

**Școala doctorală în domeniul Științe medicale**

Cu titlu de manuscris

*C.Z.U.: 616.94-037:004.8(043.2)*

**IAPĂSCURTĂ Victor**

**PREZICEREA TIMPURIE A SEPSISULUI CU  
AJUTORUL UNEI APLICAȚII PROPRII  
ELABORATE ÎN BAZA ÎNVĂȚĂRII  
AUTOMATE (INTELIGENȚĂ ARTIFICIALĂ)**

**321.19 – ANESTEZOLOGIE ȘI TERAPIE INTENSIVĂ**

**Rezumatul tezei de doctor în științe medicale**

**Chișinău, 2023**

Teza a fost elaborată la Catedra de anesteziologie și reanimatologie nr.1 “Valeriu Ghereg”, IP Universitatea de Stat de Medicină și Farmacie “Nicolae Testemițanu” din Republica Moldova.

**Conducător**

Belî Adrian,

dr. hab. șt. med, profesor universitar

\_\_\_\_\_

**Membrii comisiei de îndrumare:**

Cojocaru Svetlana,

dr. hab. informatică, prof. cercetător,  
membru corespondent al AȘ a Moldovei

Ciobanu Nellu,

dr. șt. fizico-matematice, conf. univ.

Savan Veaceslav,

dr. șt. med.

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

Susținerea va avea loc la 27 septembrie, 2023, 14:00 în incinta USMF ”Nicolae Testemițanu”, bd. Ștefan cel Mare și Sfânt, 165, biroul 205 în ședința Comisiei de susținere publică a tezei de doctorat, aprobată prin decizia Consiliului Științific al Consorțiului din 25.05.2023 (*proces verbal nr.5*).

**Componența Comisiei de susținere publică a tezei de doctorat:**

***Președinte:***

Cojocaru Victor,

dr. hab. șt. med, prof. universitar

\_\_\_\_\_

***Membrii:***

Arnaut Oleg,

dr. hab. șt. med, prof. universitar

Baltaga Ruslan,

dr. șt. med, conf. universitar

Belî Adrian,

dr. hab. șt. med, profesor universitar

Cobeț Valeriu,

dr. hab. șt. med., prof. universitar

Cojocaru Svetlana,

dr. hab. informatică, prof. cercetător, m.c. al AȘM

Cornogolub Alexandru,

dr. șt. med., conf. universitar

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

Autor

Iapăscurtă Victor

\_\_\_\_\_

© Iapăscurtă Victor, 2023

## CUPRINS

<b>Repere conceptuale ale cercetării</b> .....	4
<b>Conținutul tezei</b> .....	7
<b>1. Aplicarea tehnologiilor de învățare automată / inteligență artificială în anesteziologie-terapie intensivă și în managementul sepsisului</b> .....	8
1.1. Tehnologii de învățare automată / inteligență artificială utilizate în anesteziologie - terapie intensivă .....	8
1.2. Conceptul de sepsis. Evoluția conceptului și conținutul lui modern .....	8
1.3. Situația actuală în abordarea problemei sepsisului cu utilizarea tehnologiilor intelectuale în managementul clinic al stării de sepsis .....	8
<b>2. Materiale și metode de cercetare</b> .....	9
2.1. Designul general al cercetării .....	9
2.2. Descrierea materialului clinic .....	10
2.3. Aspecte metodologice importante pentru investigație, procesarea datelor și analiza rezultatelor. ....	11
<b>3. Etapele de procesare a datelor</b> .....	12
3.1. Reconstruirea datelor prin rezolvarea problemei valorilor lipsă .....	12
3.2. Complexitatea algoritmică Kolmogorov-Chaitin ca metrică și metodă de procesare a datelor .....	15
3.3. Procesarea datelor cu crearea caracteristicilor pentru a fi utilizate în sistemul de învățare automată .....	16
<b>4. Învățarea automată. Rezultate intermediare și discuția lor</b> .....	17
4.1. Etapa de învățare automată .....	17
4.2. Elaborarea, validarea și testarea sistemului de ÎA creat .....	18
4.3. Incorporarea sistemului creat într-o aplicație pentru utilizare clinică .....	20
4.4. Utilizarea aplicației pentru precizie continuă a riscului sepsisului .....	22
<b>5. Explicabilitatea modelelor IA. Discuții și direcții de cercetare de viitor</b> .....	23
5.1. Sumar al lucrului de cercetare efectuat. Discuția rezultatelor. Limitări .....	23
5.2. Aspecte de explicabilitate a modelelor IA .....	23
5.3. Transferabilitatea sistemelor IA. Direcții de cercetare de viitor .....	24
<b>Concluzii generale</b> .....	25
<b>Recomandări practice</b> .....	26
<b>Bibliografie</b> .....	27
<b>Lista lucrărilor științifice publicate la tema tezei</b> .....	29
<b>Adnotare (rom)</b> .....	31
<b>Adnotare (eng)</b> .....	32
<b>Adnotare (rus)</b> .....	33
<b>Glosar de termeni tehnici</b> .....	34

## REPERE CONCEPTUALE ALE CERCETĂRII

### **Actualitatea și importanța problemei abordate.**

Pot fi evidențiate două aspecte importante care conturează direcția de cercetare reflectată în lucrare: (a) Problema sepsisului ca varietate de stare critică, deseori dificil de diagnosticat la timp, iar rezultatele tratamentului depinzând strâns de timpul când se începe tratamentul (cu antibiotice) și influența directă a acestor factori asupra mortalității [1], care cu timpul a diminuat neînsemnat [2] și (b) Apariția unui nou actor – așa numitele tehnologii de inteligență artificială (IA), care deseori se declară drept panacee pentru mai multe probleme, inclusiv medicale. Cercetarea actuală încearcă să evalueze posibilitatea utilizării acestor tehnologii pentru rezolvarea unor aspecte importante în managementul acestui grup de pacienți prin precizarea timpurie (câteva ore până la debut) a stării de sepsis, ce ar diminua întârzierea începutului tratamentului.

