Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 20 No. 11 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v20i11.49673

PAPER

Harnessing Machine Learning for Quantifying Vesicoureteral Reflux: A Promising Approach for Objective Assessment

Muhyeeddin Alqaraleh[1], Mowafaq Salim Alzboon[1](✉), Mohammad Subhi Al-Batah[1], Mutaz Abdel Wahed[1], Ahmad Abuashour[2], Firas Hussein Alsmadi[3]

[1]Faculty of Science and Information Technology, Jadara University, Irbid, Jordan

[2]Faculty of Computer Studies, Arab Open University, Al-Ardiya, Kuwait

[3]Queen Rania Hospital for children Jordan Royal Medical Services, Amman, Jordan

malzboon@jadara.edu.jo

ABSTRACT

In this study, we evaluated the performance of various machine-learning models on multiple datasets labeled GR1, GR2, GR3, GR4, and GR5. We assessed the models using a range of evaluation metrics, including AUC, CA, F1, precision, recall, MCC, specificity, and log loss. The models examined were logistic regression, decision tree, kNN, random forest, gradient boosting, neural network, AdaBoost, and stochastic gradient descent. The results indicate that all models consistently demonstrated outstanding performance across all datasets, with most achieving perfect scores in all metrics. The models exhibited high accuracy and effectiveness in accurately classifying instances. Although random forests displayed slightly lower scores in some metrics, theyi still maintained an overall high level of accuracy. The findings highlight the models' ability to effectively learn the underlying patterns within the data and make accurate predictions. The low log loss values further confirmed the models' precise estimation of probabilities. Consequently, these models possess strong potential for practical applications in various domains, offering reliable and robust classification capabilities.

KEYWORDS

vesicoureteral reflux (VUR), voiding cystourethrogram (VCUG), machine learning (ML), objective grading, radiographic evaluation, pediatric urology

1 INTRODUCTION

Artificial intelligence (AI) transforms patient care [1]. AI can analyze massive amounts of data, find patterns, and provide valuable insights for medical diagnosis, treatment, and decision-making [2–4]. AI algorithms can quickly process and interpret medical images such as X-rays, MRIs, and CT scans, helping radiologists identify issues and improve diagnosis [5]. AI-powered prediction models can help doctors identify high-risk patients for specific diseases, enabling early interventions and

Alqaraleh, M., Alzboon, M.S., Al-Batah, M.S., Wahed, M.A., Abuashour, A., Alsmadi, F.H. (2024). Harnessing Machine Learning for Quantifying Vesicoureteral Reflux: A Promising Approach for Objective Assessment. *International Journal of Online and Biomedical Engineering (iJOE)*, 20(11), pp. 123–145. https://doi.org/10.3991/ijoe.v20i11.49673

Article submitted 2024-04-15. Revision uploaded 2024-05-25. Final acceptance 2024-05-25.

personalized treatment. AI-powered chatbots and virtual assistants improve healthcare information access and reduce physician workload 24/7 [6], [7]. AI equips doctors with better tools and skills, enhancing patient outcomes, efficiency, and the potential to revolutionize healthcare [8].

Vesicoureteral reflux (VUR) occurs when urine overflows into the ureter or kidneys due to a defective vesicoureteral junction. Renal scarring can result from VUR in 30% of pediatric UTIs. VUR diagnosis and grading are best done with voiding cystourethrography. Even with the international rating method, VUR grading is subjective and has up to 60% inter-rater disagreement [5–7]. VUR grading should be more objective and uniform due to this inconsistency [9]. This issue may be addressed by ML. AI and ML have enhanced healthcare diagnosis and treatment. AI-based VUR grading systems can enhance accuracy and efficiency using large datasets and powerful algorithms [10], [11].

Machine learning (ML) is used in medical imaging analysis. Baray et al. tested a ML VUR grading method using quantitative voiding cystourethrogram (VCUG) characteristics and developed a deep learning (DL)-based penile curvature measurement method that was accurate in model and patient photos. Both studies showed promise. They were limited to picture analysis and intermediate VUR grade discrimination [12], [13]. This study develops VCUG picture-based ML-based VUR grading to address past study constraints. We measure ureter and renal pelvis size and shape using voiding cystourethrograms. ML algorithms predict VUR severity using these extracted characteristics. We add elements to reflect ureter tortuosity and distinguish VUR grades [14, 15].

This study enhances the accuracy and efficiency of grading VUR using ML. The aim is to establish a dependable and unbiased grading system to assist doctors in categorizing patients, administering medications, and enhancing VUR treatment. Our strong ML classifiers and diverse set of features are expected to surpass subjective approaches. VCUG images with VUR grades were used to develop and evaluate ML models for VUR grading. VCUG were analyzed for the classification of VUR severity [16], [17].

Many image processing methods extract ureter size, renal pelvis shape, and tortuosity. These features were selected to distinguish VUR classes. Additional characteristics identified ureter tortuosity and differentiated intermediate VUR grades in this study. Various classifiers were tested to train ML models, including SVM, random forests, and CNN. The models learned the correlation between image attributes and VUR severity from the extracted features [1–3] and [7].

The ML-based VUR grading system was evaluated using accuracy, precision, recall, and the F1 score. Cross-validation made the models reliable and generalizable. The experiments demonstrated that the ML-based VUR grading system is effective. The models classified VUR grades more accurately than subjective methods. Enhancements to capture ureter tortuosity and distinguish VUR grades improved model accuracy. The clinical relevance of this study is significant. A reliable and objective VUR grading system enhances patient care and therapy. Clinicians can stratify patients based on VUR severity for more effective therapies and medication management using a reliable method. ML can also identify VUR grading issues, enhancing diagnostic quality control. This study represents a crucial step towards objective VUR measurement using ML. However, there are limitations. This study should include a larger and more diverse set of VCUG images to ensure model generalizability. The ML-based method needs to be tested in clinical settings and compared to other grading systems [10], [12], [16], and [18]. Ultimately, successful management and therapy necessitate accurate and objective VUR grading. The inter-rater discrepancy in subjective VUR grading methods underscores the necessity for a more uniform and

standardized approach. This study proposes a ML-based VUR grading system using VCUG images to enhance accuracy and efficiency compared to subjective methods.

To predict VUR severity, ML algorithms were trained on VCUG images and characteristics. The addition of ureter tortuosity characteristics enhanced the determination of VUR grade. Management and treatment decisions for VUR may benefit from accurate models [12], [18]. This study shows promise, but more significant and diverse datasets are needed to validate the ML-based strategy. The VUR grading system should undergo clinical testing and be compared to other methods. Utilizing ML in VUR grading could enhance patient outcomes and standardize diagnosis. Modern algorithms and image analysis can objectively and accurately assess VUR severity, aiding clinicians in stratifying and managing patients.

1.1 Problem statement

The radiographic assessment of VUR is currently reliant on subjective criteria, leading to variability and inconsistent diagnoses and treatment recommendations. Subjectivity poses challenges for clinicians and radiologists in assessing VUR severity through VCUG images. Therefore, developing a precise and objective grading system for VUR is essential to enhancing diagnostic precision and treatment outcomes.

1.2 Article objectives

This study aims to develop a supervised ML model to objectively grade VUR using VCUG images. ML algorithms will mitigate VUR grading subjectivity and variability. The study seeks to identify ML models that can accurately predict VUR severity across different grade levels, pinpoint key VCUG image features and patterns for precise VUR grade classification, investigate whether deformed renal calyces can predict high-grade VUR, and demonstrate how ML can enhance VUR grading objectivity and accuracy, leading to more reliable diagnostics.

1.3 Contribution of the article

The study enhances radiology and healthcare AI research by grading VUR with supervised ML. The key contributions to the article include: 1. Initial ML model creation: the study creates and tests six ML models for grading VUR using VCUG images, including logistic regression, decision tree, gradient boosting, neural network, and stochastic gradient descent. The study demonstrates that ML models accurately predict VUR severity across different grade levels without false positives or negatives, addressing the subjectivity and variability in VUR grading. 2. Identification of key features: the study highlights that deformed renal calyces indicate severe vesicoureteral reflux, aiding doctors in the faster diagnosis and treatment of VUR. The study also emphasizes the importance of enhancing ML methods, exploring new features, and expanding dataset to enhance model precision and applicability, thereby enabling advancements in ML-based VUR grading study. 5. Healthcare implications: ML-based grading systems help reduce subjectivity and unpredictability in VUR assessment, enhancing objectivity and accuracy in patient diagnoses and treatment recommendations. By utilizing ML, the study aims to enhance VUR grading accuracy, objectivity, and efficiency in VCUG image evaluation.

1.4 Article organization

Section 2 summarizes the literature on VUR prediction and classification models. Section 3 outlines the proposed methodology, which includes a description of the dataset and the ML model. Section 4 demonstrates the implementation and its outcomes. Section 5 contains the discussions. The paper concludes in Section 6.

