The Use of Machine Learning Methods to the Automated Atherosclerosis Diagnostic and Treatment System Development

Maria Demchenko and Irina Kashirina

Voronezh State University, Voronezh 394018, Russia

masha-vrn@yandex.ru

Abstract. The main objective of this study is a comprehensive analysis of an atherosclerosis disease with the use of various machine learning algorithms. One of the primary goals of this study was to develop effective diagnostic models for this disease. The real deidentified medical datasets were used for the models training, in particular, a sample of the MIMIC-III database. The use of the specialized Microsoft Azure Machine Learning platform within this study allowed us to develop highly efficient, scalable, and reusable classification models for atherosclerosis diagnostics. A significant part of the study was also devoted to working with the dataset of the Voronezh Regional Cardiology Dispensary. This dataset, containing real non-personalized data from more than 500 patients, allowed us to obtain the models for identifying the most significant predictors and markers of the atherosclerosis disease. With the purpose of the improvement of the existing models' quality, an automated system is currently being developed - the cardiologist's workplace. This application, developed using the .NET Core and the Angular framework, allows to keep track of patients' and doctors' appointments, as well as save the patients' medical data as diagnoses and prescriptions. The further extension of the medical data obtained will help us to create a comprehensive system that would allow us to identify the most optimal patient treatment strategies using the Markov Decision Process approach.

Keywords: machine learning, classification, Microsoft Azure Machine Learning, LightGBM, XGBoostClassifier, RandomForest, ExtremeRandomTrees, atherosclerosis, ROC, AUC, reinforcement learning, Markov Decision Process, automated doctor's workplace.

1 Introduction

A key aspect of successful treatment is the timely diagnosis of a disease. Often, pathology can begin to develop asymptomatically, sometimes at an early age, eventually worsening the patient's quality of life and health. In particular, the symptoms of the atherosclerosis disease can be found among 17% of people between the ages of 13-19 years, and by the age of 40 years, at least one atherosclerotic lesion is present in more

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR-WS.org/Vol-2790/paper22.pdf

than 70% of patients [1]. The high prevalence of this disease, as well as the accompanying high risk of vascular damage and ischemic lesions of organs, requires the thorough research and the development of the most effective diagnostic methods and treatment policies.

Lots of up-to-date machine learning methods have shown high efficiency in solving various medical problems. The most highly productive machine learning algorithms are often used to solve medical diagnostic problems. One of the most popular approaches is deep learning, which is actively used, in particular, in the task of diagnostics of the life-threatening diseases, which is illustrated in [2,3]. At the same time, a significant part of the studies nowadays is devoted to the application of the effective machine learning methods to the research of cardiovascular diseases. For example, in [4,5], an approach to the diagnosis of peripheral arterial diseases using supervised machine learning methods was demonstrated. Therefore, the most efficient machine learning methods have been chosen as the toolkit for the development of an automated diagnostic system and the selection of optimal atherosclerosis treatment policies.

The task of developing a system for atherosclerosis diagnostics and treatment prescriptions was set by the specialists of the Voronezh Regional Cardiological Dispensary (VOCD). This research includes the following main stages.

