A stage-oriented exploration of the prediction procedures and selection of informative features, in an application, for nephrology

Ewa Skubalska-Rafajłowicz

Wrocław University of Science and Technology, Faculty of Information and Communication Technology, Department of Computer Engineering,
Wyb. Wyspiańskiego 27, 50 370 Wrocław, Poland,
ewa.skubalska-rafajlowicz@pwr.edu.pl

Abstract: *A stage-oriented learning of the prediction procedure, along with a statistical test of feature informativeness based on the Shapley Additive explanations (SHAP) model, is proposed and investigated. Emphasizing the proper stage (rank) of each particular forecast is of special importance in medical applications, including nephrology, when a degree of disease is to be predicted. The (non-)informativeness test of each feature is based on the influence of its values on the prognosis, as measured by the SHAP values. The proposed approach was validated for predicting which class of chronic kidney disease (CKD) patients are expected to belong, after several months. A certain degree of individualized prognosis was achieved by using 12 features for each patient, from which, three were identified as non-informative or low-informative and subsequently removed. The medical aspects of this particular case study are outside the scope of this paper and will be published elsewhere. Here, we concentrate on methods and algorithms.*

Keywords: staged labels; predictor learning; informative features; Shapley values; chronic kidney disease

1 Introduction

Machine learning (ML) tools and algorithms are widely applied in prediction applications [1-3]. It seems, however, that it would be helpful to improve certain aspects of them, for more specialized applications. In particular, we have in mind, possible applications where a predictor provides the result from a finite set {1, 2, …, K} of labels, and it is essential whether an accurate prediction, k, say, is

erroneously recognized as k ± 1 or as k ± 3, say. Even more responsible is a cancer staging. This is why we advocate applying stage-oriented loss functions. Simultaneously, proper prognosis of a stage (grade, etc.) depends on selection of covariates which values are available at a moment of preparing the prognosis. As is known, in classification tasks, the larger the better rule of gaining features is frequently not true. For this reason many research have been undertaken. In particular, a number of applications based on Shapley values [4-6] is still growing.

Therefore, I propose to define a feature to be (linearly) non-informative if its values are uncorrelated with the corresponding Shapley values. For testing this property, the Sppearman's Rank Test is appropriate.

The above mentioned ingredients are embedded into more standard context of a classifier that prepares a prediction. Therefore, we describe the whole algorithm. On the other hand, one can select any reasonable classifier that allows for loss functions that react more severely when errors are larger than one stage. Thus, the selection of a classifier is also an essential step of the overall procedure, which is presented in Section 2. In Section 3, I present the results of validating all the proposed approach on historical data representing prognosis of progress of the Chronic Kidney Disease (CKD) of 183 patients. I stress that the prediction results were compared with true, but historical and anonymized observations and did not have any impact on already past decisions of physicians.

I found no published studies on the CKD stage prediction that use patient-specific laboratory measurements and ML methods for a more personalized approach. However, research on using ML for Kidney Failure (KF) prediction has been tried recently. Statistical models were tested in [7], with the patient's baseline estimated glomerular filtration rate (eGFR) as the primary input variable. The basic model included age, gender, and albuminuria, while the extended model additionally accounted for the effects of serum calcium, phosphate, bicarbonate, and albumin. Statistical tests demonstrated good KF prediction. At present, the main tool for predicting kidney failure is the Kidney Failure Risk Equation (KFRE). In [8] the accuracy of KFRE and its variants was analyzed. 59 cohorts from around the world, involving approximately three million patients were analyzed. The most important findings can be summarized as follows:

a) KFRE performance does not significantly improve globally when eGFR and Albumin-to-Creatinine Ratio (ACR) are replaced by their historical or previous values of the eGFR slope or cardiovascular comorbidities.
b) In [8], a KFRE model with an added nonlinear term was proposed, which improves the accuracy of KF prediction in the subgroup of individuals over 65 years old or when eGFR values are relatively high.
c) It is recommended to use the known KFRE for eGFR less than 45 and to apply adjustments for eGFR between 45 and 60. For eGFR above 60, KFRE is not accurate, and other approaches should be considered, such as predicting a decline in eGFR of more than 40%.

