Doctoral School in Medical Sciences

Presented as manuscript

U.D.C.: 616.94-037:004.8(043.2)

EARLY PREDICTION OF SEPSIS USING A PROPRIETARY APPLICATION DEVELOPED BASED ON MACHINE LEARNING (ARTIFICIAL INTELLIGENCE)

321.19 - ANAESTHESIOLOGY AND INTENSIVE CARE

Abstract of the Ph.D. thesis in medical sciences

Chișinău, 2023

The thesis was elaborated at the Department of Anesthesia and Intensive Care no.1 “Valeriu Ghereg” of the N. Testemițanu State University of Medicine and Pharmacy of the Republic of Moldova

Ph.D. supervisor
Belîi Adrian,
Ph.D. hab. in med. sc., university professor ____________

Members of the guidance committee:
Cojocaru Svetlana,
Ph.D. hab. in computer sc., research professor, c,m. of AS ____________
Ciobanu Nellu,
Ph.D. in phys.-math. sc., associate professor ____________
Savan Veaceslav,
Ph.D. in med. sc. ____________

The thesis defense will take place on September 27, 2023, at 14:00 bureau 205, 165 Ștefan cel Mare și Sfânt bvd., N. Testemițanu State University of Medicine and Pharmacy at the session of the PhD Commission, approved by the decision of the Scientific Council of the Consortium from 25.05.2023 (no. 5).

Ph.D. Commission:

The President of the Commission:
Cojocaru Victor, Ph.
D. hab. in med. sc., university professor ____________

Members:
Arnaut Oleg,
Ph.D. hab. in med. sc., university professor ____________
Baltaga Ruslan,
Ph.D. in med. sc., associate professor ____________
Belîi Adrian,
Ph.D. hab. in med. sc., university professor ____________
Cobeț Valeriu,
Ph.D. hab. in med. sc., university professor ____________
Cojocaru Svetlana,
Ph.D. hab. in computer sc., research professor, c.m. ____________
Cornogolub Alexandru,
Ph.D. in med. sc., associate professor ____________

Author ____________
Iapăscurtă Victor © Iapăscurtă Victor, 2023

SUMMARY

CONCEPTUAL MILESTONES OF THE RESEARCH

The topicality and importance of the problem addressed.

Two important aspects can be highlighted that outlines the research direction reflected in the paper: (a) The problem of sepsis as a variety of critical conditions, often difficult to diagnose in time, and the results of treatment depend closely on the time when treatment is started (with antibiotics) and the direct influence of these factors on mortality [1], which over time has decreased insignificantly [2] and (b) The emergence of a new player - the so-called artificial intelligence (AI) technologies, which often claim to be panaceas for many problems, including medical ones. The current research attempts to assess the possibility of using these technologies to address important issues in the management of this group of patients through early prediction (a few hours before onset) of sepsis, which would mitigate the delay in starting the treatment.

Description of the situation in the field and identification of the research problem.

Despite high associated mortality [1] and high treatment costs, sepsis [2] remains difficult to diagnose and treat. Previous research has highlighted the importance of timely recognition of sepsis to improve outcomes and reduce costs associated with treatment [3]. New definitions designed to improve clinical recognition of sepsis have recently been proposed [4], as previous use of screening based on systemic inflammatory response syndrome (SIRS) was found to be non-specific [5]. Data from the medical literature have shown that early diagnosis and treatment can reduce the risk of adverse outcomes due to sepsis [6]. Therefore, early detection of sepsis and more accurate recognition of patients at high risk of developing sepsis is essential for effective treatment. Screening tools most commonly used in clinical settings to identify septic patients include SOFA (Sequential Organ Failure Assessment), SIRS (Systemic Inflammatory Response Syndrome) criteria, and MEWS (Modified Early Warning Score). Despite some limitations, these scoring systems have established performance values and serve as important comparators for recently developed sepsis detection and prediction systems and for evaluating their effect on clinical outcomes [7]. A recent addition to this collection is artificial intelligence-based systems (AIS) [3]. Although AIS are fairly new additions to the field of clinical management of sepsis, they have the potential to improve patient outcomes by providing warning of the impending onset of sepsis thus assisting the clinician in decision making. Currently, in the field of sepsis research, machine learning-based decision support systems are a rapidly developing direction [3, 8].

Aim of the study. To assess the feasibility of AI technologies in the management of critically ill ICU patients at risk of developing sepsis, with the development of a system with discriminative abilities (sepsis vs non-sepsis) that would allow early prediction of sepsis development.

Study objectives

1. To assess the use of artificial intelligence technologies, and in particular machine learning, as one of the core AI technologies at the current stage used in anesthesia-intensive care.

2. Evaluation of the use of AIS in the management of critically ill sepsis patients.

3. Identify a data set in sufficient volume for the creation of a possible early sepsis prediction system.

4. Exploratory analysis of clinical and laboratory data and process it in the manner necessary for the creation of a discrimination/prediction system.

5. To create such a system as a practical application, which would allow the prediction of the development of sepsis in intensive care units.

Scientific research methodology.

The research is a retrospective study of a dataset containing 40336 patients/cases, of which 7.26% are sepsis patients. The rest - 92.74% represent patients admitted to intensive care units with other diagnoses (non-sepsis). Exploratory analysis and data processing are largely determined by further research steps, which involve the final creation of an artificial intelligence system with event prediction abilities (sepsis). The integrated development environment RStudio based on the R programming language (http://www.R-project.org/) was also the interaction language with the H2O platform (https://www.h2o.ai/) for the actual machine learning, was used for the data analysis and processing, including statistical data, and their further processing. High-intensity calculations were performed using Amazon Web Services (AWS EC2 https://aws.amazon.com/ec2/). The TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) guide [9] and PROBAST (Prediction model Risk Of Bias ASsessment Tool) principles [10] were used to evaluate the performance of the model created and report the results.

Scientific novelty and originality.

Based on the analysis of a large dataset (40366 cases, of which - 2932 with sepsis) and their processing with the use of a new data-lip restoration algorithm and the use of algorithmic complexity metrics a decision support system for early prediction of sepsis was created.

Important scientific problem solved in the thesis.

Sepsis is a current problem in the medical service/intensive care unit (ICU) and its early diagnosis can lead to increased treatment success rates, decreased mortality, and reduced cost of care delivered to this group of patients, especially in complex cases. The result obtained, which contributes to solving an important scientific problem, is the development of a system based on machine learning that increases the efficiency of clinical management of patients with sepsis.

The theoretical significance of the research.

The possibility of using the concepts and metrics of algorithmic dynamics in the representation of medical data, including time series data describing the patient's clinical condition, was explored and confirmed. This representation has also proved successful in the later stages of data processing - in the construction of the prediction system, and the high performance of the system is an additional argument. Aspects concerning the predictive value of some clinical parameters, which were elucidated in the study, could contribute to a better understanding of the problem of sepsis as a medical phenomenon.

The applied value of the work.

The developed software application, in which the developed early sepsis prediction system is integrated, can assist the ICU physician in the decision-making process, especially in more complex sepsis cases and especially in ambiguous situations. The proposed data reconstruction and representation methods can facilitate, diversify and enliven the work of researchers in the field.

Main scientific results submitted for defense:

1. Algorithm for reconstructing data represented by time series of physiological parameters containing missing values.
2. Data selection method for the creation of a prediction system for patient conditions (on the example of patients at risk of sepsis).
3. Method of data representation by calculating algorithmic complexity using the block decomposition method (BDM).
4. Automated system for sepsis prediction embedded in a software application for clinical use.

Implementation of scientific results.

- At the final stage, a software application was created to be transmitted to the ATI Clinic of the Institute of Emergency Medicine, Chisinau. The application has a graphical interface that allows entering data (heart rate, SpO2, temperature, systolic and diastolic BP, respiration rate) of a concrete patient over a period of three hours with the possibility of their graphical visualization and obtaining the prediction result, i.e., the risk of developing sepsis with a 4-hour horizon.

- Part of the research results are incorporated in the course "Advanced perioperative monitoring. Elements of computational medicine" (for students in the 6th year at N.Testemitanu University of Medicine and Pharmacy).
- The results of the research were presented at:
  - ✓ European Committee for Anesthesia Education Courses, 13th edition, Module 1, 7-9 February 2018.
  - ✓ European Committee for Anesthesia Education Courses, 15th edition, Module 3, 10-11 December 2020.
  - ✓ European Committee for Education in Anesthesia Courses, 16th edition, Module 5, 9-11 December 2021.
  - ✓ 5th International Conference on Nanotechnologies and Biomedical Engineering, November 3-5 2021.
  - ✓ The 12th International Conference on Electronics, Communications and Computing, 20-21 October 2022
  - ✓ The Annual Scientific Conference of the SUMPh. Research in biomedicine and health: quality, excellence and performance, Chisinau, 19-21 October 2022.
- The use of BDM for time series analysis and representation of medical data was presented at the AUTOMATA 2020 conference, Stockholm, Sweden, August 10-12, 2020.
- Patent activity: research results formalized in two invention certificates (No.18 of 12.12.2022 and No.19 of 13.12.2022) and implementing acts.

Publications on the thesis topic. Sixteen scientific papers are published on the thesis topic and are paraphrased in the thesis content, thus showing the theoretical importance and application value of the work. Three publications - in journals/collections with SCOPUS impact factor, publications as a sole author - 6, first author - 7.

Summary of thesis compartments. The thesis is set out in 107 pages of basic text, structured into 5 chapters, conclusions, and practical recommendations, followed by a list of 274 bibliographical sources and 12 appendices. The iconographic material includes 52 figures and 32 tables.

Keywords: sepsis, model, artificial intelligence, machine learning, algorithmic complexity, block decomposition method, decision support systems, prediction systems

THESIS CONTENT

The **Introduction** section reflects and argues the topicality of the topic and the opportunity of the present research based on data published in the literature. The aim and objectives of the research are formulated, the field of research, the scientific novelty of the results obtained, and the scientific problem solved are presented. The theoretical significance, the applicative value of the work, and the possibilities of implementing the scientific results obtained are described.