2 RELATED WORK

The stomach, esophagus, duodenum, small intestine, and large intestine constitute the gastrointestinal system. GI content movement is influenced by stomach dysrhythmias in many individuals globally. Common issues include dyspepsia, vomiting, abdominal pain, stomach ulcers, and GERD. Abnormalities are detected through images, endoscopies, electrogastrograms, and clinical analysis. Surface Ag and AgCl electrodes were positioned on 20 healthy stomachs for data collection and preprocessing. Signals from eight women and 12 men were used, with data provided by three women and eight men with gastrointestinal problems. Preprocessing the dataset involves eliminating signal noise. Wiener filters enhance data quality and reduce noise. Features are chosen through PSO and HGW, eliminating unnecessary signal data and accelerating the process. These features aid in classifying primary gastric lymphoma, GIST, and neuroendocrine tumors, allowing classifiers to analyze carcinoids. This classification is performed using MCFFN, which phases and categorizes the data. Performance is evaluated based on classification accuracy, sensitivity, and specificity [19].

GERD, proton pump inhibitor usage, sleep patterns, preterm birth, socioeconomic status, and depression prescription history can be analyzed using ML and population data. A retrospective cohort of 405,586 women aged 25–40, who had their first singleton pregnancy between 2015 and 2017 was used for claims from the Korea National Health Insurance Service. A total of 65 variables encompassed demographic, socioeconomic, medical, medication, and obstetric data from 2015 to 2017. Term birth was the dependent variable. Random forest variable importance identified risk factors for preterm birth and their associations with socioeconomic status, GERD, and drug history, including proton pump inhibitors, sleep aids, and antidepressants. Through random forest analysis, socioeconomic status, age, proton pump inhibitors, GERD over various years, sleep medications, and antidepressants were found to have the most impact on premature birth from 2015 to 2017. In conclusion, low socioeconomic status, GERD, proton pump inhibitors, sleep aids, and antidepressants were identified as causes of preterm birth. Preterm birth in pregnant women can potentially be reduced through medication, GERD prevention, and socioeconomic enhancement [20].

Unsystematic living can lead to GI complaints. Many diseases can be identified through screening and early diagnosis. This study proposes using ML to predict gastroesophageal reflux. SVM and logistic regression are utilized to predict symptoms. Binay trees are employed for linear representation. The algorithm assigns one attribute per core tree node and one class label per branch node. SVMs classify data using kernels and hyperplanes. ANN concepts offer more here. Additionally, there is encryption and precision. The study predicts GERD authentication can be enhanced with ECC and SHA256 [21].

GERD, periodontitis, and preterm birth were examined using ML and demographic data. Korea National Health Insurance Service claims data showed 25–40-year-old

women, 405,586 had their first singleton pregnancy between 2015 and 2017. The study focused on preterm births during the years 2015–2017. Independent variables for the period 2002–2014 included GERD (yes or no), periodontitis (yes or no), age in 2014, socioeconomic status (determined by insurance fee), and region (city). The main predictors of preterm birth and the associations between GERD and periodontitis were identified using random forest variable importance. Socioeconomic status, age, GERD in multiple years (2006, 2007, 2009, 2010, 2012, and 2013), and city in 2014 were found to be the most important random forest variables for predicting preterm birth from 2015 to 2017 periodontitis was deemed irrelevant. This concluded that. GERD contributes more to preterm birth than periodontitis. To prevent preterm birth, pregnant women should focus on GERD prevention and socioeconomic enhancement. It is crucial to address GERD symptoms, especially those that pregnant women may overlook, through active counseling [22].

Disease follow-ups always have missing observations. Across demographic, clinical, laboratory, and imaging data, hospital records rarely show VUR or rUTI. The missing ratio strategy of DL can handle high amounts of missing data. This study used MICE and FAMD imputation to evaluate DL differential diagnostics. A retrospective cross-sectional study of 611 pediatric patients found 425 with VUR, 186 with rUTI, and 26.65% missing data. Models were evaluated using 34 physical, laboratory, and imaging criteria by R 3.6.3 and CNTK. MICE-DL distinguished VUR and rUTI best with 64.05% accuracy, 64.59% sensitivity, and 62.62% specificity. During missing imputation, FAMD's accuracy, recall, and specificity yielded 61.00. DL evaluates datasets without missing values. Missing imputation improves DL prediction [13].

Gastroesophageal reflux can cause cancer. From October 2018 to December 2020, a survey was conducted on clinic-based upper gastrointestinal endoscopy patients aged 20 and older. ML models such as RF, SVM, MLP classifier, and XGBoost. High-risk esophageal erosion patients are identified before endoscopy by the risk prediction algorithm [23].

PPIs are inferior to first-generation potassium-competitive acid blockers (VPZ). Gastric, duodenal, and Helicobacter pylori ulcers are treated with faurate salt. ANN ML discovered new VPZ cocrystals. Virtual screening uses VPZ cocrystals. Eight of the 19 liquid-assisted grinding (LAG) conformers could create new solid forms with VPZ, according to the ANN model. Benzenediols and benzene triols (catechol, resorcinol, hydroquinone, and pyrogallol) were used to form phase-pure VPZ cocrystals. We synthesized and identified three new cocrystals: VPZ-RES, VPZ-CAT, and VPZ-GAL. In pH 6.8, VPZ-RES was the most soluble novel cocrystal. Fumarates were less effective than novel water-stable VPZ cocrystals [24].

Objective: Recent ML risk factor assessments. Health insurance claims data and ML were used to study asthma exacerbation. Clinically relevant risk indicators were desired. Analyzing asthma patients from May 2014 to April 2019 using MediScope® (DB), a Japanese health insurance claims database. Patients and illnesses helped us identify asthma exacerbation risk factors. Emergency medical procedures requiring transfer and IV steroid injections were identified as asthma exacerbations in the database. Out of 42,685 qualified database instances, 5.844 exacerbations occurred, accounting for 13.7%. ML extracted 25 risk assessment parameters from 3,300 illnesses. Exacerbation risk was found to be reduced by periodontitis and dyslipidemia. Conclusions: ML and in-depth claims data analysis revealed asthma exacerbation risk indicators consistent with previous studies. Further study is warranted [25].

CEVL teaches endoscopic injectable VUR therapy online. The information was created using clinical and computer skills. This technology offers staff training, assessment, and process skill recording in an online format with narration, photos,

and videos. We provide feedback, skill rehabilitation, and educational games to enhance surgical outcomes and standardize training and procedures. A digital publication like this one is ideal for knowledge sharing and collaborative research [26].

Hydronephrosis and VUR can be distinguished using DL AI. Online images of hydronephrosis and VUR are available. A DL and image analysis method were developed to analyze these images. The AI system learned to differentiate between hydronephrosis and VUR images. Discrimination was assessed by analyzing ROC curves. The analysis was conducted using Scikit-learn. Online images displayed 39 cases of hydronephrosis and 42 cases of vesicoureteral reflux. Training and validation images were randomly selected, resulting in 68 training instances and 13 validation instances. In two test cases, the AI system predicted VUR with a probability of 0.99874 and hydronephrosis with a probability of 0.00006. This study presents a method for urological image classification using a deep neural network. The AI system shows promise in distinguishing urological images [27].

Growing datasets may conflict with traditional databases. Medical study data could be used by ML to identify optimal solutions. Clinicians must accurately match clinical symptoms to renal scarring to diagnose LUTD. ML can predict renal scarring in children with lower urinary tract dysfunction. A total of 114 participants aged over three underwent urodynamic testing. The dataset includes 47 variables. The individual had a symptomatic urinary tract infection, vesicoureteral reflux, bladder trabeculation, increased bladder wall thickness, abnormal DMSA scintigraphy, and clean intermittent catheterization. ML was employed. According to confusion matrix comparisons, extreme gradient boosting (XGB) achieved the highest accuracy at 91.30%. An ANN method of SMOTE achieves 90.63% balanced data set accuracy. With the balanced (SMOTE) dataset, the ANN algorithm had 90.78% success. Conclusion: MLT's high accuracy rates lead to faster and more accurate RS estimation in LUTD patients [28].

Reliability issues hinder VUR grading. Image quantitative features may aid VUR grading. Use voiding cystourethrograms to quantify VUR. Approach An online VCUG dataset classified rental units as I-III or IV-V. UPJ, UVJ, maximal ureter width, and ureter tortuosity were quantified and standardized using image analysis and ML with three user-defined markers. Our random forest classifier classifies low-and high-grade VUR. The institutional imaging repository established external validation. Precision-recall curve and receiver-operating-characteristic analysis assessed discrimination. Shapley's additive explanations verified model predictions. An online dataset had 41 rental units, while an institutional imaging repository had 44. High- and low-grade VUR differed in UVJ, UPJ, maximal ureter width, and tortuosity. Left-one-out cross-validation yielded 0.83 accuracy, 0.90 AUROC, and 0.89 AUPRC for the random-forest classifier. The external validation had 0.84 accuracy, 0.88 AUROC, and 0.89 AUPRC. The most important was tortuosity, following by ureter, UVJ, and UPJ width. We built this web-based qVUR. https://akhondker.shinyapps.io/qVUR/ automatically evaluates VCUGs. The Authors in [10] advances an objective standard for VUR assessment, highlighting tortuosity and ureter dilatation as key predictors of high-grade VUR, and provides a user-friendly web application for automated grading.

