow Chart.



Flow diagram illustrating the patient selection, cohort formation, and machine learning model development process for predicting Extracorporeal Shock Wave Lithotripsy (ESWL) outcomes.

Supplementary Table S1.
Complete Feature List and Descriptive Statistics.

Feature Category	Feature Name	Data Type	Missing Values (%)	Descriptive Statistics	Clinical Description
Demographic	Age (years)	Continuous	0.0	Mean ± SD: 47.2 ± 14.3	Patient age at time of ESWL procedure
	Sex	Binary	0.0	Male: 918 (61.2%) Female: 583 (38.8%)	Patient biological sex
	Body Mass Index (BMI) (kg/m²)	Continuous	2.1	Mean ± SD: 26.4 ± 4.1	Weight in kilograms divided by height in meters squared
Stone Characteristics	Stone size (mm)	Continuous	0.0	Mean ± SD: 11.3 ± 3.8	Maximum axial diameter on pre-procedural non-contrast computed tomography (NCCT)
	Stone density (HU)	Continuous	0.0	Mean ± SD: 985 ± 342	Mean attenuation value in Hounsfield units on NCCT
	Stone location	Categorical	0.0	Renal: 1,127 (75.1%) Ureteral: 374 (24.9%)	Primary anatomical location of the stone
	Ureter segment*	Categorical	15.3	Upper: 158 (42.3%) Middle: 102 (27.3%) Lower: 114 (30.5%)	Specific segment for ureteral stones only
Treatment Parameters	Shock wave count	Continuous	0.0	Mean ± SD: 2,950 ± 450	Total number of shock waves delivered during ESWL session
	Maximum power (kV)	Continuous	0.0	Mean ± SD: 14.2 ± 2.1	Maximum power setting of the Dornier Compact Delta® III Pro lithotripter
	Frequency (Hz)	Continuous	0.0	Mean ± SD: 1.5 ± 0.3	Shock wave frequency in Hertz
	Double-J stent presence	Binary	0.0	Present: 623 (41.5%) Absent: 878 (58.5%)	Presence of ureteral stent at time of ESWL
Anatomical Factors	Skin-to-stone distance (mm)	Continuous	0.0	Mean ± SD: 102.3 ± 18.5	Linear distance from skin surface to stone center on NCCT
	Kidney side	Categorical	0.0	Left: 795 (53.0%) Right: 706 (47.0%)	Laterality of renal stones
Medical History	Hypertension	Binary	0.0	Present: 642 (42.8%) Absent: 859 (57.2%)	History of hypertension
	Diabetes mellitus	Binary	0.0	Present: 418 (27.8%) Absent: 1,083 (72.2%)	History of diabetes mellitus
	Previous stone surgery	Binary	32.5	Yes: 285 (19.0%) No: 731 (48.7%) Missing: 485 (32.3%)	History of previous surgical intervention for urolithiasis
Laboratory Values	Serum creatinine (μmol/L)	Continuous	8.2	Mean ± SD: 85.6 ± 24.3	Pre-procedural serum creatinine level
	Estimated glomerular filtration rate (mL/min/1.73m²)	Continuous	8.2	Mean ± SD: 78.4 ± 18.2	Calculated using CKD-EPI equation
Procedural Details	Session duration (minutes)	Continuous	0.0	Mean ± SD: 42.5 ± 12.3	Total procedure time
	Fluoroscopy time (seconds)	Continuous	0.0	Mean ± SD: 58.7 ± 21.4	Total fluoroscopy exposure time
	Analgesia requirement	Binary	0.0	Required: 892 (59.4%) Not required: 609 (40.6%)	Need for additional analgesia during procedure
Radiological Features	Hydronephrosis grade	Ordinal	0.0	0: 945 (63.0%) 1: 312 (20.8%) 2: 178 (11.9%) 3: 66 (4.4%)	Severity of hydronephrosis (0: none, 1: mild, 2: moderate, 3: severe)
	Stone burden (mm³)	Continuous	0.0	Mean ± SD: 845 ± 312	Calculated stone volume (π/6 × length × width × height)
	Multiple stones	Binary	0.0	Present: 312 (20.8%) Absent: 1,189 (79.2%)	Presence of multiple stones in the same renal unit
	Infundibulopelvic angle (degrees)	Continuous	0.0	Mean ± SD: 72.5 ± 14.3	Angle between renal infundibulum and pelvis
Excluded Features	Stone composition	Categorical	78.6	Calcium oxalate: 215 (14.3%) Uric acid: 78 (5.2%) Struvite: 31 (2.1%) Missing: 1,177 (78.4%)	Chemical composition from stone analysis (excluded due to >30% missing)
	24-hour urine chemistry	Continuous	89.3	Various	Comprehensive metabolic panel (excluded due to >30% missing)

