SPECIAL SECTION

Enhancement of COVID-19 symptom-based screening with quality-based classifier optimisation

Michał KOZIELSKI [1]*, Joanna HENZEL [1], Joanna TOBIASZ [2], Aleksandra GRUCA[1],

Paweł FOSZNER [3], Joanna ZYLA [2], Małgorzata BACH[4], Aleksandra WERNER [4],

Jerzy JAROSZEWICZ[5], Joanna POLAŃSKA [2], and Marek SIKORA [1]

[1] Department of Computer Networks and Systems, Silesian University of Technology, Gliwice, Poland
[2] Department of Data Science and Engineering, Silesian University of Technology, Gliwice, Poland
[3] Department of Graphics, Computer Vision and Digital Systems, Silesian University of Technology, Gliwice, Poland
[4] Department of Applied Informatics, Silesian University of Technology, Gliwice, Poland
[5] Department of Infectious Diseases and Hepatology, Medical University of Silesia, Katowice, Poland

Abstract. Efforts of the scientific community led to the development of multiple screening approaches for COVID-19 that rely on machine learning methods. However, there is a lack of works showing how to tune the classification models used for such a task and what the tuning effect is in terms of various classification quality measures. Understanding the impact of classifier tuning on the results obtained will allow the users to apply the provided tools consciously. Therefore, using a given screening test they will be able to choose the threshold value characterising the classifier that gives, for example, an acceptable balance between sensitivity and specificity. The presented work introduces the optimisation approach and the resulting classifiers obtained for various quality threshold assumptions. As a result of the research, an online service was created that makes the obtained models available and enables the verification of various solutions for different threshold values on new data.

Key words: classification; classification quality; screening; COVID-19.

1. INTRODUCTION

The screening tests [1] are an important tool allowing for the identification of people at risk of the disease being the subject of the analysis. Screening tests are widely used to reduce morbidity and mortality, especially for diseases such as cancer [2–4]. In the case of cancer, early diagnosis is important as it allows for an effective fight against the disease. The role of screening tests for viral diseases is to limit the spread of the pathogen by people who may be sick. This is exactly the task facing COVID-19 screening tests [5], especially when the number of infected is rapidly increasing and the health system is overloaded.

The very vigorous response of the scientific community to the SARS-CoV-2 virus pandemic resulted in numerous examples showing how the technology can help in the fight against the pandemic. These examples include the use of artificial intelligence methods [6], which can e.g., support decision-making in diagnostics in hospitals. Besides, numerous online applications were created, allowing people concerned about their health to pre-determine whether they are at risk of COVID-19 disease.

The main motivation for the conducted research was the willingness to provide technological solutions that may be useful in dealing with the progressing epidemic. Although the work [7] concludes that diagnosis based on symptoms may not be effective when distinguishing between COVID-19 and the other diseases caused by respiratory viruses, such tests can meet the requirements defined for screening tests [8] as shown in [9]. Moreover, these types of screening tests are most readily available. This is important because, as experience shows, carrying out a large number of, e.g., PCR tests in some countries can be a challenge. However, there is no work showing how the classification models used for screening tests can be tuned and what the impact of such optimisation on the quality of the classification is. The lack of such research was another motivation for the presented study.

To the best of the authors' knowledge, there is no available online tool enabling physicians or any other researchers to perform screening tests for COVID-19 based on symptoms of patients. The self-diagnosis tool presented in [9] does not pro-

M. Kozielski, J. Henzel, J. Tobiasz, A. Gruca, P. Foszner, J. Zyła, M. Bach, A. Werner, J. Jaroszewicz, J. Polańska, and M. Sikora

vide insight into the quality-based optimisation of classifiers and does not allow the user to analyse multiple records. Whereas, such a tool providing various classifiers generated in properly documented research study may be of interest to those who are looking for methods offering extended screening results.

The study aimed to assess the impact of the model optimisation on the results when it is performed towards the selected classification quality measure. The evaluation of screening tests should be performed based on a set of measures that includes: predictive values (positive – PPV and negative – NPV) and sensitivity and specificity [1, 8]. For this reason, it was important to create models optimised for these measures. In addition, the goal was to create an online application that would allow a symptom-based screening test for COVID-19 to be performed. In this way the developed methods can be made available for the wider community and analysis of further data sets is possible.

Contribution of this study consists of the analysis of the impact on the classification quality measures commonly used for screening test evaluation when a classifier is optimised for various combinations of these measures. Based on the classification quality changes and their influence on screening tests' characteristics, an expert can consciously choose the parameter thresholds, adjusting the methods to the analysis goal. cMoreover, contribution includes the presentation of the full methodology of optimisation steps enabling the creation of models for two classification algorithms: Logistic Regression and XGBoost.

The performed analysis resulted in an online service[1] available for physicians and researchers interested in symptom-based screening for COVID-19 enriched with the results for multiple models generated for various optimisation schemes.

