



HAL
open science

Using Machine Learning to Support Prostate Cancer Detection Without a Dedicated Biomarker

Hadrien Salem, Sarah Ben Othman, Marc Broucqsault, Slim Hammadi

► To cite this version:

Hadrien Salem, Sarah Ben Othman, Marc Broucqsault, Slim Hammadi. Using Machine Learning to Support Prostate Cancer Detection Without a Dedicated Biomarker. 2024 IEEE 20th International Conference on Automation Science and Engineering, Aug 2024, Bari, Italy. <hal-04729608>

HAL Id: hal-04729608

<https://hal.science/hal-04729608v1>

Submitted on 10 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Using Machine Learning to Support Prostate Cancer Detection Without a Dedicated Biomarker

Hadrien Salem¹, Sarah Ben Othman¹, Marc Broucqsault² and Slim Hammadi¹

Abstract—Early disease detection through the analysis of laboratory tests is critical for the successful treatment of certain diseases. However, current medical knowledge relies on the use of certain biomarkers in order to diagnose associated diseases. This means that some diseases cannot be detected unless the corresponding biomarker is actually measured.

In this paper, we will explore the possibility of using machine learning to detect abnormal values in biomarkers, even when they are not directly measured. This stems from the fact that, as an ensemble of features, even a set of common biomarkers contains sufficient information to suspect anomalies in a patient’s analyses.

We present a complete pipeline for processing medical laboratory data, taking into account the many challenges offered by this use case, such as data cleanliness, the abundance of missing values, and strong class imbalance. Using this framework, we perform a case study on prostate cancer and its associated biomarker, Prostate-Specific Antigen (PSA). Our implementation, based on an imbalance-aware ensemble prediction model, is tested on a real laboratory dataset of more than 8 million records. We find that basing the ensemble on Random Forest models provides the best results, with a ROC-AUC of 75.86%, indicating that it is indeed possible to detect abnormal PSA values without measuring it directly. We also show that a simple baseline of 24 commonly-measured biomarkers provides near-equivalent results to a full set of features. Our work offers new perspectives for routine disease screening.

I. INTRODUCTION

The diagnosis and prediction of diseases is one of the major use cases of big data analytics in healthcare [2]. By analyzing several types of medical data, researchers have been able to create prediction and risk stratification models for a wide array of conditions, including diabetes, Alzheimer’s disease, chronic kidney disease, tuberculosis, and various cancers.

Laboratory analyses, such as blood and urine tests, are one of the main sources of data for the study of diseases. While predictive models with good performance can be found in the literature, it can be very challenging for actors of the healthcare sector to replicate these results in practice. This is in part due to the difficulty to access data: in order to associate a patient’s laboratory analysis data with a potential disease, clinical data is required (i.e. a diagnosis

established by a physician). In practice, this association is not trivial because - with the exception of hospital stays - laboratory analyses and diagnoses are not generated by the same entities, and are hence not linked.

In particular, medical analyses laboratories have a lot of available data but are not necessarily able to leverage it for lack of associated clinical data, and because there is great variance in the number and types of analyses taken for each individual patient. This means that, even if a disease has a particular biomarker allowing its detection, it is not necessarily measured for all patients, making data labeling particularly difficult. As a solution to this issue, **the present paper explores the possibility to help with disease screening through the use of routine laboratory analyses alone.** This would allow for a form of “passive” screening: any patient getting routine laboratory results would also be able to receive recommendations on complementary analyses to get for a finer screening.

In this paper, we propose to implement a machine learning pipeline for the study of prostate cancer. Prostate cancer is one of the most frequently diagnosed cancers: it is the second most frequent cancer in men, and the fifth leading cause of death worldwide [22], [10]. The screening of prostate cancer is usually done with the Prostate Specific Antigen (PSA) blood test. However, PSA is unevenly prescribed [24]. Therefore, there is a real use for the recommendation of PSA prescription based on more commonly-measured laboratory tests. To put it simply, **our goal is the detection of elevated PSA values - which is a risk factor for prostate cancer - without using this particular marker.** We have observed that none of the existing works in the current state of the art adopt this approach.