### **Descrierea situației în domeniu și identificarea problemei de cercetare.**

În ciuda unei mari mortalități asociate [1] și a costurilor ridicate ale tratamentului, sepsisul [2] rămâne dificil de diagnosticat și de tratat. Cercetările anterioare au subliniat importanța recunoașterii în timp util a sepsisului pentru îmbunătățirea rezultatelor și reducerea costurilor asociate tratamentului [3]. Recent au fost propuse noi definiții destinate să îmbunătățească recunoașterea clinică a sepsisului [4], deoarece utilizarea anterioară a screeningului pe baza sindromului de răspuns inflamator sistemic (SIRS) a fost găsită ca fiind nespecifică [5]. Datele din literatura medicală au arătat că un diagnostic precoce și un tratament precoce poate reduce riscul rezultatelor adverse ca consecințe ale sepsisului [6]. Prin urmare, detectarea precoce a sepsisului și recunoașterea mai exactă a pacienților cu risc ridicat de dezvoltare a sepsisului este esențială pentru tratamentul eficient. Instrumentele de screening cel mai des utilizate în condiții clinice pentru identificarea pacienților septici includ SOFA (Sequential Organ Failure Assessment), criteriile SIRS (Systemic Inflammatory Response Syndrome), MEWS (Modified Early Warning Score). În pofida unor limitări, aceste sisteme de punctaj au stabilit valori de performanță și servesc ca comparatori importanți pentru sistemele de detectare și predicție a sepsisului recent dezvoltate și pentru evaluarea efectului lor asupra rezultatelor clinice [7]. Un supliment recent la aceasta colecție sunt sistemele bazate pe inteligența artificială (SIA) [3]. Deși SIA reprezintă suplimente destul de noi la domeniul managementului clinic al sepsisului, ele au potențialul de a îmbunătăți rezultatele pacienților prin avertizarea în avans a debutului iminent al sepsisului astfel asistând medicul în procesul de luare a deciziilor. În prezent în domeniul cercetării sepsisului, sistemele de asistență decizională bazate pe învățare automată reprezintă o direcție cu dezvoltare rapidă [3, 8].

**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 precizarea timpurie a dezvoltării sepsisului.

## **Obiectivele studiului**

1. Evaluarea utilizării tehnologiilor de inteligență artificială și, în special a învățării automate, ca una din tehnologiile de bază ale IA la etapa actuală, utilizate în anestezieterapie 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 precizie timpurie a sepsisului.
4. Analiza exploratorie a datelor clinice și de laborator și procesarea lor în modul necesar pentru crearea unui sistem de discriminare/precizie.
5. Crearea unui așa sistem de tipul unei aplicații practice, care ar permite precizarea dezvoltării sepsisului în secțiile de terapie intensivă.

## **Metodologia cercetării științifice.**

Cercetarea este un studiu retrospectiv al unui set de date ce conține 40336 pacienți/cazuri, dintre care 7,26% sunt pacienți cu sepsis. Restul – 92.74% reprezintă pacienți internați în secțiile de terapie intensivă cu alte diagnoze (non-sepsis). Analiza exploratorie și procesarea datelor sunt în mare măsură determinate de pașii ulteriori ai cercetării, ce presupun crearea în final a unui sistem de inteligență artificială cu abilități de precizie a unui eveniment (sepsis). Pentru analiza și prelucrarea datelor, inclusiv și cea statistică, și procesarea lor ulterioară s-a utilizat mediul integrat de dezvoltare RStudio în baza limbajului de programare R (<http://www.R-project.org/>), care a fost și limbajul de interacțiune cu platforma H2O (<https://www.h2o.ai/>) pentru învățarea automată propriu-zisă. Calculele de înaltă intensitate au fost realizate utilizând Amazon Web Services (AWS EC2 <https://aws.amazon.com/ec2/>). Pentru evaluarea performanței modelului creat și raportarea rezultatelor s-a utilizat ghidul TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) [9] și principiile PROBAST (Prediction model Risk Of Bias ASsessment Tool) [10].

## **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 precizarea timpurie a sepsisului.

## **Problema științifică importantă soluționată în teză.**

Sepsisul reprezintă o problemă actuală în serviciul medical/serviciul terapiei intensive (ATI), iar diagnosticarea lui precoce poate duce la sporirea ratei de succes a tratamentului, diminuarea mortalității și reducerea costului asistenței livrate acestui grup de pacienți, în special în cazurile complexe. Rezultatul obținut care contribuie la soluționarea unei

probleme științifice importante constă în elaborarea unui sistem în baza învățării automate, care sporește eficiența 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, inclusiv și a celor în formă de serii temporale, care descriu starea clinică a pacientului. Această reprezentare s-a dovedit a fi reușită și la etapele ulterioare de procesare a datelor – în construirea sistemului de precizie, iar performanța înaltă a sistemului este un argument suplimentar. 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 precizie 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 ambigue. Metodele propuse pentru reconstrucția și reprezentarea datelor pot facilita, diversifica și înviora activitatea cercetătorilor în domeniu.

### **Rezultatele științifice principale înaintate spre susținere:**

1. Algoritmul de reconstrucție a datelor reprezentate prin serii temporale ale parametrilor fiziologici, care conțin valori lipsă.
2. Metoda de selectare a datelor pentru crearea unui sistem de precizie a stării pacientului (pe exemplul pacienților cu risc de sepsis)
3. Metoda de reprezentare a datelor prin calcularea complexității algoritmice cu ajutorul metodei de decompoziție în blocuri (BDM).
4. Sistemul automat pentru precizarea sepsisului încorporat într-o aplicație software pentru utilizare în clinică.