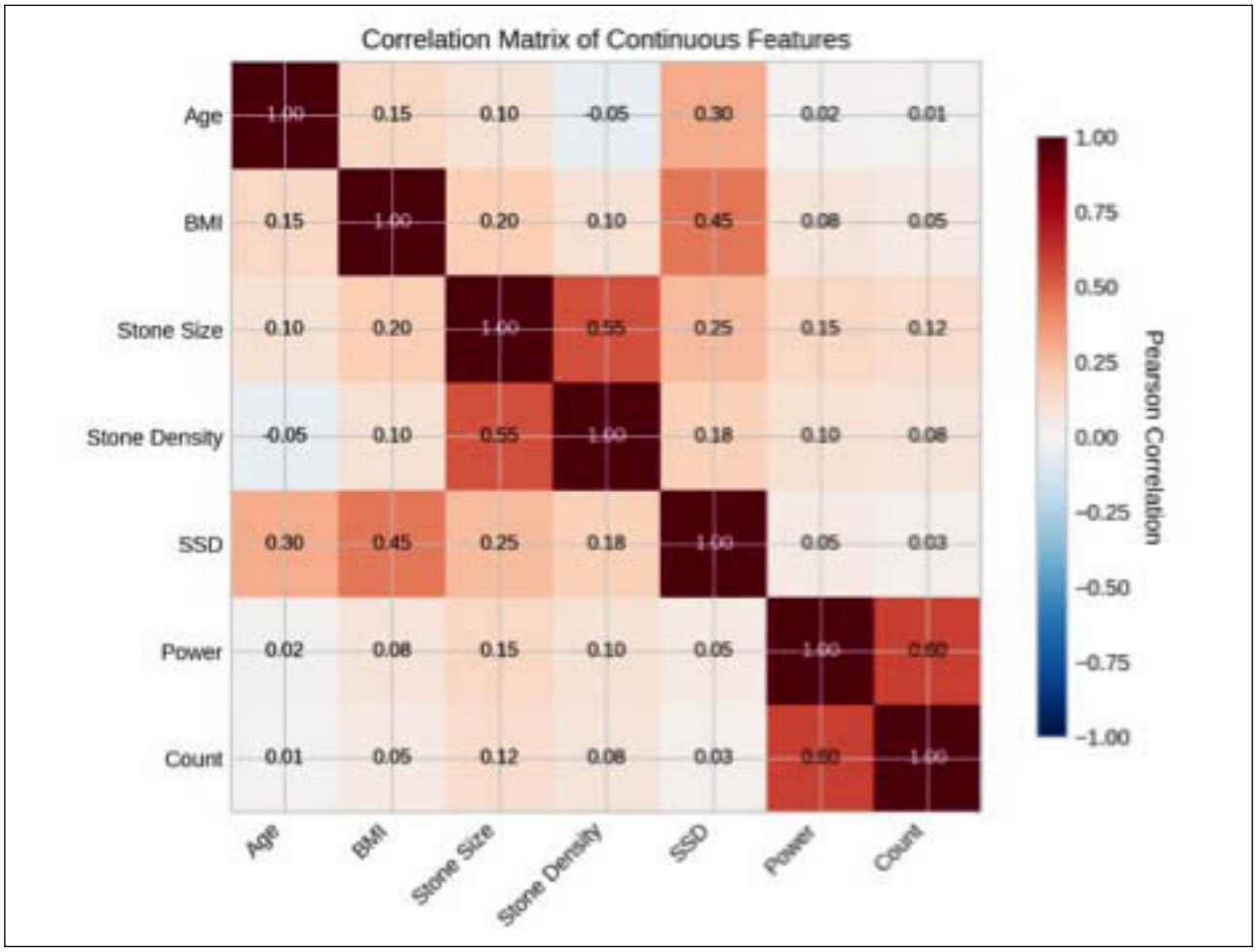
** Ureter segment missingness applies only to ureteral stones (n = 374). For renal stones, this variable was not applicable.*
SD: Standard Deviation; HU: Hounsfield Units; NCCT: Non-Contrast Computed Tomography; CKD-EPI: Chronic Kidney Disease Epidemiology Collaboration.