This use case poses many challenges, in particular related to the quality of data, its heterogeneity, and the high proportion of missing values. Our contribution is as follows:

- 1) We propose a data processing framework for the independent valorization of raw medical laboratory data, taking into account the various challenges posed by this use case.
- 2) We apply this method to the screening of patients with elevated PSA. By using only common laboratory tests (which do not include PSA), we show that it is possible to suspect abnormal values in the dedicated marker for prostate cancer screening, hence providing support in the early detection of the disease. Our work establishes a baseline for this approach.
- 3) We compare the performance achieved with different strategies for feature selection, missing data and

*This work was not supported by any organization

¹Hadrien Salem is with CRISTAL CNRS UMR 9189, Villeneuve d’Ascq, Nord, France hadrien.salem@centralelille.fr

¹Sarah Ben Othman is with CRISTAL CNRS UMR 9189, Villeneuve d’Ascq, Nord, France sara.ben-othman@centralelille.fr

²Marc Broucqsault is Altao, Lille, Nord, France mbroucqsault@altao.com

¹Slim Hammadi is with CRISTAL CNRS UMR 9189, Villeneuve d’Ascq, Nord, France slim.hammadi@centralelille.fr

machine learning, using a real-life laboratory dataset containing over 8 million biology records. We show that reducing the feature set to 24 commonly-used biomarkers hardly impacts performance, for a significant gain in practicality and reduced training cost.

II. RELATED WORK

Disease prediction and detection has received a lot of attention in the literature, especially in recent years. The type of data and the techniques used to analyze it depend on the disease that is studied. Diabetes, for example, has been extensively researched, both for risk prognosis [28] and blood glucose prediction [29]. Other diseases, such as Alzheimer’s disease [1], chronic kidney disease [30], or colorectal cancer [13] have been studied in a similar manner with various types of data: not only laboratory tests, but also imaging, and even environmental and lifestyle data [5].

In addition to “classic” machine learning approaches, such as tree-based methods, support vector machines, or artificial neural networks, other methods such as survival analysis [20], [21] and knowledge-based approaches [25], [18] have been used to establish prediction or risk-stratification models, as well as discovering semantic relationships in medical data.

In particular, prostate cancer prediction from laboratory tests has been relatively unexplored in the literature. While there have been numerous studies on the research of new biomarkers to help with its detection, such as [8] or [17], machine learning works mostly seem to focus on imaging [15]. A notable exception is Wang et al. [27], who build a prostate cancer risk prediction model based on a set of 14 analyses, including common laboratory tests, demographic information, and prostate-specific indicators such as Prostate-Specific Antigen (PSA). They obtain their best results with a random forest algorithm, and use this model for a *post-hoc* feature importance study. While prostate-related biomarkers are by far the most influential in the classification, they empirically show the correlation of prostate cancer with biomarkers that are not commonly used for its screening, such as inorganic phosphorus, triglycerides and calcium. This result shows that with the use of machine learning, the combined analysis of common biomarkers could lead to the detection of prostate cancer.

Instead of focusing on a specific condition, Razavian et al. [23] propose a prediction model for 133 diseases based on a set of 18 common laboratory tests. Using ensemble algorithms based on convolutional neural networks (CNN) and long short-term memory networks (LSTM), they manage to predict the onset of these conditions up to 15 months prior. While prostate cancer is one of the diseases included in their model, it is not extensively studied, and the set of analyses used is more adapted for renal diseases - for which they obtain the best performance. For example, no hematology results are used, despite being quite common in laboratory analyses. The present paper’s ambition is different in that, instead of proposing a one-fits-all model for multi-class prediction, we opt for a flexible framework for the study of a particular disease.

III. METHODS

In this section, we will start by defining precisely our problem and assumptions. We will then give a general overview of our proposed machine learning pipeline, before presenting each step in more detail.

