Machine learning for real-time prediction of complications induced by flexible uretero-renoscopy with laser lithotripsy

Chafik Baidada[1], Hamid Hrimech[2], Mustapha Aatila[1], Mohamed Lachgar[1], Younes Ommane[3]
[1]Information Technology Laboratory, National School of Applied Sciences, Chouaib Doukkali University, El Jadida, Morocco
[2]Analysis and Modeling of Systems and Decision Support Laboratory, National School of Applied Sciences of Berrechid, Hassan 1er University of Settat, Berrechid, Morocco
[3]Institute of Biological Sciences (ISSB), UM6P-Faculty of Medical Sciences (UM6P-FMS), Mohammed VI Polytechnic University, Ben-Guerir, Morocco

Article Info	ABSTRACT
	It is not always easy to predict the outcome of a surgery. Peculiarly, when talking about the risks associated to a given intervention or the possible complications that it may bring about. Thus, predicting those potential complications that may arise during or after a surgery will help minimize risks and prevent failures to the greatest extent possible. Therefore, the objectif of this article is to propose an intelligent system based on machine learning, allowing predicting the complications related to a flexible uretero-renoscopy with laser lithotripsy for the treatment of kidney stones. The proposed method achieved accuracy with 100% for training and, 94.33% for testing in hard voting, 100% for testing and 95.38% for training in soft voting, with only ten optimal features. Additionally, we were able to evaluted the machine learning model by examining the most significant features using the Shaply additive explanations (SHAP) feature importance plot, dependency plot, summary plot, and partial dependency plots.

Corresponding Author:

Chafik Baidada
Information Technology Laboratory, National School of Applied Sciences, Chouaib Doukkali University
El Jadida, Morocco
Email: baidada.c@ucd.ac.ma

1. INTRODUCTION

A surgical intervention has a number of risks, some of which are more likely than others depending on the circumstances. Surgical procedures such as cardiothoracic or pulmonary surgery, hepatobiliary surgery, abdominal surgery, prostate removal, and major osseous and articular surgery all have higher risks. Further, depending on their health, certain patients are at higher risk of complications: malnutrition, diabetes, obesity, and antecedents of cardiac insufficiency are all factors that increase the surgical risk.

Like all operations, kidney stone surgery by uretero-renoscopy (URS) presents an increasing rate of infectious complications. These URS related complications may increase the risk of mortality and morbidity especially in high-risk patients. Prophylactic antibiotic use, stent and operation dwell time restriction, early detection and treatment of urinary tract infections (UTIs) and urosepsis, careful planning in patients presenting with a load of substantial calculus and numerous comorbidities are some examples of these complications. With improving technology, flexible minimally invasive URS with laser lithotripsy (FURSLL) has recently been introduced as a new technic to improve the outcome of ureterorenal stone surgery. Even for individuals with proximal kidney and ureter stones smaller than 20 mm, FURSLL has become an attractive option [1], [2]. For this kind of stones, stone-free FURSLL rates can be as high as

94.1% with an overall 28% risk of preoperative complications [3], [4]. Although flexible URS lithotripsy is a less intrusive procedure, there is a possibility that complications could emerge [5]–[7].

The complications caused by URS have been divided into two categories: minor and major. complications are considered minors if they can be effectively treated without surgery. They include asymptomatic ureteral perforations, ileus, and fever. Major complications are those that necessitate surgery or present a life-threatening situation. Seldom avulsions, tears, perforations during basketing, intussusception, and infection are among the second category of serious complications that are commonly correlated to stone extractions [8]. It is noteworthy, in some cases, following URS for stone removal, Necrosis of ureteral segments is observed [9]. Major complications of URS can have serious and lasting consequences. Usually open or laparoscopic surgery is almost always required in major complications and the basic goal is to restore ureteral continuity [10].

Machine learning (ML) is attracting more and more interest as a method for analyzing statistically the complex and expanding body of medical data in order to enhance patient treatment on an individual basis. Integrating the use of ML-based decision support tools into surgical practices can have many benefits. ML refers to algorithms that can be designed to evaluate and make predictions based on new and complex features [11]. A system with ML model that incorporates the full range of patient data offers a way to stratify patients with nephrolithiasis to avoid potential complications after surgery. These systems deserve to be used as supportive in a complementary manner, and based on data inputs in a multidisciplinary, multifunctional surgical risk analysis and large reduction protocol. The necessary steps of the decision-making process should be based on the predicted risk stratification data they offer. For renal surgery teams, an equivalent process and technology integration point has the potential to be advantageous for patient safety. Thus, the aim of this work is to propose an intelligent decision support system based on ML, in order to predict the complications for the cases of the patients who underwent a surgical operation using a flexible URS.

The remainder of this article is organized as follows: section 2 presents related works. Section 3 describes the methodology used to predict the operating time of lithotripsy by flexible URS. The simulation results are presented in section 4. Section 5 discusses the experimental results. Finally, conclusions and some recommendations for research perspectives are given in section 6.

2. RELATED WORKS

By imitating cognitive capacities, artificial intelligence (AI) refers to the capacity of computer to accomplish tasks [12]. ML is a subset of AI that allows algorithms, to train models to learn from massive and complicated data then produce valuable predictive outputs [13]. New patterns and associations can be identified by using ML algorithms on new, massive data sets, and these results may have positive effects on clinical practice in medicine [14]. A large number of studies have examined at the use of ML approaches in healthcare and prove how they significantly increase healthcare quality and safety [15], [16]. Table 1 illustrates the latest applications of ML methods in predicting surgical outcomes.