Machine learning combined with biostatistics has identified significant metabolomic signatures in the plasma of pediatric patients with chronic kidney disease (CKD), offering insights into potential causes. Dysmetabolism in the sphingomyelin-ceramide axis is linked to both focal segmental glomerulosclerosis (FSGS) and the aplasia/dysplasia/hypoplasia (A/D/H) spectrum. Contrary to previous reports associating plasmalogen deficiencies with FSGS. Plasmalogen is high in FSGS kids. Oletryptophans and gut histidine cause reflux and obstructive uropathy. Metabolic trends may help us understand juvenile chronic kidney disease using ML and untargeted

plasma metabolomic analysis. Based on FSGS, OU, A/D/H, and reflux nephropathy, metabolic trends in juvenile CKD were examined. The Authors in [29] involved untargeted plasma metabolomic profiling of 702 participants from the Chronic Kidney Disease in Children study, categorized by diagnosis: FSGS (63), OU (122), A/D/H (109), and RN (86). Clinicians chose Lasso regression for extreme gradient boosting, logistic regression, SVM stratification, and random forest. After ML training on 80% of cohort subgroups, 20% of holdouts were verified. Two of the four models chose important traits. Using pathway enrichment analysis, we identified CKD-causing metabolic subpathways. Metabolic profiles by cause: ML models outperformed talent-free forecasts in holdout subsets with receiver-operator characteristics, precision-recall area-under-the-curve, F1 score, and Matthew's correlation coefficient. Ceramide FSGS had plasmalogen metabolites are and subpathways. Gut histidine metabolites linked to OU. ML models detect CKD metabolism. Gut microbiome-derived histidine metabolites, sphingomyelin-ceramide, and plasmalogen dysmetabolism were linked to OU, FSGS, and ML [29].

Global GERD status impacts community health. Chronic GERD brings many diseases. Determine GERD incidence, risk factors, and symptoms to enhance healthcare and management. ML collected the data for this study. VOSviewer created a network from ANN-categorized data. ML and AI show rising Asian GERD. According to reports, GERD is common in Pakistan. Oily food, late dinners, sedentary lifestyles, and illness diagnosis and treatment ignorance are GERD risk factors in Pakistan. The results show acid reflux and esophageal inflammation. This risk factor and symptom study will enhance GERD diagnosis. Professionals can easily monitor GERD patients to reduce related diseases. Geography and comparative data can identify disease hotspots that need more management and control [30].

We created and validated PCC models in a retrospective cohort study of chronic cough patients. Techniques In 2011–2016, specialists diagnosed 18–85-year-olds with at least three coughs. Cough is mentioned in a diagnosis, prescription, or record. Train and validate models with 400 features and two ML methods. Applied sensitivity analysis. CC or two coughs in years two and three after the index date of the specialist cohort or three for the event cohort, diagnosed with PCC. Expert cohort criteria were met by 8581 patients and event cohort criteria by 52,010, with mean ages of 60.0 and 55.5 years. 12.4% of event cohort patients and 38.2% of specialists received PCC. Baseline CC, or respiratory illness, healthcare consumption was modeled. Diagnosis-based models included age, asthma, PF, OSLD, GERD, hypertension, and bronchiectasis. Succinct models with five to seven predictors had reliable area under the curve values of 0.74 to 0.76 for utilization-based models and 0.71 for diagnosis-based models. Conclusions: Our risk prediction algorithms can identify high-risk PCC patients during clinical testing and evaluation to aid decision-making [31].

Automation of speech disorder evaluation is cutting-edge voice analysis. Eating disorders may affect voice quality, says a study. GERD and obesity are risk factors for voice, so this study examined how they affect voice. We studied illness-characteristic interactions. On 92 people, vowel phonation and phrase repetition were tested. Researched were obese GERD patients, healthy controls, and obese patients. Both Naive Bayes and SVM models extracted binary classification features well. Both GERD and obesity were detected with 0.86 and 0.82 accuracy on the validation set, respectively, showing no overfitting as performance dropped. Phonation performed worse than sentence repetition in tasks and characteristics. Mel frequency cepstral coefficients, perceptual linear prediction coefficients, bark band energy coefficients, and noise measurements matter most [32].

Pipeline measurement requires laborious drying, grinding, and refluxing of pepper samples with high-quality ethanol. Efficiency but time-and resource-intensiveness limit this method's variation resolution. Pipeline-level assessment and

prediction using ML must be faster and more accurate. ANN and fluorescence imaging are tested to enhance Javanese long pepper pipeline measurement accuracy. We propose UV-induced fluorescence imaging and ML for Javanese long pepper. Pipeline concentration was represented by colors of fluorescence from 365 nm UV LEDs. With an R2 value of 0.88025, an artificial neural network model predicted pipeline content from fluorescence picture color texture. Modeling used 10 characteristics and one R. Using'trainees' learning, 'tansig' activation, 0.1 learning rate, and 10-40-10 nodes, the ultimate ANN has a testing R2 of 0.8943 and MSE of 0.0875. Pipeline prediction by ML uses LED fluorescence. Pipeline content measurement has been enhanced [33].

The VUR correction data from the dextranomer/hyaluronic acid copolymer is inconsistent. Logistic regression analyzed injected volume and controlled for other factors affecting dextranomer/hyaluronic acid copolymer injection success. In July 2003 to June 2006, 126 patients (34 men and 92 women) with primary VUR (196 refluxing ureters) received febrile UTI injections. The mean age was 6.5 ± 3.7 years. All reflux was resolved. Age, gender, laterality, preoperative VUR grade, surgeon experience, dextranomer/hyaluronic acid copolymer volume, surgery time, and lower urinary tract symptoms were considered. Seven (3.5%), 53 (27%), and 91 (46.4%) had I-V VUR [34].

This ML and demographic study examines depression and particulate matter as preterm birth risk factors. The cohort study used Korea National Health Insurance Service claims for 405,586 (25–40-year-old) women who gave birth for the first time after a singleton pregnancy between 2015 and 2017. From 2015 to 2017, 90 independent variables affected preterm birth related to demographic, socioeconomic, environmental, health, and obstetric factors. Depression and PM caused preterm births in random forest variables. According to a random sample, the top 40 determinants of preterm birth in 2015–2017 are social class, age, proton pump inhibitor, benzodiazepine, tricyclic antidepressant, sleeping pills, progesterone, GERD for 2002–2014, particulate matter for January–December 2014, region, myoma uteri, diabetes for 2013–2014, and depression for 2011–2014. Depression and PM prenatal care should include intensive particle intervention, proactive counseling, and medication for common depression symptoms that expecting mothers ignore [35].

RST and an adaptive neuro-fuzzy inference system (ANFIS) soft sensor model predict early crude oil cutbacks to enhance oil refinery efficiency. The RST simplified ANFIS' fuzzy rule sets and decision tables. For continuous data, the best discretization algorithms were used. Expect RVP and API gravity, which affect the light naphtha cut quality. Real-time process data from the Al Doura oil refinery is analyzed to understand two crude oil sources. Response variables display the crude distillation unit's rectifying cascade controller's splitter top feedback. Separator head reflux liquid flow is controlled. A steady-state control system with a lab-tested virtual sensor was created using adaptive soft sensors. An oil refinery's cascade ANFIS controller and soft sensor model predictive control system control distillate purity. Rising and settling times are 26.65% and 84.63% faster with ANFIS-based cascade control and no over or undershoots. Comparisons of prediction and control model results with other ML methods [36].

Determining the optimal number of STING-based EI operations without simulated training. This study examined two pediatric urology fellows' EI procedures. Patients without primary vesicoureteral reflux, endoscopic injection, ureteroneocystostomy, lower urinary tract dysfunction, or duplicated ureters were excluded. Dextranomer hyaluronate and STING were used, according to O'Donnell and Puri. Statistical trials determined group size. Three combinations with 12, 24, and

36 patients and 35 EI surgeries showed significant changes. Twelve patient groups resulted. One completed 54 EIs, the other 51. Colleague 1 had three 18-eI procedure groups; colleague 2 had 17.72 participants, and 105 ureters were studied. Using both individuals' data yielded 35 three-category procedures. In the first, second, and third groups, first-person success rates were 38.3%, 66.6%, and 83.3%, p = 0.02. 41.2%, 64.7%, and 82.3% met second-person success criteria, p = 0.045. Colleagues' success rates rose. Conclusions: EI seems effective after 20 sessions and 35–40 [14].