1. APPLICATION OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN ANAESTHESIOLOGY-INTENSIVE CARE AND SEPSIS MANAGEMENT

1.1 Machine learning/artificial intelligence technologies used in anesthesiology-intensive care

Commercial applications of artificial intelligence (AI) and machine learning have recently made impressive strides, particularly in fields like picture identification, natural language processing, language translation, textual analysis, and self-learning, according to a recent article [11]. However, single-task applications where unsatisfactory outcomes and sporadic errors can be tolerated have seen the best performance from the outlined solutions. Intensive care and anesthesia (AIC) are two distinct medical specialties. Instead of any cognitive act, it adds requirements for high reliability, time restrictions on interpretation, and physical action/response. Evidence-based medicine and "Big Data" have only recently become mainstream. Contrarily, AIC doctors have long relied on patient-specific streams of quantifiable data to provide care, and improvements in monitoring and the depth of this data have been the driving force for notable advancements in patient safety in the field [12]. These professionals frequently work in "edge" situations involving the "cause-and-effect" relationship, when choices can frequently not be postponed and mistakes in judgment are frequently unavoidable. According to a review paper published in 2020 [13], the following AIC specialties employ AI technologies the most frequently: One can add mortality prediction, mechanical ventilation decision assistance, and sepsis prediction to the following: (1) anesthesia depth monitoring; (2) anesthesia control; (3) event and risk prediction; (4) ultrasonography guiding; (5) pain management; and (6) operating room logistics [14].

1.2 The concept of sepsis. Evolution of the concept and its modern content

Over the years the approach to sepsis has revolutionized starting from the concept of Sepsis-1 (1990) to Sepsis-2 (2008) and nowadays sepsis is diagnosed according to the principles of Sepsis-3 (2016), proposed in 2016, where sepsis has been defined as a dysregulated host response to infection that is life-threatening. This concept is operationalized by the Sequential Organ Failure Assessment (SOFA), where clinical diagnostic criteria for sepsis include an acute increase of at least two points in the SOFA score combined with a confirmed or suspected infection [5]. The research in this thesis uses the latest definition, "Sepsis-3".

1.3 Current situation in addressing the problem of sepsis with the use of intelligent technologies in the clinical management of sepsis

Literature data show that early detection or prediction of sepsis can lead to a decrease in antibiotic administration time [15], and early intervention in turn has been shown to reduce mortality in this group of patients [16]. For example, the use of a machine learning system (MLS) in a study [3] was associated with a 39.5% reduction in hospital mortality ($p<0.001$), a 32.3% reduction in length of stay ($p<0.001$), and a 22.7% reduction in 30-day readmissions ($p<0.001$). Although MLSs represent fairly new additions to the field of clinical sepsis management, machine learning algorithms have the potential to significantly improve patient outcomes by providing early warning of impending sepsis onset. SIAs with sepsis prediction abilities can also serve to enable clinicians to have confidence in the diagnosis of sepsis in a variety of ambiguous cases, including cases where positive culture results are not available [17] and in cases of atypical clinical presentation, especially among older patients, which comprise a majority of sepsis cases [18]. Therefore, machine learning-based decision support systems are an important area of investigation for sepsis research [8, 19].

2. RESEARCH MATERIALS AND METHODS

2.1. General research design

The research was approved by the Research Ethics Committee of N.Testemitanu State University of Medicine and Pharmacy on 18.03.2022. A retrospective multi-stage study was conducted based on an initial dataset describing 40366 patients, including 2932 sepsis patients from a public database [20]. The initial data contained missing values. After exploratory analysis, 5039 patients were selected for system creation, of which 1703 - with sepsis, with a total of 30635 hourly samples. Selection criteria included (a) the presence of at least 7 hourly observation windows (4 hours - the forecast horizon, 3 hours - for assessing the dynamics of parameters describing the patient) and (b) the maximum possible density of data (for non-sepsis patients - the presence of 7 consecutive values for each of the parameters of interest). Data from septic patients with missing values were reconstructed according to an algorithm developed in the current study. These data were used to create an early sepsis forecasting system.

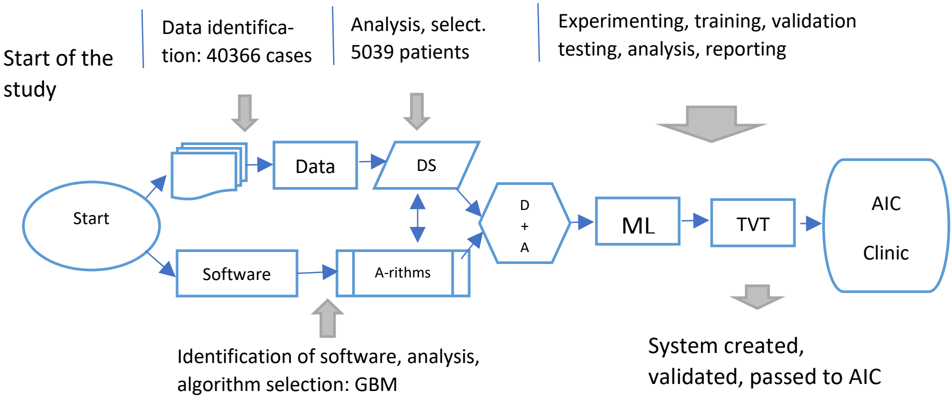

Figure 2.1. **Study design**

The development of the system followed the traditional steps in creating a machine learning system: training, cross-validation, and testing on new data, which were not used in the previous steps. Research results based on the performance of the system were analyzed, and conclusions and practical recommendations were described. The created system, in the form of an application, will be sent to the AIC Clinic of the Institute of Emergency Medicine, Chisinau for future calibration and development.

2.2 Description of clinical material

The public access data from "Early Prediction of Sepsis from Clinical Data: The PhysioNet Computing in Cardiology Challenge 2019" [20] is represented in the dataset utilized for the thesis. The data may be used after the end conference date, September 8–11, 2019 (https://physionet.org/content/challenge-2019/1.0.0), in accordance with the event rules. The Beth Israel Deaconess Medical Center (Set A) and Emory University Hospital (Set B) hospital systems in the United States, which are geographically apart, provided the data. With institutional review board consent, these data were gathered over the previous ten years, de-identified, and classified using Sepsis-3 clinical criteria [5]. The information is made up of summaries of the patient's hourly vital signs, laboratory results, and static patient descriptions. The data specifically comprise 40 clinical factors, including 26 laboratory variables, 6 demographic and logistic variables, and 9 vital signs and life support (FiO2) variables. These data total over 10.4 million data points (physiological and laboratory "non-missing" variables) and over 1.5 million hourly windows. Before formal analysis, the data that had been taken out of the individual hospitals' computer systems underwent a number of pre-processing processes. To make model building and testing easier, all patient parameters were grouped into hourly bins. For instance, the average heart rate measurement was calculated from several heart rate measurements taken within an hourly time window.

Table 2.1. **General characterization of the datasets**

Parameter	Set A	Set B
Total number of patients	20336	20000
Number of patients with sepsis	1790	1142
Age, years (mean, sd)	62.62 (16.24)	60,65 (16,67)
Age % 18-50 51-60 61-70 71-80 >80	 21.44 18.54 22.41 22.67 14.94	 25,89 20.13 24.94 18.65 10.40
Gender breakdown, women/men	8502 (41,80%) 11834 (58,2%)	9268 (46,34%) 10732 (53,66%)
Length of observation (hours in ICU, mean, sd)	38.86 (22.31)	38.10 (23,28)
Clinical parameters monitored	• 8 vital signs and 1 life support parameter • 25 laboratory parameters • demographic and logistic parameters	
Prevalence of sepsis %	8,80	5,71
Number of observations/hourly windows	790215	741952
Number of data entries/cells with 'non-missing' values	10296016	9831918

***Note**: The 'vital signs and life support' parameter group includes: heart rate (HR), peripheral blood O2 saturation (SpO_2), temperature (Temp), systolic blood pressure (SBP), mean blood pressure (MBP), diastolic blood pressure (DBP), respiratory rate (Resp), capnometry (EtCO2), inspired O_2 fraction ($EtCO_2$); laboratory parameters: base excess (BE), bicarbonate (HCO_3), pH, part. press. of CO_2 in art. blood ($PaCO_2$), O_2 saturation of art. blood (SaO_2), liver enzymes, serum ions, WBC, etc.; demographic and logistic parameters: age, gender, length of hospitalization, etc., sepsis label (SepsisLabel); sd - standard deviation*

2.3. Methodological issues important for investigation, data processing, and analysis of results

Initially, the data are in ".psv" (pipe-separated value) format. For further processing, they have been converted into ".csv" comma-separated value) format using the "rio" (R) package. The initial data are investigated using simple/standard statistical methods: mean, standard deviation, median, and percentage ratio. The analysis of the data selected for the unbiased creation of the system also includes the estimation of their distribution, which is also used to develop the algorithm for the reconstruction of missing data/values. The logic and methods of data investigation and analysis are influenced by the subsequent steps and the final goal - the development of the sepsis classification/prediction system, which in this traditional field is called "Exploratory Data Analysis". The parameters/metrics that are used to evaluate the performance of the end product - of the developed system are also traditional and include:

- Sensitivity (TPR = TP/(TP+FN), where TPR - true positive rate, TP - true positive (results), FN - false negative
- Specificity (TNR = TN/(TN+FP), where TNR - true negative rate, TN - true negative, FP - false positive
- Diagnostic accuracy = (TP+TN)/(TP+TN+FP+FN)
- ROC analysis (Receiver operating characteristic - graphical representation of the ROC curve, as the ratio of FPR (X-axis) to TPR (Y-axis), where FPR, false positive rate = FP/(FP+TN)
- Confusion matrix (showing for each class the number of correctly and incorrectly predicted cases, the error, and the error rate)
- Variable importance analysis (identification and graphical presentation of the most important predictors for the model)
- Rate and trajectory of error reduction (a graphical representation of error dynamics during model learning/training)
- Other (accuracy, negative predictive rate, false positive and negative rates, positive and negative likelihood ratios, PR analysis, odds ratio, etc.).