### **Implementarea rezultatelor științifice.**

- La etapa finală a fost creată o aplicație software pentru a fi transmisă în Clinica ATI a Institutului de medicină urgentă, Chișinău. Aplicația are o interfață grafică care permite introducerea datelor (frecvența cardiacă, SptO<sub>2</sub>, temperatura, PA sistolică și diastolică, frecvența respirației) unui pacient concret pe o perioadă de trei ore cu posibilitatea vizualizării lor grafice și obținerea rezultatului precizării, adică a riscului de dezvoltare a sepsisului cu un orizont de 4 ore.

- O parte din rezultatele cercetării sunt încorporate în cursul “Advanced perioperative monitoring. Elements of computational medicine” (pentru studenții a. VI la USMF “Nicolae Testemițanu”)
- Rezultatele de etapă ale cercetării prezentate la:
  - ✓ Cursurile Comitetului European pentru Educare în Anestezie, ediția a XIII-a, Modulul 1, 7-9 februarie, 2018,
  - ✓ Cursurile Comitetului European pentru Educare în Anestezie, ediția a XV-a, Modulul 3, 10-11 decembrie, 2020,
  - ✓ Cursurile Comitetului European pentru Educare în Anestezie, ediția a XVI-a, Modulul 5, 9-11 decembrie, 2021,
  - ✓ 5<sup>th</sup> International Conference on Nanotechnologies and Biomedical Engineering, November 3-5, 2021,
  - ✓ The 12<sup>th</sup> International Conference on Electronics, Communications and Computing, 20-21 October, 2022,
  - ✓ Conferința științifică anuală a USMF “N. Testemițanu”. Cercetarea în biomedicină și sănătate: calitate, excelență și performanță, Chișinău, 19-21 octombrie, 2022
- Utilizarea BDM pentru analiza și reprezentarea datelor medicale în formă de serii temporale a fost prezentată la conferința AUTOMATA 2020, Stockholm, Sweden, 10-12 august, 2020
- Activitate de brevetare: rezultatele cercetării oficializate în două certificate de invenție (nr.18 din 12.12.2022 și nr.19 din 13.12.2022) și acte de implementare.

**Publicații la tema tezei.** La tema tezei sunt publicate 16 lucrări științifice care sunt parafrazate în conținutul tezei, observându-se astfel importanța teoretică și valoarea aplicativă a lucrării. Trei publicații – în reviste/culegeri cu factor de impact SCOPUS, publicații în calitate de singur autor – 6, prim-autor – 7.

**Sumarul compartimentelor tezei.** Teza este expusă pe 107 pagini de text de baza, structurat în 5 capitole, concluzii și recomandări practice, urmate de lista a 274 surse bibliografice și 12 anexe. Materialul iconografic include 52 figuri și 32 tabele.

**Cuvinte cheie:** sepsis, model, inteligență artificială, învățare automată, complexitate algoritmică, metoda de decompoziție în blocuri, sisteme de suport decizional, sisteme de precizie

## CONTINUTUL TEZEI

În compartimentul **Introducere** este reflectată și argumentată actualitatea tematicii abordate și oportunitatea cercetării de față în baza datelor publicate în literatura de specialitate. Sunt formulate scopul și obiectivele cercetării, prezentate domeniul de cercetare, noutatea științifică a rezultatelor obținute și problema științifică soluționată.

Sunt redate semnificația teoretică, valoarea aplicativă a lucrării și descrise posibilitățile de implementare a rezultatelor științifice obținute.

## **1. APLICAREA TEHNOLOGIILOR DE INTELIGENȚĂ ARTIFICIALĂ ÎN ANESTEZIOLOGIE-TERAPIE INTENSIVĂ (ATI) ȘI ÎN MANAGEMENTUL SEPSISULUI**

### **1.1 Tehnologii de învățare automată/inteligență artificială utilizate în anesteziologie - terapie intensivă**

Conform unui articol recent [11], aplicațiile comerciale ale inteligenței artificiale (IA) și învățării automate au înregistrat recent progrese remarcabile, în special în domeniul precum recunoașterea imaginilor, prelucrarea limbajului natural, traducerea limbajului, analiza texturală și autoînvățarea. Cu toate acestea, tehnologiile descrise au atins cea mai mare performanță în aplicațiile cu o singură sarcină în care pot fi tolerate rezultatele imperfecte și erorile ocazionale. Practica anesteziei-terapiei intensive este diferită. Ea încorporează cerințe pentru fiabilitate ridicată, restricții temporale de interpretare, acțiune/răspuns fizic, mai degrabă decât orice act cognitiv. Tranziția la medicina bazată pe dovezi și “Big Data” este un eveniment relativ recent. În schimb, medicii ATI s-au bazat mult timp pe fluxuri personalizate de date cuantificate pentru a-și îngriji pacienții, iar progresele în monitorizare și bogăția acestor date au susținut îmbunătățirile dramatice ale siguranței pacienților în specialitate [12]. Specialiștii respectivi deseori acționează în circumstanțe “de muchie” a cuplului “cauză și efect”: de obicei, deciziile nu pot fi amânate, iar erorile de judecată pot fi adesea de neevitat. Un articol de recenzie din 2020 [13] indică următoarele domenii din specialitatea ATI cu cea mai înaltă rată de utilizare a tehnologiilor AI: (1) monitorizarea profunzimii anesteziei, (2) controlul anesteziei, (3) precizarea evenimentelor și a riscurilor, (4) ghidarea cu ultrasunet, (5) managementul durerii și (6) logistica sălii de operație, la care mai pot fi adăugate precizarea mortalității, suportul decizional în ventilația mecanică, precizarea sepsisului [14].