The structure of this paper is as follows. Works related to the topic discussed are described in Section 2. Section 3 presents the classifiers' optimisation process for the selected classification quality measures. Section 4 presents the results of the experiments performed. Section 5 outlines the developed web application. Section 6 concludes the paper.

2. RELATED WORK

Applications of machine learning (ML) methods to COVID-19 diagnostics and screening were reviewed in several studies [6, 10–12]. In these works, the available solutions were analyzed in terms of the ML methods used [10, 12] and in terms of the areas of application such as screening, contact tracing, forecasting and drugs, and vaccination [11]. The most extensive work [6] reviews 145 studies and classifies the approaches presented there into diagnostic models, prognostic models and models to predict risks of COVID-19 in population. Additionally, this review reports the data sets used in the analysis and the risks of bias for the approaches presented.

Focusing on the machine learning-based approaches for COVID-19 screening tests, it can be noticed how different char-

acteristics and data representations are taken into account in such studies. Some of the approaches require a medical background and infrastructure because they are based on blood tests [13] or X-ray image analysis [14–16]. Another reported approach requires a specialised sensor – electronic nose to collect data for further analysis by means of machine learning methods [17]. Besides, there are approaches to COVID-19 diagnosis and screening that are based solely on sound recordings that can be collected via web page [18] or telephone [19]. Finally, there is an approach presenting symptom-based classification [9] supporting screening for COVID-19. This solution, developed as an online service, is based on the information collected in the form of a questionnaire upon admission to the hospital. The screening approaches listed above used various data representations and various classification methods. For image and voice analysis, Convolutional Neural Network (CNN) was used. Moreover, the Artificial Neural Network was used as a classifier identifying COVID-19 by means of electronic nose and in case of blood analysis, where the Random Forest and Logistic Regression classifiers were additionally used. In the case of disease symptoms analysis, the Logistic Regression and XGBoost methods were used.

Some of the proposed approaches to COVID-19 diagnostics using machine learning resulted in online tools. Examples of such works are [20, 21], in which diagnostics with the use of various classifiers was performed based on blood indices. The above works are related to the tools [22, 23] enabling the use of the created classifiers. There are much more tools that are user-friendly and possibly easier accessible to a wider group of users [24–27]. These tools support self-diagnosis based on recognised symptoms. However, they are not accompanied by a description of how the approach was developed and evaluated. In addition to the above-mentioned tools, COVID-19 risk calculator [28] should be distinguished since this tool, like the presented solution, was created on the basis of cases positively diagnosed with COVID-19 in Poland. However, COVID-19 risk calculator predicts severity of COVID-19 for an individual and was not designed for screening purposes.

The presented review of the related works shows various approaches to screening for COVID-19 using machine learning methods. They take into account different representations of patient information and use different classification methods. The discussed works show the quality of the proposed approaches, and in the case of online tools, it is possible to verify what the prediction result will be for a single person. However, there is a lack of studies that show how the optimisation of the selected quality measure affects the overall results. There is also a lack of tools that allow for the use of such various models and their verification on new data sets.

3. METHODS

It was decided that one of the model generation methods would be logistic regression [29] representing statistical approaches which are preferable in medical community, while the second would be the XGBoost method [30] implementing Gradient Boosting model that is a leading data-driven machine learning

---

[1] http://decode.polsl.pl

2

Bull. Pol. Acad. Sci. Tech. Sci. 69(4) 2021, e137349

approach. For each of these two approaches a set of models optimised for different values of the quality assessment measures would be generated.

Knowing that two pairs of quality measures are used for screening, i.e., positive predictive value (PPV) and negative predictive value (NPV) on the one hand, and sensitivity and specificity on the other, it was decided to use both. Two measures were defined, described by equations (1) and (2), where M_{pv} is the weighted harmonic mean of predictive values and M_{ss} is the weighted harmonic mean of sensitivity and specificity:

$$M_{pv} = \left(\frac{\alpha}{NPV} + \frac{1-\alpha}{PPV} \right)^{-1}, \tag{1}$$

$$M_{ss} = \left(\frac{\alpha}{sensitivity} + \frac{1-\alpha}{specificity} \right)^{-1}. \tag{2}$$

The α parameter allows a user to define the significance of individual measures: NPV and PPV for M_{pv} or sensitivity and specificity for M_{ss}.

The generation of each classifier was related to the optimisation (maximisation) of one of the M measures for the selected α value. The set of α values was generated in the following way: its lower bound was set to 0.15, upper bound was set to 0.85, and the parameter values were selected with a step 0.1. Additionally, a value 0.5 was included, which resulted in the following set of 9 values: $\alpha \in \{0.15, 0.25, 0.35, 0.45, 0.50, 0.55, 0.65, 0.75, 0.85\}$.