Table 1. Latest applications of ML methods in predicting surgical outcomes

Paper	Year	Surgery	Data Set (Size)	Best Model
Zeng et al. [17]	2019	Cardiac	2308 paediatric patients	eXtreme gradient boosting classifier Boost (XGBBoost) prediction model with a SMOTE-optimized approach
Merali et al. [18]	2019	Neurosurgery	757 patients	Random forest (RF)
Mark et al. [19]	2019	Urosurgery	Not applicable	Proposed model
Khanna et al. [20]	2020	Cardiothoracic	17,433 postoperative patients	Not applicable
Zhong et al. [21]	2021	Cardiogenic	6844	XGBoost
Shah et al. [22]	2021	Orthopedic	545 patients	Novel ML-based ensemble algorithm
Lit al. [23]	2021	Orthopedic	1330 hip fracture surgical patients	Random survival forest (RSF) machine-learning algorithm
Cole et al. [24]	2021	Visceral surgery	93,024 patients undergoing exploratory laparotomy	Not applicable
Kumar et al. [25]	2021	Orthopedic	Clinical data from 2153 primary anatomic total shoulder arthroplasty (TSA) patients and 3621 primaryr everse TSA patients were analyzed	XGBoost
Devana et al. [26]	2021	Orthopedic	150,000 patients	Gradient boosting
Bunn et al. [27]	2021	Visceral surgery	Not applicable	An ensemble of the best models including, LogisticRegression (LR), RF, and XGB machines
Penny-Dimri et al. [28]	2021	Cardiac/uro surgery	Not applicable	
Shi et al. [29]	2021	Vascular surgery	672 consecutive patients who had undergone aortic arch surgery	Naïve Bayes
Fernandes et al. [30]	2021	Cardiac	5,015 patients	XGBoost

Although still in their infancy, machine learning algorithms provide surgeons an opportunity to take capitalize on the abundance of clinical data and enhance personalized patient care. Although artificial intelligence has been increasingly being utilized in healthcare, research still mostly focuses on cardiovascular, nervous system, and cancer problems since they are the main causes of disability and death. Chronic and infectious diseases like diabetes, inflammatory bowel disease, and Clostridium difficult infections have also received some attention. By increasing the quality of the extraction of clinical data and implementing that data into a properly trained and verified system, early diagnosis can now be achieved for a variety of conditions. To the best of our knowledge, no paper up to the moment has addressed the subject of predicting complications related to flexible uretero-renoscopy with laser lithotripsy. This is why we have initiated this work, in order to fill this gap.

3. METHOD

Predicting the post-operative complication risks is very important for physicians to identify patients who need intensive monitoring or any additional interventions. Risk prediction also concerns patients and their families for decision-making during surgery. However, an intelligent system for estimating the risk of complications will be of a paramount importance.

The methodology proposed in this study aims to implement a real-time prediction system for the complications risks during flexible ureteroscopy with laser lithotripsy operations, using ML. The methodology adopted is described in Figure 1. The proposed system is developed using an initial dataset consisting of 5 classes [31]. These classes correspond to the five types of complications that can occur during a flexible ureteroscopy with laser lithotripsy operation according to the clavien-dindo scale (the clavien-dindo) classification scale makes it possible to classify these complications into 5 different classes, according to the severeness and the required interventions to fix these complications, varying from the absence of complications to the death of the patient [32]. Since this dataset is imbalanced and has some classes with only one observation, it cannot be used without being preprocessed.
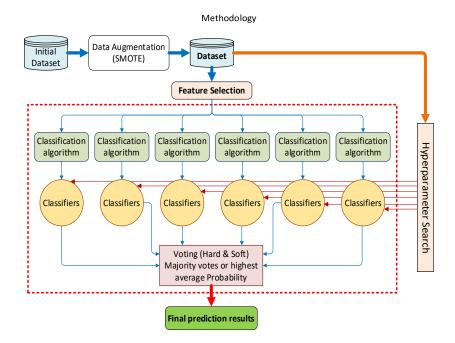

Figure 1. Description of the proposed methodology

3.1. Data preprocessing

During the data preprocessing phase, the dataset is reorganized into three main classes instead of 5. The classes (0: which corresponds to the class of no complications, 1: representing the class of complications of degree 1 and 2: which represents the class of complications of degree 2) according to the Clavien-Dindo classification scale. Classes 3 and 4 are ignored because they contain very limited number of entries that cannot be augmented in the following step. In this phase, a quality verification of the data is ensured. This verification is guaranteed by eliminating redundant values, solving the problems of missing values as well as

categorical values such as textual ones. Although we preprocessed the dataset the best we could, the size of the dataset is still scarce class wise, henceforth, it needs some augmentation as well a balancing of these classes.

3.2. Data augmentation

The dataset resulting from the preprocessing contains a small number of elements of classes 2 and 3 compared to the number of observations of class 0, the proposed prediction system attempts to predict a large number of observations during the test as being of class 0, which will influence the performance of our predictive system. To fix this problem, data augmentation is applied to balance the dataset by generating observations for classes with a very low number of observations. The data augmentation method adopted in this study is the synthetic minority over-sampling technique (SMOTE), based on the generation of synthetic observations between each sample of the class with a low number of elements and its closest observations [33].

3.3. Features selection

Once the dataset is balanced, a selection of the most relevant features is applied. This selection is performed by using six different feature selection algorithms. The features selection techniques used in this study are presented in Table 2. The most relevant features selected after the selection phase, are then used by ML models.

Table 2. List of used features selection algorithms

Features selection algorithms
F_classif [34]
Embedded (random forest) [35]
Recursive feature elimination (RFE) [35]
Hybrid recursive feature addition (H_RFA) [35]
Sequential forward selection (SFS) [35]
Sequential backward selection (SBS) [35]

3.4. Proposed classification system

In order to build a more efficient classification system, several methods were compared. First, classification is performed on the entire data set. Then, only the features retained by the best feature selection algorithm among the techniques compared are used.

The second step of classification consisted of an estimation of the risks of complications. The estimation is done by adopting a predictive model based on the vote, without and with the use of hyperparameters. This is to make a final decision while improving the performance of estimating the complication risks of the proposed system. The search for hyper-parameters is ensured by using the tuning function. This technic consists in training the system according to different possibilities for each hyper-parameter and keeping the best parameterization having generated the best score during the classification.