A separate place is given to algorithmic complexity estimation methods, which come from the field of algorithmic information dynamics. They are described in more detail in Chapter 3.

3. DATA PROCESSING STEPS

3.1. Recovering data by solving the missing values problem

Preparing the data to be delivered to a machine learning model is a crucial step, on which the performance of the model may ultimately depend. There are currently no pre-defined rules for this process. Usually, the most successful form is established by trial and error. In the present case, it would seem that the pathophysiological relevance of a particular parameter would be of some importance, but this approach has proved less productive. Several variants have been tried. Initially, parameters that would be of importance in the clinical diagnosis of sepsis were used, in particular those used for SOFA. This approach gave minor results (AUC = 0.61). Then the full set of clinical parameters (36 parameters) was tested, which conditioned a far-from-ideal model performance (AUC = 0.65). The best performers were 6 parameters and their dynamics in the form of differences in hourly values over 3 hours. The respective parameters were subsequently represented as matrices, for which the algorithmic complexity value was calculated and the difference was used as untransformed results. The final selection of the data format used for machine learning (final model training) was influenced by the principle that "along with the current value of a physiological parameter, its dynamics over time is often of major importance". It was also aimed to use as predictors some indices/parameters that are easily/standardly recorded, including in the ICU wards in the Republic of Moldova.

As previously mentioned, in the initial data there are missing values-l (Table 3.1). To avoid the consequences of the "garbage in, garbage out" situation the data were reconstructed according to a procedure described below.

Table 3.1. **Original data appearance of a patient (p000009)**

HR	SpO_2	Temp	SBP	MBP	DBP	Resp	$EtCO_2$	…	Age	Gender	…
NA	NA	NA	NA	NA	NA	NA	NA	…	27.92	1	…
117	99	NA	116	97	81	20	NA	…	27.92	1	…
NA	NA	NA	NA	NA	NA	NA	NA	…	27.92	1	…
NA	NA	NA	NA	NA	NA	NA	NA	…	27.92	1	…
NA	NA	NA	NA	NA	NA	NA	NA	…	27.92	1	…
NA	NA	NA	NA	NA	NA	NA	NA	…	27.92	1	…
120	100	36	118	84	64	30	NA	…	27.92	1	…
…	…	…	…	…	…	…	…	…	…	…	…

***Note**: NA (not available) – missing values. The omitted columns and rows are marked with "...". This is a fragment of the data set of the respective patient that includes the first 7 time windows (vertically). The full set has 258 rows (i.e., time windows/observations) and 40 columns (monitored parameters).*

In most of the similar works dealing with missing data [3, 21], their value is computed by the "last observation carried forward" (LOCF) method, which involves sorting the dataset (e.g., values of a parameter/column - see above) and creating an ordered set. Then the first missing value is determined and the previous value is used to "restore" it. The process is repeated for the next missing value until all missing values are "restored". Considering physiological processes as continuous phenomena (usually represented by time series - this is also the case for the data under study) it can be assumed that the numerical strings created from parameter values representing these processes also inherit the property of "continuity" (against the discrete variant). The LOCF method somehow ignores the "continuity principle". That is why it was decided to include the "continuity principle" by proposing the following procedure: (1) each column is evaluated for missing data (NA); (2) the first and the last value (per row) is "restored" according to the value in the most appropriate cell; (3) finally interpolation is performed, with the calculation of values between two present values respecting the trend (increasing or decreasing). The function {na.approx( )} ("zoo" package in R) is used for the interpolation. In case all values are missing, the following procedure was followed: (1) from the integral set, the values present for each parameter were extracted separately by class (e.g., septic patients in set A); (2) the number of "non-missing" values (n) was determined and the mean and standard deviation (ds) for these values were calculated. Using the function {rnorm()} in R, and "n", "mean" and "ds" as arguments to it, Gaussian distributions were generated for each of the 6 parameters. These can be viewed and compared with the initial/"true" distributions of the respective data.

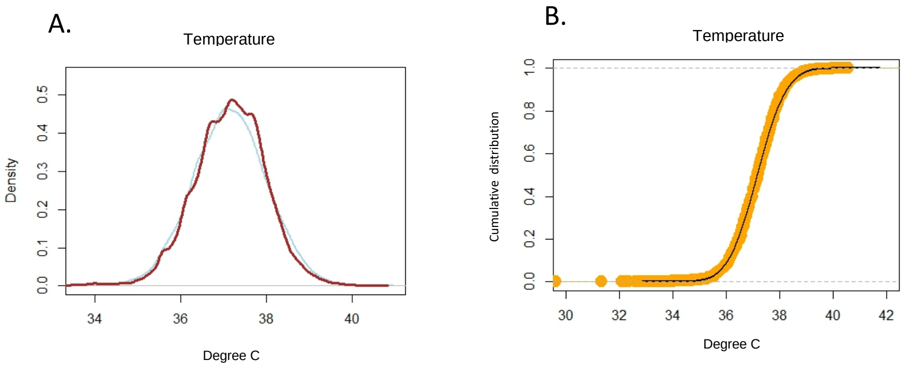

Figure 3.1. **Distribution of actual data describing the parameter "Temperature" vs. generated data. A - Gaussian distribution; B - cumulative distribution: with purple and orange - "real" data, with blue and black - generated data.**

In the case of more obvious differences between the two density/distribution curves (actual vs. generated data), it is possible to manually adjust the generated data curve so that it is as close as possible to the actual one. For this adjustment, one solution may be to use the "sn" package in R. In the case of the current research, there was no such need. Thus, the algorithm scheme used for data reconstruction is as follows:

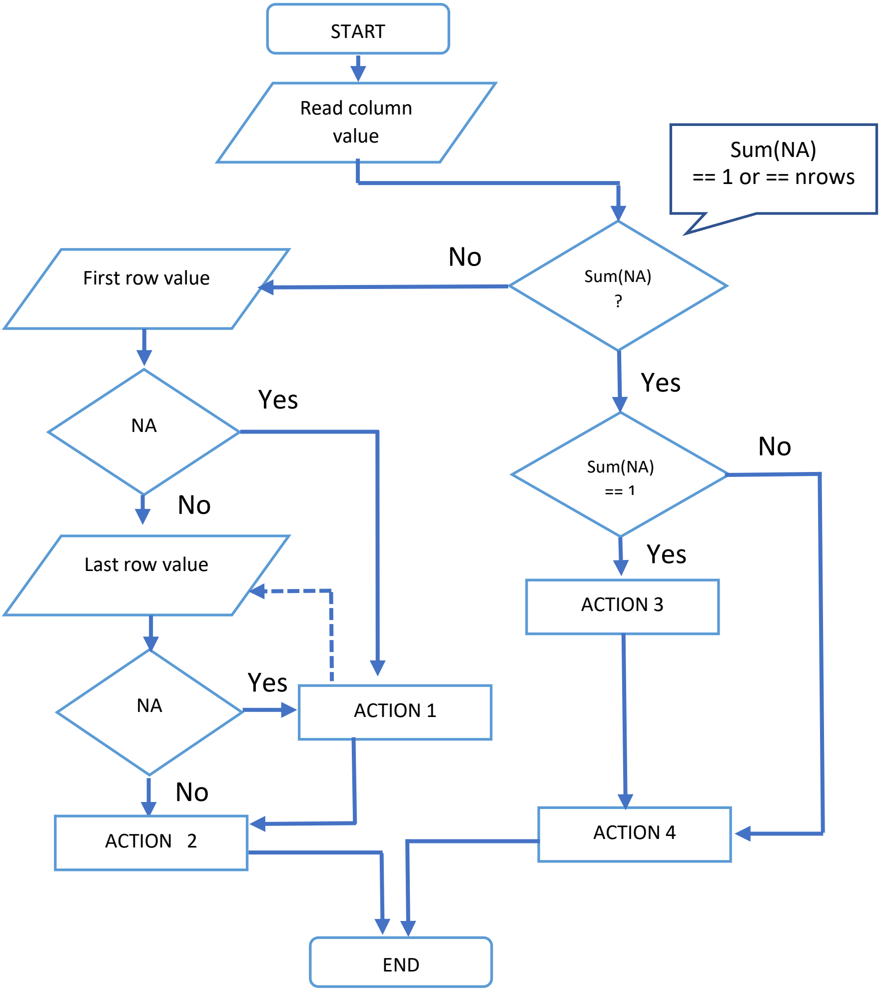

Figure. 3.2. **The algorithm for missing value recovery**

Note: *Action 1 - replacing the missing value with the nearest value (location); Action 2 - interpolating missing values by the procedure described in the text; Action 3 - deleting the value; Action 4 - filling cells with values from the generated Gaussian distribution.*

After applying the reconstruction procedure , the data looks like this:

Table 3.2. **Data appearance of the same patient (p000009) after recovery**

HR	SpO_2	Temp	SBP	DBP	Resp
117	99	36	116	81	20
117	99	36	116	81	20
118	99	36	116	78	22
119	99	36	117	74	24
119	100	36	117	71	26
120	100	36	118	67	28
120	100	36	118	64	30

Note: *Only parameter values (in columns) subject to reconstruction according to the procedure described in the text are shown. The original data set is described above in Table 3.1.*

3.2 Algorithmic Complexity as a metric and data processing method

Algorithmic complexity (AC) is approached in current research through the lens of algorithmic information dynamics (AID). AID [22] is an emerging field of complexity science based on algorithmic information theory (AIT), which encompasses literature based on the Kolmogorov - Chaitin concept of complexity and related concepts such as algorithmic probability, compression, optimal inference, universal distribution, and others. Central to TIA is the definition of algorithmic complexity (Kolmogorov - Chaitin or program size complexity - Kolmogorov, 1965; Chaitin, 1969) [23]:

$$K_T\ (s) = \{|p|,\ T_{(p)} = s\}, \tag{3.1}$$

i.e., the length of the shortest program p that generates the string s and runs on a universal Turing machine. AID tends to identify solutions to fundamental questions about causality: why a given set of circumstances leads to a given outcome. In this aspect, AC differs in essence from traditional statistics.