1. The development of the non-invasive atherosclerosis diagnostic models. The solution to this problem involves 2 main steps.
   a. Development of diagnostic models using the dataset provided by the VOCD. As a training sample, this study used data collected within a research conducted among 522 adult patients from the Bogucharsky district of the Voronezh region using multichannel volume sphygmography (MVS) – an efficient non-invasive method for atherosclerosis diagnostics [6] aimed at identifying the significant asymmetry of systolic blood pressure (SBP) on the arms or legs (ArmsIndex and LegsIndex, respectively), as well as the ankle-brachial index (ABI). The threshold value of the ABI indicator is an approved and generally accepted diagnostic marker, while the pressure asymmetry coefficients ArmsIndex and LegsIndex are still currently the subject of study. In particular, studies [7-9] have been devoted to the identifying of the SBP asymmetry parameters, which also can be considered as diagnostic features of atherosclerosis. In this study, neural network models (binary classifier of MLP architecture, self-organizing Kohonen maps) [10], as well as ensemble models (RandomForest, ExtremeRandomTrees, XGBoostClassifier) [11] were used to find the most informative values of the SBP asymmetry coefficients and use them as the basis for the calculation of the atherosclerosis markers values.
   b. Development of the diagnostic models using a sample of the international database MIMIC-III. Unfortunately, the size of the regional sample is not sufficient to obtain an expandable and generalizable solution; moreover, this sample contains only the data on atherosclerosis disease. Due to this reason, in addition to the regional sample, the current study included MIMIC-III datasets containing medical records on ICU admissions at a Boston Medical Center from 2001 to 2012. Also at this stage of the study, the Azure Machine Learning platform was

used as a tool for building diagnostic models, which allowed to develop scalable solutions with the possibility of deployment and continuous training. [12]

2. The development of the optimal patients' treatment strategies using efficient reinforcement learning models. The most recent machine learning approaches, for example, reinforcement learning, allow not only to diagnose, but also to develop optimal policies for treating patients with a specific diagnosis and health condition. In particular, in [13] the solution to the problem of treating sepsis using reinforcement learning is provided. It is also well-known that reinforcement learning is nowadays used in some mobile health systems [14], that help to continuously monitor the patient's treatment progress and adjust treatment strategies online.
3. The development of an automated system for collecting medical information, including patients' diagnoses and doctors' appointments, as well as performing administrative functions (registering patients, scheduling consultations with a doctor, etc.). According to this purpose, we have developed an application for patients and doctors, which allows to continuously track the condition, diagnoses and treatment prescriptions of patients and their appointments with the doctors, and collect new data to train and improve models. A significant advantage provided by the development of such a system is the collection of regional data, which contributes to the development of models that consider the local characteristics (climatic, social, etc.) in the most efficient way.

2 Atherosclerosis Diagnostic Models and Methods

The solution of the problem of atherosclerosis diagnostics implied the building of the classification models of high-quality, trained on medical datasets containing information about patients suffering from this disease. Within this task, we performed a comprehensive analysis of the patients dataset provided by the Voronezh Regional Cardiology Dispensary [10,11]. A model developed using this sample reflects the patterns in medical observations that are specific for particular cohorts of patients. At the same time, during this investigation, we proved the effectiveness of the main atherosclerosis markers – ArmsIndex, LegsIndex and ABI, which provide quite simple, but efficient diagnostic methods.

However, the disadvantage of such a sample is its rather small size, therefore, in order to solve this problem at the next stage of the study, it was required to use a dataset containing a larger number of patients' observations. Within this work the MIMIC-III datasets containing the arrays of patents' data who admitted to critical care units at a large tertiary care hospital were studied. Due to having a large MIMIC-III dataset sample we were able to build a model trained on many different cases and taking into account a wide range of factors and dependencies that affect the diagnostic results.

2.1 The Task of the Atherosclerosis Markers and Predictors Analysis. The Dataset of the Voronezh Regional Cardiology Dispensary

Within this study we were able to take into account the specific local properties of the atherosclerosis diagnostics problem for a particular group of patients using the dataset of patients from the Bogucharsky district of the Voronezh region, provided by the Voronezh Regional Cardiology Dispensary [6,15]. The description of the dataset features is provided in Table 1.

Table 1. Regional dataset input features.