For the voting ensemble model proposed in this study, two voting modes were compared. The soft voting based on the sum of prediction probabilities of each class among all models of proposed system and the hard voting based on the label majority of different models of the proposed system [36]. The obtained results of complication risks estimation are compared to retain the method providing the best prediction performance. Finally, a comparison of the variables according to their impacts during the classification is carried out, this classification of the variables retained according to their importance is provided by using the SHAP value. The different classifiers used for the system proposed are listed in the following Table 3.

3.5. Performance metrics

Performance metrics are an integral part of the ML model evaluation process. They are very useful for measuring and comparing model performance. The metrics used in this work are:

− The accuracy: Which represents the rate of correct predictions of the classes. The accuracy is described by the following (1) [37].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (1)$$

− The recall which represents the number of true positives compared to true positives and false negatives. The following (2) represents the recall [37].

$$Recall = \frac{TP}{TP+FN} \qquad (2)$$

− The precision representing the number of true positives compared to true positives and false positives. Precision is represented by (3) [37].

$$Precision = \frac{TP}{TP+FP} \qquad (3)$$

− The F1-score which is a metric combining precision and recall with respect to a specific positive class, it is a weighted average of precision and recall. The following (4) describes the F1-score [37].

$$F1 - score = 2 * \frac{Precision*Recall}{Precision+Recall} \qquad (4)$$

Table 3. Description of the used models

Model	Classifier
Ensemble methods	Ada boost classifier (ABC) [38]
	Bagging classifier (BC) [39]
	Extra trees classifier (ETC) [40]
	Gradient boosting classifier (GBC) [41]
	Random forest classifier (RF) [42]
Linear models	Logistic regression CV (LRC) [43]
	Passive aggressive classifier (PAC) [44]
	Ridge classifier CV (RC) [45]
	SGD classifier (SGDC) [46]
	Perceptron (PC) [47]
Naive Bayes (NB)	Bernoulli NB (BNBC) [48]
	Gaussian NB (GNBC) [49]
Trees	Decision tree classifier (DTC) [50]
	Extra tree classifier (TETC) [51]
Discriminant analysis	Linear discriminant analysis (LDAC) [35]
	Quadratic discriminant analysis (QDAC) [52]
KNN	K-Neighbors classifier (KNNC) [35]
XGB	XGB classifier [53]
Gaussian	Gaussian process classifier (GPC) [54]
Support vector machine (SVM)	SVC (SVC) [34]

4. SIMULATION RESULTS

4.1. Dataset description

In order to predict the potential postoperative complications during a flexible uretero-renoscopy with laser lithotripsy, the dataset used in this study is composed of 682 observations with 21 features representing three classes. Class 0 representing the case of no complications. Class 1 for grade 1 complications and class 2 for grade 2 complications based on the Clavien-Dindo scale. The distribution of these classes is shown in the Table 4.

We observe that the dataset is unbalanced, with a percentage of 90% the class 0 is the majority, compared to 5% for the two other classes 1 and 2. A dataset balancing was applied using the SMOTE augmentation method. Figure 2 represents the data from the original dataset and the data balanced by class.

Table 4. Description of the dataset

Dataset size	Classes	Size	Size in Percentage
682	Class 0	611	90%
	Class 1	36	5%
	Class 2	35	5%

4.2. Technical description of the used calculator

The experiments conducted were executed on a computer equipped with one processor: Intel® - Core™ i7-8550U with a frequency of 1792 MHz. This processor has 2 cores and 2 threads. The computer is equipped with 8 GB of RAM DDR3 and no GPU.

4.3. Simulation results using all dataset features

Table 5 represents the results of the different predictive models used, implemented on the dataset. The results are ranked by the accuracy as a criterion. The models generate different results, which vary between 34% and 96% for the classification of the different observations.
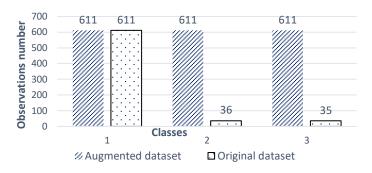
Machine learning for real-time prediction of complications induced by flexible uretero … (Chafik Baidada)

Figure 2. Imbalanced dataset description

Table 5. Obtained results using the whole dataset sorted by performance

Classifier	Accuracy	Precision	Recall	F1-score	Time	STD
RF	0.964546	0.97	0.97	0.97	2.351014	0.011473
ETC	0.962901	0.97	0.97	0.97	1.735743	0.013103
GBC	0.948174	0.95	0.95	0.95	10.324434	0.011771
XGBC	0.938358	0.94	0.94	0.94	3.858799	0.014813
BC	0.935083	0.93	0.93	0.93	0.660065	0.014287
DTC	0.890906	0.89	0.89	0.89	0.171835	0.021187
TETC	0.868511	0.86	0.86	0.86	0.062524	0.024295
ABC	0.79869	0.8	0.8	0.8	1.302251	0.028541
KNN	0.75011	0.75	0.75	0.74	0.189993	0.027075
BNBC	0.673224	0.67	0.67	0.67	0.076299	0.023681
LDAC	0.661763	0.66	0.66	0.66	0.099243	0.029559
RC	0.655756	0.66	0.66	0.66	0.079234	0.031138
QDAC	0.614819	0.59	0.61	0.59	0.05547	0.021746
GNBC	0.566886	0.61	0.57	0.54	0.050787	0.051238
LRC	0.491572	0.48	0.49	0.49	35.585023	0.030065
GPC	0.479568	0.77	0.48	0.4	36.914091	0.028373
SGDC	0.363323	0.35	0.35	0.35	0.156019	0.027466
PAC	0.348028	0.38	0.38	0.37	0.1261	0.047485
PC	0.342572	0.34	0.34	0.34	0.093727	0.049332

4.4. Results with sequential feature selector method

The previously compared algorithms were reused, except that in this case using the sequential feature selector (SFS) method. The SFS method was chosen based on the good classification performance obtained by using this technique compared to the other feature selection techniques. Table 6 represents the results obtained in decreasing order of classification accuracy. We can observe that the overall performance of the different models has improved from an accuracy that varies between 34% and 96% to an accuracy that varies between 72% and 89.6%.