3.1. XGBoost classifiers.

Classification models based on the XGBoost approach were generated using H2O machine learning platform [31]. Each XGBoost classifier was trained to maximise the selected M measure for the selected α value. Five parameters of the XGBoost method were selected for the optimisation purpose and their values were adopted using hill climbing optimisation approach. The parameters, the range in which the best value of the parameter was searched for, the step with which the search was performed and initial (default) parameter values are presented in Table 1. The parameters in Table 1 are listed in the order in which the climbing optimisation process was carried out. Therefore, at first the *learn_rate* parameter value was optimised and the other parameters were set to default values. 10×10 fold cross validation process was used to generate the

Table 1
Optimised XGBoost parameters

Parameter	Lower bound	Upper bound	Step	Def. value
learn_rate	0.05	0.6	0.05	0.3
sample_rate	0.5	1.0	0.1	1.0
col_sample_rate	0.1	1.0	0.1	1.0
max_depth	2	8	1	6
ntrees	40	130	10	50

model for consecutive parameter values within the given range and to evaluate the classifier with respect to a chosen M measure. Next, *sample_rate* parameter was optimised performing 10×10 fold cross validation for consecutive parameter values, and the values of the other parameters were set to default except *learn_rate* which value was already set in optimisation process. The process continued until all five parameters were optimised.

The XGBoost classifier returns a scoring value in a range (0, 1). In order to transform this scoring into binary decision, a threshold must be set that will determine which values correspond to class 0 and class 1. In the presented approach, a threshold with values from the following set was taken into account: *threshold* $\in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. Each time the cross validation process was executed each of the threshold values was verified to select the best one.

When the parameter values were selected, their neighbourhood was verified as additional optimisation step. Once again 10×10 fold cross validation process was applied and the selected parameters were verified in the same order. Each parameter was analysed within the ε neighbourhood of its current value, taking into account the constraints set for this parameter. The verified ε values and the constraints for each parameter are presented in Table 2.

Table 2
Neighbourhood optimisation of XGBoost parameters

Parameter	ε values	Constraints
learn_rate	±0.2, ±0.4	(0, 1]
sample_rate	±1	(0, +∞)
col_sample_rate	±0.2, ±0.5	(0, 1)
max_depth	±0.2, ±0.5	(0, 1)
ntrees	±5	(0, +∞)

Finally, having all the parameter values set, the selected threshold value was optimised within its neighbourhood. The current threshold value was modified within the ±0.06 ranges with a step set to 0.01 what resulted in 13 threshold values which were verified in 10×10 fold cross validation again.

3.2. Logistic regression classifiers.

In case of the second approach, the analysed observations were randomly divided into training and validation sets in a balanced manner with equal proportions. For the training set, a logistic regression model was built with the forward selection method: new variables were iteratively appended, starting from the null model. In each step, the model with the selected variable met 2 requirements: 1) its Bayes Factor was the highest among all models considered in a step; 2) its Bayes Factor was not lower than 1. Each logistic regression model provided the probability of belonging to the SARS-CoV-2(+) class. To identify the optimal threshold probability for classification, models were tuned to maximise the quality metrics (M_{pv} or M_{ss}) for assumed α value. Forward feature selection and tuning procedures were repeated for 100 divisions into training and validation sets.

Bull. Pol. Acad. Sci. Tech. Sci. 69(4) 2021, e137349

3

M. Kozielski, J. Henzel, J. Tobiasz, A. Gruca, P. Foszner, J. Zyła, M. Bach, A. Werner, J. Jaroszewicz, J. Polańska, and M. Sikora

The resulting set of 100 models served for feature ranking generation. Variables included in each model were sorted from the lowest Wald's test p-value. Each variable was assigned a weight w given by:

$$w = 1 - \frac{k-1}{m}. \qquad (3)$$

The k parameter is the position according to p-values and m is a maximal number of features among all models. Each weight was multiplied by model quality represented by M_{pv} or M_{ss} for assumed α. Products summed up among all models gave an importance score for each variable.

To be included in the final model, a variable had to fulfil two conditions: 1) it was supposed to appear in at least 50% of models; 2) its importance score could not be lower than 70% of the importance score for variable preceding in the ranking. Variables that met those requirements were used in the final model, which was tuned afterwards to maximise the assumed quality metrics. Therefore, the set of selected variables used in the logistic regression model may differ for each value of the α parameter.

4. EXPERIMENTS AND RESULTS

The experiments were performed on the data set of the patients admitted to the Specialised Hospital No. 1 in Bytom, Poland with COVID-19-like symptoms. Each patient was asked to fill in the survey describing the observed symptoms and then was tested with RT-PCR test to confirm SARS-CoV-2 infection. The survey data was collected between February and September 2020 and consisted of 1941 patients: 1355 patients with negative SARS-CoV-2 RT-PCR test result and 586 patients with positive SARS-CoV-2 RT-PCR test result. Each patient was initially described by 32 attributes (18 attributes describing symptoms, 7 attributes listing comorbidities, 3 attributes representing the patient's condition and 4 attributes representing other epidemiological attributes, namely: age, sex, bloodgroup and contact with infection) from which the most important features were selected depending on the classifier. The attributes used by the XGBoost classifier were adopted in accordance with the results presented in [9]. The attributes used by the Logistic regression classifier were selected in accordance with the method presented in Section 3.2.

