

Original Article

Prediction of pediatric Wilms tumor recurrence using interpretable machine learning models: insights from a 20-year real-world study and the prognostic value of Ki-67

Honggang Fang^{1,2,3,4}, Yihang Yu^{1,2,3,4}, Junjun Dong^{1,2,3,4}, Zehang Chen⁵, Fei Liu⁵, Zhong Liu⁶, Qi Li⁷, Xing Liu^{1,2,3,4}, Tao Lin^{1,2,3,4}, Dawei He^{1,2,3,4}, Guanghui Wei^{1,2,3,4}, Deying Zhang^{1,2,3,4}

¹Department of Urology, Children's Hospital of Chongqing Medical University, Chongqing 401122, China; ²National Clinical Research Center for Child Health and Disorders, Chongqing 401122, China; ³Ministry of Education Key Laboratory of Child Development and Disorders, Chongqing 401122, China; ⁴Chongqing Key Laboratory of Structural Birth Defect and Reconstruction, Chongqing 401122, China; ⁵School of Economics and Business Administration, Chongqing University, Chongqing 400030, China; ⁶Department of Urology, Jiangxi Children's Medical Center, Nanchang 330200, Jiangxi, China; ⁷School of Finance, Chongqing Technology and Business University, Chongqing 400067, China

Received June 29, 2025; Accepted September 5, 2025; Epub September 15, 2025; Published September 30, 2025

Abstract: Despite overall survival exceeding 80% in Wilms tumor (WT), approximately 15% of pediatric patients experience recurrence with poor post-relapse survival (~50%) and significant long-term complications, highlighting an unmet need for precise recurrence prediction. To address this, we developed and validated machine learning (ML) models for predicting postoperative WT recurrence using real-world clinical data. Among 476 pediatric WT patients who underwent radical surgery at our institution (June 2004-June 2024), 351 met inclusion criteria and were randomized into training (70%) and validation (30%) cohorts. Seven independent predictors - COG tumor stage, age, tumor rupture, histological subtype (COG classification), tumor thrombus, tumor volume, and Ki-67 index - were identified via feature selection intersection of Boruta, LASSO, subgroup analysis, and univariate logistic regression. Predictive models were constructed using nine ML algorithms (DT, LASSO, KNN, LightGBM, LR, MLP, RF, SVM, XGBoost), with performance evaluated using metrics such as AUC, accuracy, F1 score, specificity, positive predictive value (PPV), negative predictive value (NPV), and confusion matrix. Among the included patients, 51 (14.53%) experienced tumor recurrence (including 7 multi-site relapses), with median time to recurrence 6 months (IQR 4-16 months) and 80.4% occurring within the first postoperative year. In the validation cohort, the SVM model demonstrated the best performance (AUC = 0.851; accuracy = 0.830; specificity = 0.856; F1 = 0.550; PPV = 0.458; NPV = 0.939), and SHAP analysis highlighted unfavorable histology, COG stage IV-V, tumor thrombus, and elevated Ki-67 index as the strongest contributors to recurrence risk. This interpretable SVM-based model confirms seven key predictors - especially the Ki-67 index - and serves as a practical risk stratification tool to support individualized follow-up planning.

Keywords: Wilms tumor, tumor recurrence prediction, machine learning, support vector machine, feature selection, SHAP interpretation

Introduction

Wilms tumor (WT) is an embryonal malignancy that accounts for approximately 95% of pediatric renal neoplasms and has the potential to metastasize to the lungs, liver, bones, and lymph nodes [1]. Despite significant advancements in surgical techniques, chemotherapy, and radiotherapy - along with the integration of

multidisciplinary treatment strategies - overall survival rates for WT now exceed 80%. However, the prognosis following recurrence remains poor, with a survival rate of only around 50%. Survivors of relapse often experience long-term health complications. Accurately identifying children at high risk of recurrence remains a major unresolved clinical challenge [2-4]. The lungs are the most common site of

WT recurrence, followed by the abdomen; metastases to the liver, brain, and bones are comparatively rare. Studies have reported that approximately 15% of pediatric patients may experience relapse. Of these, 50-60% involve isolated pulmonary or pleural recurrence, 30% involve abdominal recurrence (either isolated or in combination with other sites), and the remaining 10-15% present with metastases to the brain or bone [5].

To better assess postoperative survival and recurrence risk in pediatric patients, both the Children's Oncology Group (COG) in the United States and the International Society of Paediatric Oncology (SIOP) have developed and refined risk stratification systems. These approaches aim to allow for flexible adjustments to treatment intensity based on individual recurrence risk [6]. In a retrospective analysis of patients enrolled in the United Kingdom Wilms Tumor Study 3 (UKW3), Irtan et al. [7] identified anaplastic histology and a larger tumor volume as significant risk factors for recurrence. Current evidence indicates that, even under standard treatment protocols, approximately 15% of patients with stage III disease and 25% with stage IV disease still experience relapse. These findings underscore the urgent need for more precise risk stratification and targeted interventions in high-risk patient populations [8].

Compared with traditional regression-based approaches, machine learning offers superior flexibility, scalability, computational power, and automated data-driven capabilities, enabling more accurate predictions across heterogeneous and uncorrelated datasets [9, 10]. Previous research on WT recurrence has primarily focused on post-relapse management strategies - such as surveillance, repeat surgery, and radiotherapy - while effective tools for accurately predicting the risk of recurrence remain lacking [11-13]. In this study, we leveraged clinicopathological features and routine laboratory parameters of pediatric WT patients to develop and validate nine ML-based predictive models for tumor recurrence. Our objective was to construct a clinically applicable, efficient, and accurate predictive tool based on readily accessible clinical data to identify children at high risk of relapse. Such a model could facilitate early intervention and provide an evi-

dence-based foundation for optimizing management strategies both before and after relapse in pediatric WT.

Materials and methods

Clinical and pathological data

Clinical and pathological data were retrospectively collected from pediatric patients diagnosed with Wilms tumor (WT) at the Children's Hospital of Chongqing Medical University (CHCMU) between June 2004 and June 2024. The dataset included demographic, clinical, and laboratory information such as age, sex, laterality of the tumor, low body weight, tumor volume, use of neoadjuvant chemotherapy, nephron-sparing surgery, presence of tumor thrombus, tumor rupture, COG tumor stage, number of lymph nodes dissected, lymph node status, histologic subtype (COG classification), WT1 expression, Ki-67 index, postoperative chemotherapy, and postoperative radiotherapy. Tumor volume (cm³) was calculated from preoperative contrast-enhanced CT three-dimensional images using the ellipsoid formula: $0.523 \times \text{length} \times \text{width} \times \text{height}$ [14]. Low body weight was defined as a body weight below two standard deviations (-2 SD) from the median reference of the World Health Organization (WHO) Child Growth Standards [15].

Ki-67 index was assessed by immunohistochemistry (IHC) according to the guidelines of the International Ki-67 Working Group (IKWG) [16]: specifically, formalin-fixed, paraffin-embedded sections (10% neutral buffered formalin) were cut at 4 μm, deparaffinized in xylene, rehydrated through graded alcohols, and subjected to antigen retrieval in citrate buffer (pH 6.0). Endogenous peroxidase was blocked with 3% hydrogen peroxide for 20 minutes, followed by incubation with an anti-Ki-67 rabbit monoclonal antibody (Roche Ventana, USA) for 16 minutes. After counterstaining with hematoxylin, the slides were thoroughly dried, mounted, and analyzed by light microscopy or using ImageJ software. Tumor cell nuclei were counted in four representative fields (counting at least 400 invasive tumor cells in total), using the global counting method in line with IKWG's recommendation for reproducibility. The Ki-67 proliferation index was reported as the percentage of positively stained nuclei over the total counted.

Wilms tumor recurrence prediction with ML and Ki-67

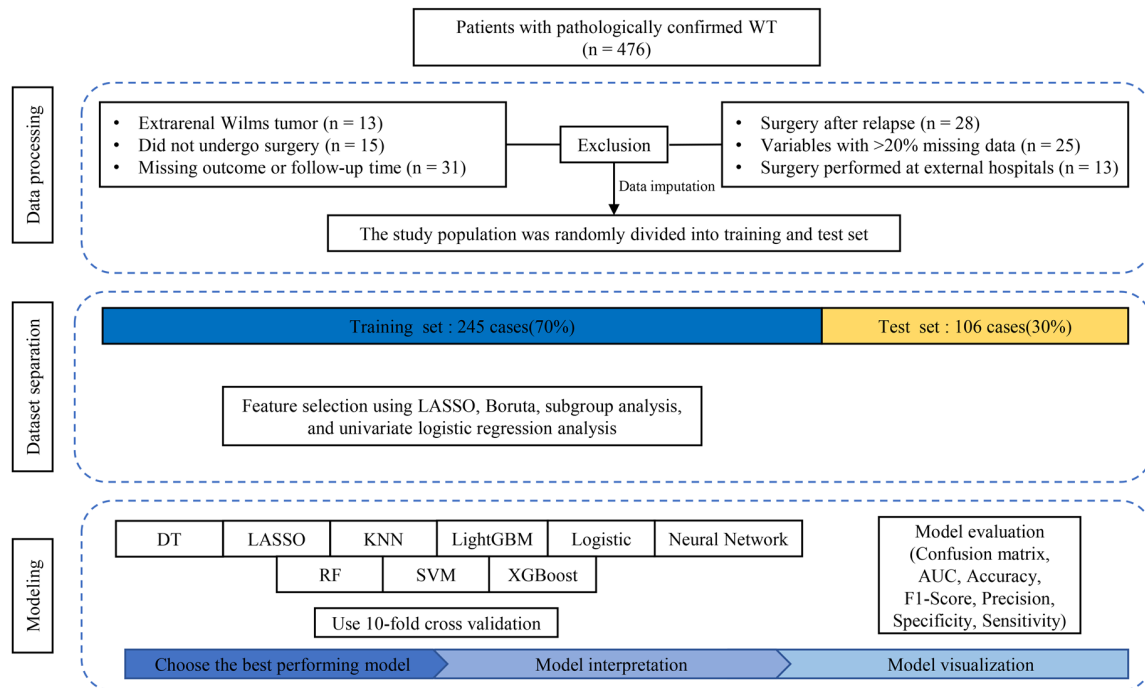


Figure 1. Visualization of data screening and clinical study design process.