For VUR and intrarenal reflux, CeVUS was used. Researchers sought IRR VUR patients' kidney locations. Methodology: The study included 70 patients with VUR and 103 URU with renal ultrasonography abnormalities for recurrent or first-time UTIs. We used GE Logiq S8 ultrasound and second-generation ultrasound contrast. In 103 cases of VUR, 49.51% had intrarenal reflux (IRR), whatever the severity ($p < 0.0001$). The median age of IRR patients was 5.5 months (IQR, 3–14.3), compared to 15.5 months (IQR, 5–41.5) for non-IRR patients ($p = 0.0069$). The highest incidence of IRR was in the superior pole (80%), followed by the inferior (62.7%), middle (37%), and all segments (27%). Conclusion: Early clinical signs in IRR-associated VUR patients. The normal distribution of composed papillae types II and III increased IRRs with VUR severity. Clinical trials may suggest adding IRR to VUR [17].

Voiding cystourethrogram (VCUG) radiographs are used to evaluate VUR patients' clinical development and treatment options. So, we created a supervised ML algorithm to rate VCUG data objectively since image-based VUR grading is subjective. The study used 113 public VCUG images. Three pediatric urologists and four pediatric radiologists assessed VUR severity for each scan. The higher severity was 4–5, and the lower was 1–3. The majority of experts graded each photo accurately. Nine features per VCUG picture were used to train, validate, and test six ML models using "leave-one-out" cross-validation. Models were trained using all analyses' best attributes. An important ML model accuracy statistic is the F1-score. The SVM and MLP classifiers have high F1 scores of 90.27% and 91.14% using the most important VCUG image attributes. SVM and MLP F1 scores were 89.37% and 90.27% with all features. Findings show severe VUR with abnormal renal calyces. ML could improve VUR objective-grading [12].

Although HRM and esophagography diagnose achalasia, esophageal motility and morphology are unknown. Uncertain POEM outcomes in new patients. There were 1,824 treatment-naïve achalasia patients in the multicenter cohort study. PoEM was given to 1,778 patients. Using age, sex, illness duration, BMI, and HRM/esophagography findings, ML clustering identified achalasia phenotypes. The ML models predicted chronic symptoms with Eckardt scores of 3 or higher and reflux esophagitis graded A to D after POEM. They classified achalasia as type I with a dilated esophagus (n = 676; 37.0%), type II (n = 203; 11.1%), and late-onset type I-III (n = 619). I and II achalasia have different clinical symptoms than phenotype 3, suggesting different HRM pathophysiologies. Persistent symptom prediction showed a 0.70 AUC. Pre-POEM Eckardt scores of 6 or higher caused the lasting symptoms. POM reflux esophagitis AUC was 0.61. Aphasia causes vary by esophageal mobility and structure. ML improved persistent symptom risk categorization by considering treatment resistance [37].

The stomach, esophagus, duodenum, and small and large intestines support human physiology. Gastric dysrhythmia, dyspepsia, unusual nausea, vomiting, abdominal pain, stomach ulcers, and GERD affect many people worldwide. Endoscopic, electrogastrographic, imaging, and clinical analyses reveal abnormalities. An electrogastrogram records stomach contractions. Electrogastrograms read stomach muscle impulses from electrodes. PCs process EGG signals. Post-meal

stomach muscle activity rises. Stomach muscle or nerve abnormalities cause postprandial ERV. Typical electrogastrograms check for bradycardia, dyspepsia, nausea, tachycardia, ulcers, and vomiting. Before and after doctor-patient meals, patients and the public are surveyed. In the MATLAB genetic method, CWT and db4 wavelets show 3D EGG signal wave patterns. Shown is the EGG signal cycle peak. Peaks classify EGG. The ARCN classifies EGG signals by attention (μ) as normal or abnormal. Using this study, doctors can diagnose stomach disorders before surgery. [38] The suggested work has 95.45% accuracy, 92.45% sensitivity, and 87.12% specificity.

Unreliable voiding cystourethrograms grade VUR subjectively. Based on VCUG ureteral tortuosity and dilatation, simple and machine-learning methods improved VUR grading reliability. We trained and validated voiding cystourethrograms. Each VCUG saw 5-7 raters agree on a VUR grade and rating reliability. The proximal, distal, and maximal ureteral tortuosity of every VCUG were checked. Labeling four traits followed. qVUR predicted the grade of VUR using specific factors. AUROC estimated the 1's performance. A total of 8,230 cystourethrograms were performed on 1,492 kidneys and ureters. Vegioureteral reflux grading had 0.71 median agreement, 0.44 internal consistency, and low rater consistency. Feature values increased vesicoureteral reflux. On all external datasets, the VUR had 0.62 accuracies (AUROC = 0.84). The strategy improved VUR grade reliability by 3.6 times over standard methods ($p < .001$). ML grades VUR in a large pediatric cohort from many institutions better than current methods. Robust, generalizable qVUR has doctor-like precision. The predictive power of quantitative indicators needs more research [39].

Antibiotic prophylaxis reduces UTIs in children with vesicoureteral reflux. Selected groups may benefit from antibiotic prophylaxis. To find these categories using ML. We randomized RIVUR data (4:1) into training and testing sets. UTIs with and without antibiotics were predicted by two models. Recurrent UTIs and antibiotic prophylaxis were confirmed. Models predicted repeated UTIs. For efficacy, continuous antibiotic prophylaxis was given at different urinary tract infection risk levels. 607 patients—558 girls and 49 boys—with a median age of 12 months were studied. Viscoureteral reflux grade, serum creatinine, race, gender, UTI history (fever or dysuria), and weight percentiles were evaluated. The AUC for recurrent urinary tract infection prediction with continuous antibiotic prophylaxis versus placebo is 0.82 (95% CI 0.74-0.87). 40% of VUR patients should receive continuous antibiotics to reduce UTIs by 10%. More than a 10% reduction in recurrent urinary tract infection risk prompted continuous antibiotic prophylaxis for 51 of the 121 test subjects. The group with model continuous antibiotic prophylaxis had fewer recurrent urinary tract infections (7.5% vs. 19.4%, $p = 0.037$). Our algorithm determines which VUR patients benefit most from antibiotics. Targeted and individualized continuous antibiotic prophylaxis improves efficacy and reduces wasteful use in unneeded individuals [40].

Most doctors evaluate and treat GERD using the Los Angeles classification. AI-DL models aid in diagnosis. The two-stage endoscopic classification of GERD uses DL and ML. Transfer learning optimizes picture feature extraction on the target dataset, and ML optimizes classification. Testing shows this study's GerdNet-RF model performs best. Enhance test accuracy from 78.8% ± 8.5% to 92.5% ± 2.1%. Enhancing AI-automated diagnostics would benefit patients [41].

The backflow of urine from the bladder into the ureters and kidneys is called VUR. It helps with urinary infections. VUR severity is assessed by VCUG. Disputed VUR surgery time and type. Regular and accurate VUR-grade recognition is crucial. Convolutional neural networks (CNN) detect and characterize VUR in VCUG images in this study. To reduce categorization disparities among observers and create a healthcare practitioner-friendly tool [16].

3 METHODOLOGY

This methodology outlines the research design, data collection methods, and data analysis techniques employed in a study focused on quantifying VUR using ML. The study utilized a dataset of VCUG images of real-world VUR cases gathered from various public sources. After eliminating images with poor quality for grading, 113 high-quality photos were selected for analysis. The severity of VUR in each image was independently graded by seven professional assessors, including three pediatric urologists and four pediatric radiologists. The chosen methods align with the research objectives of developing an automated model for VUR severity quantification [1]. The dataset used in the study is publicly available on Kaggle (https://www.kaggle.com/datasets/saidulkabir/vcug-vur-dataset), as shown in Figure 1.

3.1 Machine learning models

The evaluation was conducted using the VCUG VUR dataset. The dataset of 113 instances is divided into training and testing data. The training data consists of 80 randomly selected instances, which is about 70% of the total dataset, while the remaining 33 instances form the testing data. This selection process is deterministic, meaning it will consistently produce the same result if repeated. It consists of 113 rows and six columns. The dataset includes a categorical outcome variable with five classes representing the severity of VUR. Additionally, the dataset contains three numeric variables and two text variables as metas [3].

To perform the evaluation, multiple ML models were utilized. The following models were employed: DT, stochastic gradient descent, neural network, logistic regression, and gradient boosting. Each model was trained and evaluated using appropriate performance metrics to assess its effectiveness in predicting VUR severity [2].

Regarding data availability, the VCUG VUR dataset used in this study can be accessed through the online repository Kaggle. The dataset can be found at the following URL: [https://www.kaggle.com/datasets/saidulkabir/vcug-vur-dataset]. Researchers and interested individuals can access the dataset from this repository for further analysis and exploration [2].