Logistic regression approach does not deal with missing data. Hence, in case of this method the data set was reduced in terms of variables and patients before the analysis. The set of features with the highest percentage of lacks ($\geq 60\%$) was not considered. Observations were also excluded due to missing information concerning any of the remaining features. Consequently, for the logistic regression approach, the final data set consisted of 399 SARS-CoV-2(+) and 699 SARS-CoV-2(−) symptomatic cases with 16 features and their pair interactions. Two-feature interactions were also considered as potential model variables and reflected the co-occurrence of symptoms supporting the diagnosis.

The final parameters of the XGBoost classifiers generated on the characterised data and the final parameters of the Logistic regression classifiers generated on the reduced data set were included in the appendix available on the website[2].

In typical analyses of binary classifier performance, quality indices such as Area Under Curve (AUC), $F1$, Matthews correlation coefficient (MCC) or Balanced Accuracy ($BAcc$) are usually used to assess the ability of the classifier to detect examples both from positive and negative classes. The aim of our analysis is to show how different parameter settings can be used to optimise classifier performance for different scenarios of data analysis as usually increasing the ability of a classifier to detect one class leads to decreasing its ability to detect the other class.

By optimising the classifiers with regard to different performance measures (M_{pv} or M_{ss}) and different values of the α coefficient, it can be decided which characteristics of the classifier are important for a purpose of a particular type of analysis. Specifically, in case of population screening purposes it is usually better to optimise the classifier with its PPV and NPV quality indices, while in case of assessing the quality of particular test results, it is better to use specificity and sensitivity as performance indicators [8]. On the other hand, quality indices such as sensitivity and NPV are focused on maximising the ability of a classifier to predict positive examples (in our case patients with the positive result of RT-PCR test) while specificity and PPV are focused on the ability to distinguish examples belonging to the negative class (in our case patients that are not infected with the SARS-CoV-2 virus). Therefore, depending on the situation, we might be interested in treating different misclassifications differently.

While analysing the results presented in the Table 3 for the XGBoost classifier for M_{pv} measure with $\alpha = 0.15$ we notice high values both for NPV and PPV measures. Maximising value of the PPV measure allows us to train the classifier with the high fraction of true positive results among all positive examples. However, doing this we allow for low values of sensitivity measure which means that our classifier is not able to detect positive cases from the population of positive patients.

By increasing the value of the α coefficient we can optimise the classifiers for higher values of the sensitivity measure. While analysing obtained results, we can see that the highest values of the sensitivity are obtained for the regression classifier for the M_{pv} measure with $\alpha = 0.85$ (Table 4). However, this is a clear trade-off with regard to the values of PPV measure, as the higher sensitivity is, the lower PPV is. Such classifier is able to detect most of the positive cases from the population of positive patients, however there is less probability that a particular example assigned by a classifier to a positive class is a truly positive case.

We can also notice that optimising classifier with regard to the M_{ss} measure gives us higher values of the $BAcc$ measure for the both classifiers. This is due to the fact that $BAcc$ measure

---

[2] http://adaa.polsl.pl/wp-content/uploads/2021/03/appendix.pdf

4

Bull. Pol. Acad. Sci. Tech. Sci. 69(4) 2021, e137349

Enhancement of COVID-19 symptom-based screening with quality-based classifier optimisation

Table 3
Results of the XGBoost classifiers

Measure	α	PPV	NPV	Sensitivity	Specificity	BAcc	PPV sd	NPV sd	Sens. sd	Spec. sd	BAcc sd
M_{pv}	0.15	0.837	0.708	0.049	0.995	0.522	0.192	0.030	0.029	0.006	0.013
	0.25	0.837	0.708	0.049	0.995	0.522	0.192	0.030	0.029	0.006	0.013
	0.35	0.751	0.709	0.061	0.990	0.525	0.208	0.028	0.029	0.009	0.014
	0.45	0.722	0.713	0.084	0.986	0.535	0.163	0.028	0.038	0.011	0.018
	0.50	0.722	0.713	0.084	0.986	0.535	0.163	0.028	0.038	0.011	0.018
	0.55	0.722	0.713	0.084	0.986	0.535	0.163	0.028	0.038	0.011	0.018
	0.65	0.578	0.784	0.454	0.856	0.655	0.082	0.031	0.070	0.034	0.039
	0.75	0.585	0.786	0.460	0.858	0.659	0.081	0.031	0.067	0.033	0.037
	0.85	0.478	0.854	0.744	0.649	0.696	0.050	0.032	0.055	0.039	0.034
M_{ss}	0.15	0.566	0.786	0.466	0.845	0.656	0.081	0.032	0.076	0.035	0.040
	0.25	0.544	0.802	0.541	0.803	0.672	0.069	0.032	0.071	0.038	0.038
	0.35	0.494	0.841	0.698	0.690	0.694	0.054	0.033	0.064	0.042	0.037
	0.45	0.489	0.850	0.726	0.673	0.699	0.052	0.031	0.059	0.038	0.034
	0.50	0.488	0.849	0.723	0.672	0.697	0.050	0.031	0.056	0.039	0.032
	0.55	0.488	0.849	0.723	0.672	0.697	0.050	0.031	0.056	0.039	0.032
	0.65	0.474	0.856	0.751	0.639	0.695	0.051	0.031	0.054	0.040	0.033
	0.75	0.446	0.868	0.798	0.571	0.685	0.045	0.034	0.054	0.041	0.032
	0.85	0.410	0.884	0.859	0.464	0.662	0.044	0.034	0.044	0.048	0.029