A. Problem definition

The use case treated in this paper corresponds to that of a medical laboratory whose only available data is the results of the analyses it offers, as well as the patients’ age. In section IV, we will apply the proposed procedure to such a dataset containing more than 2 billion analyses results, originating from a French medical laboratory. The experimental dataset will be described in more detail in Section IV-A.

The objective of the proposed procedure is to detect abnormal values in a certain analysis result (which we will indifferently call “biomarker”), without using this marker in the prediction. The underlying assumption is that there exists complex correlations between biomarkers, which are poorly understood in current medical knowledge.

The relevance of this approach stems from the fact that patients are not all prescribed the same analyses. Therefore, the ultimate goal of this research is to be able to detect potential abnormalities in a patient’s biomarkers, even when these biomarkers are not directly measured during analysis.

The problem is thus formalized as a binary classification problem, where an abnormal value in a given marker is the “positive” class. The features (before proceeding to feature selection) are all of the available analyses, except for the biomarker that was used to label data.

In the rest of this paper, we will apply this generic formulation to the study of prostate cancer. We will use the Prostate-Specific Antigen (PSA) as the labeling biomarker, and consider its value abnormal if it is equal to or greater than 4 ng/mL [11].

B. Method overview

A flowchart of the proposed procedure is shown in Figure 1. Data is processed sequentially in four phases : **data extraction**, where data is filtered and formatted, **feature engineering**, where features are selected and / or introduced, **pre-processing**, which mainly entails the management of missing data, and **machine learning**, where classification models are trained and optimized for the detection of abnormal values in the selected biomarker. We note that in the proposed method, class imbalance is managed directly by the learning model.

This process is flexible and can be applied to any disease or condition falling under the problem definition described in III-A. In the rest of this section, we will detail each of these phases and explain the choices we made for the study of prostate cancer.

C. Data extraction

The first step in this data extraction phase is the choice of one (or potentially several) target biomarker(s). In our

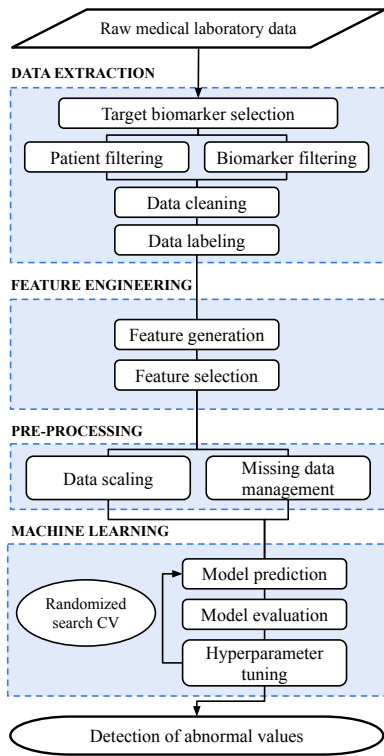


Fig. 1. Flowchart of the proposed machine learning pipeline for the detection of abnormal values in a biomarker. CV stands for “cross-validation”.

case study for prostate cancer, this marker is the Prostate-Specific Antigen (PSA). This biomarker will only be used to label data: abnormal values are marked as “positive” cases, whereas normal values are marked as “negative” cases. In the case of PSA, any value equal to or above 4ng/mL is considered abnormal [11]. Besides, we filter out female patients from the data, since prostate cancer can only affect male patients.

For this study, we use a set 72 biomarkers representative of most of the analyses medical laboratories are able to offer, selected with the help of a physician. We highlight that this list is independent of the disease that is considered.

D. Feature Engineering

The abundance of missing data is one of the main challenges in dealing with medical data - especially so with laboratory data, because different patients are prescribed different analyses. However, some biomarkers are very commonly measured compared to others, and form a “baseline” of analyses upon which other biomarkers can be added in specific cases. For example, PSA is not systematically prescribed, whereas lipid profiles (cholesterol, triglycerides, etc.) are much more common. Targeting these baseline biomarkers is fundamental to our study, since we aim to detect abnormalities in common laboratory data. As a consequence, we select all features that are present for a significant amount of patients. In Section IV-B, we will for example keep analyses realized by at least 25% of the patients in the dataset.