Supplementary Table S2.
Hyperparameter Search Spaces for Machine Learning Algorithms.

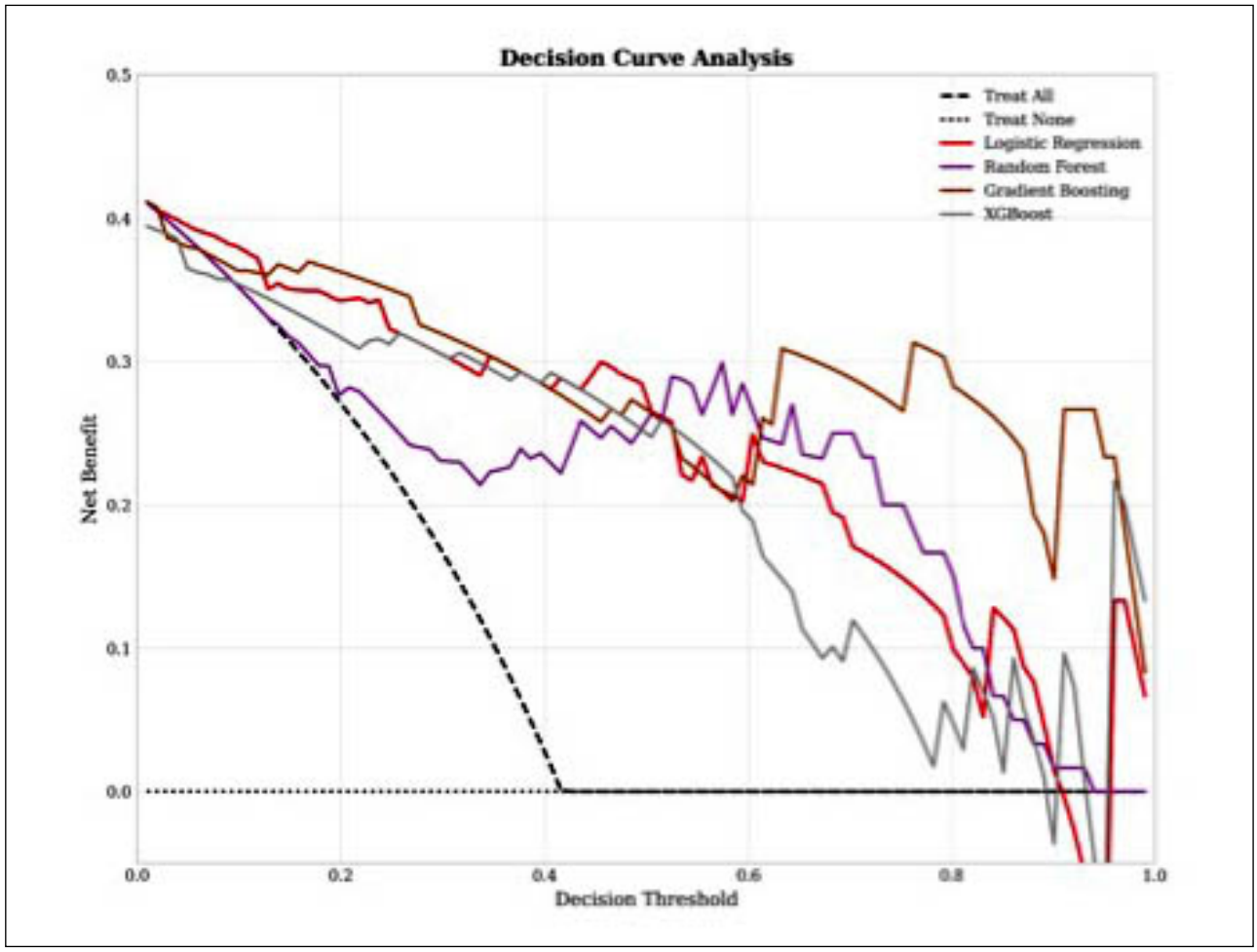
Algorithm	Hyperparameter	Search Space	Optimization Method	Cross-Validation	Primary Metric
Logistic Regression	Penalty	['l1', 'l2', 'elasticnet', 'none']	Randomized Search† (100 iterations)	5-fold stratified	AUC-ROC
	C (Regularization strength)	[0.001, 0.01, 0.1, 1, 10, 100]			
	Solver	['liblinear', 'saga'] (for l1/l2)‡ ['lbfgs', 'newton-cg'] (l2 only)‡ ['newton-cg', 'lbfgs', 'sag'] (none)‡			
	max_iter	[100, 200, 500]			
Random Forest	n_estimators	[50, 100, 200, 500]	Randomized Search† (100 iterations)	5-fold stratified	AUC-ROC
	max_depth	[3, 5, 10, 15, 20, None]			
	min_samples_split	[2, 5, 10]			
	min_samples_leaf	[1, 2, 4]			
	max_features	['sqrt', 'log2', None]			
Support Vector Machine (SVM)	C	[0.001, 0.01, 0.1, 1, 10, 100]	Randomized Search† (100 iterations)	5-fold stratified	AUC-ROC
	kernel	['linear', 'rbf', 'poly']			
	gamma	['scale', 'auto'] + [0.001, 0.01, 0.1, 1]			
	degree	[2, 3, 4] (poly kernel only)‡			
Gradient Boosting	n_estimators	[50, 100, 200, 500]	Randomized Search† (100 iterations)	5-fold stratified	AUC-ROC
	learning_rate	[0.001, 0.01, 0.1, 0.2]			
	max_depth	[3, 5, 7, 9]			
	min_samples_split	[2, 5, 10]			
	min_samples_leaf	[1, 2, 4]			
	subsample	[0.8, 0.9, 1.0]			
Extreme Gradient Boosting (XGBoost)	n_estimators	[50, 100, 200, 500]	Randomized Search† (100 iterations)	5-fold stratified	AUC-ROC
	max_depth	[3, 5, 7, 9]			
	learning_rate	[0.001, 0.01, 0.1, 0.2]			
	subsample	[0.8, 0.9, 1.0]			
	colsample_bytree	[0.8, 0.9, 1.0]			
	gamma	[0, 0.1, 0.2, 0.3]			
	scale_pos_weight	[1, 3, 6.92, 10]¶			
Neural Network (Multi-layer Perceptron)	hidden_layer_sizes	[(50,), (100,), (50, 50), (100, 50)]	Randomized Search† (100 iterations)	5-fold stratified	AUC-ROC
	activation	['relu', 'tanh']			
	alpha	[0.0001, 0.001, 0.01]			
	solver	['adam', 'sgd']			
	learning_rate	['constant', 'adaptive']			
	max_iter	[200, 300, 500]			