As the authors state in [8], this is especially important given the success of SGLT2 inhibitors in preventing or delaying eGFR decline in these populations. Therefore, efforts to identify high-risk individuals early are crucial, as they may lead to preventing or markedly delaying kidney failure over a patient's lifetime.

These conclusions are essential for us because they support directing our research toward earlier risk detection when the patient's eGFR is not yet significantly reduced, and personalizing approaches by incorporating results from routine laboratory tests. In the paper [9], a decision tree learning method was proposed to assess progress toward KF, primarily based on a patient's current and prior eGFR values. It was noted that the influence of the patient's age on medium-term prognosis is much smaller than that of the current eGFR, and that gender has almost no impact on the decision. In a recent review [10], the authors selected 16 papers for a more detailed review of machine learning methods and features used to predict CKD progression. Key conclusions include, among others, that the chosen machine learning approaches were better, or at least not worse, than KFRE, and that only half of the studies used clinical variables as inputs. This highlights opportunities for broader investigations without overshadowing well-established features such as age and sex.

I mention paper [11] because of the advanced use of SHAP values to select factors that influence classifier performance. Namely, as the global indicator of feature importance, the authors in [11] use the SHAP summary plots. Additionally, they use dependence plots to show how a feature value influences the predicted value, accounting for other factors.

I refer the reader to the survey [12], in which references to all the class of algorithms that are applied here can be found. SHAP values will be used here as the main ingredient for interpretable machine learning. The reader can find basic ideas of SHAP in [4][5] and examples of applications in [6][13]. Plots of SHAP values summarize impacts of each feature separately on the predicted class. However, I propose to apply statistical tests on more detailed plots of feature values on SHAP values for each example (e.g., patient).

2 Method

In this section are a collection of the assumptions and notations, as well as the outline of the general algorithm. Then, I introduce the notion of (non-)informative feature and discuss the related statistical testing.

2.1 Basic notions and notations

For presenting ideas, we adopt the simplest model of prognosis. Namely, an unobserved random variable Y is to be predicted. In general, Y can assume values

in an interval or a finite set. For our purposes, we allow $Y \in \{z_1, z_2, \ldots, z_n\}$, where $z_1 < z_2 < \ldots < z_n$ or {I<II<…<ZZ} or {1,2,…, K} in the simplest cases, but keeping the strict ordering to hold the interpretation of labels as stages.

Each time, the prognosis is made after observing vector $x \in R^T$ that is a realization of the random vector $X \in R^T$ which probability distribution is in practice unknown, but a learning sequence of observations is available. Elements of $X \in R^T$ are further called features. In prognostic problems, it frequently happens that one of them, X^0 say is directly linked with Y as its baseline (initial, starting) version. For example, X^0 can be a patient's eGFR level just after hospitalization, while Y is CKD stage of this patient after a fixed number of months that serves as the prediction horizon T > 0. The rest of elements of $X \in R^T$ can be lab tests of the same patient at the beginning that can help in the prognosis.

As the loss when k^{th} stage is the proper one, while a predictor suggests j^{th} one, we take the quadratic function $(k-j)^2$ as providing more loss when k,j are farther apart, e.g., 9 when |k-j|=3 . This forces a predictor to make at most one stage error. This behavior is opposite to that provided by the so-called 0-1 loss, which is most popular in learning classifiers. This loss does not make any difference whether an error occurs at one stage or three, say.

Summarizing, the aim is to find predictor $Pred^*(\mathbf{x})$ that minimizes the conditional risk r(x) defined as:

$$r(\mathbf{x}) = \sum_{k=1}^{K} (k-Pred(x))^2 P(k|x) \quad (1)$$

where $P(k|\mathbf{x})$ is the probability that k is the actual class when the vector of features x is observed, and for this reason, it is called the a posteriori probability. The overall risk R(Pred) of using Pred(**x**) is obtained by taking the expectation of r(**x**) with respect to the probability distribution of **x** i.e., $R(Pred) = \mathbf{E}_X\, r(\mathbf{X})$, where **X** is the random vector of features having the probability distribution according to which **x** is drawn. In minimizing (1), I allow for all predictors such that Pred(**X**) is a random variable.