In the evaluation process, several ML models were employed to assess their performance in predicting the severity of VUR using the VCUG VUR Dataset. The following models were utilized:

- Logistic regression: This linear classification algorithm is used for binary classification. It models the relationship between input variables and class probability. Applying a logistic function to a linear combination of input features predicts the probability of class membership. It handles linearly separable data well and is interpretable and computationally efficient.
- Decision tree: A DT is a versatile supervised learning algorithm for classification and regression. It creates a hierarchical structure of nodes representing features and splitting criteria. It classifies instances or predicts continuous values by traversing the tree based on feature values. DT can capture non-linear relationships, handle categorical and numerical features, and be easily interpretable.
- K-nearest neighbors (KNN): non-parametric instance-based classification algorithm. It classifies instances by feature space and neighbor similarity. KNN finds the k closest neighbors (based on a distance metric like Euclidean distance) and assigns the majority class label to the new instance. KNN is simple to implement and works well with small datasets and multiclass classification.

- Random forest: This ensemble learning method predicts using multiple decision trees. It trains decision trees on a random subset of data and features. The final prediction is based on majority voting (for classification) or averaging (for regression) of individual tree predictions. Random Forest estimates feature importance, handles high-dimensional data, and resists overfitting.
- Gradient Boosting: This ensemble learning method builds a predictive model by sequentially adding weak learners to correct previous errors. It creates a robust predictive model from multiple weak prediction models, usually decision trees. Previous models' residuals (the difference between actual and predicted values) are used to train each new model. Powerful gradient boosting handles complex relationships and predicts accurately.
- This versatile and powerful ML model is based on the human brain's structure and function. Interconnected neurons form layers, including input, hidden, and output layers. Each neuron transforms its weighted inputs non-linearly. Neural networks excel at image and text classification, complex non-linear relationships, and large datasets.
- An ensemble learning algorithm called AdaBoost, short for Adaptive Boosting, combines multiple weak classifiers to create a robust classifier. It gives instances misclassified by weak classifiers higher weights, so subsequent classifiers can focus on them. Averaging weak classifier predictions weighted by performance yields the final prediction. AdaBoost handles imbalanced datasets, resists overfitting, and achieves high accuracy with weak classifiers.
- Stochastic gradient descent (SGD): This optimization algorithm trains ML models on large datasets. It updates the model's parameters by randomly selecting a mini-batch of training data at each iteration. The mini-batch loss function gradients determine the updates. SGD efficiently computes, handles large datasets, and converges faster than gradient descent.

3.2 Research design

The study approach employed in this study is quantitative, as it involves the analysis of numerical data obtained from VCUG images. The study focuses on developing a machine-learning model for automated VUR severity quantification [7].

3.3 Data collection methods

This study used data from published articles, online resources, and Radiopaedia. This method includes real-world VUR cases and diverse scenarios. Low-quality grading images were excluded to ensure analysis reliability and accuracy [7].

3.4 Sample selection

The sample for this study comprises 113 VCUG images. These images were selected based on their suitability for an accurate VUR severity assessment. The sample represents a variety of VUR cases, allowing for a comprehensive analysis [8].

3.5 Data collection procedures

Each of the 113 VCUG images was independently evaluated and graded by seven professional assessors. This group included three pediatric urologists and four

pediatric radiologists. The assessors utilized their clinical expertise and knowledge to assign severity grades to each image, considering the extent and severity of VUR present [5].

3.6 Data analysis techniques

Machine learning techniques were employed to develop a model for automated VUR severity quantification. Specific feature extraction techniques were applied to capture relevant characteristics from the VCUG images. Various ML algorithms, such as CNNs, were explored and evaluated for their effectiveness in quantifying VUR severity [5].

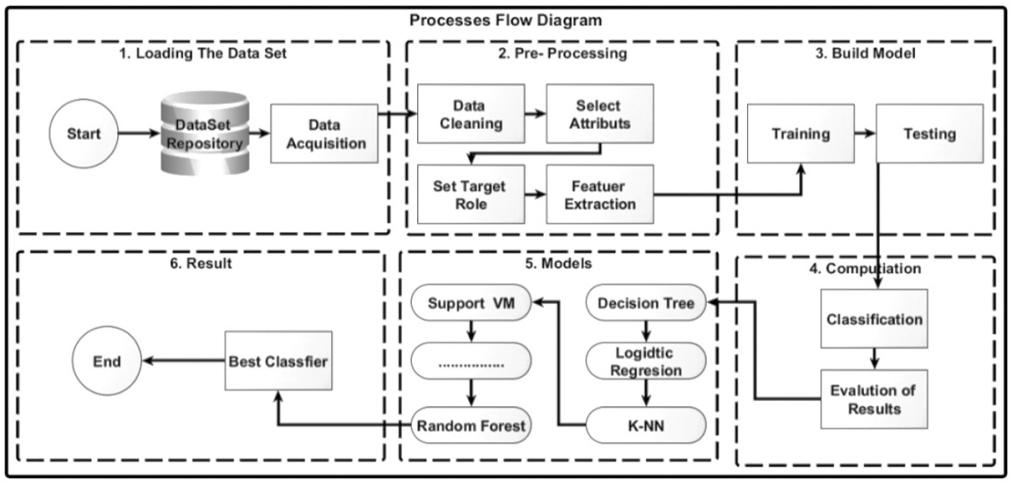

Fig. 1. The processes flow diagram

4 RESULTS

4.1 Test and score analyses

Test and score are fundamental concepts in the realm of evaluation and assessment. A test is a standardized procedure or assessment tool designed to measure a specific construct, such as knowledge, skills, or abilities, usually within a defined domain or subject area. Tests are used in various fields, including education, psychology, and research, to gather data and make informed judgments about individuals or systems. On the other hand, a score represents the numerical or qualitative result obtained from a test, indicating the performance or proficiency level of an individual or the effectiveness of a system. Scores can be expressed as raw scores, percentile ranks, standard scores, or other forms of measurement, depending on the nature of the test and the intended interpretation. The relationship between tests and scores is crucial in evaluating and comparing performances, making decisions, and providing feedback for improvement. The careful analysis and interpretation of test scores contribute to informed decision-making processes in educational, clinical, and organizational settings, as shown in Table 1 [2].

Table 1. The test and score analyses for the models logistic regression, tree, KNN, random forest, gradient boosting, neural network, AdaBoost, stochastic gradient descent 1–5

	Model	AUC	CA	F1	Prec	Recall	MCC	Spec	LogLoss
GR1	Logistic Regression	1	1	1	1	1	1	1	0
	Tree	1	1	1	1	1	1	1	0
	kNN	1	1	1	1	1	1	1	0
	Random Forest	1	0.988	0.8	1	0.667	0.811	1	0.045
	Gradient Boosting	1	1	1	1	1	1	1	0
	Neural Network	1	1	1	1	1	1	1	0
	AdaBoost	1	1	1	1	1	1	1	0
	Stochastic Gradient Descent	1	1	1	1	1	1	1	0
GR2	Logistic Regression	1	1	1	1	1	1	1	0
	Tree	0.999	0.988	0.966	0.933	1	0.959	0.985	0.027
	kNN	1	1	1	1	1	1	1	0
	Random Forest	1	1	1	1	1	1	1	0.11
	Gradient Boosting	1	1	1	1	1	1	1	0
	Neural Network	1	1	1	1	1	1	1	0.002
	AdaBoost	1	1	1	1	1	1	1	0
	Stochastic Gradient Descent	1	1	1	1	1	1	1	0
GR3	Logistic Regression	1	1	1	1	1	1	1	0
	Tree	0.998	0.975	0.96	0.96	0.96	0.942	0.982	0.048
	kNN	1	1	1	1	1	1	1	0
	Random Forest	1	1	1	1	1	1	1	0.168
	Gradient Boosting	1	1	1	1	1	1	1	0
	Neural Network	1	1	1	1	1	1	1	0.003
	AdaBoost	1	1	1	1	1	1	1	0
	Stochastic Gradient Descent	1	1	1	1	1	1	1	0
GR4	Logistic Regression	1	1	1	1	1	1	1	0
	Tree	0.997	0.975	0.929	1	0.867	0.917	1	0.041
	kNN	1	1	1	1	1	1	1	0
	Random Forest	1	0.988	0.968	0.938	1	0.961	0.985	0.145
	Gradient Boosting	1	1	1	1	1	1	1	0
	Neural Network	1	1	1	1	1	1	1	0.003
	AdaBoost	1	1	1	1	1	1	1	0
	Stochastic Gradient Descent	1	1	1	1	1	1	1	0

(*Continued*)

Table 1. The test and score analyses for the models logistic regression, tree, KNN, random forest, gradient boosting, neural network, AdaBoost, stochastic gradient descent 1–5 (*Continued*)