Inclusion and exclusion criteria

The inclusion criteria were: (1) patients who underwent surgical treatment with postoperative pathological confirmation of Wilms tumor; (2) less than 20% missing data in the key variables [17]; and (3) no prior history of other malignancies or serious systemic diseases. Exclusion criteria were: (1) Wilms tumors originating outside the kidney (n = 13); (2) unknown survival status (n = 31); (3) patients who did not undergo surgery (n = 15) or who underwent surgery again after recurrence (n = 28); (4) cases with more than 20% missing data (n = 25); and (5) patients who received chemotherapy at our hospital but had surgery performed elsewhere (n = 13) (**Figure 1**). Recurrence was defined as tumor reappearance at least one month after achieving complete remission following initial treatment, including both local recurrence at the primary site and distant metastases. This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Children's Hospital of Chongqing Medical University (Approval No. (2024) Ethics Review (Research) No. 359). As a retrospective study utilizing anonymized data, informed consent was waived.

Data preprocessing

For clinical variables with less than 20% missing values and assumed to be missing at random (WT: 14.81%, Ki-67: 17.09%), multiple imputation was performed using the mice package (version 4.5.1) in R, based on the random forest method ([Supplementary Figure 1](#)) [17]. The cohort of 351 WT patients was randomly divided into a training set (70%) and a testing set (30%). After dataset partitioning, categorical clinical and pathological variables were converted into dummy variables prior to modeling using the dummy function. This approach transforms multi-category nominal variables into numerical inputs recognizable by machine learning algorithms, preventing erroneous ordinal assumptions on category codes. Consequently, it improves model accuracy and ensures appropriate interpretation of the intrinsic properties of variables.

Feature selection

Candidate variables were screened using four methods: LASSO regression, the Boruta algorithm, subgroup analysis, and univariate logistic regression. LASSO, an L1-regularization-based regression technique, performs feature

selection by shrinking the coefficients of irrelevant variables to zero. It is well-suited for high-dimensional data and remains stable in the presence of multicollinearity, effectively identifying key predictors. The Boruta algorithm, built on a random forest framework, assesses variable importance by comparing actual features against randomized “shadow features”. This method handles high-dimensional data robustly and demonstrates strong stability in variable selection [18]. Subgroup analysis was employed to evaluate the consistency of variable performance across different clinical populations, thereby enhancing the model’s clinical generalizability. Univariate logistic regression served as an initial screening tool to preliminarily assess the statistical association between each variable and the outcome. Given the differing principles and selection criteria underlying these methods, this study used the intersection of variables identified by all four approaches as the final set for modeling. This strategy ensures consistency in statistical significance, algorithmic importance, and clinical relevance, thereby improving the predictive accuracy and interpretability of the model.

Model development and validation

Clinical and pathological features such as sex, age, and tumor volume were used as input variables, with tumor recurrence as the prediction target. Nine machine learning algorithms were employed to build predictive models: Decision Tree (DT), Least Absolute Shrinkage and Selection Operator (LASSO), K-Nearest Neighbors (KNN), Light Gradient Boosting Machine (LightGBM), Logistic Regression, Neural Network (single hidden layer), Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost). All ML analyses were conducted in the R environment (version 4.5.1).

During model training, hyperparameter tuning was performed using grid search combined with 10-fold cross-validation to identify optimal parameter configurations, with search ranges appropriately set based on sample size and feature count ([Supplementary Table 1](#)). In the testing set, key performance metrics - including accuracy, area under the curve (AUC), specificity, confusion matrix, recall (sensitivity), and F1 score - were evaluated. Additionally,

receiver operating characteristic (ROC) curves, decision curve analysis (DCA), and calibration curves were used to further assess predictive performance and model stability.

Based on these performance indicators, the best-performing predictive model was identified. To enhance interpretability, SHapley Additive exPlanations (SHAP) were applied to visualize and explain the key contributing factors in the optimal model. SHAP, a game theory-based approach, ranks the importance of input features and elucidates model outputs, effectively addressing the “black-box” nature of ML models and providing transparent interpretability for the optimal model.

Statistical analysis

Categorical variables were compared between groups using the Chi-square test or Fisher’s exact test, as appropriate. For continuous variables, the two-tailed Student’s t-test was used when data followed a normal distribution; otherwise, the Mann-Whitney U test was applied. All statistical tests were two-sided, and a *p* value <0.05 was considered statistically significant. The development, parameter specification, hyperparameter tuning, and performance evaluation of machine learning models were conducted within the tidymodels framework. SHAP-based interpretability analyses were performed using the iml and fastshap packages, and visualizations were generated with the ggbeeswarm package. All statistical analyses, model development, validation, and interpretability were carried out in the R programming environment (version 4.5.1, Vienna, Austria).

Results

Patient characteristics

Based on the inclusion and exclusion criteria, 125 patients were excluded from the initial cohort of 476 children, resulting in a final study population of 351 pediatric WT patients. Among these, 51 patients (14.53%) experienced tumor recurrence, with 7 cases (13.73%) presenting with multi-site relapse. The median time to recurrence was 6 months (IQR: 4-16 months), with 80.4% of relapses occurring within one year. Sites of recurrence included the tumor bed (17 cases), lungs (16 cases), non-tumor-bed abdominal areas (13 cases), liver (5

cases), pelvis (3 cases), mediastinum (2 cases), pancreas (1 case), and orbit (1 case). Among all recurrent patients, 29 received chemotherapy and/or radiotherapy, 21 underwent surgical resection of recurrent lesions, and 1 patient was treated with arterial embolization. The entire cohort of 351 patients was randomly split into a training set and a testing set in a 7:3 ratio.

The training set included 245 patients, comprising 112 males and 133 females, with 35 cases of WT recurrence (14.29%). Comparative analysis revealed that recurrence was significantly associated with older age, larger tumor volume, presence of tumor thrombus, tumor rupture, higher COG tumor stage, positive lymph node status, unfavorable histology, elevated Ki-67 expression, and postoperative radiotherapy (**Table 1**). No statistically significant differences were observed between the training and testing sets across demographic and clinical variables (all $P > 0.05$) (**Table 2**).

Feature selection

A total of 17 candidate variables were initially considered in this study. Their predictive importance was evaluated using four methods: LASSO regression, the Boruta algorithm, subgroup analysis, and univariate logistic regression. Based on the intersection of the selected features from all four approaches, seven key predictive variables were ultimately identified: COG tumor stage, age, tumor rupture, histologic type (COG classification), presence of tumor thrombus, tumor volume, and Ki-67 index (**Figure 2**).

ML model development and performance validation

The predictive performance of nine machine learning algorithms for pediatric WT recurrence is summarized in **Table 3**. Among them, the SVM model demonstrated the best overall performance, achieving the highest AUC of 0.851, accuracy of 0.830, specificity of 0.856, positive predictive value (PPV) of 0.458, negative predictive value (NPV) of 0.939, and F1 score of 0.550. The model also exhibited relatively high recall (0.688) and Matthews correlation coefficient (MCC) of 0.465, indicating a well-balanced ability to distinguish between recur-

rent and non-recurrent patients. The confusion matrix showed that the SVM model correctly identified the majority of recurrent cases (11 patients) and non-recurrent cases (77 patients), while maintaining reasonable false positive (13 cases) and false negative (5 cases) rates. This suggests that the model effectively balances sensitivity and high specificity, yielding strong overall discriminative capability (**Figure 3**). DCA curve further demonstrated that the SVM model provided greater net benefit across most threshold probabilities, and its calibration curve closely aligned with the ideal reference line, indicating excellent calibration (**Figure 4**).

LR and LASSO regression achieved AUC values of 0.842 and 0.839, respectively, demonstrating high specificity and NPV, although their PPV and F1 scores were lower than those of the SVM model. Both RF and MLP showed high recall rates of 0.812, indicating strong ability to identify recurrent cases, but they exhibited lower accuracy and specificity. In contrast, XGBoost and LightGBM performed relatively poorly overall, with AUCs of 0.652 and 0.762, respectively, accompanied by low PPV and F1 scores. Taken together, SVM achieved the best balance between sensitivity and specificity among the nine machine learning algorithms, demonstrating the greatest potential as a predictive model for pediatric WT recurrence.

Interpretation of the optimal ML model

The SHAP plot provides a global perspective on the contribution of each feature to the predictions made by the SVM model, elucidating how the model integrates multiple variables to inform its decision-making. SHAP analysis identified the primary influential variables in descending order as follows: unfavorable histology (20.8%), COG stage IV-V (18.9%), presence of tumor thrombus (10.1%), Ki-67 index (9.9%), intraoperative tumor rupture (8.2%), preoperative tumor rupture (7.4%), combined preoperative and intraoperative rupture (6.7%), tumor volume (6.2%), COG stage III (6.1%), and age (5.6%) (**Figure 5**).