As an applied science, AID is a new kind of discrete computing based on "computer science" that seeks to research causality by producing mechanical models to assist in identifying the fundamental concepts underlying physical occurrences, hence constructing the next iteration of machine learning [22].

The online algorithmic complexity calculator (OACC) [24], which offers the possibility of CA and algorithmic probability (AP) estimation for short and long numerical strings and two-dimensional structures better than any other conventional tool, is a special tool in the AID toolkit for providing accurate estimates for non-computable functions. Generally speaking, traditional methods are not intended to capture any algorithmic meaning beyond straightforward statistical patterns.

OACC employs the coding theorem method-based BDM approach [22,24], which is based on algorithmic probability [22]:

$$BDM = \sum_{i=1}^{n} CTM\,(block_i) + \log_2(|block_i|). \tag{3.2}$$

OACC is employed for these calculations in the current paper and is offered as an online version [25] as well as standalone packages in R and a number of other programming languages (Matlab, C, Wolfram, etc.). The employment of OACC in the given form is insufficiently productive because the intensity of these computations is rather high (due to the volume of data and the very nature of the calculations). The core of the corresponding package [25], which may be applied to two-dimensional structures (matrices), was taken out in order to accelerate these computations. This core was then included into the software handling the data flow processing in the current study [26]. Figure 3.3 provides an explanation of how the BDM value is determined.

A. Original data

82.5	84.5	80
151	147.5	146
58.5	58.5	56

88	83	74.5
95.5	100.25	90.25
51	52	50.5

B. Binarization

0	1	0
0	1	0
0	1	0

1	1	0
1	1	0
0	1	0

C. AC/BDM

16.1617 byts

17.4911 byts

Figure. 3.3. **Data format at 3 successive processing stages**

Note: *A- initial format, B - binarization (by row threshold values), C- calculated BDM value for each matrix. The initial data represent values of the circulatory parameters (HR, SBP, DBP), structured in matrices of size 3 x 3 (per row)*

3.3. Data processing with the creation of features to be used for machine learning

In the current research, based on the nature of the primary data, it was decided to use the variables with the fewest missing values. Since one of the methods of data transformation is BDM (on 2-dimensional tensors/matrices - the selected optimal variant - 2 such matrices 3 x 3) and an optimal time interval for prediction estimation - 3 hours, it was necessary to select 6 such variables (three for each matrix x 3 hours). Thus, the final selection of physiological variables/parameters includes heart rate (HR), blood oxygen saturation (SpO2), systolic (SBP) and diastolic blood pressure (DBP), body temperature (Temp), respiratory rate (Resp) – with a total of 6 parameters. The final length of the vectors used for the AI is 14 (items) plus the sample label (see Table 4.2).

4. MACHINE LEARNING. INTERMEDIATE RESULTS AND THEIR DISCUSSION

4.1 Machine learning phase

The process of creating a machine learning model is more or less standard. The following figure describes the components and steps of the process (reproduced with modifications from [27]).

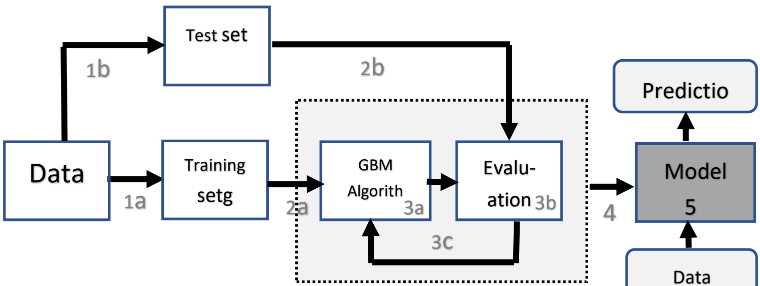

Figure 4.1. **Stages and main elements in the creation of a machine learning system: 1a, 1b, ..., 5 - stages of system creation**

The data processing steps are aligned with the process of creating an ALS and are schematically illustrated in Figure 4.2.

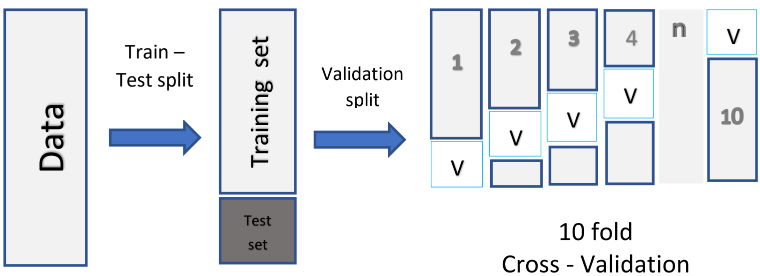

Figure 4.2. **Stages of the data processing flow**

***Note**: Note: Initially the data is divided into the training set (which will later participate in learning and cross-validation) and the test set.*

An important point is the separation at the beginning (after preliminary data processing) of the set for testing, which in the case of the current study (test set A) consists of 3052 samples (1063 sepsis and 1989 non-sepsis), which did not participate in training and validation.

4.2 Development, validation, and testing of the developed AI system

As a guiding point for the creation of the sepsis prediction system in the thesis was the InSight system, reported in the literature as one of the best performing in the field [3]. Using the rather general information from [21], this system was replicated in R programming language and tested its performance on the dataset of the current study. Results were obtained that correspond to those reported by the authors. Further, these results served as milestones in the creation and evaluation of our system. Table 4.1 shows the comparative performance metrics of these two systems.

Table 4.1. **Comparative metrics of two sepsis prediction systems (InSight vs. system created in the current research)**

Metrics	InSight	Created system
AUC	0.914 (95% IÎ: 0.902 – 0.926)	0.929 (95% IÎ: 0.919 – 0.939)
Diagnostic accuracy	0.925	0.940
Sensitivity	0.888	0.916
Specificity	0.941	0.942

Note: *AUC – area under the ROC curve.*

The performance of these two systems is quite close. A higher sensitivity in the system created would indicate a more balanced system that detects sepsis cases slightly better than the InSight system. The created system differs from InSight in the initial data used for machine learning, the data processing methods, and the format of the final data, which is ultimately delivered for learning and validation. For example, the final data in the created system are vectors of length 14 (numeric strings consisting of 14 numbers/values) compared to 30 in InSight, i.e., a more than 2-fold reduction in dimensionality. The layout of the final data used by the system created is illustrated in Table 4.2.

The final data organized in the described format was ultimately delivered to a model based on the Gradient Boosting Machine (GBM) algorithm, which performed prediction by classification (sepsis vs. non-sepsis). Optimal hyperparameters (learning rate, number of trees, depth, no. of branches) were initially identified by automated machine learning

(AutoML) using the H2O platform. These hyperparameters were then justified in the context of the present data by a coarse grid search with multiple 10 folds cross-validation).

Table 4.2. **Format of data passed to ML algorithm**

CAT	BDM-MR	BDM-CIRC	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12
0	16.956	17.491	-4.5	1.5	0	-1.5	-2.5	-3	2	-1.5	0.28	-3.5	0	0.5
0	18.328	14.815	-8.5	0	0.5	-10	-1.5	0	-5	0	0.1	4.75	1	0
1	18.456	15.942	1	0	0.2	28	10	9	4	-1	0.1	-13	-3	0
0	17.855	16.900	6	0	0.2	-15	-6.5	-0.5	-16	0	0.2	21	8.5	0.5
1	16.956	16.956	3	-1	0	-5	-2	0	2	-1	0.1	4	1	0

Note: *These are the final data describing 5 patients: 2 with sepsis and 3 - with other pathologies. "CAT"- is the label of each subset/patient (1 - sepsis, 0 - other pathologies); "BDM-MR/BDM-CIRC" - represents the BDM values for the respiratory-metabolic and circulatory data groups; "V1 - V12" represent the dynamics of the 6 physiological parameters (HR, SpO2, Temp, SBP, DBP, Resp) over 3 hours.*

According to traditional statistical terminology, the system created is a multivariate prediction model (with 14 predictors) based on binomial classification. According to PROBAST (2019) principles [10] - it is a prognostic prediction model (where as predictors serve 6 physiological parameters and their dynamics, generating 14 features), and according to TRIPOD (2015) guidelines [9] - the current research is of "Type 3" which involves developing a predictive model with the use of a dataset and evaluating its performance using separate data, the latter is also called "external validation".

The system was developed according to the principles and following the steps described above. It was validated (10 folds cross-validation) and tested on the test set. The performance of the created system is illustrated in the following figures and tables:

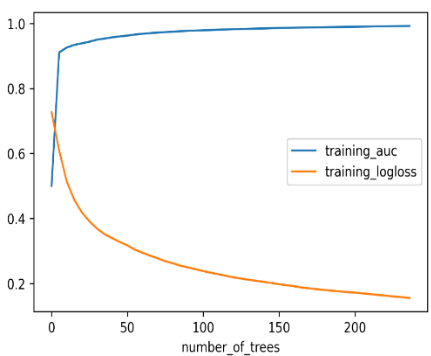

Figure 4.3. **Logloss during taining (logloss with orange) - Y-axis logloss value. X-axis - number of trees. GBM algorithm.**

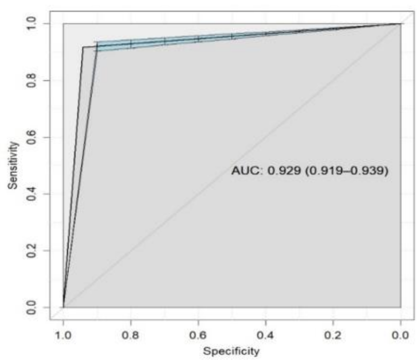

Figure 4.4. **ROC curve on the test set (A): on X-axis - false positive rate (FPR), on Y-axis - true positive rate (TPR).**

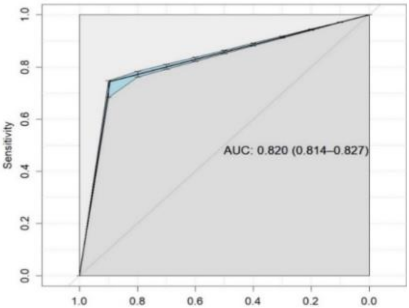

Figure 4.5. **ROC curve on the test set (B): on X-axis - false positive rate (FPR), on Y-axis - true positive rate (TPR).**