E. Pre-processing

We realize a 75%-25% stratified train-test split, meaning that we ensure there is roughly the same proportion of each class in both groups. All features are simply standardized by removing the mean and scaling to unit variance. To avoid data leak, the mean and standard deviation are obviously computed from the training set.

Missing data also needs to be managed at this step. As explained in Section III-D, the selection of common biomarkers as our features reduces the amount of missing data considerably. We will be considering three strategies for the management of the remaining missing values in the dataset: **the removal of rows with missing values**, **mean imputation**, and **iterative imputation** [26], which consists in iteratively estimating each feature using a regression on other features.

F. Machine Learning

1) *Imbalance-aware machine learning model*: In health-care, and more specifically when studying diseases, it is common for classification problems to be imbalanced - meaning that one class (usually the negative class) is much more represented than the other(s). In our example, less than 15% of patients have an abnormal PSA value (above 4ng/mL). If machine learning models are trained naively with such a difference in class proportions, they will tend to be biased toward the majority class. In particular, there would be a lot of false negatives in the prediction, which is not desirable in the context of disease screening.

Several strategies have been employed in the literature to deal with class imbalance, such as over-sampling the minority class, under-sampling the majority class, or Synthetic Minority Over-sampling Technique (SMOTE) [6] and its variants.

In this paper, we will use a method based on Liu et al.’s EasyEnsemble algorithm [16]. We train a number N of models on balanced datasets, and use these models as an ensemble to perform predictions through a majority vote. Each of the balanced datasets is formed by concatenation of all minority class samples, and an equal number of samples from the majority class selected at random. For clarity, a schematic representation of this model is shown in Figure 2. Training models on balanced datasets prevents the pitfalls mentioned above, and using them as an ensemble allows for minimal loss of information despite the undersampling of the majority class. For these reasons, we decide to use this method in our proposed procedure.

The nature of the individual machine learning models in the ensemble can be changed to optimize the results. For our implementation, we have trained and compared three models within the ensemble algorithm presented in the previous section: **Random Forests** [4], **AdaBoost** [9] (short for Adaptive Boosting) and **XGBoost** [7] (or Extreme Gradient Boosting). The results obtained with each of these algorithms will be presented in Section IV.

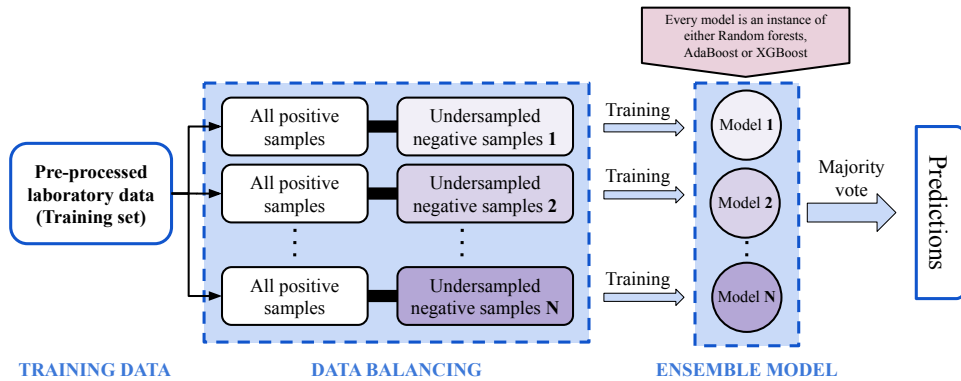


Fig. 2. Diagram of the imbalance-aware ensemble prediction model

2) *Model evaluation*: When it comes to evaluating a classification model for an imbalanced dataset, accuracy (defined as the proportion of correctly classified examples) can be misleading. If the proportion of samples in the majority class is significantly higher than in the minority class, then good accuracy can be achieved by only predicting the majority class. This must be avoided in real-life applications, especially when it comes to disease screening. For this reason, we will also consider other performance metrics that take into account the quality of prediction on both the majority and minority class:

- **Sensitivity** (also called recall or true positive rate) measures the proportion of actual positive samples that are correctly identified by a classifier. A high sensitivity is desirable in our context of disease suspicion.
- **Specificity** is the counterpart of sensitivity for negative samples. It measures the proportion of actual negative samples correctly identified by a classifier.
- **Area under ROC curve (AUC)**. The Receiver Operating Characteristic curve plots the true positive rate against the false positive rate at different classification thresholds. A higher area under this curve indicates better performance.
- **F1-score** is the harmonic mean of sensitivity and precision, where precision is the proportion positive predictions that are actually correct. As such, the F1-score is a balanced measure of the classifier's performance, making it a popular metric to study imbalanced datasets.

3) *Hyperparameter tuning*: Hyperparameter tuning is performed by Randomized search cross-validation [3]. This technique consists in randomly selecting hyperparameters within a given range and evaluating their performance with cross-validation.

IV. RESULTS

In this section, we will apply the proposed procedure to a real dataset with the goal of detecting patients with high PSA, thus providing support in the screening of prostate cancer.

A. Dataset description

The experiments were realized on a dataset containing 8 million anonymized biology records (one record being one

set of measurements for one person) from a French medical laboratory, gathered between 2018 and 2022, for a total of 2 billion individual analyses results. After data extraction, we are left with about 150,000 records containing PSA analyses that can be used for labeling, for a total of nearly 78,000 patients. Once data is labeled (positive cases being patients with a PSA superior or equal to 4ng/mL), the PSA value is no longer used in the classification task.

B. Selected features

As explained in Section III-D, we select all features that are present for more than 25% of patients. This threshold was set empirically to keep a sufficient number of features without significantly dropping in prediction performance. The list of 24 analyses kept with this process is shown in Table I, along with the corresponding LOINC codes. The percentage of missing data goes down from 62% to 15%, with 56% of the rows having no missing data at all (instead of 0% when no feature selection is performed). These statistics highlight the existence and importance of the biomarker baseline mentioned in Section III-D.

C. Comparison of prediction results for the models

In accordance to what was exposed throughout Section III, we train a number of models to perform the binary classification task of detection of abnormal PSA values. The results of these experiments are shown in Table II. We realize two sets of tests: one with the complete set of 72 analyses filtered during data extraction ("Without feature selection"), and one with the 24 selected features listed in table I ("With feature selection"). The age of the patients is also included.

For both of these data sources, several strategies are considered for missing data management: removal of missing values ("Removal"), mean imputation and iterative imputation. Likewise, we test three different estimators within the ensemble prediction model: Random forests, AdaBoost and XGBoost. We use the performance metrics presented in Section III-F.2: sensitivity, specificity, F1-score, and area under ROC curve (noted ROC AUC), as well as accuracy. In order to reduce the influence of randomness in our results, we realize each test five times and plot the averaged results.

TABLE I

LIST OF ANALYSES KEPT IN THE STUDY AFTER FEATURE SELECTION

Biomarker	LOINC
Creatinine	2160-0
Estimated Glomerular Filtration Rate	62238-1
Mean Corpuscular Hemoglobin Concentration	28540-3
Red blood cell distribution width	30385-9
Platelets in plasma	26516-5
Leukocytes	26464-8
Hemoglobin	718-7
Neutrophils	30448-5
Lymphocytes	26474-7
Monocytes	26484-6
Hematocrit	20570-8
Erythrocytes	26453-1
Mean Corpuscular Hemoglobin	28539-5
Mean Corpuscular Volume	30428-7
Cholesterol	2093-3
Triglyceride	2571-8
Cholesterol in HDL	2085-9
Cholesterol non HDL	43396-1
Glucose	2345-7
Potassium	2823-3
Alanine aminotransferase	1742-6
Aspartate aminotransferase	1920-8
Sodium	2951-2
Chloride	2075-0

We note that in Table II, removal of missing values is not considered without feature selection. This is due to the fact that, when removing rows with missing values in the dataset without feature selection, no data remains: every row has at least one missing value.