Moreover, the SHAP summary bar plot highlights that unfavorable histology and COG stage IV-V contribute most substantially to predicting Wilms tumor recurrence risk, significantly

Wilms tumor recurrence prediction with ML and Ki-67

Table 1. Demographic and clinical characteristics of children with Wilms tumor in the training set

Variables	Total (n = 245)	No Recurrence (n = 210)	Recurrence (n = 35)	Statistic	P
Age, M (Q ₁ , Q ₃)	27.00 (15.00, 46.00)	25.50 (13.25, 43.00)	45.00 (24.00, 65.50)	Z = -3.35	<.001
Gender, n (%)				$\chi^2 = 0.00$	1
Female	133 (54.29)	114 (54.29)	19 (54.29)		
Male	112 (45.71)	96 (45.71)	16 (45.71)		
Laterality, n (%)				$\chi^2 = 2.98$	0.226
Left	116 (47.35)	102 (48.57)	14 (40.00)		
Right	120 (48.98)	99 (47.14)	21 (60.00)		
Bilateral	9 (3.67)	9 (4.29)	0 (0.00)		
Low weight for age, n (%)				$\chi^2 = 2.54$	0.111
No	193 (78.78)	169 (80.48)	24 (68.57)		
Yes	52 (21.22)	41 (19.52)	11 (31.43)		
Tumor volume, M (Q ₁ , Q ₃)	502.08 (244.53, 789.08)	449.09 (212.87, 750.42)	662.04 (405.67, 1227.96)	Z = -2.92	0.003
Neoadjuvant chemotherapy, n (%)				$\chi^2 = 2.16$	0.142
No	179 (73.06)	157 (74.76)	22 (62.86)		
Yes	66 (26.94)	53 (25.24)	13 (37.14)		
Nephron sparing surgery, n (%)				$\chi^2 = 1.25$	0.264
No	222 (90.61)	188 (89.52)	34 (97.14)		
Yes	23 (9.39)	22 (10.48)	1 (2.86)		
Tumor thrombus, n (%)				$\chi^2 = 9.70$	0.002
No	221 (90.20)	195 (92.86)	26 (74.29)		
Yes	24 (9.80)	15 (7.14)	9 (25.71)		
Tumor rupture, n (%)				-	<.001
No rupture	179 (73.06)	165 (78.57)	14 (40.00)		
Pre-op rupture	30 (12.24)	22 (10.48)	8 (22.86)		
Intra-op rupture	26 (10.61)	17 (8.10)	9 (25.71)		
Pre- & intra-op rupture	10 (4.08)	6 (2.86)	4 (11.43)		
COG tumor stage, n (%)				$\chi^2 = 36.03$	<.001
Stage I-II	139 (56.73)	135 (64.29)	4 (11.43)		
Stage III	76 (31.02)	56 (26.67)	20 (57.14)		
Stage IV-V	30 (12.24)	19 (9.05)	11 (31.43)		
Number of lymph nodes examined, M (Q ₁ , Q ₃)	1.00 (0.00, 5.00)	1.00 (0.00, 4.75)	3.00 (0.00, 7.00)	Z = -1.48	0.138
Lymph node status, n (%)				$\chi^2 = 16.03$	<.001
Not examined	112 (45.71)	99 (47.14)	13 (37.14)		
Negative	116 (47.35)	102 (48.57)	14 (40.00)		
Positive	17 (6.94)	9 (4.29)	8 (22.86)		

Wilms tumor recurrence prediction with ML and Ki-67

Histology type (COG), n (%)				$\chi^2 = 9.17$	0.002
Favorable Histology	216 (88.16)	191 (90.95)	25 (71.43)		
Unfavorable Histology	29 (11.84)	19 (9.05)	10 (28.57)		
WT1, n (%)				$\chi^2 = 0.23$	0.628
Negative	42 (17.14)	35 (16.67)	7 (20.00)		
Positive	203 (82.86)	175 (83.33)	28 (80.00)		
Ki67, M (Q ₁ , Q ₃)	0.50 (0.30, 0.70)	0.40 (0.30, 0.70)	0.60 (0.40, 0.70)	Z = -2.33	0.02
Adjuvant chemotherapy, n (%)				-	1
No	4 (1.63)	4 (1.90)	0 (0.00)		
Yes	241 (98.37)	206 (98.10)	35 (100.00)		
Adjuvant radiotherapy, n (%)				$\chi^2 = 6.81$	0.009
No	210 (85.71)	185 (88.10)	25 (71.43)		
Yes	35 (14.29)	25 (11.90)	10 (28.57)		

Z: Mann-Whitney test, χ^2 : Chi-square test, M: Median, Q₁: 1st Quartile, Q₃: 3st Quartile.

Table 2. Demographic and clinical characteristics of children with Wilms tumor

Variables	Total (n = 351)	Train (n = 245)	Test (n = 106)	P	No Recurrence (n = 300)	Recurrence (n = 51)	P
Age, M (Q ₁ , Q ₃)	27.00 (14.00, 46.00)	27.00 (15.00, 46.00)	26.00 (13.00, 45.75)	0.863	26.00 (13.00, 43.25)	42.00 (20.50, 64.50)	0.002
Gender, n (%)				0.368			0.971
Female	185 (52.71)	133 (54.29)	52 (49.06)		158 (52.67)	27 (52.94)	
Male	166 (47.29)	112 (45.71)	54 (50.94)		142 (47.33)	24 (47.06)	
Laterality, n (%)				0.159			0.648
Left	178 (50.71)	116 (47.35)	62 (58.49)		154 (51.33)	24 (47.06)	
Right	161 (45.87)	120 (48.98)	41 (38.68)		135 (45.00)	26 (50.98)	
Bilateral	12 (3.42)	9 (3.67)	3 (2.83)		11 (3.67)	1 (1.96)	
Low weight for age, n (%)				0.616			0.184
No	279 (79.49)	193 (78.78)	86 (81.13)		242 (80.67)	37 (72.55)	
Yes	72 (20.51)	52 (21.22)	20 (18.87)		58 (19.33)	14 (27.45)	
Tumor volume, M (Q ₁ , Q ₃)	462.47 (250.23, 792.62)	502.08 (244.53, 789.08)	446.16 (263.97, 813.34)	0.701	448.11 (242.78, 751.71)	636.39 (353.74, 912.54)	0.01
Neoadjuvant chemotherapy, n (%)				0.248			0.148
No	250 (71.23)	179 (73.06)	71 (66.98)		218 (72.67)	32 (62.75)	
Yes	101 (28.77)	66 (26.94)	35 (33.02)		82 (27.33)	19 (37.25)	
Nephron sparing surgery, n (%)				0.989			0.234
No	318 (90.60)	222 (90.61)	96 (90.57)		269 (89.67)	49 (96.08)	
Yes	33 (9.40)	23 (9.39)	10 (9.43)		31 (10.33)	2 (3.92)	
Tumor thrombus, n (%)				0.095			<.001
No	310 (88.32)	221 (90.20)	89 (83.96)		276 (92.00)	34 (66.67)	
Yes	41 (11.68)	24 (9.80)	17 (16.04)		24 (8.00)	17 (33.33)	

Wilms tumor recurrence prediction with ML and Ki-67

Tumor rupture, n (%)				0.183			<.001
No rupture	248 (70.66)	179 (73.06)	69 (65.09)		229 (76.33)	19 (37.25)	
Pre-op rupture	46 (13.11)	30 (12.24)	16 (15.09)		35 (11.67)	11 (21.57)	
Intra-op rupture	37 (10.54)	26 (10.61)	11 (10.38)		24 (8.00)	13 (25.49)	
Pre- & intra-op rupture	20 (5.70)	10 (4.08)	10 (9.43)		12 (4.00)	8 (15.69)	
COG tumor stage, n (%)				0.345			<.001
Stage I-II	194 (55.27)	139 (56.73)	55 (51.89)		188 (62.67)	6 (11.76)	
Stage III	117 (33.33)	76 (31.02)	41 (38.68)		88 (29.33)	29 (56.86)	
Stage IV-V	40 (11.40)	30 (12.24)	10 (9.43)		24 (8.00)	16 (31.37)	
Number of lymph nodes examined, M (Q ₁ , Q ₃)	1.00 (0.00, 4.00)	1.00 (0.00, 5.00)	1.00 (0.00, 3.00)	0.423	1.00 (0.00, 4.00)	2.00 (0.00, 5.50)	0.176
Lymph node status, n (%)				0.235			0.002
Not examined	158 (45.01)	112 (45.71)	46 (43.40)		141 (47.00)	17 (33.33)	
Negative	173 (49.29)	116 (47.35)	57 (53.77)		147 (49.00)	26 (50.98)	
Positive	20 (5.70)	17 (6.94)	3 (2.83)		12 (4.00)	8 (15.69)	
Histology type (COG), n (%)				0.127			<.001
Favorable Histology	303 (86.32)	216 (88.16)	87 (82.08)		267 (89.00)	36 (70.59)	
Unfavorable Histology	48 (13.68)	29 (11.84)	19 (17.92)		33 (11.00)	15 (29.41)	
WT1, n (%)				0.485			0.768
Negative	57 (16.24)	42 (17.14)	15 (14.15)		48 (16.00)	9 (17.65)	
Positive	294 (83.76)	203 (82.86)	91 (85.85)		252 (84.00)	42 (82.35)	
Ki67, M (Q ₁ , Q ₃)	0.50 (0.30, 0.70)	0.50 (0.30, 0.70)	0.50 (0.30, 0.70)	0.148	0.45 (0.30, 0.70)	0.60 (0.40, 0.75)	0.004
Adjuvant chemotherapy, n (%)				0.438			1
No	4 (1.14)	4 (1.63)	0 (0.00)		4 (1.33)	0 (0.00)	
Yes	347 (98.86)	241 (98.37)	106 (100.00)		296 (98.67)	51 (100.00)	
Adjuvant radiotherapy, n (%)				0.974			0.004
No	301 (85.75)	210 (85.71)	91 (85.85)		264 (88.00)	37 (72.55)	
Yes	50 (14.25)	35 (14.29)	15 (14.15)		36 (12.00)	14 (27.45)	

Z: Mann-Whitney test, χ^2 : Chi-square test, M: Median, Q₁: 1st Quartile, Q₃: 3st Quartile.

[illegible]

exceeding the predictive impact of other clinical and pathological features. The distribution of SHAP values for individual variables is illustrated in **Figure 6**.