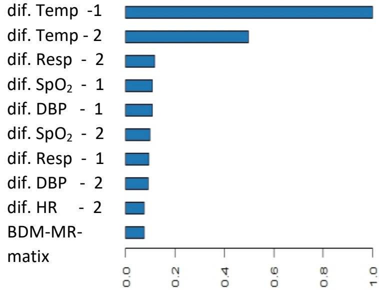

Figura 4.6. **Variable importance determined during learning and cross-validation. X-axis – scara importance (max = 1,0)**

Table. 4.3. **Summary of the performance of the GBM-based system**

Performance metrics	**Cross-validation**	**Test set -1/ internal validation**	**Test set-2/ external validation**
Recall/ True Positive Rate (TPR)	0.885 (0.016)	0.916 (0.898 – 0.932)	0.713 (0.701 – 0.725)
Specificity/ True Negative Rate (TNR)	0.951 (0.009)	0.942 (0.931 – 0.952)	0.913 (0.908 – 0.918)
Diagnostic accuracy	0.928 (0.006)	0.940 (0.931 – 0.948)	0.902 (0.897 – 0.906)
Positive predictive value (PPV)	0.915 (0.91-0.92)	0.605 (0.561 – 0.646)	0.332 (0.320 – 0.344)
Negative predictive value (NPV)	0.931 (0.93-0.93)	0.992 (0.987 – 0.993)	0.981 (0.981 – 0.982)
False Positive Rate (FPR)	0.048	0.058	0.087
False Negative Rate (FNR)	0.121	0.084	0.287
Positive likelihood ratio (LR+)	18.31 (17.1-19.6)	15.85 (13.26 – 18.94)	8.20 (7.76 – 8.66)
Negative likelihood ratio (LR-)	0.13 (0.12-0.13)	0.09 (0.07 – 0.11)	0.31 (0.30 – 0.33)
Diagnostic odds ratio (DOR)	140.85	176.11	26.45
ROC analysis (AUC)	0.970 (0.004)	0.929 (95% IÎ:0.919 – 0.939)	0.813 (95% IÎ: 0.807 - 0,820)
PR analysis (prAUC)	0.959 (0.004)	0.864	0.647

4.3. Embedding of the developed system into an application for clinical use.

For clinical use of the system, it has been developed as a software application [28], which is convenient to use by the ATI physician. Figures 4.8 and 4.9 illustrate the graphical interface of this application.

The convenience of the application for the user was aimed at entering data at one-hour intervals (which can be downloaded as a ".csv" file if necessary), and in case of previously accumulated data - it is possible to import the data with its visualization.

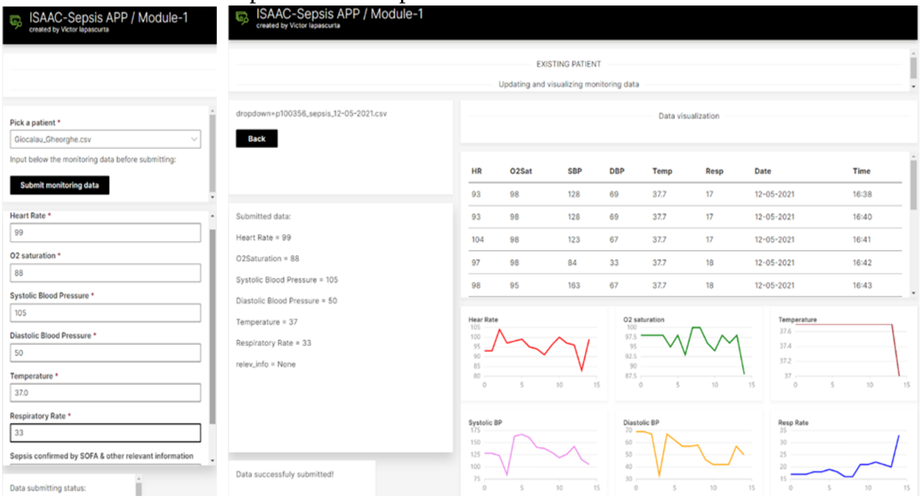

Figure 4.7 **Graphical interface of the application for early prediction of sepsis: imputing and visualizing the data**

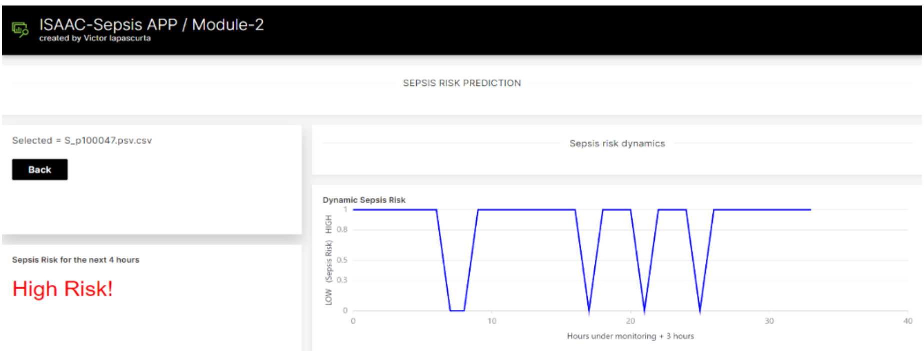

Figure. 4.8. **Graphical interface of the application for early prediction of sepsis: obtaining the prediction result and visualization of risk dynamics**

This application can be used in the following way: (1) data describing the case/patient (the 6 parameters) is entered - it can be entered hourly or in tabular form (when data for several hourly intervals); (2) data can be visualized graphically in the form of time series. With the accumulation of 3 hourly windows the risk of sepsis in the patient concerned is automatically determined ("High risk" vs "Low risk"); (3) as time progresses after each hour the patient's condition (in terms of risk of sepsis) is reassessed with the degree of risk displayed for each assessment.

4.4. Use of the application for continuous sepsis risk prediction

To evaluate the performance of the developed system in terms of tracking the dynamics of sepsis risk over time, 10 patients from the test set with sepsis (confirmed by traditional methods) and from the set with other pathologies were selected by random sampling.

Table 4.4. **Continuous prediction of the sepsis risk**

Pacient ID	hour 1	hour 2	hour 3	hour 4	hour 5	hour 6	hour 7	hour 8	hour 9
Patients with sepsis									
p002706	1	1	1	1	1	1	1	1	1
p002808	1	1	1	1	1	1	1	1	0
p004197	1	1	1	1	1	1	1	1	1
p011056	1	1	1	1	1	1	1	1	1
p011676	1	1	1	1	1	1	1	1	1
p013361	1	1	1	1	1	1	1	1	1
p014518	0	0	1	0	1	1	1	1	1
p015892	1	0	0	1	1	1	0	0	0
p016487	1	1	1	1	1	1	1	1	1
p020538	1	1	1	1	1	1	1	1	1
Non-sepsis patients									
p008368	0	0	0	0	0	0	1	0	0
p012801	0	0	0	0	0	0	0	0	0
p013050	0	0	0	0	0	0	0	0	0
p013661	0	0	0	0	1	0	0	0	0
p014981	1	0	0	0	0	0	0	0	0
p015223	0	0	0	0	0	0	0	0	0
p015886	0	0	0	0	0	0	0	0	0
p017073	0	0	0	0	0	0	0	0	0
p017567	0	0	1	0	0	0	0	0	0
p020600	0	0	0	0	0	0	0	0	0

***Note**: Prediction result: 1 - sepsis, 0 - non-sepsis. Sepsis in sepsis - diagnosed by traditional methods at hour 7.*

For uniformity for each selected case, an observation period of 11 hours was drawn. For patients with sepsis, this period includes 8 hours until clinical confirmation of sepsis and 3 hours - after. For non-septic patients, 11 hours were extracted without missing values of parameters of interest. The result of the application of the system is illustrated in Tab. 4.4.

5. EXPLAINABILITY OF ML MODELS. DISCUSSION AND FUTURE RESEARCH DIRECTIONS

5.1. Summary of the research work carried out. Discussion of results. Limitations

Previously the performance of the model created in this study was compared with the performance of the InSight system (see Tab. 4.1). A credible source for comparison may be the "International Guidelines for the Management of Sepsis and Septic Shock" of November 2021 [29], where the potential role of machine learning, which "may improve the performance of screening tools", is first mentioned, based on a meta-analysis of 42623 patients from seven studies for the prediction of hospital sepsis, reporting an area under the ROC curve of 0.89 (95% IÎ, 0.86-0.92); sensitivity - 81%; (95% CI, 80-81) and specificity - 72% (95% CI, 72-72). The performance of the system created in the current study is superior by each of the performance metrics listed.

One of the main limitations of the prediction system created in the current research is related to a topic that is currently under debate. A recent meta-analysis [30] indicated that the majority of severely ill COVID-19 patients (78%) met the Sepsis 3.0 criteria for sepsis/septic shock with acute respiratory distress syndrome (ARDS) as the most common organ dysfunction (88%). The data upon which the system was created includes only cases of bacterial sepsis.

Other foreseeable limitations at this stage may be related to the difference between the data from the current study and the ATI ward where the system could be used in the future, caused by a different spectrum of septic patients, and different monitoring and treatment techniques. A potential solution, in this case, would be to create a new system using the experience gained.

5.2. Explainability aspects of AI models

The explainability of AI models is of major importance for understanding how the model works, which influences its adoption by the medical community. This area is called XAI (eXplainable Artificial Intelligence) and although it is in its infancy it is developing rapidly. An example of an approach is the estimation of the importance of variables for prediction, which was presented earlier (Fig. 4.6). For similar purposes PD (Partial

Dependence), ICE (Individual Conditional Expectations), LIME (Local Interpretable Model-agnostic Explanations), etc. methods are used, which are described in the basic text of the thesis.