Table 6. Results obtained by applying the SFS method of features selection

Classifier	Accuracy	Precision	Recall	F1-score	Time	STD
ABC	**0.895908**	0.8	0.9	0.85	0.621365	0.057029
GPC	**0.895908**	0.8	0.9	0.85	1.784786	0.057029
LRC	**0.895908**	0.8	0.9	0.85	2.404244	0.057029
RC	**0.895908**	0.8	0.9	0.85	0.082582	0.057029
BNBC	**0.895908**	0.8	0.9	0.85	0.040798	0.057029
LDAC	**0.895908**	0.8	0.9	0.85	0.054753	0.057029
TETC	0.894459	0.85	0.89	0.86	0.04488	0.050158
ETC	0.891539	0.84	0.89	0.86	0.896618	0.048591
XGBC	0.891496	0.8	0.89	0.84	0.373503	0.057713
RF	0.887127	0.85	0.89	0.86	1.1961	0.048557
KNN	0.887106	0.83	0.89	0.85	0.064682	0.060538
DTC	0.885656	0.84	0.89	0.85	0.047142	0.046637
GBC	0.884207	0.81	0.88	0.84	0.546534	0.053396
SGDC	0.881223	0.8	0.89	0.84	0.097078	0.053173
BC	0.878303	0.84	0.88	0.85	0.178523	0.036519
QDAC	0.847506	0.82	0.85	0.83	0.047974	0.034877
GNBC	0.806351	0.84	0.81	0.82	0.048964	0.168637
PAC	0.735656	0.81	0.62	0.7	0.075248	0.270387
PC	0.723487	0.82	0.72	0.77	0.053181	0.284594

4.5. Results using voting ensemble and SFS method without models hyperparameter tuning

In this section, we use the hard and soft voting-based classification without models' hyper-parameter tuning. This classification is performed on the features selected by the SFS method. The results of this experiment are summarized in Table 7. To better compare results, the ensemble prediction system's performances, based on the vote (hard and soft), during the training and the test, using the whole dataset is presented on Figure 3.

Table 7. Results by voting without hyper-parameters tuning and applying the SFS features selection algorithm

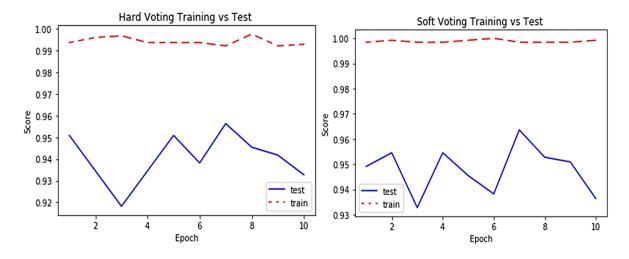| Selection method	Hard voting accuracy			Soft voting accuracy		
	Training w/bin mean	Test w/bin mean	Test w/bin 3*std	Training w/bin mean	Test w/bin mean	Test w/bin 3*std
Mutual Inf	98.93	92.31	+/- 2.50	99.91	93.73	+/- 2.12
fclassif	95.65	88.35	+/- 3.77	97.81	89.47	+/- 3.89
Embedded	97.83	89.65	+/- 3.21	99.49	90.62	+/- 3.74
RFE_Features	98.68	91.75	+/- 2.65	99.87	92.93	+/- 3.07
h_RFA_Features	81.05	58.69	+/- 7.37	94.28	59.58	+/- 6.04
SFS	**99.35**	**92.33**	**+/- 2.96**	**99.95**	**94.38**	**+/- 2.25**
SBS	98.27	90.96	+/- 4.07	99.6	92.4	+/- 3.27
All Balanced Dataset	**99.43**	**94.04**	**+/- 3.19**	**99.88**	**94.78**	**+/- 2.74**

Figure 3. Algorithms comparison for all balanced datasets using votes without hyper-parameter search

4.6. Results using voting ensemble, SFS method and models hyperparameter tuning

This section represents the results obtained by adopting a hard and soft voting based on classification with hyper-parameter tuning, and SFS method. Table 8 represents the obtained results. The Figure 4 represent as example performances of predictions, in terms of accuracy estimation of post-operatory complication risks, obtained both hard and soft voting over 10 epochs via the ensembling predictive model using SFS feature selection.

Table 8. Results by voting without hyper-parameter fine-tuning and using SFS algorithm for feature selection

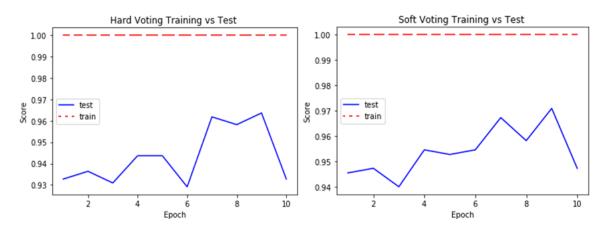| Selection Method	Hard voting accuracy			Soft voting accuracy		
	Training w/bin mean	Test w/bin mean	Test w/bin 3*std	Training w/bin mean	Test w/bin mean	Test w/bin 3*std
Mutual Inf (MI)	100	93.45	+/- 2.79	100	94.49	+/- 2.65
fclassif	98.3	88.73	+/- 5.08	99.65	89.95	+/- 4.65
Embedded	99.98	90.75	+/- 3.62	100	92.45	+/- 3.54
RFE_Features	100	93.36	+/- 2.99	100	94.38	+/- 2.72
h_RFA_Features	90.76	57.93	+/- 5.81	68.58	45.05	+/- 8.21
SFS	**100**	**94.33**	**+/- 3.80**	**100**	**95.38**	**+/- 2.75**
SBS	99.94	92.27	+/- 2.62	100	93.38	+/- 3.04

Figure 4. Algorithms comparison for SFS feature selection using votes with hyper-parameter search

5. DISCUSSION

The results obtained by the different classifiers trained on the complete dataset (21 variables), without features selection, are represented in Table 5. Based on these results, the random forest, extra tree classifier, gradient boosting classifier, XGBClassifier and BaggingClassifier algorithms constitutes the top five classifications. These algorithms provide an estimation the estimation of the risk of complications with an accuracy of 96.45%, 96.29%, 94.82%, 93.83%, and 93.51% respectively. The result generated by RF is relatively better with a standard deviation of 0.011473 compared to 0.013103 for ETC. Other algorithms are less efficient with a classification accuracy varying between 34.26% and 89.10%. The PC classifier is the least efficient among all the models used. The Figure 5 represents a comparative performance of the different classifiers used on the whole dataset.