Wilms tumor is the most common renal malignancy in children and ranks as the fourth most

Table 3. Comparison of parameters among 9 machine learning algorithms for predicting recurrence of pediatric Wilms tumor

Model	AUC	Accuracy	Specificity	PPV	NPV	MCC	Balanced Accuracy	Recall	F1 Score
LR	0.842	0.802	0.833	0.400	0.926	0.387	0.729	0.625	0.488
LASSO	0.839	0.792	0.822	0.385	0.925	0.372	0.724	0.625	0.476
DT	0.711	0.792	0.856	0.350	0.895	0.268	0.647	0.438	0.389
RF	0.820	0.736	0.722	0.342	0.956	0.399	0.767	0.812	0.481
XGBoost	0.652	0.679	0.722	0.219	0.878	0.125	0.580	0.438	0.292
SVM	0.851	0.830	0.856	0.458	0.939	0.465	0.772	0.688	0.550
MLP	0.821	0.679	0.656	0.295	0.952	0.340	0.734	0.812	0.433
LightGBM	0.762	0.623	0.622	0.227	0.903	0.180	0.624	0.625	0.333
KNN	0.747	0.717	0.744	0.281	0.905	0.239	0.653	0.562	0.375

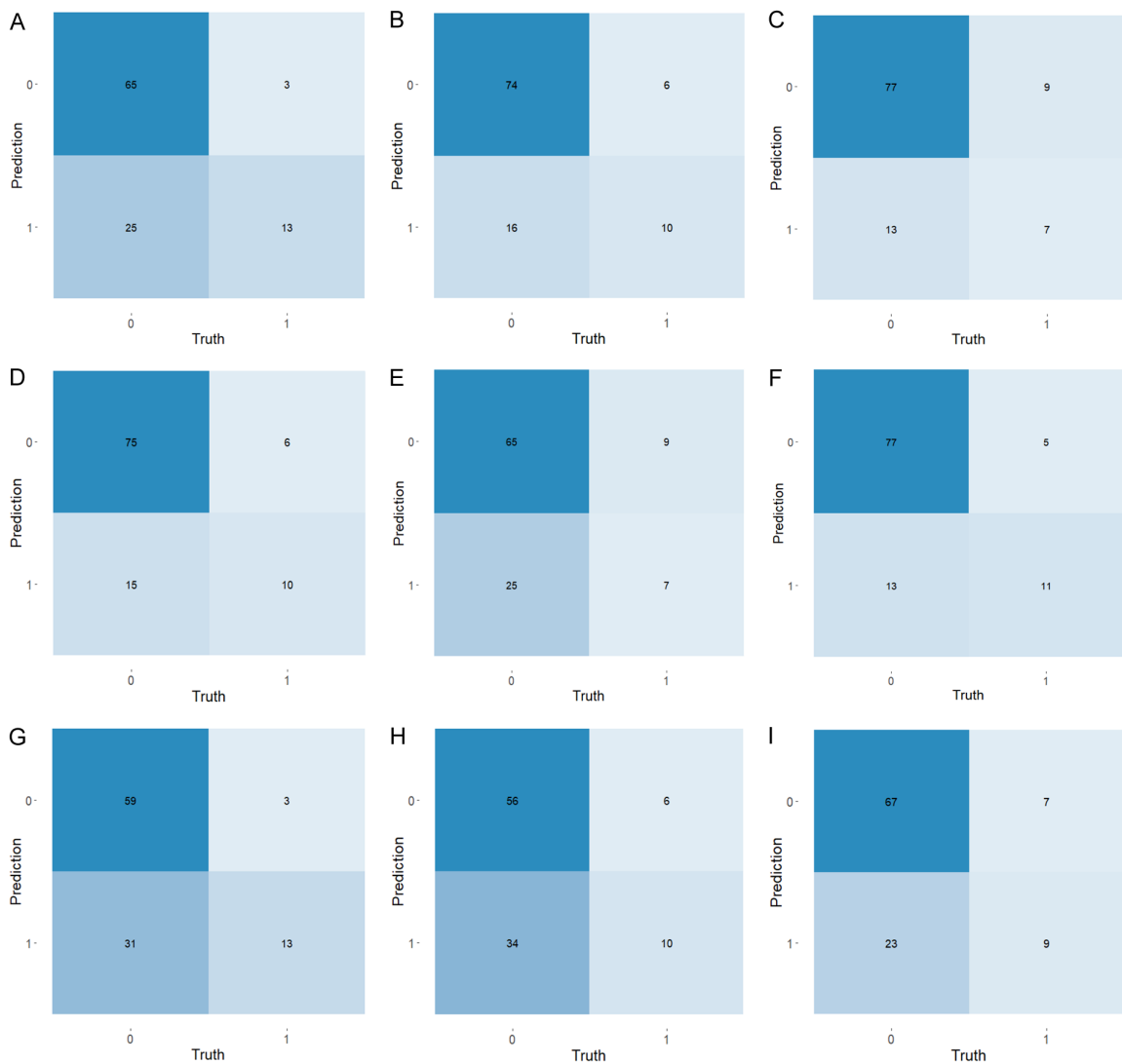


Figure 3. Confusion matrices of nine machine learning models on the test set. (A) Logistic Regression, (B) Least Absolute Shrinkage and Selection Operator, (C) Decision Tree, (D) Random Forest, (E) Extreme Gradient Boosting, (F) Support Vector Machine, (G) Single-Layer Neural Network, (H) Light Gradient Boosting Machine, and (I) K-Nearest Neighbors.

Wilms tumor recurrence prediction with ML and Ki-67

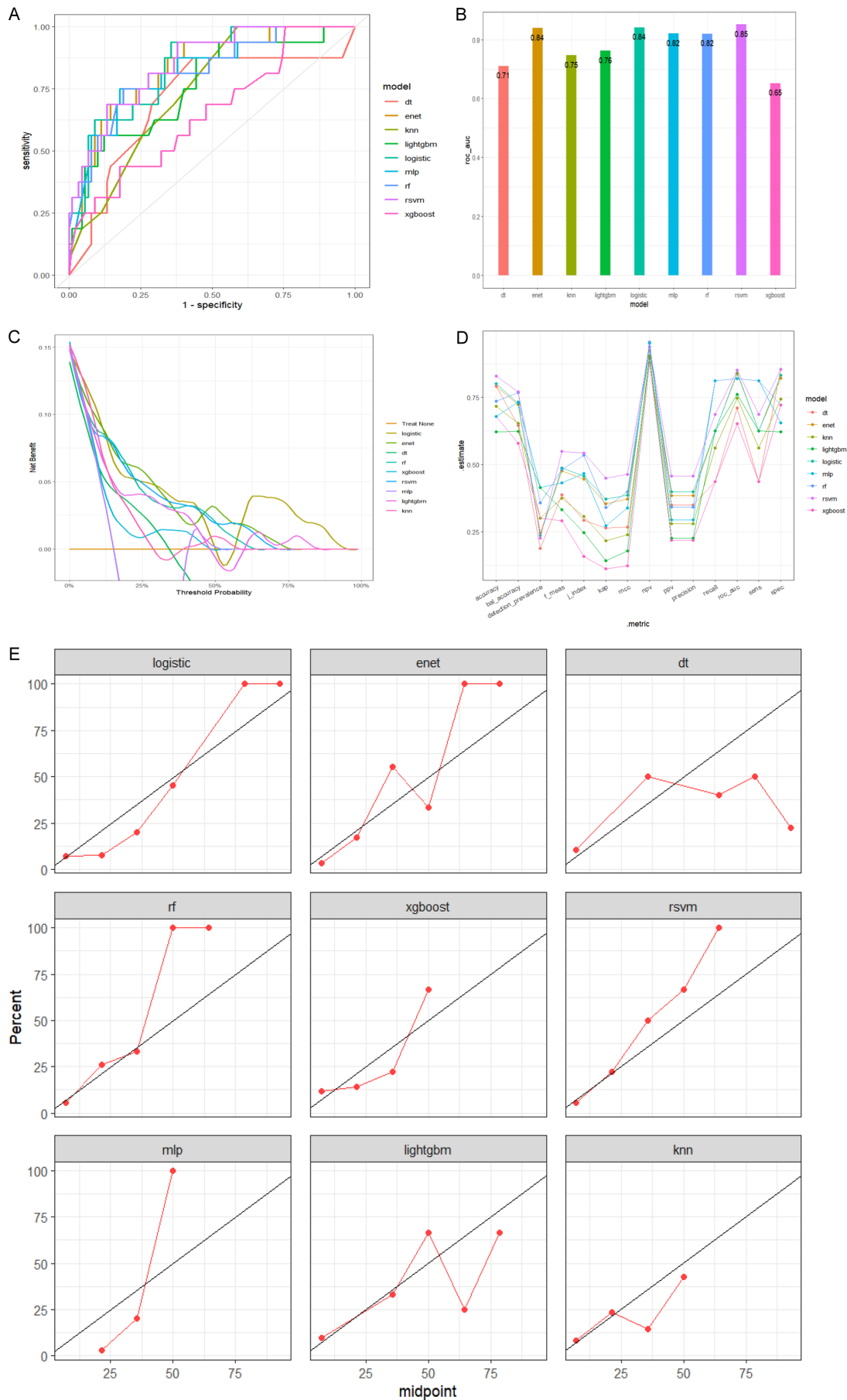


Figure 4. Performance comparison of nine machine learning models for predicting pediatric Wilms tumor recurrence. A. Receiver Operating Characteristic curves; B. Bar plot of Area Under the Curve values; C. Clinical Decision Curve Analysis; D. Line plots of various performance metrics; E. Calibration curves.