### **1.2. Conceptul de sepsis. Evoluția conceptului și conținutul lui modern**

Pe parcursul anilor abordarea sepsisului a revoluționat începând de la conceptul Sepsis-1 (1990) la Sepsis-2 (2008) și în prezent sepsisul este diagnosticat conform principiilor Sepsis-3 (2016), propus în 2016, unde sepsisul a fost definit ca un răspuns dereglat al gazdei la infecție, care pune viața în pericol. Acesta concept este operaționalizat prin evaluarea insuficienței secvențiale de organ (SOFA), unde criteriile clinice de diagnostic ale sepsisului cuprind o creștere acută cu cel puțin două puncte a scorului SOFA combinat cu o infecție confirmată sau suspectată [5]. Cercetarea din cadrul tezei utilizează cea mai recentă definiție, “Sepsis-3”.

### **1.3. Situația actuală în abordarea problemei sepsisului cu utilizarea tehnologiilor intelectuale în managementul clinic al stării de sepsis**

Datele din literatura arată că depistarea precoce sau predicția sepsisului pot duce la o scădere a timpului de administrare a antibioticelor [15], iar intervenția timpurie la rândul său s-a dovedit a reduce rata mortalității în acest grup de pacienți [16]. Spre exemplu,



utilizarea unui sistem de învățare automată (SIA) din studiul [3] a fost asociat cu o reducere cu 39,5% a mortalității spitalicești ( $p < 0,001$ ), o reducere cu 32,3% a duratei internării ( $p < 0,001$ ) și o reducere cu 22,7% în readmisii la 30 de zile ( $p < 0,001$ ). Deși SIA reprezintă suplimentul destul de noi la domeniul managementului clinic al sepsisului, algoritmi de învățare automată au potențialul de a îmbunătăți în mod semnificativ rezultatele pacienților prin avertizarea în avans a debutului iminent al sepsisului. SIA-urile cu abilități de predicție a sepsisului pot servi, de asemenea, pentru a permite medicilor clinici să aibă încredere în diagnosticul de sepsis într-o varietate de cazuri ambigui, inclusiv cazuri în care rezultatele culturii pozitive nu sunt disponibile [17] și în cazurile de prezentare clinică atipică, în special în rândul pacienților mai în vârstă, care cuprind o majoritate de cazuri de sepsis [18]. Prin urmare, sistemele de asistență decizională bazate pe învățare automată reprezintă un domeniu important de investigare pentru cercetarea sepsisului [8, 19].

## 2. MATERIALE ȘI METODE DE CERCETARE

### 2.1. Designul general al cercetării

Cercetarea a fost aprobată de Comitetul de Etică a Cercetării USMF „Nicolae Testemițanu”, aviz favorabil nr. 3 din 18.03.2022. A fost efectuat un studiu retrospectiv cu câteva etape în baza unui set inițial de date ce descriu 40366 pacienți, inclusiv 2932 de pacienți cu sepsis dintr-o bază de date publică [20]. Datele inițiale conțin valori lipsă. După analiza exploratorie au fost selectați pentru crearea sistemului 5039 pacienți, dintre care 1703 – cu sepsis, cu un total de 30635 eșantioane orare. Criteriile de selectare au inclus: (a) prezența a cel puțin 7 ferestre de observare orare (4 ore – orizontul de prognoză, 3 ore – pentru evaluarea dinamicii parametrilor care descriu pacientul) și (b) densitatea maxim posibilă a datelor (pentru pacienții non-septici – prezența a 7 valori consecutive pentru fiecare dintre parametrii de interes). Datele pacienților septici cu valori lipsă au fost reconstruite conform unui algoritm elaborat în cadrul studiului curent. Aceste date au fost utilizate pentru crearea unui sistem de prognoză timpurie a sepsisului.

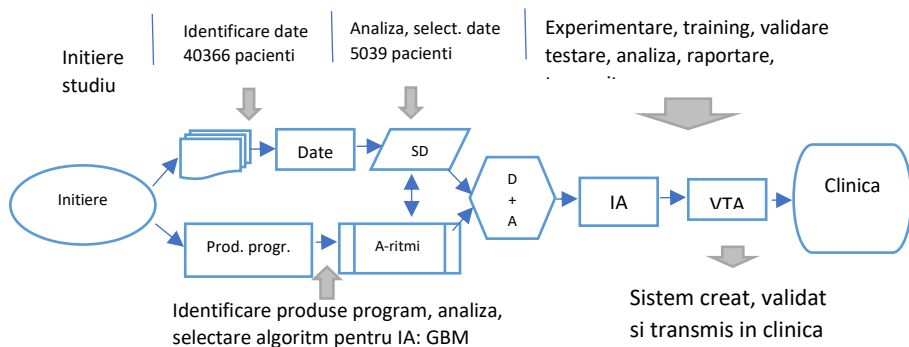


Figura 2.1. Designul studiului

Elaborarea sistemului a urmat etapele tradiționale în crearea unui sistem de învățare automată: training, cross-validare, testare pe date noi, care nu au fost utilizate la etapele precedente. Au fost analizate rezultatele cercetării bazate pe performanța sistemului, descrise concluziile finale și recomandări practice. Sistemul creat, sub forma de aplicație, va fi transmis în Clinica ATI a Institutului de medicină urgentă, Chișinău pentru calibrare și dezvoltare de viitor.

## 2.2. Descrierea materialului clinic

Setul de date utilizat în teză reprezintă date de acces public din cadrul “Early Prediction of Sepsis from Clinical Data: The PhysioNet Computing in Cardiology Challenge 2019” [20].