Table 4
Results of the Logistic Regression classifiers

Measure	α	PPV	NPV	Sensitivity	Specificity	BAcc	PPV sd	NPV sd	Sens. sd	Spec. sd	BAcc sd
M_{pv}	0.15	0.567	0.646	0.057	0.977	0.517	0.223	0.014	0.084	0.036	0.025
	0.25	0.566	0.648	0.067	0.973	0.520	0.222	0.018	0.102	0.042	0.031
	0.35	0.564	0.649	0.075	0.970	0.523	0.222	0.022	0.118	0.047	0.036
	0.45	0.560	0.655	0.103	0.958	0.531	0.219	0.030	0.156	0.063	0.047
	0.50	0.558	0.656	0.109	0.956	0.532	0.219	0.031	0.159	0.065	0.048
	0.55	0.559	0.659	0.124	0.949	0.536	0.218	0.036	0.181	0.078	0.052
	0.65	0.548	0.700	0.294	0.860	0.577	0.210	0.074	0.317	0.158	0.082
	0.75	0.498	0.832	0.791	0.526	0.659	0.047	0.053	0.114	0.163	0.040
	0.85	0.411	0.925	0.959	0.204	0.582	0.034	0.054	0.047	0.131	0.044
M_{ss}	0.15	0.582	0.720	0.439	0.820	0.630	0.027	0.015	0.054	0.029	0.019
	0.25	0.566	0.745	0.541	0.762	0.652	0.029	0.019	0.059	0.039	0.021
	0.35	0.557	0.767	0.612	0.722	0.667	0.024	0.022	0.059	0.035	0.022
	0.45	0.545	0.785	0.672	0.679	0.675	0.023	0.024	0.056	0.040	0.022
	0.50	0.541	0.792	0.690	0.665	0.678	0.023	0.024	0.057	0.044	0.020
	0.55	0.536	0.795	0.703	0.651	0.677	0.022	0.026	0.060	0.045	0.020
	0.65	0.519	0.811	0.751	0.601	0.676	0.024	0.025	0.053	0.052	0.020
	0.75	0.502	0.823	0.790	0.551	0.670	0.023	0.028	0.050	0.053	0.021
	0.85	0.481	0.839	0.836	0.484	0.660	0.020	0.029	0.041	0.050	0.023