I assume that learning sequence $L_n = \{(\mathbf{x}(i), y(i)), i=1,2,\ldots,n\}$ is available, where $x(i) \in R^T$ is i^{th} vector of features, while $y(i) \in \{1,2,\ldots,K\}$ is the corresponding actual response after time T that is available as historical data only for training and validating purposes only.

When L_n is available, an empirical counterpart of (1) is minimized directly, for given x. In the considered case, it is possible to find the minimizer directly, namely, it is given by the a posteriori mean of weighted $P(x|\mathbf{k})$, k=1,2,…,K rounded to the nearest integer. Replacing $P(x|\mathbf{k})$, k=1,2,…,K by their empirical counterparts, denoted as $P\hat{}(x|\mathbf{k})$, k=1,2,…,K, we obtain the following empirical predictor:

$$Pred(x) = Round[\sum_{k=1}^{K} k P\hat{}(k|\mathbf{x})] \quad (2)$$

Where $P\hat{}(k|\mathbf{x})$, k=1,2,…,K are estimated, e.g., by a neural network (NN) or a classifier.

2.2 A General Algorithm

We present a general approach to learning, indicating points where we propose to elaborate, while simultaneously leaving flexibility in other areas:

Step 1 Choose predicted variable Y and its stages k=1,2,…K as well as vector of features $x \in R^T$ that are covariates for predicting Y. At the beginning of the learning phase, $X \in R^T$ may contain potentially useful features.

Step 2 Choose also the prediction horizon, T > 0, taking into account data availability after T units of time. Collect learning sequence $L_n=\{(\mathbf{x}(j),y(j), j=1,2,\ldots,n\}$ a statistically independent training sequence $L_n=\{(\mathbf{x}'(j),y'(j), j=1,2,\ldots,n'\}$ that has the same structure as L_n but possibly different length n'.

Step 3 Select the loss function either in the form (1) or in another one, but still sensitive to stage errors, e.g., (1) with the power two replaced by 4^{th} one

Step 4 Consider a family of classifiers, e.g., those in Table 1, and algorithms of their learning that are able to minimize the loss function selected in Step 3. Perform their learning and testing and select the best according to the loss function, but taking into account also other quality measures such as the accuracy, specificity, precision etc. If the results of validation are satisfactory, then STOP and provide the learned classifier together with the loss function and feature vector $X \in R^T$ as the selected predictor. Otherwise, consider extending the family of admissible classifiers and repeat Step 4. If it does not provide satisfactory results, go to Step 5.

Step 5 Compute Shapley values for all components of vector $X \in R^T$ and prepare the standard feature impact plots (see Fig. 1 and 2 for examples of such plots). For each feature prepare also data points and feature value impact plots (see Fig. 3 and 4 for examples of such plots). For each pair consisting of a given feature values examples and the corresponding Shapley values, state and test the hypothesis that this feature is noninformative in a way described below. If the list of noninformative features is not empty remove the corresponding features from vector $X \in R^T$ and from the corresponding positions of L_n and L'_n and run the learning and validating the predictor again. If the result is satisfactory, then STOP, otherwise, also stop and try other approaches, e.g., extending the list of features or reducing the prediction horizon etc.

Usage: After successful termination of the above algorithm, the predictor Pred^(x) is ready for use, after possible reduction of features in the input vector.

Most of the steps of the above procedure are standard ones. We have already explained the role of applying stage sensitive loss functions, especially in medicine applications. The only ingredient that require explanations is the one in Step 5, where we introduced the notion of (linearly) noninformative features, basing on the notions from the SHAP approach.

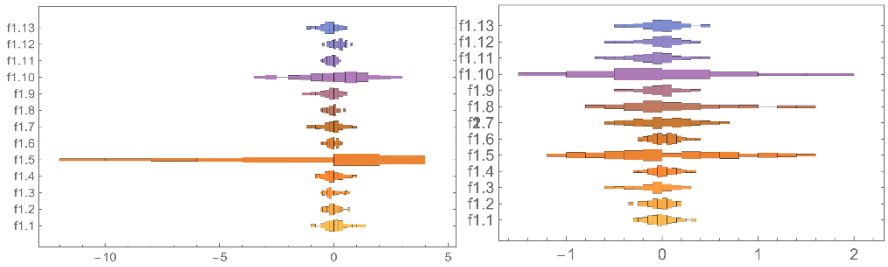


Figure 1

Feature impact plots for the nephrology data when the SVM classifier is used and class I (left panel) and class II (right panel) are considered. The explanations of coding features are provided in Figure 2.