	Model	AUC	CA	F1	Prec	Recall	MCC	Spec	LogLoss
GR5	Logistic Regression	1	1	1	1	1	1	1	0
	Tree	0.998	0.988	0.979	0.958	1	0.97	0.982	0.035
	kNN	1	1	1	1	1	1	1	0
	Random Forest	0.998	0.975	0.957	0.957	0.957	0.939	0.982	0.11
	Gradient Boosting	1	1	1	1	1	1	1	0
	Neural Network	1	1	1	1	1	1	1	0.001
	AdaBoost	1	1	1	1	1	1	1	0
	Stochastic Gradient Descent	1	1	1	1	1	1	1	0
Average over classes	Logistic Regression	1	1	1	1	1	1	1	0.001
	Tree	0.998	0.963	0.962	0.964	0.963	0.951	0.987	0.076
	kNN	1	1	1	1	1	1	1	0
	Random Forest	1	0.975	0.974	0.976	0.975	0.967	0.992	0.307
	Gradient Boosting	1	1	1	1	1	1	1	0
	Neural Network	1	1	1	1	1	1	1	0.005
	AdaBoost	1	1	1	1	1	1	1	0
	Stochastic Gradient Descent	1	1	1	1	1	1	1	0

- Grade 1: Table 1 shows that various ML models (LR, Tree, KNN, RF, GB, NN, AdaBoost, and SGD) performed excellently in classifying Grade 1 instances. Most models achieved perfect scores in metrics such as AUC, classification accuracy (CA), F1-score, precision, recall, Matthew's correlation coefficient (MCC), and specificity, except RF, which had slightly lower scores.
- Grade 2: The same models were analyzed for Grade 2 instances, with similar impressive results. All models performed well across all metrics, with LR, KNN, GB, NN, AdaBoost, and SGD achieving perfect scores. Tree and RF had slightly lower scores in some metrics.
- Grade 3: For Grade 3, the models again showed strong performance. LR, KNN, GB, NN, AdaBoost, and SGD achieved perfect scores across all metrics, while tree and RF had slightly lower scores.
- Grade 4: In the Grade 4 classification, the models maintained high performance. LR, KNN, GB, NN, AdaBoost, and SGD attained perfect scores in all metrics, with tree and RF scoring slightly lower in some areas.
- Grade 5: The models' performance in Grade 5 was also outstanding. Most models achieved perfect scores, with tree and RF again showing slightly lower scores in some metrics.
- Overall performance: When evaluating the average performance across all grades, all models demonstrated exceptional classification capabilities. LR, KNN, GB, NN, AdaBoost, and SGD achieved perfect scores across all metrics, while tree and RF had slightly lower scores. These consistently high scores highlight the effectiveness of these models in accurately classifying VUR instances, as detailed in Table 1.

4.2 Confusion matrix analyses

The confusion matrix is a crucial ML and statistical analysis tool that shows a classification model's performance. This matrix-like Table 2 shows the model's prediction accuracy by comparing predicted and actual class labels. The confusion matrix's rows are the actual class labels and the predicted columns. Each matrix cell shows the count or percentage of instances in a specific predicted and actual class. This matrix details the model's classification abilities, showing correct and incorrect predictions. A comprehensive evaluation of the model's effectiveness is possible by calculating accuracy, precision, recall, and F1 score using the confusion matrix. This quantitative assessment of the model's performance aids medical, financial, and social science classification task decision-making and optimization, as shown in Table 2.

Table 2. The confusion matrix analyses for the models logistic regression, tree, KNN, random forest, gradient boosting, neural network, AdaBoost, and stochastic gradient descent

		Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Σ
Logistic Regression	Grade 1	3	0	0	0	0	3
	Grade 2	0	14	0	0	0	14
	Grade 3	0	0	25	0	0	25
	Grade 4	0	0	0	15	0	15
	Grade 5	0	0	0	0	23	23
	Σ	3	14	25	15	23	80
Tree	Grade 1	3	0	0	0	0	3
	Grade 2	0	14	0	0	0	14
	Grade 3	0	0	24	0	1	25
	Grade 4	0	1	1	13	0	15
	Grade 5	0	0	0	0	23	23
	Σ	3	15	25	13	24	80
kNN	Grade 1	3	0	0	0	0	3
	Grade 2	0	14	0	0	0	14
	Grade 3	0	0	25	0	0	25
	Grade 4	0	0	0	15	0	15
	Grade 5	0	0	0	0	23	23
	Σ	3	14	25	15	23	80
Random Forest	Grade 1	2	0	0	0	1	3
	Grade 2	0	14	0	0	0	14
	Grade 3	0	0	25	0	0	25
	Grade 4	0	0	0	15	0	15
	Grade 5	0	0	0	1	22	23
	Σ	2	14	25	16	23	80

(*Continued*)

Table 2. The confusion matrix analyses for the models logistic regression, tree, KNN, random forest, gradient boosting, neural network, AdaBoost, and stochastic gradient descent (*Continued*)

		Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Σ
Gradient Boosting	Grade 1	3	0	0	0	0	3
	Grade 2	0	14	0	0	0	14
	Grade 3	0	0	25	0	0	25
	Grade 4	0	0	0	15	0	15
	Grade 5	0	0	0	0	23	23
	Σ	3	14	25	15	23	80
Neural Network	Grade 1	3	0	0	0	0	3
	Grade 2	0	14	0	0	0	14
	Grade 3	0	0	25	0	0	25
	Grade 4	0	0	0	15	0	15
	Grade 5	0	0	0	0	23	23
	Σ	3	14	25	15	23	80
AdaBoost	Grade 1	3	0	0	0	0	3
	Grade 2	0	14	0	0	0	14
	Grade 3	0	0	25	0	0	25
	Grade 4	0	0	0	15	0	15
	Grade 5	0	0	0	0	23	23
	Σ	3	14	25	15	23	80
Stochastic Gradient Descent	Grade 1	3	0	0	0	0	3
	Grade 2	0	14	0	0	0	14
	Grade 3	0	0	25	0	0	25
	Grade 4	0	0	0	15	0	15
	Grade 5	0	0	0	0	23	23
	Σ	3	14	25	15	23	80

The confusion matrices show the performance of various models in classifying instances across different grades. The LR, KNN, GB, NN, AdaBoost, and SGD models all accurately predicted three instances of Grade 1, 14 of Grade 2, 25 of Grade 3, 15 of Grade 4, and 23 of Grade 5 out of 80 predictions, demonstrating their strong performance. The tree model had minor misclassifications but generally performed well. The RF model showed slight inaccuracies, particularly with one misclassified instance in Grade 1. Overall, all models demonstrated their proficiency in classifying instances, with most achieving near-perfect predictions, highlighting their effectiveness and accuracy in the task.

5 DISCUSSION

The evaluation metrics for different ML models were applied to multiple datasets labeled GR1, GR2, GR3, GR4, and GR5. The metrics assessed include AUC, CA,

F1, precision, recall, MCC, specificity, and logloss. The models evaluated, such as LR, tree, KNN, RF, GB, NN, AdaBoost, and SGD, consistently demonstrate outstanding performance across all datasets. Most models achieve perfect scores (1) in all metrics, indicating their high accuracy and effectiveness in classifying instances. While RF displays slightly lower scores in some metrics, it still maintains an overall high level of accuracy. These results highlight the models' ability to effectively learn the underlying patterns in the data and make accurate predictions. The low LogLoss values further confirm the models' precise estimation of probabilities. Consequently, these models exhibit strong potential for practical applications in various domains, providing reliable and robust classification capabilities, as shown in Table 1.

The performance of different ML models (LR, tree, KNN, RF, GB, NN, AdaBoost, and SGD) on a classification task across five different grades (Grade 1, Grade 2, Grade 3, Grade 4, and Grade 5). Each cell in Table 2 indicates the number of instances in which each model correctly classified a specific grade. The last row and column (Σ) display the total number of cases correctly classified for each grade and the overall total. Upon analyzing the results, several observations can be made: All models achieve perfect classification accuracy (indicated by the numbers in each row summing up to the total in the last column) across all grades. This suggests that the models were highly influential in correctly predicting the class labels for the instances in the dataset. The models consistently perform well across all grades, with no cases misclassified. This indicates that the models successfully captured the underlying patterns and relationships in the data for each grade. The performance of each model is consistent across the different grades. For example, in LR, three instances were correctly classified for Grade 1, 14 for Grade 2, 25 for Grade 3, 15 for Grade 4, and 23 for Grade 5. Similar patterns can be observed for the other models as well. Overall, the results demonstrate the strong performance of the ML models in accurately classifying instances across different grades. The consistent perfect scores across all models and grades suggest that the models have effectively learned the patterns in the data, making them reliable for classification tasks in this context, as shown in Table 2.