** All algorithms: Hyperparameter optimization was conducted exclusively on the training set (n = 1.125) to prevent data leakage. The test set (n = 376) was used only for final model evaluation.*
† Randomized Search: 100 random hyperparameter combinations were evaluated per algorithm. The best combination was selected based on the highest mean AUC-ROC score across 5-fold stratified cross-validation.
‡ Solver/kernel-specific parameters: Parameters are listed for specific penalty types or kernels only, as indicated in parentheses.
¶ scale_pos_weight (XGBoost): The value 6.92 corresponds to the inverse class ratio in the training set (success:failure ≈ 13:1), specifically addressing the class imbalance for the minority failure class.

Supplementary Figure S4.
Feature Correlation Matrix. A heatmap displaying Pearson correlation coefficients between the continuous variables used in the model. The absence of strong correlations (all $|r| < 0.7$) confirms that multicollinearity was not a significant concern, supporting the stability of the feature importance rankings.



Supplementary Figure S5.
Decision Curve Analysis. Plots the net clinical benefit of using the XGBoost model across a range of probability thresholds (0.01 to 0.99). The model (red line) provides a higher net benefit than the default strategies of treating all patients or no patients within a clinically relevant threshold range, supporting its potential utility for guiding treatment decisions.



Supplementary Figure S6.
Learning Curves. Plots model performance (ROC-AUC) for the training and cross-validation sets as a function of training set size. The convergence of the two curves indicates that the model was adequately trained without substantial overfitting.