A more complex approach that includes elements of the methods described is the SHAP (SHapley Additive exPlanations) method. The SHAP diagram for a GBM model in the current study is shown in Fig. 5.1. It combines the importance of features (position of the feature on the Y-axis; the higher - the more important) with the effects of features (impact on prediction - on the X-axis; to the right of the vertical line "0" - in favor of sepsis, and to the left - against sepsis, and the further away from "0" - the greater the impact). The color represents the normalized value of the feature from low to high. Thus, in the case of temperature, a difference in magnitude as large (e.g., when body temperature is essentially and rapidly increasing) translates to a low risk of sepsis, a fact known (and apparently self-evident) to the clinician. Going the other way round, modest temperature dynamics could denote the situation when the body's defense forces of the septic patient are exhausted, not being able to provide an adequate response (to the pathogen), expressed in this case by the "attenuated" pyretic reaction.

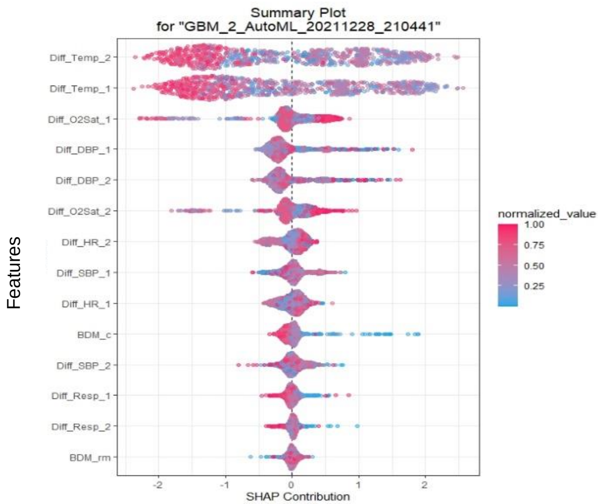

Figure 5.1. **Summary SHAP diagram for a GBM models**

5.3 Transferability of AI systems. Future research directions

A separate issue discussed in the literature [3, 21] is the transferability of AI systems, i.e., the use of a system built on data from one hospital to another. Usually, in this case, the performance is lower. In the current study the system created based on set "A" was tested on data from set "B" with a performance of 0.813 (95% IÎ: 0.807 - 0.820, by AUC ROC). Solutions to the performance problem, in this case, are proposed at least two: (a) creation of a new model based on new data, using knowledge gained from developing similar models in other studies (i.e., replication of a study already performed on new data); (b) transfer learning, where an already created model is additionally trained on new data.

Possible future research directions would include: (a) calibrating the sepsis prediction system under clinical conditions in the Republic of Moldova; (b) improving the performance of the system by transfer learning on local data; (c) studying the interaction of the ATI specialist with the system under clinical conditions.

CONCLUSIONS

1. The phenomenon of the penetration of AI technologies in ITA is relatively new, but of a scale and with prospects that are difficult to predict. The aspects of the intensivist's work that can benefit from these technologies are numerous, starting from the diagnosis/monitoring/prediction of different facets of the critical condition. According to the literature, the most commonly used AI technology at this stage is supervised machine learning, which is found in over 90% of papers.
2. The algorithm developed and used in the current research for restoring continuous biomedical data with missing values received from sepsis patients provides plausible results, which is confirmed by the high performance of the prediction system. Based on the results obtained (insignificant change in the performance of the system when tested on data-restored with the algorithm vs. complete cases) it can be stated that if there is a bias introduced by the algorithm, it is minor.
3. The method of representing and processing data in the form of time series of physiological parameters in septic patients using the Kolmogorov - Chaitin algorithmic complexity metric, which was first used for this purpose in the current research, is a successful method of data processing for machine learning, which is demonstrated by the final results obtained.
4. The six physiological parameters, represented by the data - values over time of them, were sufficient to create a powerful prediction system using the processing methods described above. Of importance is also the fact that these parameters are routinely monitored in ATI, including in the Republic of Moldova.

5. High performance (higher than 92% vs 89% in the latest Guidelines for the Management of Sepsis and Septic Shock, 2021) of the created prediction system makes further research with calibration of the system in the target ICU in Moldova and the subsequent possibility of use in clinical practice for early prediction of an imminent risk of developing sepsis with timely initiation of treatment (primarily infection-related aspects) rational. This can have a positive impact on treatment outcomes while providing decision support for the intensivist.

RECOMMENDATIONS

1. When the research is dealing with missing data and especially in the case of time series data, which in ATI are ubiquitous, the use of the data reconstruction-restoration algorithm could facilitate the researcher's work and contribute to boosting his performance.
2. Algorithmic complexity as a possible metric of the processes running in the subject of study and the representation of the data obtained using it (e.g., using the block decomposition method - BDM) has a potential that is still little explored. Therefore, experimentation with these tools even in studies outside the field of AI is welcome. It could successfully complement and even outperform traditional statistical methods.
3. Even though the clinician's reasoning process probably differs essentially from computational logic and computational performance cannot be compared to that of a computer, drawing attention to the dynamics of the factors with the highest predictive value determined by the model (e.g. subtle dynamics of temperature, peripheral blood oxygen saturation - SpO_2, diastolic blood pressure, etc.) could strengthen the (subconscious) prognostic abilities of the human specialist in the clinical management of sepsis patients. In this context, the explicability of the models is also important.
4. When developing similar decision support systems for practical use in ATI it is advisable to use the TRIPOD guidelines and PROBAST principles, which are also applicable for non-AI studies, in line with clinical reasoning and good medical practices.
5. The number of technologies and devices based on machine learning and AI will increase and, in this context, a correct attitude of the practitioner towards these devices is necessary. A successful one seems to be their acceptance as assistants, and not - rivals, all the more so as the final decision maker and implementer, at least at the moment and in the foreseeable future, remains the human specialist.

BIBLIOGRAPHY

1. Strandberg, G. et al. Mortality after Severe Sepsis and Septic Shock in Swedish Intensive Care Units 2008-2016 - A nationwide observational study. In: *Acta Anaesthesiol Scand.* 2020, nr (00), pp.1–9. ISSN 1399-6576.
2. Stevenson, E. et al. Two decades of mortality trends among patients with severe sepsis: a comparative meta-analysis. In: *Crit Care Med*. 2014, 1(42), pp.625–631. ISSN 0090-3493.
3. Burdick, H. et al. Effect of a sepsis prediction algorithm on patient mortality, length of stay and readmission: a prospective multicentre clinical outcomes evaluation of real-world patient data from US hospitals. In: *BMJ Health Care Inform.* 2020, nr 27(e100109). ISSN 2632-1009.
4. Singer, M. et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). In: *JAMA.* 2016, nr 1(315), pp.801–810. ISSN 0098-7484.
5. Shankar-Hari, M. et al. Developing a new definition and assessing new clinical criteria for septic shock: for the third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA.* 2016, nr 1(315), pp.775–787. ISSN 0098-7484.
6. Damiani, E. et al. Effect of performance improvement programs on compliance with sepsis bundles and mortality: a systematic review and meta-analysis of observational studies. In: *PLoS One.* 2015, nr 1(10), e0125827–24. ISSN 1932-6203.
7. Finkelsztein, E. et al. Comparison of qSOFA and SIRS for predicting adverse outcomes of patients with suspicion of sepsis outside the intensive care unit. In: *Crit Care.* 2017, nr 1(21), p.73. ISSN 1364-8535.
8. Delahanty, R. et al. Development and evaluation of a machine learning model for the early identification of patients at risk for sepsis. In: *Ann Emerg Med*. 2019, 1(73), pp.334–344. ISSN 0196-0644.
9. Collins, G. et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD Statement. In: *BMC Medicine*. 2015, nr 13(1). ISSN 1741-7015.
10. Wolff, R. et al. PROBAST: ATool to Assess the Risk of Bias and Applicability of Prediction Model. In: *Ann Intern Med*. 2019, nr 170, pp. 51-58. ISSN 0003-4819.
11. Connor, W. Artificial Intelligence and Machine Learning in Anesthesiology. In: *Anesthesiology.* 2019, nr 131, pp. 1346–1359. ISSN 1365-2044.
12. Eichhorn, J. et al. Standards for patient monitoring during anesthesia at Harvard Medical School. In: *JAMA.* 1986, nr 256, pp. 1017–1020. ISSN 0098-7484.
13. Hashimoto, D. Artificial Intelligence in Anesthesiology: Current Techniques, Clinical Applications, and Limitations. In: *Anesthesiology.* 2020, nr 132(2), pp.379-394. ISSN 1365-2044.
14. Lovejoy, C. et al. Artificial intelligence in the intensive care unit. In: *Critical Care*. 2019, 23(7), ISSN 1364-8535.
15. Shimabukuro, D. et al. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. In: *BMJ Open Respir Res.* 2017, nr 4(e000234). ISSN 2052-4439.