6 CONCLUSION

This study demonstrates the significant potential of ML models for accurately grading VUR using VCUG images. Among the various models tested, including LR, DTs, KNN, RF, GB, NNs, AdaBoost, and SGD, the GB model consistently showed superior performance across multiple evaluation metrics such as accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUROC). The inclusion of additional features, such as ureter tortuosity, further enhanced the model's ability to distinguish between different VUR grades, surpassing the traditional subjective methods, which suffer from high inter-rater variability. Our ML-based approach not only provides a more reliable and objective grading system but also holds promise for improving patient stratification and treatment decisions in clinical settings. However, to fully realize the clinical applicability and generalizability of this model, further validation with larger and more diverse datasets is necessary. Additionally, testing the model in real-world clinical environments will be crucial to ensuring its robustness and effectiveness in improving diagnostic accuracy and patient outcomes.

7 REFERENCES

[1] M. S. Alzboon, M. S. Al-Batah, M. Alqaraleh, A. Abuashour, and A. F. H. Bader, "Early diagnosis of diabetes: A comparison of machine learning methods," *Internation Journal of Online Biomedical Engineering (iJOE)*, vol. 19, no. 15, pp. 144–165, 2023. https://doi.org/10.3991/ijoe.v19i15.42417

[2] M. S. Alzboon, M. Al-Batah, M. Alqaraleh, A. Abuashour, and A. F. Bader, "A comparative study of machine learning techniques for early prediction of prostate cancer," in *2023 IEEE Tenth International Conference on Communications and Networking, (ComNet),* 2023, pp. 1–12. https://doi.org/10.1109/ComNet60156.2023.10366703

[3] M. S. Alzboon, S. Qawasmeh, M. Alqaraleh, A. Abuashour, A. F. Bader, and M. Al-Batah, "Machine learning classification algorithms for accurate breast cancer diagnosis," in *2023 3rd International Conference on Emerging Smart Technologies and Applications (eSmarTA)*, Taiz, Yemen, 2023, pp. 1–8. https://doi.org/10.1109/eSmarTA59349.2023.10293415

[4] M. S. Al-Batah, M. S. Alzboon, and R. Alazaidah, "Intelligent heart disease prediction system with applications in Jordanian Hospitals," *International Journal of Advanced Computer Science Appllications (IJACSA)*, vol. 14, no. 9, 2023. https://doi.org/10.14569/IJACSA.2023.0140954

[5] M. S. Alzboon and M. S. Al-Batah, "Prostate cancer detection and analysis using advanced machine learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 8, pp. 388–396, 2023. https://doi.org/10.14569/IJACSṁA.2023.0140843

[6] M. Al-Batah, B. Zaqaibeh, S. A. Alomari, and M. S. Alzboon, "Gene microarray cancer classification using correlation based feature selection algorithm and rules classifiers," *Internation Journal of Online Biomedical Engineering (iJOE)*, vol. 15, no. 8, pp. 62–73, 2019. https://doi.org/10.3991/ijoe.v15i08.10617

[7] M. Alzboon, "Semantic text analysis on social networks and data processing: Review and future directions," *Inf. Sci. Lett.*, vol. 11, no. 5, pp. 1371–1384, 2022. https://doi.org/10.18576/isl/110506

[8] M. S. Alzboon, "Survey on patient health monitoring system based on internet of things," *Inf. Sci. Lett.*, vol. 11, no. 4, pp. 1183–1190, 2022. https://doi.org/10.18576/isl/110418

[9] F. O'Kelly, "Commentary to quantification of vesicoureteral reflux using machine learning," *Journal of Pediatric Urology*, vol. 20, no. 2, pp. 265–266, 2023. https://doi.org/10.1016/j.jpurol.2023.10.043

[10] A. Khondker *et al.*, "A machine learning-based approach for quantitative grading of vesicoureteral reflux from voiding cystourethrograms: Methods and proof of concept," *J. Pediatr. Urol.*, vol. 18, no. 1, pp. 78.E1–78.E7, 2022. https://doi.org/10.1016/j.jpurol.2021.10.009

[11] S. Ganapathy, H. K. T. B. Jindal, P. S. Naik, and N. S. Nair, "Comparison of diagnostic accuracy of models combining the renal biomarkers in predicting renal scarring in pediatric population with vesicoureteral reflux (VUR)," *Irish Journal Medical Science*, vol. 192, pp. 2521–2526, 2023. https://doi.org/10.1007/s11845-023-03275-z

[12] S. Kabir, J. L. Pippi Salle, M. E. H. Chowdhury, and T. O. Abbas, "Quantification of vesicoureteral reflux using machine learning," *Journal of Pediatric Urollogy*, vol. 20, no. 2 pp. 257–264, 2023. https://doi.org/10.1016/j.jpurol.2023.10.030

[13] T. Köse, S. Özgür, E. Coşgun, A. Keskinoğlu, P. Keskinoğlu, and D. Mrozek, "Effect of missing data imputation on deep learning prediction performance for vesicoureteral reflux and recurrent urinary tract infection clinical study," *BioMed Research International*, vol. 2020, no. 1, 2020. https://doi.org/10.1155/2020/1895076

[14] A. Dalkiliç, G. Bayar, H. Demirkan, and K. Horasanli, "The learning curve of sting method for endoscopic injection treatment of vesicoureteral reflux," *Internationl Braz J Urol*, vol. 44, no. 6, pp. 1200–1206, 2018. https://doi.org/10.1590/s1677-5538.ibju.2017.0465

[15] M. Escolino *et al.*, "Endoscopic injection of bulking agents in pediatric vesicoureteral reflux: A narrative review of the literature," *Pediatr. Surg. Int.*, vol. 39, no. 1, 2023. https://doi.org/10.1007/s00383-023-05426-w

[16] O. Ergün, T. A. Serel, S. A. Öztürk, H. B. Serel, S. Soyupek, and B. Hoşcan, "Deep-learning-based diagnosis and grading of vesicoureteral reflux: A novel approach for improved clinical decision-making," *J. Surg. Med.*, vol. 8, no. 1, pp. 12–16, 2024. https://doi.org/10.28982/josam.8020

[17] A. Simicic Majce *et al.*, "Intrarenal reflux in the light of contrast-enhanced voiding urosonography," *Front. Pediatr.*, vol. 9, 2021. https://doi.org/10.3389/fped.2021.642077

[18] I. Selvi and N. Baydilli, "Selecting children with vesicoureteral reflux who are most likely to benefit from antibiotic prophylaxis: Application of machine learning to RIVUR. letter," *Journal of Urology*, vol. 206, no. 5, pp. 1337–1338, 2021. https://doi.org/10.1097/JU.0000000000002191

[19] G. Gurumoorthy and S. Ganesh Vaidyanathan, "Gastric disorder analysis using hybrid optimization with machine learning," *Journal of Biomatererilas and Tissue Engineering*, vol. 13, no. 3, pp. 453–462, 2023. https://doi.org/10.1166/jbt.2023.3269

[20] K. S. Lee, I. S. Song, E. S. Kim, H. I. Kim, and K. H. Ahn, "Association of preterm birth with medications: Machine learning analysis using national health insurance data," *Arch. Gynecol. Obstet.*, vol. 305, pp. 1369–1376, 2022. https://doi.org/10.1007/s00404-022-06405-7

[21] B. V. Srividya and S. Sasi, "Early detection of gastroesophageal reflux disease using logistic regression and support vector machine," *International Journal Organization and Collectoctive Intelligence (IJOCI)* vol. 11, no. 2, pp. 75–90, 2021. https://doi.org/10.4018/IJOCI.2021040104

[22] K. S. Lee, E. S. Kim, D. Y. Kim, I. S. Song, and K. H. Ahn, "Association of gastroesophageal reflux disease with preterm birth: Machine learning analysis," *J. Korean Med. Sci.*, vol. 36, no. 43, 2021. https://doi.org/10.3346/jkms.2021.36.e282

[23] P. H. Lin, J. G. Hsieh, P. C. Wu, H. C. Yu, and J. H. Jeng, "Machine learning approach for risk prediction of erosive esophagitis in a health check-up population in Taiwan," in *2021 3rd IEEE Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)*, 2021, pp. 208–210. https://doi.org/10.1109/ECBIOS51820.2021.9510751

[24] M. J. Lee *et al.*, "Novel cocrystals of vonoprazan: Machine learning-assisted discovery," *Pharmaceutics*, vol. 14, no. 2, p. 429, 2022. https://doi.org/10.3390/pharmaceutics14020429

[25] S. Hozawa, S. Maeda, A. Kikuchi, and M. Koinuma, "Exploratory research on asthma exacerbation risk factors using the Japanese claims database and machine learning: A retrospective cohort study," *J. Asthma*, vol. 59, no. 7, pp. 1328–1337, 2022. https://doi.org/10.1080/02770903.2021.1923740