Feature category	Variables
Hemodynamic	Systolic/diastolic/ pulse arterial pressure on the right/left arm /leg (SBPra, DBPra, PPra, SBPla, DBPla, PPla, SBPrl, SBPll), heart rate (HR), pulse wave velocity (cfPWVV, baPWV)
Socio-Demographic	Gender, age, smoker
Anthropometric	Height, weight, body mass index (BMI)
Laboratory	Glucose, cholesterol
Clinical	Arterial hypertension (AH), stenocardia, infarction, acute disorder of cerebral circulation (ADCC), coronary artery bypass grafting/percutaneous intervention (CABG/PCI), diabetes, chronic heart failure (CHF), atrial flutter and atrial fibrillation (AF), obesity

Within this research we were able to identify and prove the diagnostic efficiency of the main atherosclerosis markers (represented by equations (1)-(3), where ABIr/ABIl is right/left ankle-brachial index, SBPa/SBPl is arms/legs systolic blood pressure).

$$ArmsIndex = \begin{cases} 1, & if \quad |\Delta SBPa| \geq 14 \\ 0, & otherwise \end{cases}. \tag{1}$$

$$LegsIndex = \begin{cases} 1, & if \quad |\Delta SBPl| \geq 15 \\ 0, & otherwise \end{cases}. \tag{2}$$

$$ABI = \begin{cases} 1, & if \quad ABIl \leq 0.9 \quad or \quad ABIr \leq 0.9 \\ 0, & otherwise \end{cases}. \tag{3}$$

As a result of this research we were able to develop the efficient, though quite simple method of atherosclerosis diagnostics, which doesn't assume the execution of any complex laboratory or clinical tests or hospitalizations. We also managed to identify the set of atherosclerosis predictors, highly associated with this diagnosis (heart rate, arterial hypertension, diabetes mellitus, age, height, weight, chronic heart failure). The details and results of this research can be found in [10,11]

The models trained on the dataset under consideration reflect local and regional dependences with high accuracy. However, often (for example when the purpose is delivering and integrating the developed solution in the live environment) it is also important to take into account the most general dependencies and data samples of a larger size, which are highly representative and include many observations.

2.2 Automated Experiments Based on Microsoft Azure Machine Learning Platform Using the MIMIC-III Data

MIMIC-III sample of a patients having atherosclerosis. MIMIC-III database [16] («Medical Information Mart for Intensive Care») is a large, single-center database comprising information relating to patients admitted to critical care units at a large tertiary care hospital in Boston (Massachusetts). Data includes vital signs, medications, laboratory measurements, observations and notes charted by care providers, fluid balance, procedure codes, diagnostic codes, imaging reports, hospital length of stay, survival data, and more.

Within this study, aimed at atherosclerosis disease investigation, we prepared a dataset containing a sample of laboratory (glucose, cholesterol, creatinine) and clinical (diabetes, history of myocardial infarction) features. A fragment of the original sample is shown in Table 2.

Table 2. MIMIC-III sample fragment.

	Glucose	Cholesterol	Creatinine	Diabetes	Infarction	Atherosclerosis
1	125.6875	129	1.9444	0	0	0
2	148.5	208	1.58	0	1	0
3	152.5714	252	0.575	1	1	1
4	122	153	1.0667	0	0	1
5	103.6667	173	1.5	0	0	1

The graph displaying the pairwise laboratory features distribution considering the atherosclerosis diagnosis approved or not is shown in Fig. 1, along with 7 different kinds of atherosclerosis specified.

Classification models building using Azure ML. Based on this sample, the task of the patients' classification based on the laboratory and clinical features was solved. In this case, the "Atherosclerosis" field of the dataset, representing whether the patient has atherosclerosis diagnosis or not, was chosen as a target (output) variable.

Usually the solution of the machine learning problems involves time-consuming steps of the models setting up and selection, as well as the calibration of hyperparameters, which sometimes is a complex and long-term process. However, the development of the current technologies allows us to automate these steps, which can significantly reduce costs and optimize the obtaining of models that have the highest quality metrics.

In particular, in this study, we examined the capabilities of the Microsoft Azure Machine Learning platform, which allowed us to conduct many experiments (iterations of building and training models) on different versions of data sets.