16. Yealy, D. et al. A randomized trial of protocol-based care for early septic shock. In: *N Engl J Med.* 2014, nr 370, pp.1683– 1693. ISSN 0028-4793.
17. Phua, J. et al. Characteristics and outcomes of culture-negative versus culture-positive severe sepsis. In: *Crit Care.* 2013, nr 17(R202). ISSN 1364-8535.
18. Lamantia, M. et al. Predictive value of initial triage vital signs for critically ill older adults. In: *West J Emerg Med.* 2013, nr 14, pp.453–460. ISSN 1936-9018.
19. Horng, S. et al. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. In: *PLoS One.* 2017, nr 12(e0174708). ISSN 1932-6203.
20. Reyna, M. et al. (2019) 'Early Prediction of Sepsis from Clinical Data: the PhysioNet Computing in cardiology Challenge 2019' (version 1.0.0), *PhysioNet*. [citat: 24.12. 2022]. Disponibil: https://doi.org/10.13026/v64v-d857
21. Mao, Q. et al. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. In: *BMJ Open.* 2018, nr 8(e017833). ISSN 2398-8703.
22. Zenil, H., Hernández-Orozco, S., Kiani, N. et al. A Decomposition Method for Global Evaluation of Shannon Entropy and Local Estimations of Algorithmic Complexity, In: *Entropy.* 2018, 20(8), p. 605. ISSN 1099-4300.
23. Soler-Toscano, F., Zenil, H., Delahaye, J.-P., Gauvrit, N. Calculating Kolmogorov Complexity from the Output Frequency Distributions of Small Turing Machines. In: *PLoS ONE.* 2014, 9(5) e96223. ISSN 1932-6203.
24. Zenil, H., Soler-Toscano, F., Dingle, K., Louis, A. Correlation of Automorphism Group Size and Topological Properties with ProgramSize Complexity Evaluations of Graphs and Complex Networks. In: *Physica A: Statistical Mechanics and Its Applications.* nr 404, 2014, pp. 341–358. ISSN 03784371.
25. Zenil, H., Soler-Toscano, F., Gauvrit, N. et al. The Online Algorithmic Complexity Calculator (OACC) v3.0, Algorithmic Dynamics Lab, Science for Life Laboratory (SciLifeLab), Unit of Computational Medicine, Center for Molecular Medicine at the Karolinska Institute in Stockholm, Sweden [software]. [citat: 24.12.2022]. Disponibil: www.algorithmicdynamics.net/software.html.
26. **Iapăscurtă**, **V**. A less traditional approach to biomedical signal processing for sepsis prediction, In: *5th International Conference on Nanotechnologies and Biomedical Engineering, November 3-5, 2021, Springer IFMBE Proceedings Series,* 2022, pp. 215-222, ISBN 978-3-030-92327-3.
27. Pant, A. Workflow of a Machine Learning project [online]. [citat 24.12.2022]. Disponibil: https://towardsdatascience.com/workflow-of-a-machine-learning-project-ec1dba419b94
28. **Iapăscurtă**, **V**., Belîi, A. Preclinical stage of building a machine learning system for sepsis prediction: a comparative study of four algorithms, In: *5th International Conference on Nanotechnologies and Biomedical Engineering, November 3-5, 2021, Springer IFMBE Proceedings Series,* 2022, pp. 448-455, ISBN 978-3-030-92327-3.
29. Evans, L. et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 2021. In: *Intensive Care Medicine*, ISSN 0342-4642.
30. Karakike, E. et al. Coronavirus disease 2019 as cause of viral sepsis: a systematic review and meta-analysis. *Crit Care Med.* 2021;49(12):2042–2057. ISSN 0090-3493.

THE LIST OF SCIENTIFIC PAPERS PUBLISHED ON THE THESIS TOPIC

- **Articles in scientific journals abroad**

✓ **articles in ISI journals, SCOPUS, and other international databases**

1. **Iapăscurtă**, **V**. Detection of Movement toward Randomness by Applying the Block Decomposition Method to a Simple Model of the Circulatory System. In: *Complex Systems Journal*. 2019, 28(3), pp. 59-77. ISSN 0891-2513. IF 0.79
2. **Iapăscurtă**, **V**. A less traditional approach to biomedical signal processing for sepsis prediction, In: *5th International Conference on Nanotechnologies and Biomedical Engineering, November 3-5, 2021, Springer IFMBE Proceedings Series,* 2022, pp. 215-222, ISBN 978-3-030-92327-3. IF 0.38
3. **Iapăscurtă**, **V**., Belîi, A. Preclinical stage of building a machine learning system for sepsis prediction: a comparative study of four algorithms, In: *5th International Conference on Nanotechnologies and Biomedical Engineering, November 3-5, 2021, Springer IFMBE Proceedings Series,* 2022, pp. 448-455, ISBN 978-3-030-92327-3. IF 0.38

- **Articles in accredited national scientific journals**

✓ **articles in category B journals**

4. **Iapăscurtă**, **V**. Managementul pacientului în stare critică actualități și perspective. In: *Curierul Medical*, Chișinău, 1/1997, pp.15-18. ISSN 1875-0666
5. Belîi, A., Crivorucica, V., Severin, Gh., Manastărschi, S., **Iapăscurtă**, **V**., Fortuna, E. Tratamentele medicamentoase, utilizate în cadrul terapiei intensive a pacienților cu SARS-CoV-2: revistă critică de literatură. In: *Moldovan Journal of Health Sciences.* 2020, 23(1), pp. 90-100. ISSN 2345-1467

- **Papers in scientific conferences:**

✓ **international held in the Republic of Moldova**

6. **Iapăscurtă V**., Ghereg V., Popescu A. Sisteme intelectuale în managementul stării critice. In: *Materialele conferinței practico-științifice moldo-americane "Parteneriat în sănătatea publică", Chișinău,* 1996, pp. V-5 – V-8.
7. **Iapăscurtă V**., Ghereg V. Alterarea transportului și utilizării oxigenului – axa conceptului modern al patogeniei și tratamentului stării de șoc (Revistă a literaturii). In: *Materialele conferinței practico-științifice moldo-americane "Parteneriat în sănătatea publică", secț. Anesteziologie-reanimatologie, Chișinău,* 1998, pp. 5 – 8.
8. **Iapăscurtă V.** Dealing with Missing Continuous Biomedical Data: a Data Recovery Method for Machine Learning purposes. *The 12th International Conference on Electronics, Communications and Computing, October 20-21, 2022, Chișinău*, pp. 1-5.

- **Abstracts/thesis in the proceedings of national and international scientific conferences**

9. **Iapăscurtă V**., Ghereg V., Revițchi V. Tehnologii informaționale în terapia intensivă. In: *Materialele conferinţei ştiinţifice anuale a colaboratorilor şi studenţilor, Universitatea de Stat de Medicină și Farmacie, Chișinău*, 1995, pp. 92-92.
10. Sevastianov E., Ghereg V., **Iapascurtă V.** Automatizarea procesului de diagnostic diferențial cu utilizarea indicelui leucocitar de intoxicație la pacienții cu procese toxico-septice. In: *Materialele conferinţei ştiinţifice anuale a colaboratorilor şi studenţilor, Universitatea de Stat de Medicină și Farmacie, Chișinău,*, 1995, pp. 110-110.
11. **Iapăscurtă V**., Ghereg V. Utilizarea bazelor de date medicale în anestezie-terapie intensivă. In: *Materialele conferinței practico-științifice moldo-americane "Parteneriat în sănătatea publică", Chișinău,* 1996, pp. V-10 – V-10.
12. Ghereg V., **Iapăscurtă V**., Popescu A. Analiza complexă și algoritmul clinic în anestezie-terapie intensivă. In: *Materialele conferinței practico-științifice moldo-americane "Parteneriat în sănătatea publică", Chișinău,* 1996, pp. V-10 – V-10.
13. **Iapăscurtă V.** Unele principii ale terapiei de perfuzie IV în cadrul tratamentului pacienților cu traumă craniocerebrală asociată cu șoc hemoragic. In: *Materialele conferinţei ştiinţifice anuale a colaboratorilor şi studenţilor, Universitatea de Stat de Medicină și Farmacie, Chișinău*, 1997, pp. 223-223.
14. **Iapăscurtă V**., Belîi A. Sepsisul: provocări curente și soluții noi în baza tehnologiilor moderne. O variantă de management clinic, *Conferința științifică anuală a USMF „N. Testemițanu". Cercetarea în biomedicină și sănătate: calitate, excelență și performanță, 20-22 octombrie, 2021, Chișinău, Abstract Book*, pp. 302-302. ISBN 978-9975-82-223-7.
15. **Iapăscurtă V**. Gestionarea valorilor lipsă în date biomedicale cu caracter continuu. *Conferința științifică anuală a USMF „N. Testemițanu". Cercetarea în biomedicină si sănătate: calitate, excelentă și performanță, 19-21 octombrie, 2022, Chișinău. MJHS*, 29(3), 2022, pp. 250-250. ISSN 2345-1467.
16. Monastîrschi S., **Iapăscurtă V**. Modele sistemice dinamice pentru anestezia clinică (pe exemplul propofolului). *Conferința științifică anuală a USMF „N. Testemițanu". Cercetarea în biomedicină si sănătate: calitate, excelentă și performanță, 19-21 octombrie, 2022, Chișinău. MJHS,* 29(3), 2022, pp. 314-314. ISSN 2345-1467.

- **Certificates for innovation**

17. **IAPĂSCURTĂ, V.**, BELÎI, A. Implementarea aplicației software ISAAC-Sepsis în baza învățării automate/inteligenței artificiale pentru prezicerea timpurie a sepsisului, Certificat nr. 5965, Ministerul Sănătății, USMF "Nicolae Testemițanu"; Certificat nr. 18, IMSP Instutul de medicină urgentă, 12.12.2022
18. **IAPĂSCURTĂ, V.** Implementarea aplicației software Sistem-expert acido-bazic ABB 1.1.1 pentru managementul pacienților în stare critică, Certificat nr. 5968, Ministerul Sănătății, USMF "Nicolae Testemițanu"; Certificat nr. 19, IMSP Instutul de medicină urgentă, 13.12.2022

ADNOTARE

Iapăscurtă Victor "Prezicerea timpurie a sepsisului cu ajutorul unei aplicații proprii elaborate în baza învățării automate (inteligență artificială)", teza de doctor în științe medicale, Chișinău, 2023

Teza este expusă pe 170 pagini și include: introducere, 5 capitole, concluzii, bibliografie din 278 de surse, 12 anexe, 52 figuri și 32 tabele. Rezultatele obținute sunt publicate în 16 lucrări științifice, dintre care 6 în calitate de singur autor, 7 – prim autor, 3 publicații în reviste cu factor de impact.

Cuvinte cheie: sepsis, model, inteligență artificiala, învățare automată, complexitate algoritmică, metoda de decompoziție în blocuri, sisteme de suport decizional, sisteme de prezicere.

Scopul studiului: Evaluarea fezabilității tehnologiilor IA în managementul pacientului critic din unitatea de terapie intensivă cu risc de a dezvolta sepsis, cu elaborarea unui sistem cu abilități discriminative (sepsis vs non-sepsis), care ar permite prezicerea timpurie a dezvoltării sepsisului.