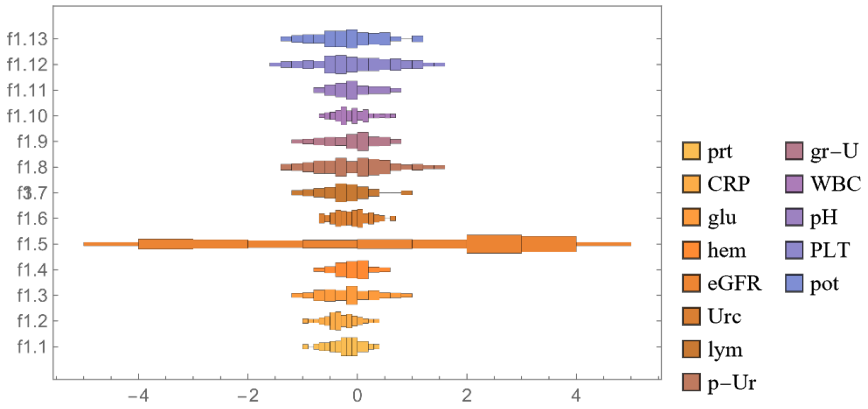
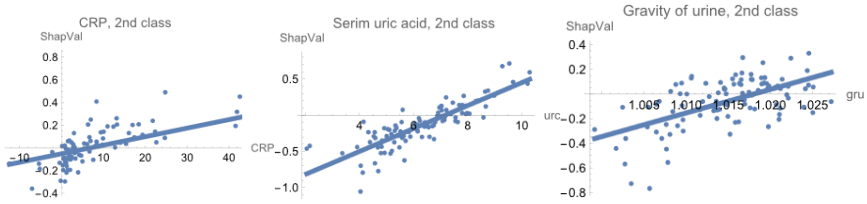


Figure 2

Feature impact plots for the kidney data when the SVM classifier is used and class III is a candidate to be predicted. Right panel contains explanations of the correspondence between the colors and the abbreviations of the feature names (in agreement with Table 1).



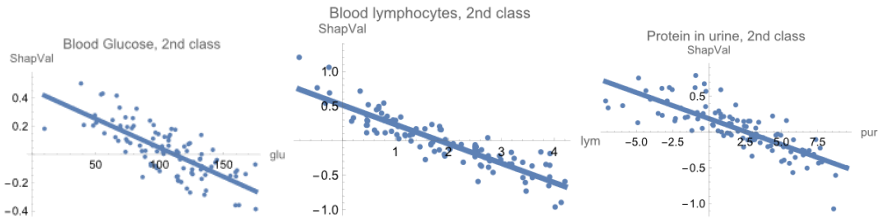


Figure 3

Examples of feature value impact plots of features that were tested as informative. Straight lines were obtained by the ordinary least squares as patterns along which the Shapley values vs the feature values are grouping. The dots correspond to data of patients for that are used for predicting class II CKD by the SVM classifier.

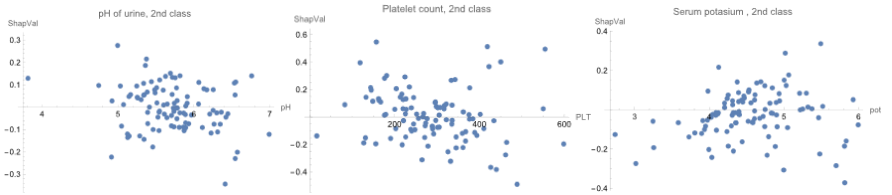


Figure 4

Examples of feature value impact plots of features that were tested as noninformative for the SVM classifier for predicting class II CKD

Table 1

The quadratic loss function and the accuracy of classifiers achieved in our experiments for selecting an initial version of the predictor

Classifier type	**Accuracy %**	**Quadratic loss**
Support vector machine (SM)	77	21
Logistic regression (LogReg)	74	25
Neural Network (4 Layers NN)	75	23
Decision Tree	71	28
RandomForest (70 forests)	64	34