consists of specificity and sensitivity components which are also components of the M_{ss} measure. Another observation is that with regard to M_{pv} optimisation we do not notice changes in the values of different quality indices in the lower ranges of the α coefficient. This can be seen by analysing the graphical representation of the results presented in the Fig. 1 where we can notice that while optimising with regard to the M_{ss} measure changes of quality indices have linear characteristic, while in case of optimising with regard to M_{pv} measure, changes of their values have sigmoidal characteristic.
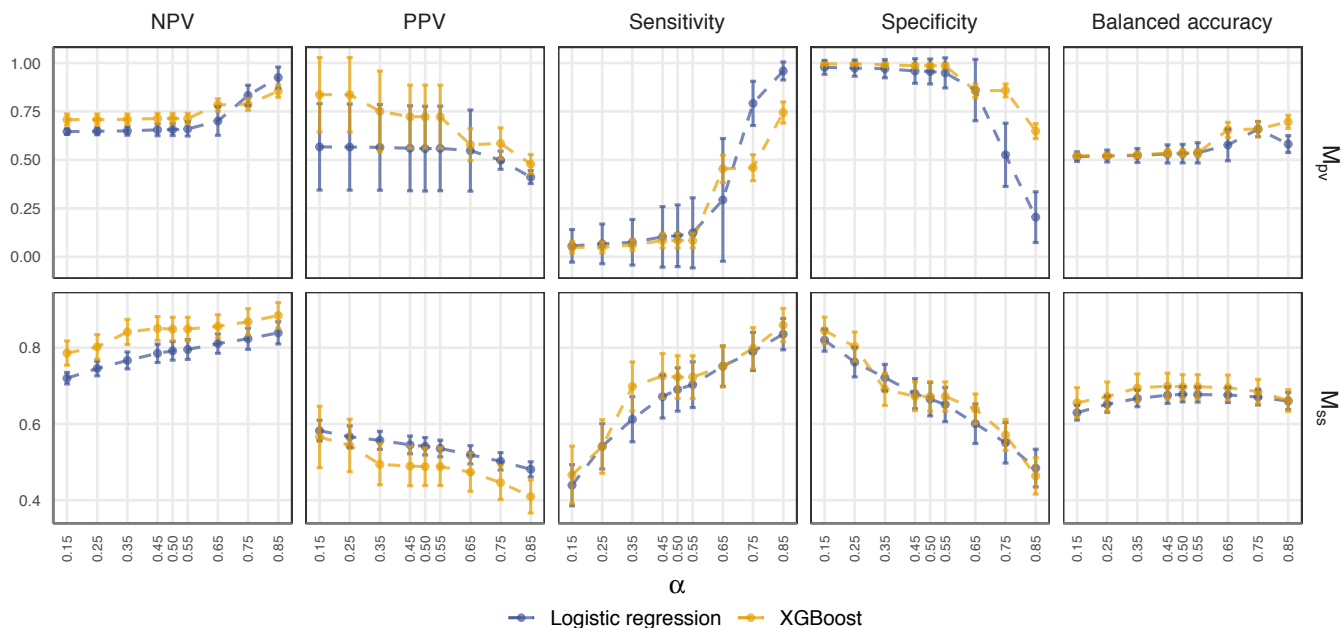
Bull. Pol. Acad. Sci. Tech. Sci. 69(4) 2021, e137349

5

M. Kozielski, J. Henzel, J. Tobiasz, A. Gruca, P. Foszner, J. Zyła, M. Bach, A. Werner, J. Jaroszewicz, J. Polańska, and M. Sikora

Fig. 1. Dependence of the value of the classifiers assessment measures on the α parameter

5. CREATED SOLUTION

The solution created was implemented as a web application that extends the existing DECODE service supporting self-verification in terms of COVID-19 [32]. The DECODE service has the form of a questionnaire that collects the information needed for the assessment of the possibility of being sick with COVID-19. After completing and sending the questionnaire, the patient receives suggestion in the form of Negative/Positive statement, rating value and a visualisation positioning the results in relation to the training data set.

The DECODE service extension introduced in this study is available after signing in as a new user. From the user's perspective (see Fig. 2), the service offers a choice of one of two classifiers: Logistic regression or XGBoost, one of two evaluation
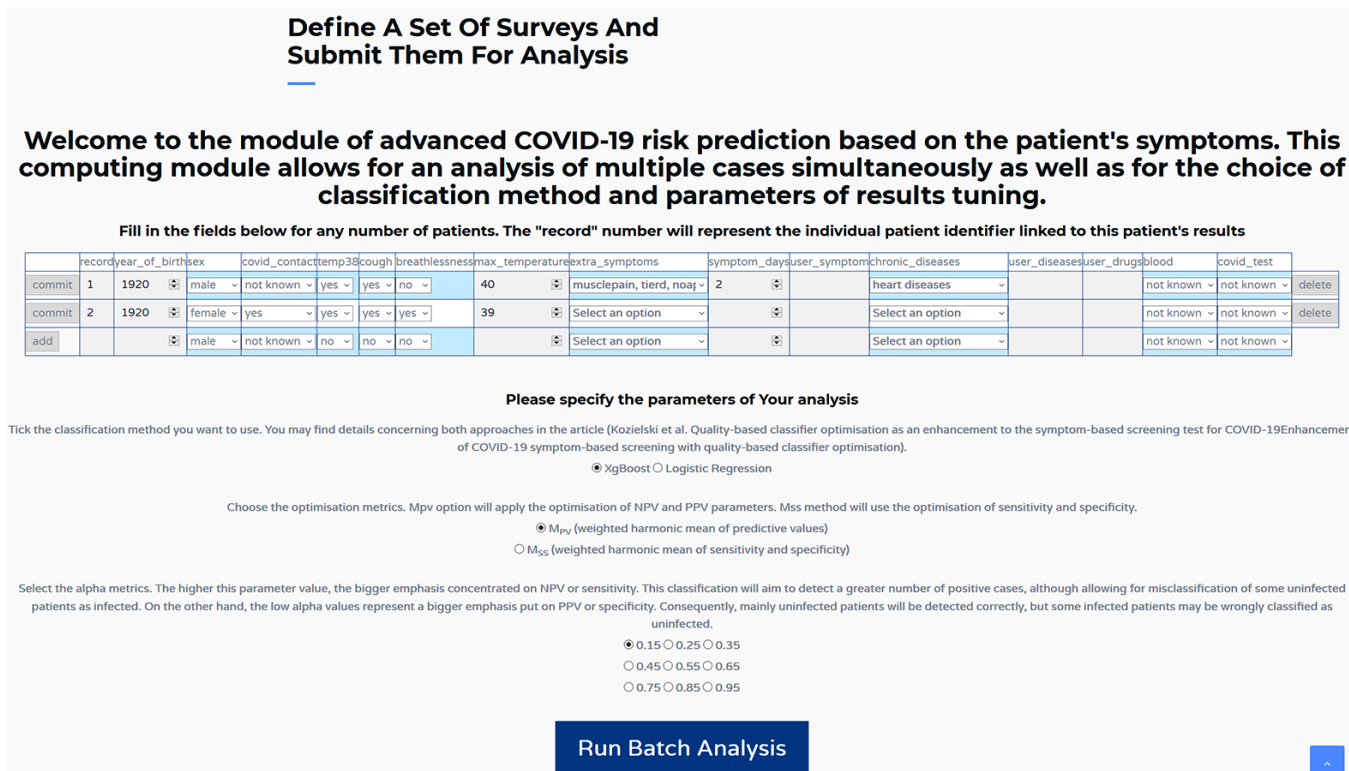