In terms of execution time, the RF and ETC algorithms provide estimations in 2.351014 s and 1.735743 s respectively. Nonetheless, the algorithm that consumes more time is the GPC with 36.914091 s and an accuracy which does not exceed 47%. The SFS method was selected as the feature selection technique. Analyzing the results in Table 6, generated by the different algorithms proposed in this study using the SFS method, the ABC, GPC, LRC, RC, BNBC and LDAC algorithms allow a prediction with an accuracy of around 90%, and the execution time varying between 0.04 s and 2.4 s. The PC algorithm represents the least-performing classifier with a classification accuracy of about 70%. The application of the SFS variable selection method has significantly improved the performance of the lowest performing models by over 36%. Therefore, the accuracy of PC classifier increased from 34% to 70% and the GPC increased from 48% to 89.6%, hence the interest of using the most important features. Figure 6 represents a comparison of the classification results of the different models used in this section.
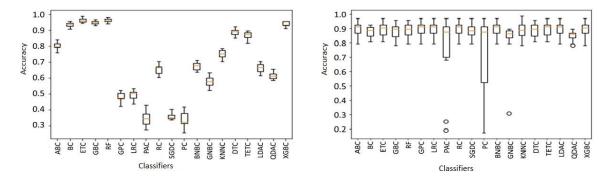

Figure 5. Algorithm's comparison using all dataset

Figure 6. Algorithm's comparison using SFS method

The use of a hard and soft voting-based ensemble prediction system could improve the performance of individual models. Decision making by combining the estimates of several models would be of great interest. The training of the proposed hard and soft vote-based ensemble prediction model without the use of hyper-parameters, and using the different features selection methods, generated the results shown in Table 7.

These results show that the use of the variables selected by the SFS method allows an estimation with the highest accuracy. The accuracy in this case was respectively 99.35% and 92.33% for training and testing in the hard vote and 99.95% and 94.38% respectively for training and testing in the soft vote.

The execution of the same proposed model, but with the use of hyper-parameters, provides the results shown in Table 8 for the different features selection methods. The use of hyper-parameters significantly increases the estimation performance of the proposed model and the SFS method is also the best feature selection methods in this case. The accuracy is 100% for training and, 94.33% for testing in hard voting and 100% for testing and 95.38% for training in soft voting. The soft vote generates a better prediction in both cases, without and with the use of hyper-parameters. The study of the impact of the different variables on the performance of the predictive model, during the classification, is accomplished by calculating the SHAP value for the different variables. Figure 7 represents the SHAP value of the selected features.
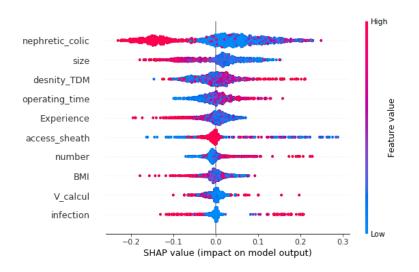

Figure 7. SHAP value of the features selected by the SFS technique during the classification

The most important feature is nephretic_colic, high values for this value have a significant negative impact on the estimation of the classification system, while low values of this variable have a positive impact on the performance of the system. The other variables that have a significant impact on the prediction are size and density_TDM. Figure 8 presents the impact of the different features on the different classes of the dataset. Figure 8 shows that the variables nephretic_colic, size and density_TDM have a significant impact on all classes of the dataset. The three classes of the dataset use the most relevant variables almost equally, except that the size variable is used more by classes 1 and 2.
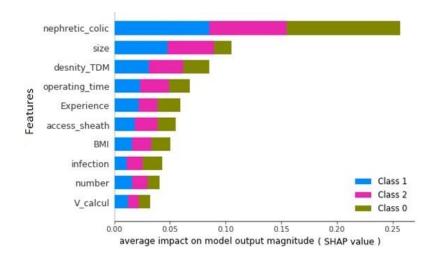

Figure 8. SHAP value of the features selected by the SFS technique by class

6.   CONCLUSION

Despite their power, AI systems cannot replace medical and surgical teams. But they can be used in certain cases to help these teams make better decisions. For example, these systems can be used to predict the likelihood of complications during surgery. Kidney surgery is a high-risk field that seeks to reduce problems and enhance surgical methods and results. The contemporary surgeon operates in a technology setting and has access to a wealth of data input before, during, and after surgery. ML has the potential to be employed in more creative ways to enhance patient safety during kidney surgery and raise the likelihood of successful outcomes.

The use of ML in urologic surgery was highly heterogeneous, and this, together with the preponderance of single studies, shows that there is a significant potential for further refining the studies in this field. The ensemble approach proposed in this work appeared to be successful in predicting postoperative complications in urological surgery, by predicting with precising a variety of surgical results. Ongoing use of ML through well-developed clinical decision support systems is probably going to reduce complication rates and improve the safety and quality of surgery. Our future work consists on developing a system that integrates. Our future work consists of developing operational decision support software for surgeons that incorporates these ML models.

REFERENCES

[1]   A. Srisubat, S. Potisat, B. Lojanapiwat, V. Setthawong, and M. Laopaiboon, "Extracorporeal shock wave lithotripsy (ESWL) versus percutaneous nephrolithotomy (PCNL) or retrograde intrarenal surgery (RIRS) for kidney stones," *Cochrane Database of Systematic Reviews*, vol. 2014, no. 11, Nov. 2014, doi: 10.1002/14651858.CD007044.pub3.