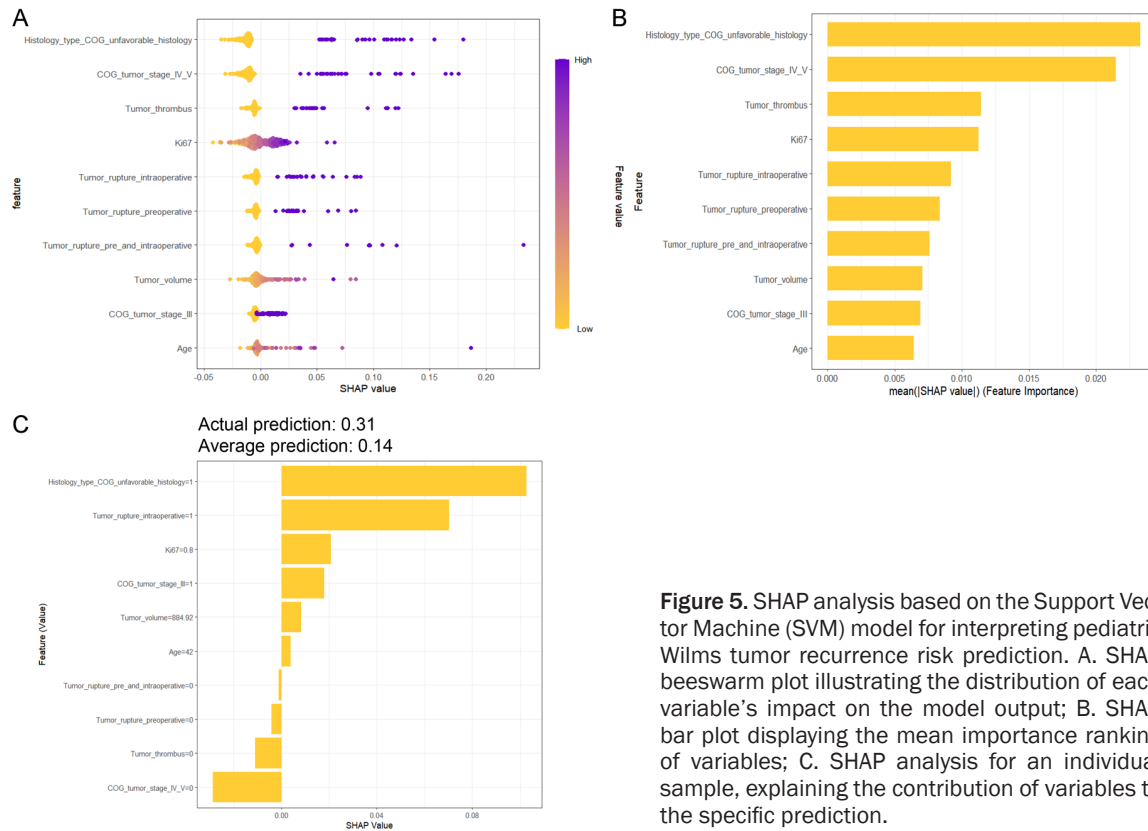


Figure 5. SHAP analysis based on the Support Vector Machine (SVM) model for interpreting pediatric Wilms tumor recurrence risk prediction. A. SHAP beeswarm plot illustrating the distribution of each variable's impact on the model output; B. SHAP bar plot displaying the mean importance ranking of variables; C. SHAP analysis for an individual sample, explaining the contribution of variables to the specific prediction.

frequent pediatric malignancy overall. Despite significant improvements in overall survival rates achieved through ongoing clinical trials and efforts by cooperative groups such as NWTSG, SIOP, and COG, approximately 15% of patients still experience disease recurrence [19-21]. The majority of relapses occur within two years after surgery and are associated with multiple high-risk factors, including bilateral WT, unfavorable histologic subtypes, and advanced tumor staging [3, 5, 22]. Studies have demonstrated that over 60% of WT recurrences are not detectable by physical examination alone, necessitating reliance on imaging modalities for diagnosis. Detecting a single sub-clinical relapse within the first two years post-operatively may require an average of 112 imaging scans, increasing to as many as 500 scans beyond two years, imposing a substantial burden on both patients' families and healthcare systems [5]. Therefore, developing precise predictive models and identifying key

risk factors for recurrence are crucial for optimizing surveillance and management of pediatric WT patients. Such models can help avoid excessive imaging in low-risk patients, thereby reducing unnecessary radiation exposure, and alleviate psychological stress experienced by patients and families during the uncertain phases of disease progression [23].

To more accurately predict WT recurrence and identify the optimal predictive model, this study employed four complementary feature selection methods - univariate logistic regression, LASSO regression, Boruta feature importance analysis, and subgroup analysis - to enhance the robustness of variable selection and mitigate the limitations inherent to any single method, especially considering that logistic regression alone cannot fully address multicollinearity among predictors [24]. Variables consistently identified across these methods were regarded as the most reliable predictors. Ultimately,

Wilms tumor recurrence prediction with ML and Ki-67

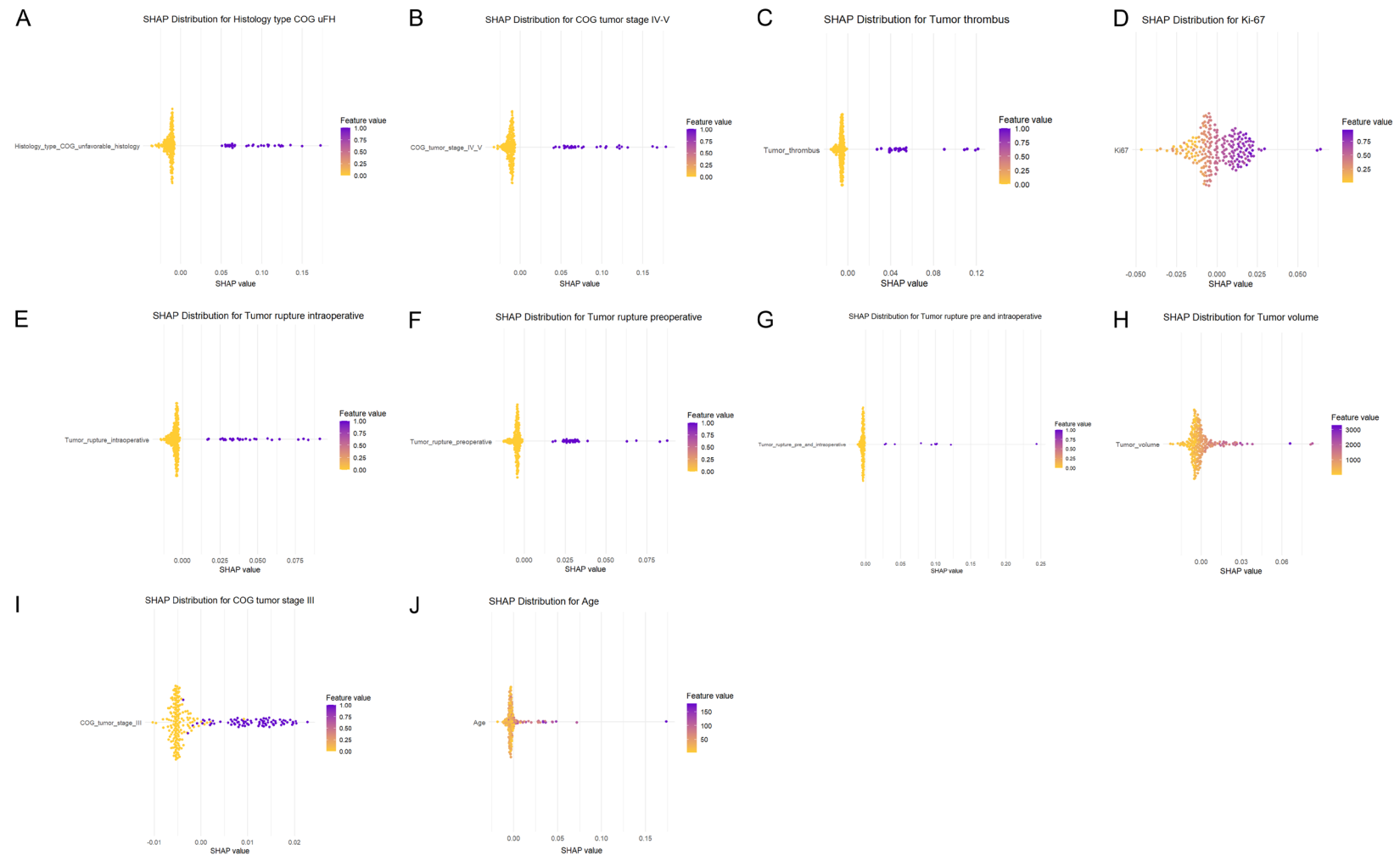


Figure 6. Distribution of SHAP values for individual variables ranked by their impact strength, including (A) unfavorable histologic, (B) COG stage IV-V, (C) presence of tumor thrombus, (D) Ki-67 index, (E) intraoperative tumor rupture, (F) preoperative tumor rupture, (G) combined preoperative and intraoperative rupture, (H) tumor volume, (I) COG stage III, and (J) age.

among 17 candidate variables, seven independent risk factors were identified: COG tumor stage, age, tumor rupture, unfavorable histologic subtype, tumor thrombus, tumor volume, and Ki-67 expression level. COG staging serves as a central reference for WT treatment and prognosis, with its correlation to recurrence risk well established in multiple studies [25-27]. Age is widely recognized as a critical prognostic factor; older children generally have poorer outcomes and higher recurrence risk, which is also associated with more advanced tumor stages and a higher prevalence of high-risk histologic subtypes - likely reflecting differences in tumor biology and therapeutic response [8, 25, 28]. Studies report that the incidence of spontaneous or traumatic tumor rupture detected by preoperative imaging ranges from 3% to 23%, with up to 10% risk of intraoperative tumor spillage in patients not receiving preoperative chemotherapy [26]. Tumor rupture can facilitate dissemination of tumor cells and is a known risk factor for local recurrence. Evidence suggests that any form of intraoperative rupture increases the risk of recurrence approximately threefold [29, 30].

Unfavorable histologic subtype, specifically anaplastic Wilms tumor, carries a recurrence risk three- to fivefold higher than that of favorable histology and is frequently associated with MYCN and TP53 mutations. These mutations are found in approximately 5%-10% of WT cases and are indicative of increased tumor aggressiveness and chemoresistance [31, 32]. Irtan et al. [7], analyzing patients from the UKW3 trial, reported that in univariate analysis, biopsy status, unfavorable histology, and tumor size were associated with an increased risk of local recurrence, while age, unfavorable histology, tumor size, and lymph node metastasis significantly elevated the risk of distant relapse. The presence of tumor thrombus is also a critical risk factor for WT recurrence. Tumor extension into the renal vein or inferior vena cava occurs in 4% to 10% of cases, with 1% to 3% extending into the right atrium or ventricle, reflecting vascular invasion that serves as an important marker for systemic metastasis and recurrence [33, 34]. Because preoperative imaging does not always reliably detect venous tumor thrombi, intraoperative palpation and/or intraoperative ultrasound are typically employed to assess renal

vein involvement. Tumor thrombus removal is generally performed concurrently with tumor resection; failure to do so results in automatic upstaging to stage III pathologically [35]. Consequently, NWTS/COG protocols recommend neoadjuvant chemotherapy when the thrombus extends to the hepatic veins or beyond [36, 37]. Additionally, tumor volume is an important risk factor for WT recurrence. Larger tumor volumes are associated with increased cellular heterogeneity, which may raise the likelihood of residual disease and tumor rupture [25, 38].