Fig. 2. Application window after loading sample data

6

Bull. Pol. Acad. Sci. Tech. Sci. 69(4) 2021, e137349

measures ((1) or (2)), and a value of parameter α. The chosen configuration defines which of the available 36 classifiers will be applied. The methods of generating the classifiers and selecting their parameter values are presented in Section 3 and their performance is discussed in Section 4.

The data that is to be analysed by the system is loaded as a batch and it has to be provided by a user. The user has to ensure proper formatting of the loaded data set and its compatibility with the features of training data set [9].

As a result of its operation the system returns the original data set extended with combinations of features used by the classifiers and both scoring values and binary classification results of each example.

6. CONCLUSIONS

This paper presents research on symptom-based screening tests for COVID-19 using machine learning methods. Since screening tests should be assessed taking into account several classification quality measures, the aim of the conducted research was to verify how the results of the used classifiers are affected by the model optimisation in relation to the selected measures.

Therefore, a scheme of activities was proposed based on the selection of thresholds defining the weights of classification quality measures used for screening tests, optimisation of classifier parameters and finally, generation and quality assessment of the obtained models in the cross-validation process. The adopted scheme and the analysis performed constitute a contribution of this work.

Results presented in this study show that by changing the value of the α parameter in formulas (1) and (2) the classifiers can be optimised for different purposes. However, there is always a trade-off between the ability of the classifier to detect positive cases from the population and the confidence in the classifier results. Understanding the dependencies between different types of the classifier performance measures is crucial in order to train the model according to the specific needs of medical doctors or researchers. For example, in the case of patient screening for a highly contagious disease such as COVID-19, it is more important to be able to detect all the positive cases even at the cost of being oversensitive and classifying negative examples into positive class.

The tool created as a result of the research allows physicians and researchers for flexible analysis of the collected data representing the symptoms of sick patients. The analysis may include the selection of a properly calibrated classifier or the verification of different classifiers for different thresholds determining the weight of the classification quality measure.

REFERENCES

[1] L.D. Maxim, R. Niebo, and M.J. Utell, "Screening tests: a review with examples", *Inhal. Toxicol.*, vol. 26, no. 13, pp. 811–828, 2014.

[2] D. Ardila *et al.*, "End-to-end lung cancer screening with threedimensional deep learning on low-dose chest computed tomography", *Nat. Med.*, vol. 25, no. 6, pp. 954–961, 2019.

[3] R. Landy, F. Pesola, A. Castañón, and P. Sasieni, "Impact of cervical screening on cervical cancer mortality: estimation using stage-specific results from a nested case–control study", *Br. J. Cancer*, vol. 115, no. 9, pp. 1140–1146, 2016.

[4] E.F. Conant *et al.*, "Breast cancer screening using tomosynthesis in combination with digital mammography compared to digital mammography alone: a cohort study within the prospr consortium", *Breast Cancer Res. Treat.*, vol. 156, no. 1, pp. 109–116, 2016.

[5] K. Gostic, A.C. Gomez, R.O. Mummah, A.J. Kucharski, and J.O. Lloyd-Smith, "Estimated effectiveness of symptom and risk screening to prevent the spread of COVID-19", *eLife* 9, vol. 9, p. e55570, 2020.

[6] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of COVID-19: systematic review and critical appraisal", *BMJ*, vol. 369, p. m1328, 2020.

[7] A. Callahan *et al.*, "Estimating the efficacy of symptom-based screening for COVID-19", *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–3, 2020.