[2]   E. S. Hyams *et al.*, "A prospective, multi-institutional study of flexible ureteroscopy for proximal ureteral stones smaller than 2 cm," *Journal of Urology*, vol. 193, no. 1, pp. 165–169, Jan. 2015, doi: 10.1016/j.juro.2014.07.002.

[3]   B. Javanmard, A. H. Kashi, M. M. Mazloomfard, A. A. Jafari, and S. Arefanian, "Retrograde intrarenal surgery versus shock wave lithotripsy for renal stones smaller than 2 cm: A randomized clinical trial," *Urology journal*, vol. 13, no. 5, pp. 2823–2828, Oct. 2016.

[4]   H. Kilicarslan *et al.*, "Unfavorable anatomical factors influencing the success of retrograde intrarenal surgery for lower pole renal calculi.," *Urology journal*, vol. 12, no. 2, pp. 2065–2068, Apr. 2015.

[5]   W. Zhong, G. Leto, L. Wang, and G. Zeng, "Systemic inflammatory response syndrome after flexible ureteroscopic lithotripsy: a study of risk factors," *Journal of Endourology*, vol. 29, no. 1, pp. 25–28, Jan. 2015, doi: 10.1089/end.2014.0409.

[6]   C. Senocak *et al.*, "Risk factors of infectious complications after flexible uretero-renoscopy with laser lithotripsy," *Urology journal*, vol. 15, no. 4, pp. 158–163, Jul. 2018, doi: 10.22037/uj.v0i0.3967.

[7]   J. P. Blackmur, N. U. Maitra, R. R. Marri, F. Housami, M. Malki, and C. Mcilhenny, "Analysis of factors' association with risk of postoperative urosepsis in patients undergoing ureteroscopy for treatment of stone disease," *Journal of Endourology*, vol. 30, no. 9, pp. 963–969, Sep. 2016, doi: 10.1089/end.2016.0300.

[8]   A. Gaizauskas, M. Markevicius, S. Gaizauskas, and A. Zelvys, "Possible complications of ureteroscopy in modern endourological era: Two-point or 'scabbard' avulsion," *Case Reports in Urology*, vol. 2014, pp. 1–6, 2014, doi: 10.1155/2014/308093.

[9]   V. Gupta *et al.*, "Complete ureteral avulsion.," *TheScientificWorldJournal*, vol. 5, pp. 125–127, 2005, doi: 10.1100/tsw.2005.13.

[10]   K. Kijvikai, G. E. Haleblian, G. M. Preminger, and J. de la Rosette, "Shock wave lithotripsy or ureteroscopy for the management of proximal ureteral calculi: an old discussion revisited," *Journal of Urology*, vol. 178, no. 4, pp. 1157–1163, Oct. 2007, doi: 10.1016/j.juro.2007.05.132.

[11]   M. Kohli, L. M. Prevedello, R. W. Filice, and J. R. Geis, "Implementing machine learning in radiology practice and research," *American Journal of Roentgenology*, vol. 208, no. 4, pp. 754–760, Apr. 2017, doi: 10.2214/AJR.16.17224.

[12]   Z. Obermeyer and E. J. Emanuel, "Predicting the future-big data, machine learning, and clinical medicine," *New England Journal of Medicine*, vol. 375, no. 13, pp. 1216–1219, Sep. 2016, doi: 10.1056/nejmp1606181.

[13]   S. Raschka et V. Mirjalili, *Python machine learning: Machine learning and deep learning with Python*. Scikit-Learn TensorFlow Second Ed. Ed, 2017.

[14]   G. D. Brusko, J. P. G. Kolcun, and M. Y. Wang, "Machine-learning models: the future of predictive analytics in neurosurgery," *Clinical Neurosurgery*, vol. 83, no. 1, pp. E3--E4, Jun. 2018, doi: 10.1093/neuros/nyy166.

[15]   D. Lopez and G. Manogaran, "A survey of big data architectures and machine learning algorithms in healthcare," *International Journal of Biomedical Engineering and Technology*, vol. 25, no. 2/3/4, p. 182, 2017, doi: 10.1504/ijbet.2017.10008616.

[16]   R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Briefings in Bioinformatics*, vol. 19, no. 6, pp. 1236–1246, May 2017, doi: 10.1093/bib/bbx044.

[17]   X. Zeng *et al.*, "Prediction of complications after paediatric cardiac surgery," *European Journal of Cardio-thoracic Surgery*, vol. 57, no. 2, pp. 350–358, Jul. 2020, doi: 10.1093/ejcts/ezz198.

[18]   Z. G. Merali, C. D. Witiw, J. H. Badhiwala, J. R. Wilson, and M. G. Fehlings, "Using a machine learning approach to predict outcome after surgery for degenerative cervical myelopathy," *PLoS ONE*, vol. 14, no. 4, Apr. 2019, doi: 10.1371/journal.pone.0215133.

[19]   E. Mark, D. Goldsman, B. Gurbaxani, P. Keskinocak, and J. Sokol, "Using machine learning and an ensemble of methods to predict kidney transplant survival," *PLoS ONE*, vol. 14, no. 1, Jan. 2019, doi: 10.1371/journal.pone.0209068.

[20]   A. K. Khanna *et al.*, "A nomogram to predict postoperative pulmonary complications after cardiothoracic surgery," *Journal of Thoracic and Cardiovascular Surgery*, vol. 165, no. 6, pp. 2134–2146, Jun. 2023, doi: 10.1016/j.jtcvs.2021.08.034.

[21]   Z. Zhong, X. Yuan, S. Liu, Y. Yang, and F. Liu, "Machine learning prediction models for prognosis of critically ill patients after open-heart surgery," *Scientific Reports*, vol. 11, no. 1, Feb. 2021, doi: 10.1038/s41598-021-83020-7.

[22]   A. A. Shah, S. K. Devana, C. Lee, R. Kianian, M. van der Schaar, and N. F. SooHoo, "Development of a novel, potentially universal machine learning algorithm for prediction of complications after total hip arthroplasty," *Journal of Arthroplasty*, vol. 36, no. 5, pp. 1655--1662, May 2021, doi: 10.1016/j.arth.2020.12.040.