Ki-67 is a nuclear protein expressed during all active phases of the cell cycle (G1, S, G2, and M phases) but absent during the resting phase (G0). Its expression is regulated by phosphorylation and it is widely recognized as a classical marker of cellular proliferation [39]. A high Ki-67 index reflects a large proportion of tumor cells in active proliferation, indicating increased tumor aggressiveness and malignancy, which correlates with a higher risk of recurrence. High Ki-67 expression has been demonstrated in numerous studies of adult and pediatric solid tumors to be closely associated with poor prognosis [40-42]. In WT, Ki-67 has been employed to assess tumor proliferative activity and staging; however, previous clinical studies have not definitively established its independent association with recurrence, possibly due to small sample sizes and nonlinear data relationships.

In this study, machine learning algorithms demonstrated the independent prognostic value of Ki-67 in predicting WT recurrence, providing important clues for subsequent precision treatment [43]. For example, M. Atwa et al. [44] conducted a retrospective analysis of 75 WT patients and found that cyclin A immunopositivity correlated with higher recurrence rates, whereas Ki-67 showed no statistically significant association with adverse outcomes. Similarly, Jurić et al. [45] used immunohistochemical analysis of Ki-67 expression in paraffin-embedded tissues from 48 pediatric WT cases, finding correlations between Ki-67 levels, histologic subtype, and stage, but emphasized the need for larger sample sizes and deeper analyses to clarify the prognostic significance of Ki-67 in WT. Additionally, in our study, neither lymph node status nor the number of lymph node biopsies emerged as independent

risk factors for WT recurrence. This result may be attributed to two key factors. First, lymph node biopsy rates are inherently low in pediatric populations. Second, the younger age of patients at our center makes intraoperative lymph node identification and dissection technically challenging, resulting in a limited number of harvested nodes and potentially reducing the likelihood of detecting positive nodes [46]. Therefore, lymph node-related variables may differ significantly across age groups and institutions, limiting their generalizability as predictors of recurrence.

Based on the seven selected variables, we developed nine machine learning-based predictive models and conducted a systematic comparison. The SVM model emerged as the top performer, outperforming the traditional LR model - achieving higher AUC (0.851 vs. 0.842), accuracy (0.830 vs. 0.802), specificity (0.856 vs. 0.833), and F1 score (0.550 vs. 0.488) - and surpassing all other algorithms. This superiority reflects SVM's intrinsic strengths: by selecting support vectors via the maximum-margin criterion, it effectively mitigates overfitting in high-dimensional, small-sample settings and enhances generalizability [47]; unlike LR's reliance on a log-linear relationship, SVM makes no strict linearity assumptions and remains stable even when the number of variables approaches sample size [48]; and its focus on boundary samples confers natural robustness to noise and outliers [49]. Moreover, as a purely data-driven method, SVM retains strong discriminative power in limited-sample contexts [50], highlighting its potential as a risk-stratification tool for WT recurrence rather than merely a "black-box" classifier.

In this study, the SVM model combined with the SHAP analysis clearly illustrated the global contribution of each feature to recurrence prediction, demonstrating inherent interpretability and a solid theoretical foundation. SHAP results revealed that unfavorable histology and COG stage IV-V had the greatest impact on model output, significantly exceeding other variables. This finding aligns with previous conclusions from the NWTs-4 and SIOP studies, indirectly validating the model's predictive accuracy [29, 51]. Additionally, factors such as tumor thrombus formation and Ki-67 index ranked highly, reflecting the combined effects of multidimen-

sional biological mechanisms on recurrence risk and further complementing existing research findings [26]. Moreover, the quantitative assessment provided by SHAP, which accounts for both positive and negative effects and their magnitudes, offers a more comprehensive basis for variable selection compared to traditional methods that rely solely on *p*-values, thereby better capturing the true contribution of variables within complex models and enhancing clinical trustworthiness.

WT commonly recurs within two years following nephrectomy; in this study, the median time to recurrence was 6 months, with approximately 80% occurring within the first postoperative year, consistent with previous literature. Brok et al. [5] retrospectively analyzed the International Society of Paediatric Oncology Renal Tumor Study Group (RTSG-SIOP) database from 2001 and found that among 4,271 eligible WT patients, 538 (13%) experienced recurrence, with about 80% of recurrences occurring within two years after nephrectomy; notably, 70% of these recurrences were detected through scheduled surveillance imaging. Similarly, Fernandez et al. [52] analyzed 116 patients with very low-risk WT and reported a median time to first recurrence of 4.3 months, with 91.6% of recurrences occurring within two years after treatment. Therefore, for patients at low risk of recurrence, it is recommended to perform abdominal ultrasound and chest X-ray every three months during the first two years following radical surgery, then every 4 to 6 months during years three to four. Follow-up frequency beyond five years should be tailored to clinical needs. For high-risk patients, shorter intervals between imaging studies are advised, utilizing higher-sensitivity modalities such as CT or MRI for monitoring [23].

This study has several limitations. First, the use of a retrospective cohort for model development may introduce selection bias; prospective datasets are generally more suitable for enhancing predictive accuracy. Second, although we aimed to include all relevant variables, some early cases lacked critical genetic information, such as 1p and 11q status. These are recognized prognostic biomarkers in Wilms tumor, and their absence may have limited the accuracy and generalizability of the model [26]. Third, chemotherapy regimens varied across

different periods, and differences in treatment intensity may have introduced confounding effects on recurrence risk. Finally, this study was based on single-center data with a limited sample size. Although internal validation was performed, the lack of external or multicenter cohort validation limits the generalizability of the results. Future work will focus on multicenter, large-scale prospective cohort studies to validate and improve the model's stability, generalizability, and clinical utility.

Conclusions

This study identified key clinical features by integrating four feature selection methods and developed nine machine learning models to predict postoperative relapse in Wilms' tumor. The results demonstrate that machine learning can effectively predict Wilms' tumor recurrence, with the Ki-67 index showing strong independent prognostic value. Among the models, the SVM model exhibited superior performance. This model represents a reliable tool for forecasting postoperative relapse and can assist clinicians in risk stratification and planning individualized follow-up strategies for pediatric WT patients.

Acknowledgements

The authors would like to thank the all the team members for their great assistance in this study. This study was supported by the Chongqing key project of technology innovation and application (2024TIAD-KPX0035) and Joint project of Chongqing Health Commission and Science and Technology Bureau (2025ZDXM038).

Project Viva mothers provided written informed consent for their own participation and that of their children.

Disclosure of conflict of interest

None.

Abbreviations

AUC, Area under the receiver operating characteristic curve; COG, Children's Oncology Group; CHCMU, Children's Hospital of Chongqing Medical University; CT, Computed tomography; DCA, Decision curve analysis; DT, Decision tree;

IHC, Immunohistochemistry; IKWG, International Ki-67 Working Group; IQR, Interquartile range; KNN, K-nearest neighbors; LASSO, Least absolute shrinkage and selection operator; LightGBM, Light gradient boosting machine; LR, Logistic regression; MCC, Matthews correlation coefficient; ML, Machine learning; MRI, Magnetic resonance imaging; NN, Neural network; NPV, Negative predictive value; PPV, Positive predictive value; RF, Random forest; ROC, Receiver operating characteristic; RTSG, Renal Tumor Study Group; SHAP, SHapley additive exPlanations; SIOP, International Society of Pediatric Oncology; SVM, Support vector machine; WT, Wilms tumor; XGBoost, Extreme gradient boosting.

Address correspondence to: Deying Zhang, Department of Urology, Children's Hospital of Chongqing Medical University, No. 20 Jinyu Avenue, Liangjiang New Area, Chongqing 401122, China. Tel: +86-13594136249; E-mail: zdy@hospital.cqmu.edu.cn