[8] R. Trevethan, "Sensitivity, specificity, and predictive values: Foundations, pliabilities, and pitfalls in research and practice", *Front. Public Health*, vol. 5, p. 307, 2017.

[9] J. Henzel *et al.*, "Classification supporting COVID-19 diagnostics based on patient survey data", arXiv:2011.12247, 2020.

[10] H. Swapnarekha, H.S. Behera, J. Nayak, and B. Naik, "Role of intelligent computing in COVID-19 prognosis: A state-of-theart review", *Chaos Solitons Fractals*, vol. 138, p. 109947, 2020.

[11] S. Lalmuanawma, J. Hussain, and L. Chhakchhuak, "Applications of machine learning and artificial intelligence for COVID-19 (SARS-CoV-2) pandemic: A review", *Chaos Solitons Fractals*, vol. 139, p. 110059, 2020.

[12] Y. Mohamadou, A. Halidou, and P.T. Kapen, "A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19", *Appl. Intell.*, vol. 50, no. 11, pp. 3913–3925, 2020.

[13] A. Banerjee *et al.*, "Use of machine learning and artificial intelligence to predict SARS-CoV-2 infection from full blood counts in a population", *Int. Immunopharmacol.*, vol. 86, p. 106705, 2020.

[14] T. Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim, and U. Rajendra Acharya, "Automated detection of COVID-19 cases using deep neural networks with x-ray images", *Comput. Biol. Med.*, vol. 121, p. 103792, 2020.

[15] R.M. Pereira, D. Bertolini, L.O. Teixeira, C.N. Silla, and Y.M. Costa, "COVID-19 identification in chest x-ray images on flat and hierarchical classification scenarios", *Comput. Meth. Programs Biomed.*, vol. 194, p. 105532, 2020.

[16] D. Singh, V. Kumar, V. Yadav, and M. Kaur, "Deep neural network-based screening model for COVID-19-infected patients using chest x-ray images", *Int. J. Pattern Recogn. Artif. Intell.*, vol. 25, no. 3, p. 2151004, 2021.

[17] A.G. Wintjens *et al.*, "Applying the electronic nose for preoperative SARS-CoV-2 screening", *Surg. Endosc.*, 2020, doi: 10.1007/s00464-020-08169-0.

[18] J. Laguarta, F. Hueto, and B. Subirana, "COVID-19 artificial intelligence diagnosis using only cough recordings", *IEEE Eng. Med. Biol. Mag.*, vol. 1, pp. 275–281, 2020.

Bull. Pol. Acad. Sci. Tech. Sci. 69(4) 2021, e137349

7

M. Kozielski, J. Henzel, J. Tobiasz, A. Gruca, P. Foszner, J. Zyła, M. Bach, A. Werner, J. Jaroszewicz, J. Polańska, and M. Sikora

[19] P. Bagad *et al*., "Cough against covid: Evidence of covid-19 signature in cough sounds", arXiv:2009.08790 (2020).

[20] C. Feng *et al*., "A novel triage tool of artificial intelligence assisted diagnosis aid system for suspected COVID-19 pneumonia in fever clinics", medRxiv 2020.03.19.20039099.

[21] D. Brinati, A. Campagner, D. Ferrari, M. Locatelli, G. Banfi, and F. Cabitza, "Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study", *J. Med. Syst.*, vol. 44, p. 135, 2020.

[22] "Suspected COVID-19 pneumonia diagnosis aid system", [Online]. Available: https://intensivecare.shinyapps.io/COVID19/, (Accessed on 28/12/2020).

[23] "ML-based COVID-19 test from routine blood test", [Online]. Available: https://covid19-blood-ml.herokuapp.com/, (Accessed on 28/12/2020).

[24] Symptomate, "Symptomate COVID-19 risk assessment tool", [Online]. Available: https://symptomate.com/covid19/checkup, (Accessed on 12/28/2020).

[25] CDC, "Testing for COVID-19", [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/testing.html, (Accessed on 28/12/2020).

[26] Apple Inc., "Apple covid-19", [Online]. Available: https://covid19.apple.com/screening, (Accessed on 28/12/2020).

[27] "COVID-19 Risk Assessment", [Online]. Available: https://covid.preflet.com/en, (Accessed on 28/12/2020).

[28] M. DataLab, M. Groups, "COVID-19 risk calculator". [Online]. Available: https://crs19.pl/ (Accessed on 28/12/2020).

[29] P. McCullagh, J.A. Nelder, *Generalized Linear Models*, 2nd ed. Chapman & Hall, 1989.

[30] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System", *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[31] "Open source machine learning platform". [Online]. Available: https://www.h2o.ai/, (Accessed on 20/12/2020).

[32] "Data drivEn COVID19 DEtection with machine learning"). [Online]. Available: https://decode.polsl.pl (Accessed on 30/12/2020).

8

Bull. Pol. Acad. Sci. Tech. Sci. 69(4) 2021, e137349