[23] Y. Li, M. Chen, H. Lv, P. Yin, L. Zhang, and P. Tang, "A novel machine-learning algorithm for predicting mortality risk after hip fracture surgery," *Injury*, vol. 52, no. 6, pp. 1487–1493, Jun. 2021, doi: 10.1016/j.injury.2020.12.008.

[24] J. Cole, S. Hughey, A. Metzger, P. Geiger, L. Fluke, and G. J. Booth, "Machine learning to predict fascial dehiscence after exploratory laparotomy surgery," *Journal of Surgical Research*, vol. 268, pp. 514–520, Dec. 2021, doi: 10.1016/j.jss.2021.06.068.

[25] V. Kumar *et al.*, "Using machine learning to predict clinical outcomes after shoulder arthroplasty with a minimal feature set," *Journal of Shoulder and Elbow Surgery*, vol. 30, no. 5, pp. e225--e236, May 2021, doi: 10.1016/j.jse.2020.07.042.

[26] S. K. Devana, A. A. Shah, C. Lee, A. R. Roney, M. van der Schaar, and N. F. SooHoo, "A novel, potentially universal machine learning algorithm to predict complications in total knee arthroplasty," *Arthroplasty Today*, vol. 10, pp. 135–143, Aug. 2021, doi: 10.1016/j.artd.2021.06.020.

[27] C. Bunn *et al.*, "Application of machine learning to the prediction of postoperative sepsis after appendectomy," *Surgery (United States)*, vol. 169, no. 3, pp. 671–677, Mar. 2021, doi: 10.1016/j.surg.2020.07.045.

[28] J. C. Penny-Dimri, C. Bergmeir, C. M. Reid, J. Williams-Spence, A. D. Cochrane, and J. A. Smith, "Machine learning algorithms for predicting and risk profiling of cardiac surgery-associated acute kidney injury," *Seminars in Thoracic and Cardiovascular Surgery*, vol. 33, no. 3, pp. 735–745, 2021, doi: 10.1053/j.semtcvs.2020.09.028.

[29] S. Shi *et al.*, "Using machine learning to predict postoperative liver dysfunction after aortic arch surgery," *Journal of Cardiothoracic and Vascular Anesthesia*, vol. 35, no. 8, pp. 2330–2335, Aug. 2021, doi: 10.1053/j.jvca.2021.02.046.

[30] M. P. B. Fernandes, M. Armengol de la Hoz, V. Rangasamy, and B. Subramaniam, "Machine learning models with preoperative risk factors and intraoperative hypotension parameters predict mortality after cardiac surgery," *Journal of Cardiothoracic and Vascular Anesthesia*, vol. 35, no. 3, pp. 857–865, Mar. 2021, doi: 10.1053/j.jvca.2020.07.029.

[31] F. Richard *et al.*, "Evaluation and comparison of scoring systems for predicting stone-free status after flexible ureteroscopy for renal and ureteral stones," *PLoS ONE*, vol. 15, no. 8 August, p. e0237068, Aug. 2020, doi: 10.1371/journal.pone.0237068.

[32] D. Wang *et al.*, "Associations of postoperative complications assessed by clavien–dindo classification and comprehensive complication index with long-term overall survival in elderly patients after radical crc resection," *Clinical Interventions in Aging*, vol. 15, pp. 1939–1949, Oct. 2020, doi: 10.2147/CIA.S271969.

[33] A. Fernández, S. García, F. Herrera, and N. V Chawla, "SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary," *Journal of Artificial Intelligence Research*, vol. 61, pp. 863–905, Apr. 2018, doi: 10.1613/jair.1.11192.

[34] C. Yan, J. Zhang, X. Kang, Z. Gong, J. Wang, and G. Zhang, "Comparison and evaluation of the combinations of feature selection and classifier on microarray data," in *2021 IEEE 6th International Conference on Big Data Analytics, ICBDA 2021*, Mar. 2021, pp. 133–137, doi: 10.1109/ICBDA51983.2021.9403151.

[35] M. Aatila, M. Lachgar, H. Hamid, and A. Kartit, "Keratoconus severity classification using features selection and machine learning algorithms," *Computational and Mathematical Methods in Medicine*, vol. 2021, pp. 1–26, Nov. 2021, doi: 10.1155/2021/9979560.

[36] X. Yu *et al.*, "Deep ensemble learning for human action recognition in still images," *Complexity*, vol. 2020, pp. 1–23, Jan. 2020, doi: 10.1155/2020/9428612.

[37] R. Shaaque, A. Mehmood, G. S. Choi, R. Shafique, and S. Ullah, "Cardiovascular disease prediction system using extra trees classiier cardiovascular disease prediction system using extra trees classifier," Sep. 2019, doi: 10.21203/rs.2.14454/v1.

[38] S. K. Prabhakar and H. Rajaguru, "Adaboost classifier with dimensionality reduction techniques for epilepsy classification from EEG," in *IFMBE Proceedings*, vol. 66, Springer Singapore, 2018, pp. 185–189.

[39] A. Subasi, B. Kadasa, and E. Kremic, "Classification of the cardiotocogram data for anticipation of fetal risks using bagging ensemble classifier," *Procedia Computer Science*, vol. 168, pp. 34–39, 2020, doi: 10.1016/j.procs.2020.02.248.

[40] L. Abhishek, "Optical character recognition using ensemble of SVM, MLP and extra trees classifier," *2020 International Conference for Emerging Technology, INCET 2020*, Jun. 2020, doi: 10.1109/INCET49848.2020.9154050.

[41] N. Chakrabarty, T. Kundu, S. Dandapat, A. Sarkar, and D. K. Kole, "Flight arrival delay prediction using gradient boosting classifier," in *Advances in Intelligent Systems and Computing*, vol. 813, 2019, pp. 651–659.