2.3 On Testing Non-Informativeness of Features

The idea of applying Shapley values is to assist the predictor designer in selecting features that support the prediction by forming their ranking. Computing or even approximating the Shapley values is a complicated process and we refer the reader to [14] for a survey of such methods and to github/shapley for details. Let $S \subset \{0,1,\ldots,r\}$ be a subset of features indexes. Then, roughly speaking, a partial

contribution of i^{th} feature is the difference of the prediction goal function obtained by applying features $SU\{i\}$ minus the goal function obtained by using features from S only. The result is averaged over all possible subsets S (see formula (6) in [14] and its description). Hence, we can utilize the Shapley values and related plots that enable us to assess the influence of specific features on the overall prognosis of many patients. As observed in our computational experiments, the analysis of the Shapley values correctly identifies the most influential feature. However, it can be insufficient to distinguish the impact of less influential features reliably. Fortunately, it is possible to analyze the effect of each feature values on the Shapley values and we explore this way as more detailed than the most popular feature impact plots.

To this end, define by $\{\sim x\}^1$ a vector of i^{th} feature observations that consists of all elements $x^i(j)$ j=1,2,…,n. The corresponding Shapley values will be denoted by $s_i(j)$, j = 1,2,…,n, while their vector by $\{\sim s\}^1$, i=1,2,…,r. Pairs $(x^i(j),s^i(j))$, j = 1,2,…,n form a plot called feature value impact plot of i^{th} feature (see Fig. 3 and Fig. 4) for examples of such plots). Each point of such a plot corresponds to one learning example (e.g., a patient). Examinations of feature value impact plots reveal that some of them exhibit regular patterns, mainly grouping around an interval, while some of them remain scattered.

This kind of the behavior suggests to introduce the notion of (linearly) noninformative feature if its Shapley values randomly react on changes values of this feature.

Definition. Feature i^{th} is (linearly) noninformative if $\{\sim x\}^1$ and $\{\sim s\}^1$ are uncorrelated.

Now, we are at the position to quantify this notion, just by stating the hypothesis:

H_0: $\{\sim x\}^1$ and $\{\sim s\}^1$ are uncorrelated at a prescribed significance level $0 < \alpha < 1$

If H_0 is rejected, then we shall consider i^{th} feature to be informative. As a tool for testing H_0 one can select any nonparametric statistical test among those that are well verified in many statistical packages, e.g., Goodman-Kruskal, Kendall, Wilks. For our further purposes we take the Spearman Rank test and $\alpha = 0.05$ as the significance level.

Notice that I confined our attention to the cases of testing an existence of linear dependence between a feature and its Shapley values. As in general case, it happens that a linear correlation is low, but points(xi(j),si(j)), j = 1,2,…,n form another regular pattern, e.g., a quadratic one. Then, one can repeat the above reasoning and testing the significance of a quadratic model.

Data structure for learning the above algorithm are typical for pattern recognition problems (see [14] for discussion). The user has only to keep in mind, that historical data about predicted events, should be collected near the prediction horizon T. Hints concerning possible generalizations to other loss functions can be found in [15]

3 The Results And Validation

The proposed approach was validated using real-life medical data concerning 183 patients suffering from kidney diseases. Here, I provide a summary of this verification.

3.1 Available Data

Our aim is to learn a predictor of the Chronic Kidney Disease (CKD) after about eight months after the first hospitalization of a given patient. In more detail, CKD stages are standardized as 1 to 5, where the fifth stage is the most severe. For many medical reasons and data availability, we group these stages into three classes, denoted here as I, II and III. Class I comprises of original CKD stages 1 and 2. Class II is the same as the original CKD stage 3. As class III we grouped the CKD stages 4 and 5. This split resulted in the following empirical frequencies of class I, II and III: 0.34, 0.36 and 0.30 , respectively.

The prediction of a patient's class is done basing on his/her eGFR level at the beginning of the first hospitalization (baseline level). The estimated glomerular filtration rate is directly linked with the CKD stages and its level at the beginning is expected to have an essential influence on the prognosis, but information that it contains is worth be enriched by those contained in other features that are based on standard laboratory tests. For our purposes, we selected additional twelve features. Their names and acronyms are listed in Table 2 below. I start learning from initial eGFR and all twelve of them, but later I will check their informativity.