The multiple experiments (classification models training, testing and validation iterations) were run using the MIMIC-III dataset sample and such classification methods as RandomForest, ExtremeRandomTrees, XGBoostClassifier (the detailed description of these algorithms are provided in [11]), as well as gradient boosting framework LightGBM.

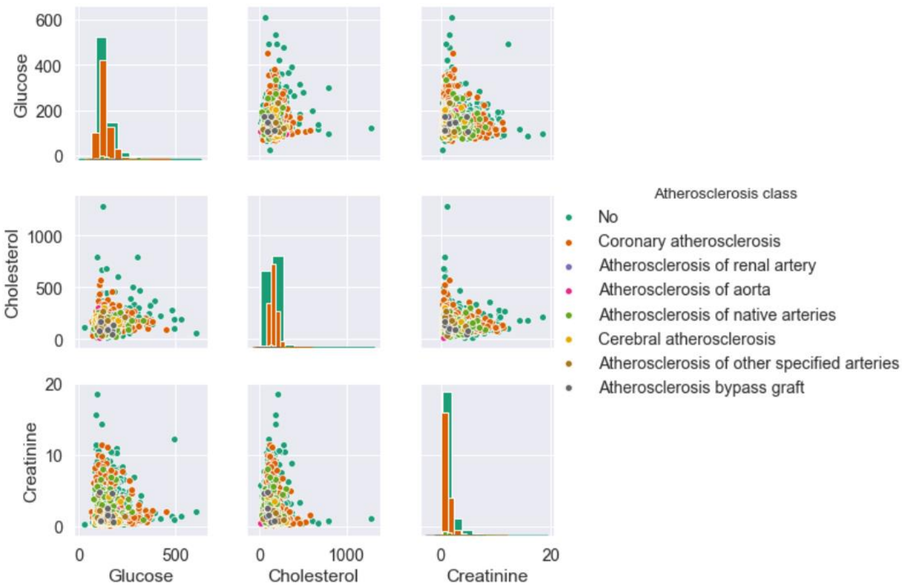

Fig. 1. The distribution of laboratory features. MIMIC-III.

The advantages of this framework are its high performance and training speed, low memory costs, and also the high calculation accuracy.

The high efficiency of this framework is a result of the optimized tree growth algorithm. In opposite to the majority of the decision tree learning algorithms, that grow trees by level (depth-wise), LightGBM grows trees leaf-wise (best-first), which is displayed in Fig. 2. [17]

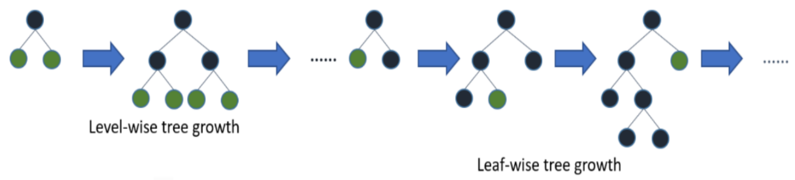

Fig. 2. Lever-wise and leaf-wise tree growth algorithms comparison.

Metrics and models' performance. The initial sample is a large sample containing 7021 records of patient diagnoses. Moreover, this sample is balanced, because the proportion of patients with diagnosed atherosclerosis in the entire sample was 53% (3722 cases, 3299 healthy patients).

The Azure Machine Learning platform provides a number of built-in quality metrics for classification models, including Accuracy, Weighted accuracy, Average precision score macro / micro / weighted, AUC macro / micro / weighted, etc. An important step of the classification task is to determine the metrics for the current analysis with the highest priority.

Often the optimal approach for assessing the accuracy of medical diagnostic models is ROC-curves, and the area under the ROC-curve AUC is the primary aggregated quality metric. The Weighted AUC criterion, in its turn, is an extension to AUC that takes into account the weights of each class in the dataset target variable. Consequently, the performance of the final classifiers was evaluated, primarily, using the weighted AUC criterion.