[42] N. Dogru and A. Subasi, "Traffic accident detection using random forest classifier," in *2018 15th Learning and Technology Conference, L and T 2018*, Feb. 2018, pp. 40–45, doi: 10.1109/LT.2018.8368509.

[43] G. P. Mallya, A. Hiremani, G. Gajakosh, D. Badiger, and M. P. G. Sunitha Hiremath, "Framework for automatic diagnosis of psychological disorders," in *Proceedings - IEEE 2018 International Conference on Advances in Computing, Communication Control and Networking, ICACCCN 2018*, Oct. 2018, pp. 929–934, doi: 10.1109/ICACCCN.2018.8748450.

[44] S. R. K. Indarapu, J. Komalla, D. R. Inugala, G. Reddy Kota, and A. Sanam, "Comparative analysis of machine learning algorithms to detect fake news," in *2021 3rd International Conference on Signal Processing and Communication, ICPSC 2021*, May 2021, pp. 591–594, doi: 10.1109/ICSPC51351.2021.9451690.

[45] U. Soni, N. Gupta, and Sakshi, "An artificial intelligence approach for forecasting ebola disease," *Journal of Physics: Conference Series*, vol. 1950, no. 1, Aug. 2021, doi: 10.1088/1742-6596/1950/1/012038.

[46] N. Deepa *et al.*, "An AI-based intelligent system for healthcare analysis using Ridge-Adaline Stochastic Gradient Descent Classifier," *Journal of Supercomputing*, vol. 77, no. 2, pp. 1998–2017, May 2021, doi: 10.1007/s11227-020-03347-2.

[47] S. Ahmadi, H. Beyhaghi, A. Blum, and K. Naggita, "The strategic perceptron," in *EC 2021 - Proceedings of the 22nd ACM Conference on Economics and Computation*, Jul. 2021, pp. 6–25, doi: 10.1145/3465456.3467629.

[48] B. Meindl, I. Ott, and U. Zierahn, "Binary patent classification methods for few annotated samples," *Proceedings of the 1st Workshop on on Patent Text Mining and Semantic Technologies/Editors: Linda Andersson, Hidir Aras, Florina Piroi, Allan Hanbury*, 2019, doi: 10.34726/PST2019.4.

[49] C. H. Chen, S. J. H. Yang, J. X. Weng, H. Ogata, and C. Y. Su, "Predicting at-risk university students based on their e-book reading behaviours by using machine learning classifiers," *Australasian Journal of Educational Technology*, vol. 37, no. 4, pp. 130–144, Jun. 2021, doi: 10.14742/ajet.6116.

[50] A. Trabelsi, Z. Elouedi, and E. Lefevre, "Decision tree classifiers for evidential attribute values and class labels," *Fuzzy Sets and Systems*, vol. 366, pp. 46–62, Jul. 2019, doi: 10.1016/j.fss.2018.11.006.

[51] A. Sharaff and H. Gupta, "Extra-tree classifier with metaheuristics approach for email classification," in *Advances in Intelligent Systems and Computing*, vol. 924, Springer Singapore, 2019, pp. 189–197.

[52] B. Ghojogh and M. Crowley, "Linear and quadratic discriminant analysis: Tutorial," *arXiv preprint arXiv:1906.02590*, Jun. 2019.

[53] T. Zheng, L. Ning, Q. Ye, and F. Jin, "An XGB-based reliable transmission method in the mMTC scenarios," *Security and Communication Networks*, vol. 2021, pp. 1–12, Dec. 2021, doi: 10.1155/2021/9929051.

[54] A. Devonport and M. Arcak, "Data-driven reachable set computation using adaptive gaussian process classification and monte carlo methods," in *Proceedings of the American Control Conference*, Jul. 2020, vol. 2020-July, pp. 2629–2634, doi: 10.23919/ACC45564.2020.9147918.

BIOGRAPHIES OF AUTHORS

Chafik Baidada holds a PhD in computer sciences in software engineering from Moulay Ismail University, Morocco. Currently, he is a professor at the National school of Applied Sciences, Chouaib Doukkali, El jadida, Morocco. His research interests include software engineering, reverse engineering, model transformation, micro service architecture, iot systems, machine and deep learning. He can be contacted at email: baidada.c@ucd.ac.ma.

Hamid Hrimech in 2009, he received a PhD in computer science from Art et Métiers ParisTech in France. He is an associate professor at University Hassan 1st at Morocco's ENSA, Department of Computer Sciences and Mathematics. Artificial intelligence, collaborative virtual environment and driving simulation are among his research interest. He can be contacted at email: hamid.hrimech@uhp.ac.ma.

Mustapha Aatila is a Ph.D. student at the LTI Laboratory, ENSA, Chouaib Doukkali University, El Jadida, Morocco. He received the M.S degree in software quality from Cadi Ayyad University, Morocco in 2009. His research focuses on artificial intelligence, machine learning and deep learning applied to the health sector, particularly in ophthalmology field. He can be contacted at email: aatila.m@ucd.ac.ma.

Mohamed Lachgar received Ph. D in computer science at the Cadi Ayyad University in 2017. He is a professor in computer science at the National School of Applied Sciences, Chouaib Doukkali University of El Jadida, Morocco. His research interests are in the areas of automation tools development in embedded software, software modeling and design, metamodel design, model transformation, model verification and validation method, machine learning and deep learning. He can be contacted at email: lachgar.m@ucd.ac.ma.

Younes Ommane he is a Professor of Biostatistics at the Faculty of Medical Sciences, Mohammed VI Polytechnic University. He spent his postdoctoral fellowship in machine learning and data analysis at the College of Computing - UM6P. He obtained his bachelor in fundamental and applied mathematics from the University of Cadi Ayyad, then he moved to France to obtain his master's degree in informatics, statistics and mathematics applied to production management from the University of Toulouse Jean-Jaures, then he went back home to receive his Ph.D. in Bayesian statistical decision theory from University of Cadi Ayyad. His research interests are ML/AI for healthcare and clinical research, and Bayesian statistical decision theory. He can be contacted at email: Younes.ommane@um6p.ma.