Tabelul 2.1. Caracterizarea generală a seturilor de date utilizate

Parametru	Setul A	Setul B
Numărul total de pacienți	20336	20000
Numărul de pacienți cu sepsis	1790	1142
Vârsta, ani (media, ds)	62.62 (16.24)	60,65 (16,67)
Vârsta %		
18-50 ani	21.44	25,89
51-60 ani	18.54	20.13
61-70 ani	22.41	24.94
71-80 ani	22.67	18.65
>81 ani	14.94	10.40
Repartizare pe sexe, femei/bărbați	8502 (41,80%) 11834 (58,2%)	9268 (46,34%) 10732 (53,66%)
Durata observării (ore în ATI) (media, ds)	38.86 (22.31)	38.10 (23,28)
Parametrii clinici monitorizați	<ul style="list-style-type: none"> <li>• 8 semne vitale și 1 parametru de suport vital</li> <li>• 25 parametri de laborator</li> <li>• 6 parametri demografici și de logistica</li> </ul>	
Prevalența sepsisului %	8,80	5,71
Numărul de observații/ferestre orare	790215	741952
Numărul de intrări de date/celule cu valori ‘non-missing’	10296016	9831918

**Notă:** Grupul de **parametri “semne vitale și suport vital”** include: frecvența cardiac (FCC), satur. sânge perif. cu O<sub>2</sub> (SpO<sub>2</sub>), temperatura (Temp), presiunea arter. sistolică (PAS), presiunea arter. medie (PAM), presiunea arter. diastolică (PAD), frecv. respirației (Resp), capnometrie (EtCO<sub>2</sub>), fracția O<sub>2</sub> inspir. (FiO<sub>2</sub>); **parametrii de laborator:** exces de baze (BE), bicarbonat (HCO<sub>3</sub>), pH, pres.part. CO<sub>2</sub> sânge art. (P<sub>a</sub>CO<sub>2</sub>), sat. O<sub>2</sub> sânge art. (S<sub>a</sub>O<sub>2</sub>), fermenți hepatici, ioni serici, hemoleucograma, etc.; **parametrii demografici și de logistica:** vârstă, sex, timp internare, durata spitalizării, etc., marcaj sepsis (SepsisLabel); ds – deviația standard

Conform regulilor evenimentului, aceste date pot fi utilizate în cadrul unor eventuale publicații după data conferinței de finalizare, 8-11 septembrie, 2019 (<https://physionet.org/content/challenge-2019/1.0.0>). Datele provin din două sisteme de spital din SUA distincte din punct de vedere geografic: Centrul medical Beth Israel Deaconess (setul A) și Spitalul universitar Emory (setul B). Aceste date au fost colectate în ultimul deceniu cu aprobarea comisiilor de revizuire instituționale, au fost de identificate și etichetate folosind criteriile clinice Sepsis-3 [5]. Datele constau dintr-o combinație de rezumate ale semnelor vitale pe oră, valori de laborator și descrieri statice ale pacientului. În special, datele conțin 40 de variabile clinice: 9 variabile de semne vitale și suport vital ( $\text{FiO}_2$ ), 26 de variabile de laborator și 6 variabile demografice și logistice. În total, aceste date includ peste 1,5 milioane de ferestre orare și peste 10.4 milioane de puncte de date (variabile fiziologice și de laborator “non-missing”). Datele extrase din sistemele electronice ale spitalelor respective au trecut printr-o serie de etape de preprocesare înainte de analiza formală. Toate caracteristicile pacientului au fost condensate în coșuri orare, simplificând dezvoltarea și testarea modelului; de exemplu, mai multe măsurări ale frecvenței cardiace într-o fereastră orară de timp au fost rezumate ca măsurarea ritmului cardiac mediu.

### **2.3. Aspecte metodologice importante pentru investigație, procesarea datelor și analiza rezultatelor**

Inițial datele sunt în format “.psv” (pipe separated value). Pentru procesarea lor ulterioară ele au fost convertite în format “.csv” (comma separated value) cu ajutorul pachetului “rio” (R). Datele inițiale sunt cercetate cu ajutorul metodelor statistice simple/standard: medie, deviație standard, mediană, raport procentual. Analiza datelor selectate pentru crearea nemijlocită a sistemului mai include estimarea distribuției lor, care servește și la elaborarea algoritmului de reconstrucție a datelor/valorilor lipsă. Logica și metodele de investigare și analiză a datelor sunt influențate de etapele ulterioare și scopul final – elaborarea sistemului de clasificare/precizare a sepsisului, ceea ce în acest domeniu tradițional poartă denumirea “Analiza exploratorie a datelor”. Parametrii/ metricile care sunt utilizate pentru evaluarea performanței produsului finit – al sistemului elaborat sunt, de asemenea, tradiționale și includ:

- Sensibilitate ( $\text{TPR} = \text{TP}/(\text{TP}+\text{FN})$ ), unde TPR – true positive rate, TP – true positive (rezultate), FN – false negative
- Specificitate ( $\text{TNR} = \text{TN}/(\text{TN}+\text{FP})$ ), unde TNR – true negative rate, TN – true negative, FP – false positive
- Acuratețe diagnostică =  $(\text{TP}+\text{TN})/(\text{TP}+\text{TN}+\text{FP}+\text{FN})$
- Analiza ROC (Receiver operating characteristic – reprezentare grafică a curbei ROC, ca raport dintre FPR (axa X) și TPR (axa Y), unde FPR, false positive rate =  $\text{FP}/(\text{FP}+\text{TN})$ )

- Matricea de confuzie (cu indicarea pentru fiecare clasă a numărului de cazuri prezise corect și incorect, eroarea și rata erorii)
- Analiza importanței variabilelor (identificarea și prezentarea grafică a celor mai importanți pentru model predictorii)
- Rata și traiectoria diminuării erorii (reprezentarea grafică a dinamicii erorii în cadrul învățării/training-ului modelului)
- Altele (precizia, rata predictivă negativă, ratele fals pozitivă și negativă, raporturile de probabilitate pozitiv și negativ, analiza PR, raportul șanselor).