Table 2

The list of features, their acronyms and the results of applying the Spearman rank test for verifying their (non-)informativity, according to the way proposed in Section 2.3.

Acronym	Description	Spearman Rank Test Statistics	p-Value	Correlation/Conclusion
prt	Total protein (g/l)	-0.66	5.2×10^{-14}	0.68 /Informative
CRP	CRP (mg/l)	0.81	3.8×10^{-25}	0.72/Informative

hem	Blood hemoglobin (g/l)	-0.56	0.4×10^{-8}	0.64/ Informative
urc	Serum uric acid (micromol/l)	0.87	8.4×10^{-3}	0.85/Informative
lym	Blood lymphocytes (/ul)	-0.91	1.3×10^{-3}	-0.89/Informative
pur	Protein in urine (g/l)	-0.85	7.1×10^{-3}	-0.81/Informative
Gru	Gravity of urine (g/l)	0.44	1.5×10^{-5}	0.54/Informative
WBC	Leukocyte in urine (WBC, /ul)	-0.92	0.0	-0.89/Informative
pH	pH of urine	-0.17	0.11	-0.22/NonInformative
Pot	Serum potassium (mmol/l)	0.04	0.66	0.11/Non-Informative
Glu	Blood glucose (mmol/l)	-0.76	9.5×10^{-20}	-0.75/ Informative
PLT	Platelet count (/ul)	-0.17	0.1	-0.19/Non-nformative

3.2 Selecting a Basic Predictor

According to Step 3 of the above algorithm, we select the quadratic loss function (1) as the basic measure of validating the predictor. According to Step 4, we choose a family of classifiers that are able to learn the minimization of this loss function. The elements of this family are listed in Table 1. In fact, I examined much longer list of possible classifiers, but in Table 1 only those providing a reasonable accuracy are listed. Notice that when a classification is to three classes, then a purely random classifier would give 33% accuracy.

The analysis of Table 1 suggests to select the support vector machine (SVM) as the basic classifier for our predictor. Observe that it provides not only the smallest quadratic loss but also the largest accuracy among all the tested algorithms. Its other advantages are also well known. From my point of view it is important that after

learning the SVM it is portable and does not require to have a learning sequence at ones disposal.

It is also worth noticing that the confusion matrix of the SVM predictor (Figure 5, right panel) indicates that the errors committed by it are plus or minus one class. The left panel of this figure shows that the deeply learned neural network with four hidden layers also has this property, due to the proper choice of the loss function. I have observed this kind of functioning of other predictors that we were trained with the quadratic loss.

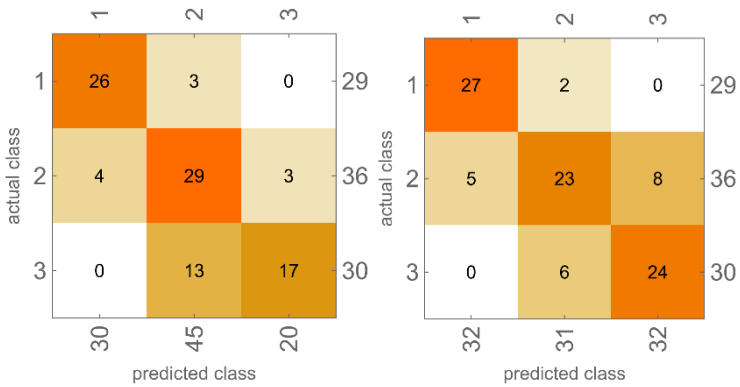


Figure 5

Left panel – the confusion matrix plot when a neural network with four hidden layers is trained on eGFR data. Right panel – the confusion matrix of the SVM classifier trained on the same eGDR data with all the covariates (see also Table 2).

3.3 Detecting Noninformative Features, Their Rejection And Final Validation

According to Step 5, we firstly compute feature impact plots for the SVM predictor when classes I, II (Figure 1) and class III (Figure 2) are predicted. As expected, the analysis of these figures indicate that the initial eGFR level (feature 5) has the most essential impact on the class I, II and III predictions. In the other hand, impacts of other features is not unequivocal. Even feature 10 that has noticeable impact when classes I and II are predicted (see Figure 1) present a low impact on predicting class III (see Figure III).