References

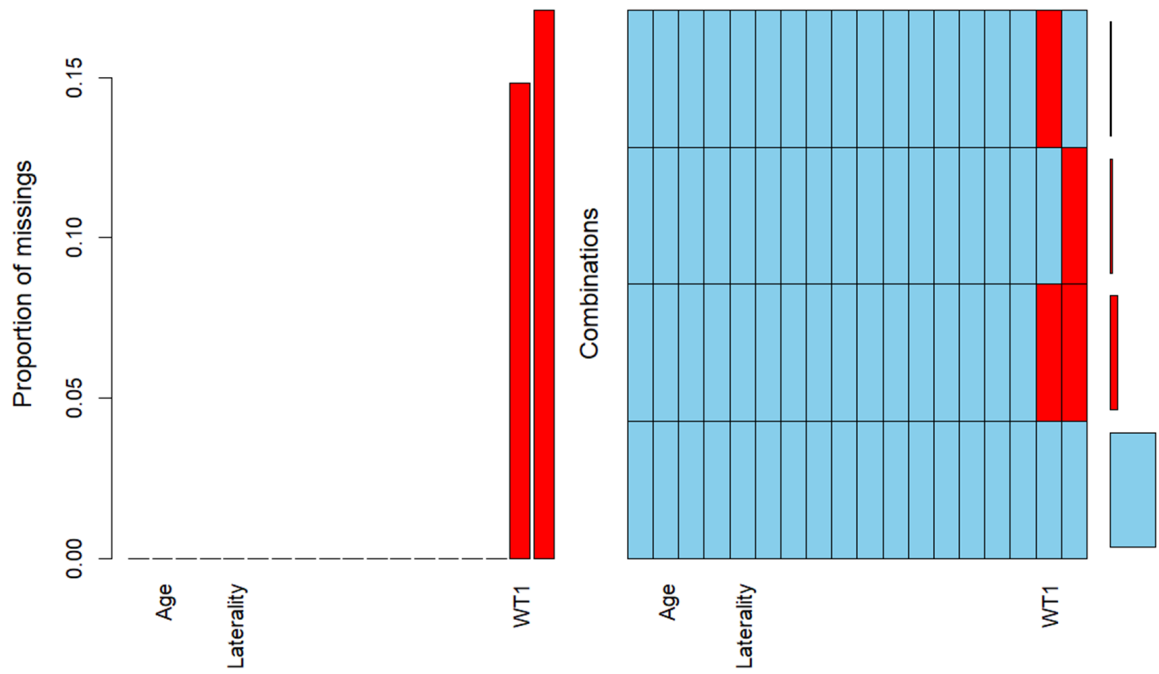
- [1] Spreafico F, Fernandez CV, Brok J, Nakata K, Vujanic G, Geller JI, Gessler M, Maschietto M, Behjati S, Polanco A, Paintsil V, Luna-Fineman S and Pritchard-Jones K. Wilms tumour. *Nat Rev Dis Primers* 2021; 7: 75.
- [2] Palmisani F, Kovar H, Kager L, Amann G, Metzelder M and Bergmann M. Systematic review of the immunological landscape of Wilms tumors. *Mol Ther Oncolytics* 2021; 22: 454-467.
- [3] Malogolowkin M, Spreafico F, Dome JS, van Tinteren H, Pritchard-Jones K, van den Heuvel-Eibrink MM, Bergeron C, de Kraker J and Graf N; COG Renal Tumors Committee and the SIOP Renal Tumor Study Group. Incidence and outcomes of patients with late recurrence of Wilms' tumor. *Pediatr Blood Cancer* 2013; 60: 1612-1615.
- [4] Spreafico F, Pritchard Jones K, Malogolowkin MH, Bergeron C, Hale J, de Kraker J, Dallorso S, Acha T, de Camargo B, Dome JS and Graf N. Treatment of relapsed Wilms tumors: lessons learned. *Expert Rev Anticancer Ther* 2009; 9: 1807-1815.
- [5] Brok J, Lopez-Yurda M, Tinteren HV, Treger TD, Furtwängler R, Graf N, Bergeron C, van den Heuvel-Eibrink MM, Pritchard-Jones K, Olsen ØE, de Camargo B, Verschuur A and Spreafico F. Relapse of Wilms' tumour and detection methods: a retrospective analysis of the 2001 Renal Tumour Study Group-International Society of Paediatric Oncology Wilms' tumour pro-

- tocol database. *Lancet Oncol* 2018; 19: 1072-1081.
- [6] Nelson MV, van den Heuvel-Eibrink MM, Graf N and Dome JS. New approaches to risk stratification for Wilms tumor. *Curr Opin Pediatr* 2021; 33: 40-48.
- [7] Irtan S, Jitlal M, Bate J, Powis M, Vujanic G, Kelsey A, Walker J, Mitchell C, Machin D and Pritchard-Jones K; Renal Tumours Committee of the Children's Cancer and Leukaemia Group (CCLG). Risk factors for local recurrence in Wilms tumour and the potential influence of biopsy - the United Kingdom experience. *Eur J Cancer* 2015; 51: 225-232.
- [8] Grundy PE, Breslow NE, Li S, Perlman E, Beckwith JB, Ritchey ML, Shamberger RC, Haase GM, D'Angio GJ, Donaldson M, Coppes MJ, Malogolowkin M, Shearer P, Thomas PR, Macklis R, Tomlinson G, Huff V and Green DM; National Wilms Tumor Study Group. Loss of heterozygosity for chromosomes 1p and 16q is an adverse prognostic factor in favorable-histology Wilms tumor: a report from the National Wilms Tumor Study Group. *J Clin Oncol* 2005; 23: 7312-7321.
- [9] Ngiam KY and Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol* 2019; 20: e262-e273.
- [10] Wu WT, Li YJ, Feng AZ, Li L, Huang T, Xu AD and Lyu J. Data mining in clinical big data: the frequently used databases, steps, and methodological models. *Mil Med Res* 2021; 8: 44.
- [11] Kalapurakal JA, Li SM, Breslow NE, Beckwith JB, Macklis R, Thomas PR, D'Angio GJ, Kim T, de Lorimier A, Kelalis P, Shochat S, Ritchey M, Haase G, Hrabovsky E, Otherson HB, Grundy P and Green DM; National Wilms' Tumor Study Group. Influence of radiation therapy delay on abdominal tumor recurrence in patients with favorable histology Wilms' tumor treated on NWTs-3 and NWTs-4: a report from the National Wilms' Tumor Study Group. *Int J Radiat Oncol Biol Phys* 2003; 57: 495-499.
- [12] Kieran K, Williams MA, McGregor LM, Dome JS, Krasin MJ and Davidoff AM. Repeat nephron-sparing surgery for children with bilateral Wilms tumor. *J Pediatr Surg* 2014; 49: 149-153.
- [13] Firoozi F and Kogan BA. Follow-up and management of recurrent Wilms' tumor. *Urol Clin North Am* 2003; 30: 869-879.
- [14] Joseph LL, Boddu D, Srinivasan HN, Regi SS, Antonisamy B, John R, Mathew LG and Totadri S. Postchemotherapy tumor volume as a prognostic indicator in Wilms tumor: a single-center experience from South India. *Pediatr Blood Cancer* 2022; 69: e29454.
- [15] Madewell ZJ, Keita AM, Das PM, Mehta A, Akeilo V, Oluoch OB, Omoro R, Onyango D, Sagam CK, Cain CJ, Chukwuegbo C, Kaluma E, Luke R, Ogbuanu IU, Bassat Q, Kincardett M, Mandomando I, Rakislova N, Varo R, Xerinda EG, Dangor Z, du Toit J, Lala SG, Madhi SA, Mahtab S, Breines MR, Degefa K, Heluf H, Madrid L, Scott JAG, Sow SO, Tapia MD, El Arifeen S, Gurley ES, Hossain MZ, Islam KM, Rahman A, Mutevedzi PC, Whitney CG, Blau DM, Suchdev PS and Kotloff KL; Child Health and Mortality Prevention Surveillance Network. Contribution of malnutrition to infant and child deaths in Sub-Saharan Africa and South Asia. *BMJ Glob Health* 2024; 9: e017262.
- [16] Leung SCY, Nielsen TO, Zabaglio L, Arun I, Badve SS, Bane AL, Bartlett JMS, Borgquist S, Chang MC, Dodson A, Enos RA, Fineberg S, Focke CM, Gao D, Gown AM, Grabau D, Gutierrez C, Hugh JC, Kos Z, Lænkholm AV, Lin MG, Mastropasqua MG, Moriya T, Nofech-Mozes S, Osborne CK, Penault-Llorca FM, Piper T, Sakatani T, Salgado R, Stanczyński J, Viale G, Hayes DF, McShane LM and Dowsett M. Analytical validation of a standardized scoring protocol for Ki67: phase 3 of an international multicenter collaboration. *NPJ Breast Cancer* 2016; 2: 16014.
- [17] Li J, Guo S, Ma R, He J, Zhang X, Rui D, Ding Y, Li Y, Jian L, Cheng J and Guo H. Comparison of the effects of imputation methods for missing data in predictive modelling of cohort study datasets. *BMC Med Res Methodol* 2024; 24: 41.
- [18] Zhou H, Xin Y and Li S. A diabetes prediction model based on Boruta feature selection and ensemble learning. *BMC Bioinformatics* 2023; 24: 224.
- [19] Balis F, Green DM, Anderson C, Cook S, Dhillon J, Gow K, Hiniker S, Jasty-Rao R, Lin C, Lovvorn H, MacEwan I, Martinez-Agosto J, Mullen E, Murphy ES, Ranalli M, Rhee D, Rokitka D, Tracy EL, Vern-Gross T, Walsh MF, Walz A, Wickiser J, Zapala M, Berardi RA and Hughes M. Wilms Tumor (Nephroblastoma), Version 2.2021, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* 2021; 19: 945-977.
- [20] Fernandez CV, Mullen EA, Chi YY, Ehrlich PF, Perlman EJ, Kalapurakal JA, Khanna G, Paulino AC, Hamilton TE, Gow KW, Tochner Z, Hoffer FA, Withycombe JS, Shamberger RC, Kim Y, Geller JI, Anderson JR, Grundy PE and Dome JS. Outcome and prognostic factors in stage III favorable-histology Wilms tumor: a report from the Children's Oncology Group Study AREN0532. *J Clin Oncol* 2018; 36: 254-261.
- [21] Libes J, Hol J, Neto JCA, Vallance KL, Tinteren HV, Benedetti DJ, Villar GLR, Duncan C and Ehrlich PF. Pediatric renal tumor epidemiology: global perspectives, progress, and challenges. *Pediatr Blood Cancer* 2023; 70: e30006.