Un loc separat îl au metodele de estimare a complexității algoritmice, care vin din domeniul dinamicii informației algoritmice. Ele sunt mai detaliat descrise în capitolul 3.

### 3. ETAPELE DE PROCESARE A DATELOR

#### 3.1. Reconstruirea datelor prin rezolvarea problemei valorilor lipsă

Pregătirea datelor pentru a fi livrate unui model de învățare automată este o etapă crucială, de care în ultimă instanță poate depinde performanța modelului. La moment nu sunt reguli prestabilite pentru acest proces. De obicei, forma cea mai reușită se stabilește prin încercări. În cazul curent s-ar părea că o oarecare importanță ar avea-o relevanța patofiziologică a unui parametru anumit, însă această abordare s-a dovedit a fi mai puțin productivă. Au fost încercate o serie de variante. La început au fost utilizați parametri care ar avea importanță în diagnosticarea clinică a sepsisului, și în special cei utilizați pentru SOFA. Această abordare a dat rezultate minore ( $AUC = 0.61$ ). Apoi s-a încercat setul integral de parametri clinici (36 parametri), ce a condiționat o performanță a modelului departe de ideală ( $AUC = 0.65$ ). Cei mai reușiți s-au dovedit a fi 6 parametri și dinamica lor sub formă de diferența a valorilor orare timp de 3 ore. Parametrii respectivi ulterior au fost reprezentați sub formă de matrice, pentru care a fost calculată valoarea complexității algoritmice, iar diferența a fost utilizată sub formă de rezultate netransformate. La selecția finală a formatului datelor utilizate pentru învățarea automată (trainingul modelului final) a contribuit principiul “de rând cu valoarea la moment a unui parametru fiziologic, de importanță majoră, deseori, este și dinamica lui în timp”. De asemenea s-a tins spre utilizare în calitate de predictorii a unor indici/parametri care se înregistrează cu ușurință/standard, inclusiv și în secțiile ATI din Republica Moldova.

După cum a fost menționat anterior, în datele inițiale sunt valori-lipsă (Tabelul 3.1). Pentru a evita consecințele situației "garbage in, garbage out" datele au fost reconstruite conform unei proceduri descrise în continuare.

În majoritatea lucrărilor similare ce au de a face cu date-lipsă [3, 21], valoarea lor se calculează prin metoda “last observation carried forward” (LOCF), care presupune sortarea setului de date (ex. valorile unui parametru/coloană – vezi mai sus) și crearea unui set ordonat. Ulterior se determină prima valoare lipsă și se utilizează valoarea precedentă pentru a o “restabili”. Procesul se repetă pentru următoarea valoare lipsă până când toate valorile lipsă sunt “restabilite”.

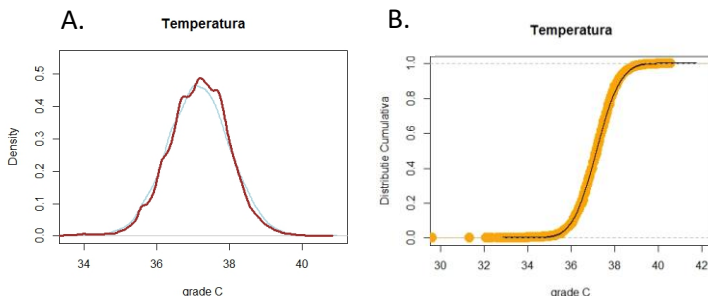
**Tabelul 3.1. Aspectul inițial al datelor unui pacient (p000009)**

FCC	SpO <sub>2</sub>	Temp	PAS	PAM	PAD	Resp	EtCO <sub>2</sub>	...	Vârsta	Sex	...
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	...
...	...	...	...	...	...	...	...	...	...	...	...

*Notă: NA (not available) – valori lipsă. Cu “...” sunt marcate coloanele și rândurile omise.*

*Acesta este un fragment al setului de date ale pacientului respectiv ce include primele 7 ferestre orare (pe verticala). Setul complet are 258 rânduri (adică ferestre orare/observații) și 40 coloane (parametrii monitorizați).*

Considerând procesele fiziologice drept fenomene cu caracter continuu (de obicei, reprezentate prin serii temporale – acesta este și cazul datelor studiate) se poate de presupus că și șirurile numerice create din valori ale parametrilor care reprezintă aceste procese moștenesc proprietatea de “continuitate” (contra variantei discrete). Metoda LOCF cumva ignoră “principiul de continuitate”. De aceea s-a decis de a include principiul “de continuitate”, propunând următoarea procedură: (1) fiecare coloană este evaluată în privința datelor lipsă (NA); (2) prima și ultima valoare (pe rânduri) este “restabilită” conform valorii din cea mai apropiată celulă; (3) în final se efectuează interpolarea, cu calcularea valorilor între două valori prezente cu respectarea tendinței (creștere sau descreștere).



**Figura 3.1. Distribuția datelor reale care descriu parametrul “Temperatura” vs cea a datelor generate. A. - distribuția Gaussiană; B. – distribuția cumulativă: cu culoare violetă și orange – datele “reale”, cu albastru și negru – datele generate.**

În cazul unor deosebiri mai evidente dintre cele două curbe ale densității/distribuției (datele reale vs cele generate) se poate manual de ajustat curba datelor generate, astfel încât ea să fie maximal posibil de aproape de cea reală. Pentru această ajustare o soluție poate fi utilizarea pachetului “sn” în R. În cazul cercetării curente nu a fost o așa

necesitate. Astfel, schema algoritmului utilizat pentru reconstrucția datelor este următoarea:

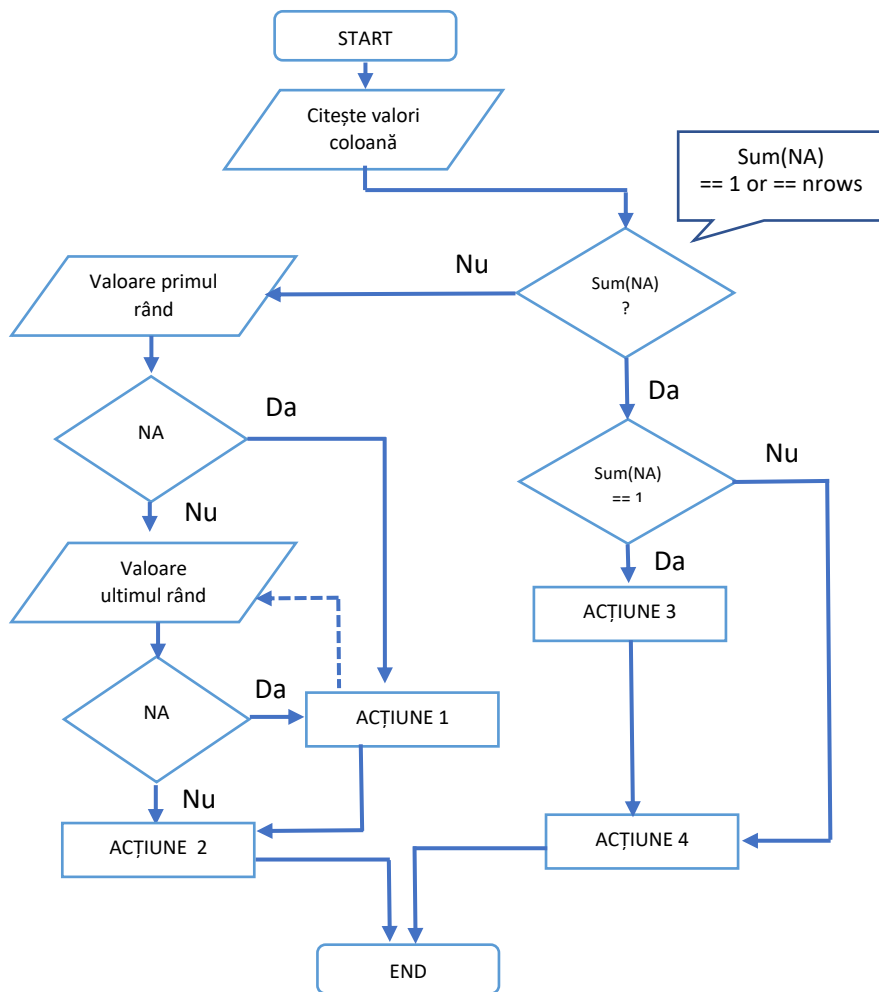


Figura. 3.2. Algoritmul pentru reconstrucția datelor în cazul lipsei valorilor

*Notă:* Acțiunea 1 – înlocuirea valorii-lipsă cu cea mai apropiată valoare (locație);  
 Acțiunea 2 – interpolarea valorilor lipsă prin procedura descrisă în text; Acțiune 3 – ștergerea  
 valorii; Acțiune 4 – umplerea celulelor cu valori din distribuția Gaussiană generată.

După aplicarea procedurii de reconstrucție datele capătă următorul aspect:

**Tabelul 3.2. Aspectul datelor aceluiași pacient (p000009) după reconstrucție**

FCC	SpO <sub>2</sub>	Temp	PAS	PAD	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

*Notă: Sunt prezentate numai valorile parametrilor (în coloane) supuși reconstrucției conform procedurii descrise în text. Setul de date este acel descris anterior în Tabelul 3.1.*

### 3.2. Complexitatea algoritmică ca metrică și metodă de procesare a datelor.

Complexitatea algoritmică (CA) este abordată în cercetarea curentă prin prisma dinamicii informației algoritmice (AID, algorithmic information dynamics). AID [22] este un domeniu emergent al științei complexității bazat pe teoria informației algoritmice (TIA), care cuprinde literatura de specialitate bazată pe conceptul de complexitate Kolmogorov – Chaitin și concepte conexe precum probabilitatea algoritmică, compresia, inferența optimă, distribuția universală și altele. Centrală pentru TIA este definiția complexității algoritmice (complexitatea Kolmogorov – Chaitin sau a mărimii programului - Kolmogorov, 1965; Chaitin, 1969) [23]:

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

adică lungimea celui mai scurt program  $p$  care generează șirul  $s$  și care rulează pe o mașină Turing universală. AID tinde să identifice soluții la întrebări fundamentale despre cauzalitate: de ce un anumit ansamblu de circumstanțe duce la un anumit rezultat. În acest aspect CA, diferă în esență de statisticile tradiționale.

Ca știință aplicată, AID este un nou tip de calcul discret bazat pe “computer science” și care vizează studierea cauzalității prin generarea de modele mecanice care să ajute la găsirea primelor principii ale fenomenelor fizice, construind următoarea generație de învățare automată [22].

În setul de instrumente AID, există un instrument special pentru furnizarea de estimări fiabile pentru funcțiile necomputabile, și anume calculatorul on-line al complexității algoritmice (OACC) [24], care oferă posibilitatea estimării CA și a probabilității algoritmice (PA) pentru șiruri numerice scurte și lungi și pentru structuri bidimensionale mai bine decât oricare alt instrument tradițional. Metodele tradiționale în esența lor nu sunt concepute pentru a capta careva conținut algorithmic dincolo de tiparele statistice simple.