Therefore, according to the second part of Step 5, prepare feature value impact plots for each feature. Examples of such plots are shown in Figs. 3 and 4. In those cases when these plots form clear patterns along straight lines, they have a medical interpretation. For example, consider the middle plot in the upper row of Fig. 3 that represents the impact of values of the Serum uric acid on the prediction to class II. Points with low values of this acid that are still below the y-axis, being negative, tend to make the class prediction lower, i.e., tending to class I, which is an optimistic

prognosis for these patients, since class I corresponds to CKD stages 1 or 2. Conversely, the points in the same figure that have larger values of serum uric acid and simultaneously positive Shapley values (above the y-axis) show a tendency for the class to transition from II to III, corresponding to worse kidney functioning. Analogous interpretations can be given to other subplots in Fig. 3. In contrast, we are not able to provide clear interpretations to subplots in Fig. 4.

Thus, according to the description in Section 2.3 (just after the Algorithm), we state the hypothesis that features shown in Fig. 4 are noninformative for prediction. In more detail, the following features: pH of urine, Serum potassium, and the Platelet count were first tested for being noninformative. The Spearman rank test was used, and its p-values (0.1, 0.11, 0.66) are shown in Table 2. They are all larger than 0.05. Thus, there are no reasons to reject the hypothesis that these features are uncorrelated with their Shapley values, which means that we consider them to be noninformative. On the contrary, the p-values of the rest of the features in Table 2 are much smaller than 0.05, which means that we reject the hypothesis that they are uncorrelated and we interpret them as informative.

According to the last step of the Algorithm, we remove these three noninformative features from the list, then learn and validate the SVM predictor. As a result, we obtained noticeable improvements, specifically a reduction in the loss function from 21 to 19 (see the confusion matrix in Fig. 6). Furthermore, the global accuracy increased from 80% to 88 %. The remaining quality measures of the final predictor are at a reasonable level, as evident from Table 3 and Figure 7, which presents the ROC curves for all three classes. Notice that these improvements were obtained in one trial only, which was based on the proposed approach to testing feature informativeness. This is in contrast to most feature selectors used, which are based on cumbersome step-by-step removals, checks, and possible returns of particular features.

Table 3

The quality measures of the SVM predictor re-trained after removing three noninformative features (see Table 2)

	Stage I	**Stage II**	**Stage III**
Specificity	0.94	0.80	0.95
Precision	0.87	0.71	0.88
Area ROC	0.96	0.82	0.95

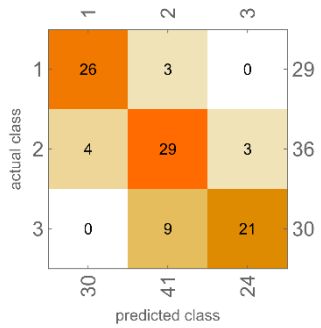


Figure 6
The confusion matrix of the SVM predictor re-trained after removing three noninformative features (see Table 2).

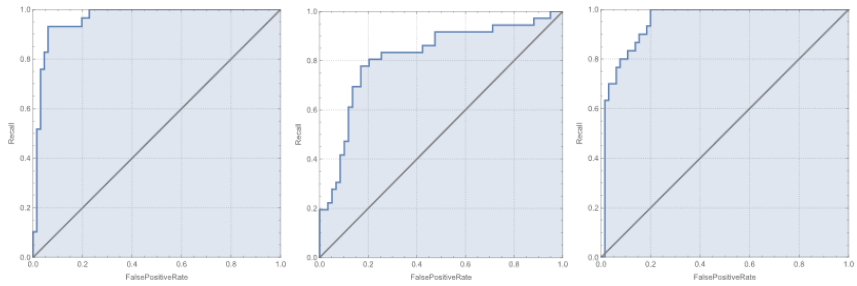


Figure 7

The ROC curves of the trained eGFR classes predictor. Left panel – class I 1 versus all the rest; middle panel – class II versus all the rest. Similarly, the right panel represents class III

4 Conclusions

The algorithm of prediction, based on ML is proposed and tested on a nontrivial case study, for the prediction of staged classes in Chronic Kidney Disease (CKD), in an individualized way, based on the individual's laboratory tests.