3 The Results of Atherosclerosis Diagnostics Task Solution

As a result of the research using the data of clinical patients' examination in the Bogucharsky district, diagnostic models, described in Table 3, were built. This sample was unbalanced (the rate of patients with atherosclerosis was only 14%), therefore, in particular, ensemble models with built-in class balancing algorithms were considered. The model of the highest quality contained the ABI marker as a diagnostic feature and reflected its dependency on the patient's hemodynamic parameters. [11]

Table 3. The best diagnostic models for each of the atherosclerosis markers.

Marker	Features set	Classifier	AUC
ABI	SBPra, SBPll, HR, DBPra, cfPWV_calc> 10 m/s	Blagging	0.89
ArmsIndex	SBPla, SBPrl, SBPll, PPla, cfPWV, HR, Gender, AH, diabetes, CCF, Age, Height, Weight	ExtraTrees	0.73
LegsIndex	SBPll, SBPla, SBPra, DBPra, HR, cfPWV, DBPla, Gender, AH, diabetes, CCF, Age, Height, Weight	Blagging	0.71

The building of the atherosclerosis classification models based on the MIMIC-III dataset using the Microsoft Azure Machine Learning cloud-based platform allowed to carry out a series of iterations for building the classification models using such frameworks and methods as LightGBM, XGBoostClassifier, RandomForest, and ExtremeRandomTrees. The models were compared primarily based on the weighted AUC metric. The list of 10 best models is shown in Table 4.

The model with the maximum weighted AUC (0.85746) was built using the LightGBM classifier and the Max-AbsScaler standardization algorithm. Moreover, in addition to AUC, other metrics were calculated. Quality metrics of the best classifier are given in Table 5.

Also, the Azure Machine Learning platform provides tools for visualization and interpretation of the generated models. For example, Fig. 3 illustrates the graph of ROC curves, and Fig. 4 shows an automatically generated confusion matrix.

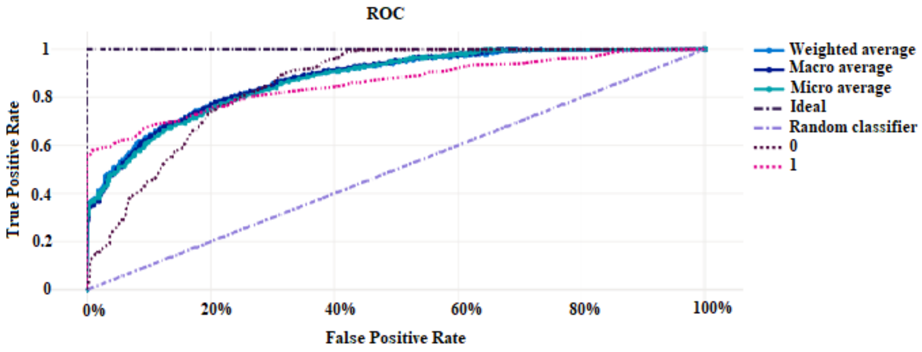

Fig. 3. ROC curves of the best classifier.

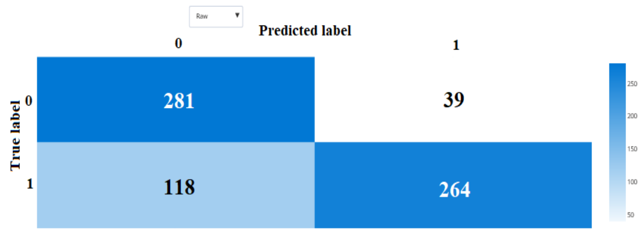

Fig. 4. Confusion matrix.

Table 4. Classification models built using Azure ML.