Overall, performance with and without feature selection appears to be very close. The best results for accuracy, specificity, F1-score and ROC AUC are all obtained without feature selection, using mean imputation with the random forests estimator. Nevertheless, the counterpart of this model with feature selection achieves very similar results for all these metrics, and has better sensitivity - which, is arguably the most important in the context of disease screening. Further, using a lower amount of features makes this second model much faster during training.

Regarding missing values handling strategies, iterative imputation performs slightly worse than mean imputation, while removing all missing values seems to impact performance negatively (and is not exploitable at all without feature selection).

As for estimators, results obtained with AdaBoost and XGBoost are mostly equivalent, and Random Forests perform nearly systematically better than both of them. However, XGBoost is significantly faster than the other two estimators, which may be preferable in some use cases.

D. Discussion

The results of our study show the feasibility of using common biomarkers to predict abnormal values in a patient's laboratory results, in particular Prostate-Specific Antigen (PSA) - which is to date the most used analysis for prostate cancer screening. Our findings also reveal that the use of a large number of biomarkers does not significantly improve the results over the use of a select number of common analy-

ses. This outcome suggests that the baseline of 24 analyses is sufficient for the classification task. In other words, disease screening can be performed almost systematically when laboratory tests are taken. Besides, this feature selection improves computational time and generalizability. Finally, the relative balance between sensitivity and specificity in our results demonstrates the suitability of the prediction model to manage the imbalance in the data.

In order to further improve performance and reduce the relatively high number of false positives, an interesting avenue of future research would be to study the temporal evolution of biomarkers. This would add another dimension of information to the data.

V. CONCLUSION

In this paper, we have presented a flexible data processing and machine learning pipeline for the study of diseases through laboratory analyses. We have then applied it to the case of prostate cancer and its dedicated marker, the Prostate-Specific Antigen (PSA). We have shown that a baseline of common biomarkers carries enough information for machine learning algorithms to detect abnormalities in laboratory results. These findings open up many perspectives for disease screening: the data collected by medical laboratories, which is widely underused for lack of associated diagnoses, can be independently leveraged for the predictive study of diseases. Ultimately, this approach could lead to a systematic screening whenever patients get laboratory analyses, even when disease-specific biomarkers are not measured. Determining which diseases are most efficiently predicted in this context is a possible avenue of future research.

Future research should also aim to improve the classification performance by adding a temporal dimension to the data, providing additional information for the machine learning algorithms and allowing for longer-term disease screening.

REFERENCES

- [1] Tamer Abuhmed, Shaker El-Sappagh, and Jose M Alonso. 2021. Robust hybrid deep learning models for Alzheimer's progression detection. *Knowledge-Based Systems* 213 (2021), 106688.
- [2] Kornelia Batko and Andrzej Ślezak. 2022. The use of Big Data Analytics in healthcare. *Journal of big Data* 9, 1 (2022), 3.
- [3] James Bergstra and Yoshua Bengio. 2012. Random search for hyperparameter optimization. *Journal of machine learning research* 13, 2 (2012).
- [4] Leo Breiman. 2001. Random forests. *Machine learning* 45 (2001), 5–32.
- [5] Victor Chang, Meghana Ashok Ganatra, Karl Hall, Lewis Golightly, and Qian- wen Ariel Xu. 2022. An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators. *Healthcare Analytics* 2 (2022), 100118.
- [6] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16 (2002), 321–357.
- [7] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 785–794.
- [8] Liuxi Chu, Xin Shu, Yao Huang, Tong Chu, Meina Ge, and Qin Lu. 2022. Sex steroid hormones in urinary exosomes as biomarkers for the prediction of prostate cancer. *Clinica Chimica Acta* 531 (2022), 389–398.
- [9] Yoav Freund and Robert E Schapire. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences* 55, 1 (1997), 119–139.

TABLE II

MEAN RESULTS FOR 5 PERFORMANCE INDICATORS (ACCURACY, SENSITIVITY, SPECIFICITY, F1-SCORE AND ROC AUC) COMPUTED OVER 5 DIFFERENT TRAIN-TEST SPLITS. THE BEST RESULTS FOR A GIVEN PERFORMANCE METRIC ARE WRITTEN IN **BOLD**.