[26] M. Bauschard *et al.*, "Computer-enhanced vsual learning method to teach endoscopic correction of vesicoureteral reflux: An invitation to residency training programs to utilize the CEVL method," *Advances in Urology*, vol. 2012, no. 1, 2012. https://doi.org/10.1155/2012/831384

[27] A. Serel, S. A. Ozturk, S. Soyupek, and H. B. Serel, "Deep learning in urological images using convolutional neural networks: An artificial intelligence study," *Turkish J. Urol.*, vol. 48, no. 4, pp. 299–302, 2022. https://doi.org/10.5152/tud.2022.22030

[28] Ö. Çelik, A. F. Aslan, U. Ö. Osmanoğlu, N. Cetın, and B. Tokar, "Estimation of renal scarring in children with lower urinary tract dysfunction by utilizing resampling technique and machine learning algorithms," *J. Surg. Med.*, vol. 4, no. 7, pp. 573–577, 2020. https://doi.org/10.28982/josam.691768

[29] A. M. Lee *et al.*, "Using machine learning to identify metabolomic signatures of pediatric chronic kidney disease etiology," *J. Am. Soc. Nephrol.*, vol. 33, no. 2, pp. 375–386, 2022. https://doi.org/10.1681/ASN.2021040538

[30] M. Kamal Pasha, "Machine learning and artificial intelligence based identification of risk factors and incidence of gastroesophageal reflux disease in Pakistan," *Int. J. Educ. Manag. Eng.*, vol. 11, no. 5, pp. 23–31, 2021. https://doi.org/10.5815/ijeme.2021.05.03

[31] W. Chen *et al.*, "Prediction of persistent chronic cough in patients with chronic cough using machine learning," *ERJ Open Res.*, vol. 9, 2023. https://doi.org/10.1183/23120541.00471-2022

[32] F. Amato *et al.*, "Obesity and gastroesophageal reflux voice disorders: A machine learning approach," in *2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Messina, Italy, 2022, pp. 1–6. https://doi.org/10.1109/MeMeA54994.2022.9856574

[33] Sandra *et al.*, "Predicting piperine content in javanese long pepper using fluorescence imaging and machine learning model," in *BIO Web Conf.*, 2024, vol. 90. https://doi.org/10.1051/bioconf/20249002003

[34] S. Dave *et al.*, "Learning from the learning curve: Factors associated with successful endoscopic correction of vesicoureteral reflux using dextranomer/hyaluronic acid copolymer," *Journal of Urology*, vol. 180, no. 4S, pp. 1594–1600, 2008. https://doi.org/10.1016/j.juro.2008.03.084

[35] K. S. Lee *et al.*, "Association of preterm birth with depression and particulate matter: Machine learning analysis using national health insurance data," *Diagnostics*, vol. 11, no. 3, p. 555, 2021. https://doi.org/10.3390/diagnostics11030555

[36] A. J. A. Hussein, M. L. Othman, A. Ishak, B. S. M. Noor, and A. H. M. S. Sajitt, "Optimization of distribution control system in oil refinery by applying hybrid machine learning techniques," *IEEE Access*, vol. 10, pp. 3890–3903, 2022. https://doi.org/10.1109/ACCESS.2021.3134931

[37] K. Takahashi *et al.*, "Achalasia phenotypes and prediction of peroral endoscopic myotomy outcomes using machine learning," *Dig. Endosc.*, vol. 36, no. 7, pp. 789–800, 2024. https://doi.org/10.1111/den.14714

[38] C. Gandhi *et al.*, "Biosensor-assisted method for abdominal syndrome classification using machine learning algorithm," *Comput. Intell. Neurosci.*, vol. 2022, no. 1, 2022. https://doi.org/10.1155/2022/4454226

[39] A. Khondker *et al.*, "Multi-institutional validation of improved vesicoureteral reflux assessment with simple and machine learning approaches," *J. Urol.*, vol. 208, no. 6, pp. 1314–1322, 2022. https://doi.org/10.1097/JU.0000000000002987

[40] H. Wang, M. Li, D. Bertsimas, C. Estrada, and C. Nelson, "MP64-03 selecting children with VUR who are most likely to benefit from antibiotic prophylaxis: Application of machine learning to RIVUR data," *Journal of Urology*, vol. 201, Supplement no. 4, p. e941, 2019. https://doi.org/10.1097/01.JU.0000556895.20387.16

[41] H. H. Yen, H. Y. Tsai, C. C. Wang, M. C. Tsai, and M. H. Tseng, "An improved endoscopic automatic classification model for gastroesophageal reflux disease using deep learning integrated machine learning," *Diagnostics*, vol. 12, no. 11, p. 2827, 2022. https://doi.org/10.3390/diagnostics12112827

8 AUTHORS

Muhyeeddin Alqaraleh is an Assistant Professor at Jadara University. Dr. Alqaraleh's areas of expertise include Computer Engineering, Information and Communication Technology, Computer Networking, Digital Signal Processing, Electronics and Communication Engineering, Signal, Image and Video Processing, Information Technology, Network Communication, Communication & Signal Processing, Networking, Cloud Computing, Network Security, Network Architecture, Wireless Computing, Network technology, Signal Processing, Signal Processing for Communication, Radio Communication, Information Theory, Discrete-Time Signal Processing, Computer Technology, IT Infrastructure, Hardware Troubleshooting, Computer Networks Security, Security, Network Management, IT Security, Network Administration, Network Configuration, Network Simulation, and Information Security. He is fluent in English, Arabic, and Russian (E-mail: m.qaralleh@jadara.edu.jo).

Mowafaq Salim Alzboon is an Assistant Professor at the Science and Information Technology Faculty at Jadara University, Jordan. He holds a PhD degree in computer science from the ersity of Utara Malaysia. Dr. Alzboon's research interests center around cloud computing, autonomic computing, visualization technology, load balancing, overlay networks, mobile application development, and the Internet of Things. He has expertise in a range of disciplines, including computer architecture, computer communications networks, and distributed computing. Dr. Alzboon's skill-set encompasses overlay network, computer networks, load balancing, cloud computing, parallel and distributed computing, networking, virtualization, virtualization technology, distributed computing, autonomic computing, and grid computing (E-mail: malzboon@jadara.edu.jo).

Mohammad Subhi Al-Batah holds a PhD in Computer Science with a specialization in Artificial Intelligence, which he received from the University of Science Malaysia in 2009. He currently serves as a Lecturer at the Faculty of Sciences and Information Technology at Jadara University, Jordan. In 2018, he also served as the Director of the Academic Development and Quality Assurance Center at Jadara University. Dr. Al-Batah's research interests span a range of topics, including image processing, artificial intelligence, real-time classification, and software engineering (E-mail: albatah@jadara.edu.jo).

Mutaz Abdel Wahed is an Assistant Professor at the Department of Computer Networks and Cybersecurity at Jadara University in Irbid, Jordan. With a solid academic background, he earned his PhD in computer networks from HSE University in Moscow, Russian Federation, in 2007. His current research interests include network security networks security, networks quality of service, and artificial intelligence. In addition to his academic pursuits, he brings invaluable industrial experience to his role, having worked extensively in the fields of networking and cybersecurity. His practical insights enrich his teaching methodologies and research perspectives, bridging the gap between theory and real-world applications. He holds certifications from industry leaders such as Fortinet and Cisco, underscoring his dedication to professional development and ensuring his expertise remains current and relevant. He continues to inspire and educate the next generation of cybersecurity professionals, fostering innovation and excellence in the dynamic field of computer networks and cybersecurity (E-mail: mutaz@jadara.edu.jo).

Ahmad Abuashour is a Professor Assistant in the Department of Information Technology and Computing at Arab Open University in Kuwait. He received his

Bachelor of Engineering degree in Computer Engineering from Jordan University of Science and Technology and his Master of Engineering degree in Electrical Engineering from Concordia University. Currently, he is pursuing his PhD in Electrical Engineering at the Ecole de Technologie Superieure (ETS), University of Quebec, Canada. His current research interests are focused on Intelligent Transportation Systems (ITS), Vehicular Ad-Hoc Networks (VANETs), cluster-based routing protocols, network management and monitoring, and quality of service. Specifically, he is concentrating on VANET routing protocols. The Faculty of Computer Studies at Arab Open University is in the Ardiya Industrial Area, Farwanya, with the Information Technology and Computing (ITC) department being assigned the postal code 13033. Postal correspondence may be addressed to P.O. Box 3322, Kuwait (E-mail: aabuashour@aou.edu.kw).

Firas Hussein Alsmadi is a Pediatric Radiology Specialist at Queen Rania Hospital for children, which is a part of the Jordan Royal Medical Services. He holds the Jordanian Board in Diagnostic Radiology from the Jordan Medical Council. Dr. Alsmadi's research interests center on leveraging artificial intelligence and machine learning to enhance diagnostic accuracy and streamline workflows. He aims to improve patient outcomes and contribute to the advancement of pediatric radiology (E-mail: firassmadi24@gmail.com).