Algorithm name	Weighted AUC
MaxAbsScaler, LightGBM	0.85746
MaxAbsScaler, XGBoostClassifier	0.85473
StandartScalerWrapper, XGBoostClassifier	0.85411
MaxAbsScaler, LightGBM	0.85270
StandardScalerWrapper, LightGBM	0.85219
MinMaxScaler, RandomForest	0.85125
StandardScalerWrapper, LightGBM	0.85029
StandardScalerWrapper, XGBoostClassifier	0.84944
StandardScalerWrapper, ExtremeRandomTrees	0.84933
MinMaxScaler, LightGBM	0.84933

4 The Task of Obtaining Optimal Treatment Strategies

The importance of timely atherosclerosis diagnostics can be approved by many medical experts. However, an equally important task is the development and prescription of a most efficient treatment strategies, which can be solved successfully using the up-to-date reinforcement learning methods.

Table 5. Metrics of the best classifiers built.

Metric	Value
Accuracy	0.75783
Weighted accuracy	0.74548
Average precision score macro	0.85105
Average precision score micro	0.87781
Average precision score weighted	0.85537
AUC macro	0.85437
AUC micro	0.86882
AUC weighted	0.85437

Reinforcement learning is one of the most actively developing areas in artificial intelligence. It is a computational approach to learning whereby an agent tries to maximize the total amount of reward it receives while interacting with a complex, uncertain environment. [18]

4.1 Reinforcement Learning Model in the Clinical Trials Problems

The task of conducting clinical trials and prescribing optimal strategies for treating patients is one of the many areas of reinforcement learning application. In this case, the model under consideration is a Markov Decision Process with a defined set of states, actions and rewards.

1. We define sets of patient’s parameters at each stage of treatment (for example, systolic blood pressure, pulse, blood cholesterol, etc.) as a set of states.
2. This problem can be formulated as a series of treatment episodes, where each episode represents one of the possible treatment outcomes. An episode is considered as completed if the patient was successfully discharged from the hospital, transferred to another department, or in case of the patient's death - we define these outcomes as terminal states.
3. We define actions as the chosen treatment strategies at each treatment step (medications, procedures etc.).
4. We define rewards as zero if the treatment is being continued, positive in case of the patient’s discharge or getting better, or negative in the case of patient’s death or getting worse.
5. Thus, the agent’s goal will be to identify and further follow the treatment strategy leading to maximum total reward.

   A schematic example of this process is illustrated in Fig. 5.

   Among the key features of the model, the following should be specified:

1. An array of data is a collection of large amounts of medical information.
2. It is necessary to take into account the change in the patient's state, and, consequently, the agent's reward, not only at the end of the episode, but also at each step of the episode (continuous treatment process).

3. It is necessary not only to train the agent to act according to the optimal treatment strategy, but also to identify the most efficient treatment policies leading to the maximum reward for each of the possible conditions.

The solution of the reinforcement learning problems is technically the solution of optimization problems with respect to the objective function Q(s,a), which approximates the total rewards of the agent who chooses action a, being in state s (where a is one of the possible actions of the treatment strategy). Therefore, in order to approximate this function on a large amount of data in the most accurate way, deep learning neural networks will be used.

In order to take into account properties 2 and 3 of this problem, the Q-Learning algorithm is applied.

The development of a reinforcement learning model is the current step of this study. The most suitable dataset for such a solution is the MIMIC-III dataset containing a history of medical treatments, procedures and prescribed drugs.

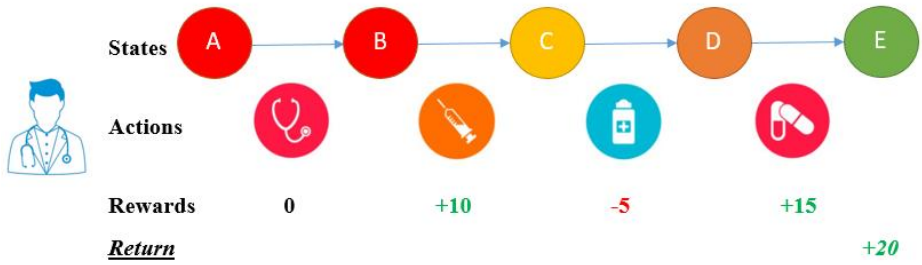

Fig. 5. Markov Decision Process. Clinical trial problem.