Data source	Missing Values	Estimator	Accuracy	Sensitivity	Specificity	F1-Score	ROC AUC
Without feature selection	Mean Imputation	Random Forests	0.6835	0.7032	0.6800	0.3970	0.7586
		AdaBoost	0.6715	0.6847	0.6691	0.3818	0.7438
		XGBoost	0.6657	0.7117	0.6577	0.3868	0.7546
	Iterative Imputation	Random Forests	0.6778	0.6981	0.6743	0.3910	0.7497
		AdaBoost	0.6597	0.6888	0.6547	0.3749	0.7396
		XGBoost	0.6675	0.6993	0.6619	0.3839	0.7495
With feature selection	Removal	Random Forests	0.6061	0.7424	0.5890	0.2955	0.7258
		AdaBoost	0.5956	0.7311	0.5787	0.2870	0.7125
		XGBoost	0.6092	0.7170	0.5958	0.2900	0.7147
	Mean Imputation	Random Forests	0.6724	0.7189	0.6643	0.3940	0.7578
		AdaBoost	0.6677	0.6824	0.6651	0.3783	0.7403
		XGBoost	0.6617	0.7095	0.6534	0.3832	0.7507
	Iterative Imputation	Random Forests	0.6679	0.7145	0.6597	0.3893	0.7531
		AdaBoost	0.6582	0.6903	0.6526	0.3745	0.7370
		XGBoost	0.6579	0.7078	0.6492	0.3800	0.7455