Obiectivele studiului: 1. Evaluarea utilizării tehnologiilor de inteligența artificială și, în special, a învățării automate, ca una din tehnologiile de bază ale IA la etapa actuală, utilizate în anestezie-terapie intensivă. 2. Evaluarea utilizării SIA în managementul pacienților cu sepsis, în stare critică. 3. Identificarea unui set de date în volum suficient pentru crearea unui eventual sistem de prezicere timpurie a sepsisului; 4. Analiza exploratorie a datelor clinice și de laborator și procesarea lor în modul necesar pentru crearea unui sistem de discriminare/prezicere; 5. Crearea unui sistem de tipul unei aplicații practice, care ar permite prezicerea sepsisului la pacienții din secțiile de terapie intensivă.

Noutatea și originalitatea științifică: În baza analizei unui set larg de date (40366 cazuri, dintre care – 2932 cu sepsis) și prelucrare a lor cu utilizarea unui nou algoritm de restabilire a datelor-lipsă și utilizarea metricii complexității algoritmice s-a creat un sistem de suport decizional pentru prezicerea timpurie a sepsisului.

Problema științifică importantă soluționată în teză: Sepsisul reprezintă o problemă actuală în anesteziologie-terapie intensivă, iar diagnosticarea lui precoce este crucială pentru tratamentul eficient. Rezultatul obținut care contribuie la soluționarea unei probleme științifice importante constă în elaborarea unui sistem în baza învățării automate ce are ca efect îmbunătățirea managementului clinic al pacienților cu sepsis.

Semnificația teoretică a cercetării: A fost explorată și confirmată posibilitatea utilizării conceptelor și metricii dinamicii algoritmice în reprezentarea datelor medicale, care descriu starea clinică a pacientului. Această reprezentare este reușită și la elaborarea sistemului de prezicere a sepsisului. Aspectele ce se referă la valoarea predictivă a unor parametri clinici, care au fost elucidate în studiu, ar putea contribui la o înțelegere mai bună a problemei sepsisului ca fenomen medical.

Valoarea aplicativă a lucrării: Aplicația software creată, în care este integrat sistemul de prezicere timpurie a sepsisului elaborat, poate asista medicul din ATI în procesul de luare a deciziilor, în special în cazurile de sepsis mai complexe și în deosebi în situațiile ambigui. Metodele propuse pentru reconstrucția si reprezentarea datelor pot facilita, diversifica și înviora activitatea cercetătorilor în domeniu.

Implementarea rezultatelor științifice: Rezultatele studiului au fost implementate în activitatea didactică, curativă și de cercetare, confirmate prin două certificate de inovație și acte de implementare.

SUMMARY

Iapăscurtă Victor "Early prediction of sepsis using a proprietary application developed based on machine learning (artificial intelligence)", doctoral thesis in medical sciences, Chisinau, 2023

The thesis is presented on 170 pages and includes: introduction, 5 chapters, conclusions, bibliography from 278 sources, 12 appendices, 52 figures and 32 tables. The results obtained are published in 16 scientific works, of which 6 as sole author, 7 – first author, 3 publications in journals with an impact factor.

Keywords: sepsis, model, artificial intelligence, machine learning, algorithmic complexity, block decomposition method, decision support systems, prediction systems.

The purpose of the study. Evaluation of the use of AI technologies in anesthesia-intensive care, in particular for the management of patients in critical condition and especially in patients who may develop sepsis with the creation of a system with discriminatory abilities (sepsis vs non-sepsis), which allows early prediction of sepsis development.

The objectives of the study: 1. Evaluation of the use of machine learning (ML), as one of the basic technologies of artificial intelligence (AI) at the current stage, used in anesthesia and intensive care; 2. Evaluation of the use of AI in the management of critically ill patients, particularly in the case of sepsis; 3. Identification of a data set with sufficient volume for the creation of an early sepsis prediction system; 4. Exploratory analysis of this data and its processing in the manner necessary to create a discrimination/prediction system; 5. Creating such a system in the form of a practical software application to predict sepsis in intensive care units.

Scientific novelty and originality. Based on the analysis of a large set of data (40366 cases, of which – 2932 with sepsis) and their processing with the use of a new algorithm for restoring missing data and the use of the algorithmic complexity metric, a decision support system was created for the early prediction of sepsis.

The important scientific problem solved in the thesis. Sepsis is a current problem in anesthesiology and intensive care, and its early diagnosis is crucial for effective treatment. The result that contributes to the solution of an important scientific problem consists in developing an ML-based system with potentially improving effects on the clinical management of patients with sepsis.

The theoretical significance of the research. The possibility of using the concepts and metrics of algorithmic dynamics in the representation of medical data, which describe the clinical condition of the patient, was explored and confirmed. This representation is also successful in developing the sepsis prediction system. The aspects related to the predictive value of some clinical parameters, which were elucidated in the study, could contribute to a better understanding of the problem of sepsis as a medical phenomenon.

The applicative value of the work. The created software application, in which the developed sepsis early prediction system is integrated, can assist the intensivist in the decision-making process, especially in more complex sepsis cases and especially in ambiguous situations. The proposed methods for data reconstruction and representation can facilitate, diversify and invigorate the activity of researchers in the field.

Implementation of scientific results. The results of the study were implemented in teaching, curative and research activities, confirmed by two innovation certificates and implementation documents.

АННОТАЦИЯ

Япэскуртэ Виктор «Раннее прогнозирование сепсиса с помощью собственного приложения, разработанного на основе машинного обучения (искусственного интеллекта)», докторская диссертация, Кишинев, 2023 г.

Диссертация представлена на 170 страницах и включает: введение, 5 глав, выводы, библиографию из 278 источников, 12 приложений, 52 рисунков и 32 таблиц. Полученные результаты опубликованы в 16 научных работах, из них 6 как единственный автор, 7 – первый автор, 3 публикации в журналах с импакт-фактором.

Ключевые слова: сепсис, модель, искусственный интеллект, машинное обучение, алгоритмическая сложность, метод блочной декомпозиции, системы поддержки принятия решений, системы прогнозирования.

Цель исследования: Оценка использования технологий ИИ в анестезиологии-реаниматологии, в частности для ведения пациентов в критическом состоянии и особенно у пациентов, у которых возможно развитие сепсиса, с созданием системы с дискриминационными способностями (сепсис vs не-сепсис), которая позволяет раннее прогнозирование развития сепсиса.

Задачи исследования: 1. Оценка использования технологий искусственного интеллекта и особенно машинного обучения, как одной из базовых технологий ИИ на современном этапе, применяемых в анестезиологии-реаниматологии; 2. Оценка использования ИИ при лечении пациентов в критическом состоянии, особенно в случае сепсиса; 3. Выявление набора данных достаточного объема для создания возможной системы раннего прогнозирования сепсиса; 4. Исследовательский анализ этих данных и их обработка в порядке, необходимом для создания системы различения/прогнозирования; 5. Создание такой системы в виде практического приложения, которое можно использовать с целью прогнозирования сепсиса в отделениях интенсивной терапии.

Научная новизна и оригинальность: На основе анализа большого набора данных (40366 случаев, из них — 2932 с сепсисом) и их обработки с применением нового алгоритма восстановления отсутствующих данных и применением метрики алгоритмической сложности была создана система поддержки принятия решений для раннего прогнозирования сепсиса.

Важная научная проблема, решенная в диссертации: Сепсис является актуальной проблемой в анестезиологии-реаниматологии и его ранняя диагностика имеет решающее значение для эффективного лечения. Полученный результат, способствующий решению важной научной задачи, заключается в разработке системы на основе машинного обучения, что позволяет улучшить клиническое ведение больных с сепсисом.

Теоретическая значимость исследования: Исследована и подтверждена возможность использования понятий и метрик алгоритмической динамики в представлении медицинских данных, описывающих клиническое состояние пациента. Это представление также успешно используется при разработке системы прогнозирования сепсиса. Выясненные в исследовании аспекты, связанные с прогностической ценностью некоторых клинических параметров, могли бы способствовать лучшему пониманию проблемы сепсиса как медицинского явления.

Прикладное значение исследования: Созданное программное приложение, в которое интегрирована разработанная система раннего прогнозирования сепсиса, может помочь врачу в процессе принятия решения, особенно в более сложных случаях сепсиса и особенно в неоднозначных ситуациях. Предлагаемые методы реконструкции и представления данных могут облегчить, разнообразить и активизировать деятельность исследователей в этой области.

Внедрение научных результатов: Результаты исследования внедрены в педагогическую, лечебную и научно-исследовательскую деятельность, что подтверждено двумя инновационными свидетельствами и документами на внедрение.

GLOSSARY OF TECHNICAL TERMS

Algorithm - a set of reasoning or operations aimed at solving problems.

Machine learning algorithm - program used to learn a machine learning model from data.

Features - inputs used for prediction or classification. A feature is a column in the data set. They are assumed to be interpretable, which means it is easy to understand what they mean. But if it is hard to understand the input features, it is even harder to understand what the model is doing.

Gradient boosting machine (GBM) - machine learning algorithm based on decision trees using the gradient lowering (decreasing) boosting method.

Artificial intelligence - the science of developing computer systems that can perform tasks that normally require human intelligence.

Machine learning (ML) - a set of methods that allow computers to learn from data to make and improve predictions.

AutoML – a set of methods by which the computer applies multiple machine learning algorithms to the delivered data and automatically determines the best-performing machine learning models.

Machine learning model - the learned program that maps inputs to predictions.

Black Box Model - a system that does not reveal its internal mechanisms. In machine learning, a "black box" describes patterns that cannot be understood by analyzing their parameters.

Glass Box Model - an interpretable model.

Prediction - what the machine learning model "guesses" should be the target value based on the given characteristics.

Artificial Neural Networks (ANN) - a machine learning algorithm that uses similarity to the functioning of brain neurons.

IAPĂSCURTĂ Victor

EARLY PREDICTION OF SEPSIS USING A PROPRIETARY APPLICATION DEVELOPED BASED ON MACHINE LEARNING (ARTIFICIAL INTELLIGENCE)

321.19 - ANAESTHESIOLOGY AND INTENSIVE CARE

Abstract of the Ph.D. thesis in medical sciences

Approved for printing 15.06.2023	A5 format
Offset paper, digital printing	Print run 10 copies
Print units: 2.0	Order no. 21

Printed at “SRL Sirius”
Chisinau, 2 Lapușneanu str., phone: 022 23 23 52