5 The Development of the Automated Cardiologist's Workplace

The development and further performance improvement of machine learning models is inseparably linked with the continuous updating of a data array with the most relevant incoming information.

Usually, this process can be set up using the automated systems and applications that control the data flow and history and optimize many processes that the specialists (doctors) should perform on a daily basis.

The aim of this study is to develop an automated application to manage and support the work of a specialist (cardiologist) in examining and treating patients with a specialized pathology (diseases of the cardiovascular system).

The application is considered as a tool which helps to reduce the time to perform the doctor's auxiliary operations, related to medical data input, storage, search and analysis. The application, implementing the artificial intelligence features, should manage the decision making process tasks, executed by doctors, and support the remote patients' monitoring.

5.1 Key Application Features

A list of some of the specific features of the application in development:

1. Compliance with legal requirements and regulations on the protection of personal and medical data (medical databases), as well as the storage of medical information.
2. Fast medical data entry using the contextual dictionaries.
3. Facilitated input (transfer) of digital, textual and graphic data from various third-party medical protocols and conclusions into the application.
4. Automatic integration and merging of various types ("blocks") of medical information for a particular patient at a specific point in time or at a specific time interval.
5. An automatic patient scheduler with a list of tasks and their implementations.
6. The module for notifying patients (SMS, e-mail, Viber, WhatsApp) about assigned tasks and scheduled appointments.
7. Interactive module for the remote monitoring via SMS or a separate mobile application of the patient's condition according to the selected parameters and scenarios.
8. Evaluation of a short-term prognosis based on data obtained using the module for remote monitoring of the patient's condition.
9. The presence of flexibly customizable queries for data search both by patient groups and by individual patients, including the generation of various statistical reports.
10. Digital Signature Support.
11. Support for the medical decision making process in individual problems of diagnosis and treatment assignment using artificial intelligence.

5.2 The Main Components of the Application

The developed automated system includes such components as:

- Web application for doctors, which is used to enter and search for medical information, appointments and consultations scheduling;
- A mobile application for patients, used to track patient health condition and schedule appointments with a doctor;
- A database containing medical information, including patients' data.

5.3 Patient Information Module

The application's module responsible for the basic logic of registration and subsequent maintenance of the patients' information includes such features as:

- The creation of a medical card with the minimum necessary information about the patient;
- Scheduling and tracking the patients' appointments with the doctors for the purpose of consultation, observation, diagnostic or other medical procedures;
- The storage of a list of medical prescriptions and diagnoses for the patient.

Thus, having the minimum necessary set of basic features, this application supports the functions of collecting and storing the medical information (the diagnoses and prescriptions of patients, scheduled appointments with doctors).

The main goal of the application is to provide the support for the primary atherosclerosis diagnostics, as well as intended to support the automatic selection of optimal treatment policies.

6 Conclusion

As a result of the current research, a complete analysis of the atherosclerosis disease was carried out. This research can be considered as a sequence of the following stages.
1. Identification of the most significant markers and predictors of atherosclerosis and building the diagnostic models of atherosclerosis using the data set of the Voronezh Regional Cardiological Dispensary. This stage was implemented within the research [10,11].
2. The development of a generalized model for the atherosclerosis diagnostics using a large dataset. At the current stage of the study, using a sample of MIMIC-III data, it was possible to develop highly efficient, flexible and scalable diagnostic models , available for the further deployment.
3. Within the current study, a system for identifying the optimal atherosclerosis treatment strategies based on the reinforcement learning methods was modelled. Optimization and deployment of this model is planned as the next stage of the study.
4. In order to support and improve the functioning of the created models, an automated cardiologist's workplace is also being developed – which is a system that allows you to collect and save the relevant medical information: prescriptions and diagnoses made by doctors, patient visits and health condition. Further collection of the medical information will continuously provide the most relevant data for training the system and for achieving the most accurate diagnostic results and developing optimal treatment policies for the disease.