- [10] David C Grossman, Susan J Curry, Douglas K Owens, Kirsten Bibbins-Domingo, Aaron B Caughey, Karina W Davidson, Chyke A Doubeni, Mark Ebell, John W Epling, Alex R Kemper, et al. 2018. Screening for prostate cancer: US Preventive Services Task Force recommendation statement. *Jama* 319, 18 (2018), 1901–1913.
- [11] Ove Gustafsson, Essam Mansour, Ulf Norming, Adelaida Carlsson, Magnus Törnblom, and Claes R Nyman. 1998. Prostate-specific antigen (PSA), PSA density and age-adjusted PSA reference values in screening for prostate cancer: A study of a randomly selected population of 2,400 men. *Scandinavian journal of urology and nephrology* 32, 6 (1998), 373–377.
- [12] Alistair Johnson, Lucas Bulgarelli, Tom Pollard, Steven Horng, Leo Anthony Celi, and Roger Mark. 2020. Mimic-iv. PhysioNet. Available online at: <https://physionet.org/content/mimiciv/1.0/> (accessed August 23, 2021) (2020).
- [13] Yaron Kinar, Nir Kalkstein, Pinchas Akiva, Bernard Levin, Elizabeth E Half, Inbal Goldshtein, Gabriel Chodick, and Varda Shalev. 2016. Development and validation of a predictive model for detection of colorectal cancer in primary care by analysis of complete blood counts: a binational retrospective study. *Journal of the American Medical Informatics Association* 23, 5 (2016), 879–890.
- [14] Andrew S Levey, Lesley A Stevens, Christopher H Schmid, Yaping Zhang, Alejandro F Castro III, Harold I Feldman, John W Kusek, Paul Eggers, Frederick Van Lente, Tom Greene, et al. 2009. A new equation to estimate glomerular filtration rate. *Annals of internal medicine* 150, 9 (2009), 604–612.
- [15] Geert Litjens, Robert Toth, Wendy Van De Ven, Caroline Hoeks, Sjoerd Kerkstra, Bram van Ginneken, Graham Vincent, Gwenael Guillard, Neil Birbeck, Jindang Zhang, et al. 2014. Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge. *Medical image analysis* 18, 2 (2014), 359–373.
- [16] Tian-Yu Liu. 2009. Easysensemble and feature selection for imbalance data sets. In 2009 international joint conference on bioinformatics, systems biology and intelligent computing. IEEE, 517–520.
- [17] Shweta J. Malode and Nagaraj P. Shetti. 2022. New generation biomarkers for the detection of prostate cancer. *Biosensors and Bioelectronics: X* 12 (2022), 100250. <https://doi.org/10.1016/j.biosx.2022.100250>
- [18] Tobias Mayer, Elena Cabrio, and Serena Villata. 2020. Transformer-based argument mining for healthcare applications. In ECAI 2020. IOS Press, 2108–2115.
- [19] Clement J McDonald, Stanley M Huff, Jeffrey G Suico, Gilbert Hill, Dennis Leavelle, Raymond Aller, Arden Forrey, Kathy Mercer, Georges DeMoor, John Hook, et al. 2003. LOINC, a universal standard for identifying laboratory observations: a 5-year update. *Clinical chemistry* 49, 4 (2003), 624–633.
- [20] Ghazal Mirabnazar, Da Ma, Cédric Beaulac, Sieun Lee, Kartek Popuri, Hyun-woo Lee, Jiguo Cao, James E Galvin, Lei Wang, Mirza Faisal Beg, et al. 2023. Predicting time-to-conversion for dementia of Alzheimer’s type using multi-modal deep survival analysis. *Neurobiology of Aging* 121 (2023), 139–156.
- [21] Arturo Moncada-Torres, Marissa C van Maaren, Mathijs P Hendriks, Sabine Siesling, and Gijs Geleijnse. 2021. Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. *Scientific reports* 11, 1 (2021), 6968.
- [22] Prashanth Rawla. 2019. Epidemiology of prostate cancer. *World journal of oncology* 10, 2 (2019), 63.
- [23] Narges Razavian, Jake Marcus, and David Sontag. 2016. Multi-task prediction of disease onsets from longitudinal laboratory tests. In Machine learning for healthcare conference. PMLR, 73–100.
- [24] Martina Taborelli, Federica Toffolutti, Ettore Bidoli, Luigino Dal Maso, Stefania Del Zotto, Elena Clagnan, Michele Gobatto, Diego Serraino, and Silvia Franceschi. 2022. The use of PSA testing over more than 20-years: A population-based study in North-Eastern Italy. *Tumori Journal* (2022), 03008916221128343.
- [25] Kateryna Tymoshenko, Swapna Somasundaran, Vinodkumar Prabhakaran, and Vinay Shet. 2012. Relation mining in the biomedical domain using entity-level semantics. In ECAI 2012. IOS Press, 780–785.
- [26] Stef Van Buuren and Karin Groothuis-Oudshoorn. 2011. mice: Multivariate imputation by chained equations in R. *Journal of statistical software* 45 (2011), 1–67.
- [27] Chun Wang, Qinxue Chang, Xiaomeng Wang, Keyun Wang, He Wang, Zhuang Cui, and Changping Li. 2022. Prostate Cancer Risk Prediction and Online Calculation Based on Machine Learning Algorithm. *Chinese Medical Sciences Journal* 37, 3 (2022), 210–218.
- [28] Zidian Xie, Olga Nikolayeva, Jiebo Luo, and Dongmei Li. 2019. Building risk prediction models for type 2 diabetes using machine learning techniques. *Preventing chronic disease* 16 (2019).
- [29] Tao Yang, Ruijun Wu, Rui Tao, Shuang Wen, Ning Ma, Yuhang Zhao, Xia Yu, and Hongru Li. 2020. Multi-Scale Long Short-Term Memory Network with Multi-Lag Structure for Blood Glucose Prediction. *KDH@ ECAI* 45 (2020), 136–140.
- [30] Wen Ye, Xuemei Ding, Nathaniel Putnam, Ryan Farej, Rakesh Singh, Di Wang, Shihchen Kuo, Sheldon X Kong, Jay C Elliott, Jason Lott, et al. 2022. Development of clinical prediction models for renal and cardiovascular outcomes and mortality in patients with type 2 diabetes and chronic kidney disease using time-varying predictors. *Journal of Diabetes and its Complications* 36, 5 (2022), 108180.