In addition to obtaining an algorithm that emphasizes a proper ordering of grouped CKD stages, I equipped it with statistical tests for assessing usefulness (or not useful) for each feature, based on their Shapley values. As a result, I was able to remove the noninformative features, which provided a noticeable improvement in the predictor's performance.

The approach was validated using medical data of a moderate size, which was sufficient for verifying the overall performance as a computer application. A larger dataset would be necessary for drawing truly meaningful medical conclusions.

Acknowledgement

The author expresses a sincere “thank you” to Professor Sławomir Zmonarski, PhD, DSc, from Wrocław Medical University, for allowing the use of anonymized medical data for testing the proposed approach.

References

[1] Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2014). Machine learning applications in cancer prognosis and prediction. Computational and Structural Biotechnology Journal. 2014; 13, 8-17.

[2] Chen, J. H., & Asch, S. M., Machine learning and prediction in medicine—beyond the peak of inflated expectations. The New England journal of medicine, 2017; 376(26), 2507.

[3] Gerse, Á., Dineva, A., & Fleiner, R., Comparative assessment of physical and machine learning models for wind power estimation: A case study for Hungary. Acta Polytechnica Hungarica, 2024; 21(10), 205-222.

[4] Burkart N, Huber MF. A Survey on the Explainability of Supervised Machine Learning. J Artif Int Res. 2021;70:245–317.

[5] Li M, Sun H, Huang Y, Chen H. Shapley value: from cooperative game to explainable artificial intelligence. Autonomous Intelligent Systems. 2024;4(1):2.

[6] Ponce-Bobadilla AV, Schmitt V, Maier CS, Mensing S, Stodtmann S. Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. Clinical and Translational Science. 2024;17(11):e70056.

[7] Tangri N, Stevens LA, Griffith J, Tighiouart H, Djurdjev O, Naimark D, et al. A Predictive Model for Progression of Chronic Kidney Disease to Kidney Failure. JAMA. 2011;305(15):1553-9.

[8] Grams ME, Brunskill NJ, Ballew SH, Sang Y, Coresh J, Matsushita K, et al. The Kidney Failure Risk Equation: Evaluation of Novel Input Variables including eGFR Estimated Using the CKD-EPI 2021 Equation in 59 Cohorts. Journal of the American Society of Nephrology. 2023;34(3):482-94.

[9] Reddy S, Roy S, Choy KW, Sharma S, Dwyer KM, Manapragada C, et al. Predicting chronic kidney disease progression using small pathology datasets

and explainable machine learning models. Computer Methods and Programs in Biomedicine Update. 2024;6:100160.

[10] Miller ZA, Dwyer K. Artificial Intelligence to Predict Chronic Kidney Disease Progression to Kidney Failure: A Narrative Review. Nephrology. 2025;30(1):e14424.

[11] Stompór T, Adamczak M, Kurnatowska I, Naumnik B, Nowicki M, Tylicki L, et al. Pharmacological Nephroprotection in Non-Diabetic Chronic Kidney Disease-Clinical Practice Position Statement of the Polish Society of Nephrology. J Clin Med. 2023;12(16).

[12] Omar ED, Mat H, Abd Karim AZ, Sanaudi R, Ibrahim FH, Omar MA, et al. Comparative Analysis of Logistic Regression, Gradient Boosted Trees, SVM, and Random Forest Algorithms for Prediction of Acute Kidney Injury Requiring Dialysis After Cardiac Surgery. Int J Nephrol Renovasc Dis. 2024;17:197-204..

[13] Qi X, Wang S, Fang C, Jia J, Lin L, Yuan T. Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with dietary antioxidants. Redox Biology. 2025;79:103470.

[14] Rafajłowicz, E. Data structures for pattern and image recognition and application to quality control. Acta Polytechnica Hungarica, 2018. 15(4), 233-262.

[15] Rafajłowicz, E., & Krzyżak, A. Pattern recognition with ordered labels. Nonlinear Analysis: Theory, Methods & Applications, 71(12), 2009, e1437-e1441