References

1. Tuzcu, E.M., Kapadia, S.R., Tutar, E. et al.: High prevalence of coronary atherosclerosis in asymptomatic teenagers and young adults: evidence from intravascular ultrasound. Circulation 103(22), 2705-2710 (2001).
2. Reddy, A.V.N., Krishna, C.P., Mallick, P.K. *et al.* Analyzing MRI scans to detect glioblastoma tumor using hybrid deep belief networks. *J Big Data* **7,** 35 (2020).
3. Jaiswal, A.K., Tiwari, P., Kumar, S., Gupta, D.R., Khanna, A., & Rodrigues, J.J. (2019). Identifying pneumonia in chest X-rays: A deep learning approach. *Measurement*, *145*, 511-518.
4. Ross EG, Shah NH, Dalman RL, Nead KT, Cooke JP, Leeper NJ. The use of machine learning for the identification of peripheral artery disease and future mortality risk. *J Vasc Surg*. 2016;64(5):1515-1522.e3. doi:10.1016/j.jvs.2016.04.026.
5. Duval S, Massaro JM, Jaff MR, et al. An evidence-based score to detect prevalent peripheral artery disease (PAD). *Vasc Med*. 2012;17(5):342-351. doi:10.1177/1358863X12445102.

6. Khokhlov, R.A., Gaydashev, A.E., Ostroushko N.I., Gaydashev A.E.: Multi-channel volume sphygmography in cardioangiological screening of the adult population. Rational Pharmacotherapy in Cardiology 11(4), 371-379 (2015).
7. Weinberg I., et al. The Systolic Blood Pressure Difference Between Arms and Cardiovascular Disease in the Framingham Heart Study. The American Journal of Medicine 2014:127; 209-15.
8. Clark CE, Taylor RS, Shore AC, et al. Association of a difference in systolic blood pressure between arms with vascular disease and mortality: a systematic review and meta-analysis. Lancet 2012:380 (9838):218.
9. Mitchell E. Noninvasive diagnosis of arterial disease. Up-ToDate URL: https://www.uptodate.com/contents/noninvasive-diagnosis-of-arterial-disease.
10. Lvovich Y.E., Kashirina I.L., Demchenko M.V. The Use of Machine Learning Methods to Study Markers of Atherosclerosis of the Great Arteries. Information technology 26(1), 46-55 (2020).
11. Demchenko, M.V., Kashirina, I.L.: The development of the atherosclerosis diagnostic models under conditions of unbalanced classes. J. Phys.: Conf. Ser. 1479 012026 (2020).
12. Azure Machine Learning documentation https://docs.microsoft.com/en-us/azure/machine-learning/, last accessed 2020/5/31.
13. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med*. 2018;24(11):1716-1720. doi:10.1038/s41591-018-0213-5.
14. Istepanian RSH, Al-Anzi T. m-Health 2.0: New perspectives on mobile health, machine learning and big data analytics. *Methods*. 2018;151:34-40. doi:10.1016/j.ymeth.2018.05.015.
15. Khokhlov, R.A., Gaydashev, A.E., Akhmedzhanov N.M.: Predictors of atherosclerotic lesions of limb arteries according to cardioangiological screening of the adult population. Rational Pharmacotherapy in Cardiology 11(5), 470-476 (2015).
16. Johnson, A. E. W. et al. MIMIC-III, a freely accessible critical care database. Sci. Data 3, 160035 (2016).
17. LightGBM documentation, https://lightgbm.readthedocs.io/en/latest/, last accessed 2020/5/31.
18. Sutton, R., Barto, A.: Reinforcement learning. 2nd edn. The MIT Press, Cambridge (2018).