- [22] Pritchard-Jones K, Moroz V, Vujančić G, Powis M, Walker J, Messahel B, Hobson R, Levitt G, Kelsey A and Mitchell C. Treatment and outcome of Wilms' tumour patients: an analysis of all cases registered in the UKW3 trial. *Ann Oncol* 2012; 23: 2457-2463.
- [23] Lovvorn HN 3rd. Optimising surveillance for relapse of Wilms' tumour. *Lancet Oncol* 2018; 19: 1006-1007.
- [24] Zhang YB, Yang G, Bu Y, Lei P, Zhang W and Zhang DY. Development of a machine learning-based model for predicting risk of early post-operative recurrence of hepatocellular carcinoma. *World J Gastroenterol* 2023; 29: 5804-5817.
- [25] Hol JA, Lopez-Yurda MI, Van Tinteren H, Van Grotel M, Godzinski J, Vujanic G, Oldenburger F, De Camargo B, Ramírez-Villar GL, Bergeron C, Pritchard-Jones K, Graf N and Van den Heuvel-Eibrink MM. Prognostic significance of age in 5631 patients with Wilms tumour prospectively registered in International Society of Paediatric Oncology (SIOP) 93-01 and 2001. *PLoS One* 2019; 14: e0221373.
- [26] Groenendijk A, Spreafico F, de Krijger RR, Drost J, Brok J, Perotti D, van Tinteren H, Venkatramani R, Godziński J, Rube C, Geller JI, Graf N, van den Heuvel-Eibrink MM and Mavinkurve-Groothuis AMC. Prognostic factors for Wilms tumor recurrence: a review of the literature. *Cancers (Basel)* 2021; 13: 3142.
- [27] Verschuur A, Van Tinteren H, Graf N, Bergeron C, Sandstedt B and de Kraker J. Treatment of pulmonary metastases in children with stage IV nephroblastoma with risk-based use of pulmonary radiotherapy. *J Clin Oncol* 2012; 30: 3533-3539.
- [28] Breslow N, Sharples K, Beckwith JB, Takashima J, Kelalis PP, Green DM and D'Angio GJ. Prognostic factors in nonmetastatic, favorable histology Wilms' tumor. Results of the Third National Wilms' tumor study. *Cancer* 1991; 68: 2345-2353.
- [29] Shamberger RC, Guthrie KA, Ritchey ML, Haase GM, Takashima J, Beckwith JB, D'Angio GJ, Green DM and Breslow NE. Surgery-related factors and local recurrence of Wilms tumor in National Wilms tumor study 4. *Ann Surg* 1999; 229: 292-297.
- [30] Green DM, Breslow NE, D'Angio GJ, Malogolowkin MH, Ritchey ML, Evans AE, Beckwith JB, Perlman EJ, Shamberger RC, Peterson S, Grundy PE, Dome JS, Thomas PR and Kalapurakal JA. Outcome of patients with Stage II/ favorable histology Wilms tumor with and without local tumor spill: a report from the National Wilms tumor study group. *Pediatr Blood Cancer* 2014; 61: 134-139.
- [31] Williams RD, Chagtai T, Alcaide-German M, Apps J, Wegert J, Popov S, Vujanic G, van Tinteren H, van den Heuvel-Eibrink MM, Kool M, de Kraker J, Gisselsson D, Graf N, Gessler M and Pritchard-Jones K. Multiple mechanisms of MYCN dysregulation in Wilms tumour. *Oncotarget* 2015; 6: 7232-7243.
- [32] Bardeesy N, Falkoff D, Petruzzi MJ, Nowak N, Zabel B, Adam M, Aguiar MC, Grundy P, Shows T and Pelletier J. Anaplastic Wilms' tumour, a subtype displaying poor prognosis, harbours p53 gene mutations. *Nat Genet* 1994; 7: 91-97.
- [33] Dzhuma K, Powis M, Vujanic G, Watson T, Olsen O, Shelmerdine S, Oostveen M, Brok J, Irtan S, Williams R, Tugnait S, Smeulders N, Mushtaq I, Chowdhury T, Al-Saadi R and Pritchard-Jones K. Surgical management, staging, and outcomes of Wilms tumours with intravascular extension: results of the IMPORT study. *J Pediatr Surg* 2022; 57: 572-578.
- [34] Elayadi M, Hammad M, Sallam K, Ahmed G, Ahmed S, Ibrahim A, Refaat A, Elkinaai N, Younes A, Graf N and Zekri W. Management and outcome of pediatric Wilms tumor with malignant inferior Vena cava thrombus: largest cohort of single-center experience. *Int J Clin Oncol* 2020; 25: 1425-1431.
- [35] Theilen TM, Braun Y, Bochennek K, Rolle U, Fiegel HC and Friedmacher F. Multidisciplinary treatment strategies for Wilms tumor: recent advances, technical innovations and future directions. *Front Pediatr* 2022; 10: 852185.
- [36] Kieran K and Ehrlich PF. Current surgical standards of care in Wilms tumor. *Urol Oncol* 2016; 34: 13-23.
- [37] Shamberger RC, Ritchey ML, Haase GM, Bergemann TL, Loecheit-Yoshioka T, Breslow NE and Green DM. Intravascular extension of Wilms tumor. *Ann Surg* 2001; 234: 116-121.
- [38] Gow KW, Barnhart DC, Hamilton TE, Kandel JJ, Chen MK, Ferrer FA, Price MR, Mullen EA, Geller JI, Gratias EJ, Rosen N, Khanna G, Naranjo A, Ritchey ML, Grundy PE, Dome JS and Ehrlich PF. Primary nephrectomy and intraoperative tumor spill: report from the Children's Oncology Group (COG) renal tumors committee. *J Pediatr Surg* 2013; 48: 34-38.
- [39] Menon SS, Guruvayoorappan C, Sakthivel KM and Rasmi RR. Ki-67 protein as a tumour proliferation marker. *Clin Chim Acta* 2019; 491: 39-45.
- [40] Lu L, Wan X, Xu Y, Chen J, Shu K and Lei T. Prognostic factors for recurrence in pituitary adenomas: recent progress and future directions. *Diagnostics (Basel)* 2022; 12: 977.
- [41] Kitsel Y, Cooke T, Sotirchos V and Sofocleous CT. Colorectal cancer liver metastases: genom-

- ics and biomarkers with focus on local therapies. *Cancers (Basel)* 2023; 15: 1679.
- [42] Ferràu F, Giuffrida G, Casablanca R, Alessi Y, Tuccari G, Cotta OR, Angileri FF and Cannavò S. Clinical and prognostic implications of pituitary macroadenomas (PitNets) grading: a monocentric experience. *Pituitary* 2025; 28: 41.
- [43] Krishna OH, Kayla G, Abdul Aleem M, Malleboyina R and Reddy Kota R. Immunohistochemical expression of Ki67 and p53 in Wilms tumor and its relationship with tumor histology and stage at presentation. *Patholog Res Int* 2016; 2016: 6123951.
- [44] Atwa AM, Hafez AT, Abdelhameed M, Dawaba M, Nabeeh A and Helmy TE. Does immunohistochemical staining of P53, Ki 67 and cyclin A accurately predict Wilms tumor recurrence and survival? *Arab J Urol* 2022; 20: 107-114.
- [45] Jurić I, Pogorelić Z, Kuzmić-Prusac I, Biocić M, Jakovljević G, Stepan J, Zupancić B, Culić S and Kruslin B. Expression and prognostic value of the Ki-67 in Wilms' tumor: experience with 48 cases. *Pediatr Surg Int* 2010; 26: 487-493.
- [46] Tan X, Wang J, Tang J, Tian X, Jin L, Li M, Zhang Z and He D. A nomogram for predicting cancer-specific survival in children with Wilms tumor: a study based on SEER database and external validation in China. *Front Public Health* 2022; 10: 829840.
- [47] Quéllec G, Kowal J, Hasler PW, Scholl HPN, Zweifel S, Konstantinos B, de Carvalho JER, Heeren T, Egan C, Tufail A and Maloca PM. Feasibility of support vector machine learning in age-related macular degeneration using small sample yielding sparse optical coherence tomography data. *Acta Ophthalmol* 2019; 97: e719-e728.
- [48] Zhou X, Li X, Zhang Z, Han Q, Deng H, Jiang Y, Tang C and Yang L. Support vector machine deep mining of electronic medical records to predict the prognosis of severe acute myocardial infarction. *Front Physiol* 2022; 13: 991990.
- [49] Son YJ, Kim HG, Kim EH, Choi S and Lee SK. Application of support vector machine for prediction of medication adherence in heart failure patients. *Healthc Inform Res* 2010; 16: 253-259.
- [50] Yu W, Liu T, Valdez R, Gwinn M and Khoury MJ. Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. *BMC Med Inform Decis Mak* 2010; 10: 16.
- [51] Reinhard H, Schmidt A, Furtwängler R, Leuschner I, Rübe C, Von Schweinitz D, Zoubek A, Niggli F and Graf N. Outcome of relapses of nephroblastoma in patients registered in the SIOP/GPOH trials and studies. *Oncol Rep* 2008; 20: 463-467.
- [52] Fernandez CV, Perlman EJ, Mullen EA, Chi YY, Hamilton TE, Gow KW, Ferrer FA, Barnhart DC, Ehrlich PF, Khanna G, Kalapurakal JA, Bocking T, Huff V, Tian J, Geller JI, Grundy PE, Anderson JR, Dome JS and Shamberger RC. Clinical outcome and biological predictors of relapse after nephrectomy only for very low-risk Wilms tumor: a report from Children's Oncology Group AREN0532. *Ann Surg* 2017; 265: 835-840.

Wilms tumor recurrence prediction with ML and Ki-67



Supplementary Figure 1. Bar chart showing missing data patterns for clinical variables in children with Wilms tumor.

Wilms tumor recurrence prediction with ML and Ki-67

Supplementary Table 1. Grid search range and optimal parameters for each machine learning model

Model	Range		Optimal parameters
	Lower	Upper	
Decision Tree			
tree_depth	3	8	7
min_n	5	15	6
cost_complexity	1E-6	1E-1	1.89E-5
Random Forest			
mtry	2	8	2
trees	800	1000	1000
min_n	10	30	10
XgBoost			
mtry	2	8	2
min_n	5	20	
tree_depth	3	6	8
learn_rate	1E-3	1E-1	7.59E-2
loss_reduction	-3	0	0.515
sample_prop	0.8	1	0.903
LASSO			
penalty	1E-5	1E-1	1E-1
SVM			
cost	-2	3	0.25
rbf_sigma	1E-3	1E-1	4.64E-3
MLP			
hidden_units	5	15	5
penalty	1E-3	0	1E-3
epochs	50	100	100
LightGBM			
tree_depth	1	4	1
trees	100	700	245
learn_rate	1E-3	1E-1	7.4E-2
mtry	2	8	5
min_n	5	20	13
loss_reduction	-3	0	0.521
KNN			
neighbors	3	12	11