OACC utilizează metoda BDM [22,24], care se bazează pe probabilitatea algoritmică definită prin metoda teoremei de codificare (CTM) [22]:

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

OACC este disponibil ca versiune online [25], precum și pachete autonome în R și o serie de alte limbaje de programare (Matlab, C, Wolfram, etc.) și este utilizat pentru calculele respective pentru scopul lucrării curente. Deoarece intensitatea acestor calcule este destul de înaltă (datorită volumului de date și însăși esența calculelor) utilizarea OACC în forma descrisă este insuficient de productivă. Pentru accelerarea acestor calcule a fost extras nucleul pachetului respective [25], care poate fi aplicat pe structuri bidimensionale (matrice) și integrat în programul care efectuează procesarea fluxului de date în cercetarea curentă [26]. Procesul calculării valorii BDM este ilustrat în Figura 3.3.

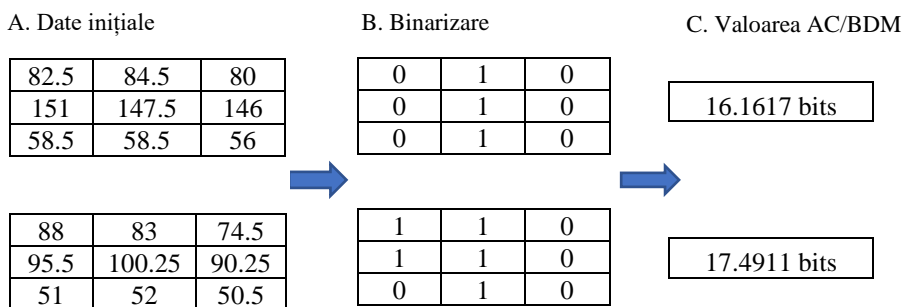


Figura. 3.3. **Formatul datelor la 3 etape succesive de prelucrare a lor**

*Notă:* A- formatul inițial, B – binarizare (după valorile-prag pe rânduri), C- valoarea BDM calculată pentru fiecare matrice. Datele inițiale reprezintă valori ale parametrilor circulatori (HR, SBP, DBP), structurate în matrice cu dimensiunile 3 x 3 (pe rânduri).

### 3.3. Procesarea datelor cu crearea caracteristicilor pentru a fi utilizate în sistemul de învățare automată

În cazul cercetării curente, reieșind din caracterul datelor primare, s-a decis de utilizat variabilele cu cele mai puține valori lipsă. Deoarece una din metodele de transformare a datelor este BDM (pe tensori bidimensionali/matrice – varianta optimă selectată – 2 asemenea matrice 3 x 3) și un interval de timp optim pentru estimarea prezicerii – 3 ore, a fost necesară selectarea a 6 asemenea variabile (cate trei pentru fiecare matrice x 3 ore). Astfel, selecția finală a variabilelor/parametrilor fiziologici include: frecvența cardiacă (FCC), saturația sângelui cu oxigen (SpO<sub>2</sub>), presiunea arterială sistolică (PAS) și diastolică (PAD), temperatura corpului (Temp), frecvența respirației (Resp) – în total 6 la număr. Lungimea finală a vectorilor utilizați pentru ÎA este de 14 (elemente), plus eticheta eșantionului (vezi Tab. 4.2).



## 4. ÎNVĂȚAREA AUTOMATĂ. REZULTATE INTERMEDIARE ȘI DISCUȚIA LOR

### 4.1. Etapa de învățare automată

Procesul de creare a unui model de învățare automată este mai mult sau mai puțin standard. Figura care urmează descrie componentele și etapele procesului (reprodus cu modificări din [27]).

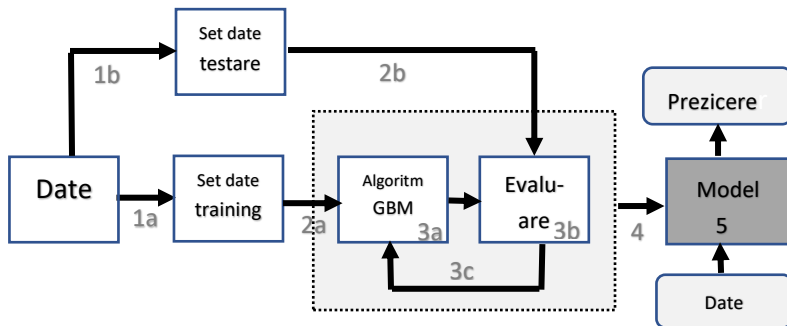


Figura 4.1. Etapele și elementele principale în crearea unui sistem de învățare automată: 1a, 1b, ..., 5 – etapele creării sistemului

Etapele procesării datelor sunt aliniate procesului de creare a unui SÎA și sunt schematic ilustrate în figura 4.2.

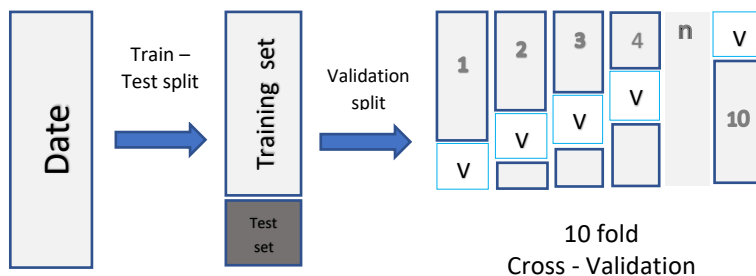


Figura 4.2. Etapele fluxului de procesare a datelor la crearea unui SÎA

*Notă:* Inițial datele sunt împărțite în setul pentru training (care ulterior va participa la învățare și cross-validare) și setul – test.

Un moment important este separarea de la început (după procesarea preliminară a datelor) a setului pentru testare, care în cazul studiului curent (setul A, test) este format din 3052 eșantioane (1063 sepsis și 1989 non-sepsis), care nu au participat la training și validare.

## 4.2. Elaborarea, validarea și testarea sistemului de ÎA creat

Ca punct de orientare pentru crearea sistemului de prezicere a sepsisului în cadrul tezei a fost sistemul InSight, raportat în literatura de specialitate ca fiind unul dintre cele mai performante în domeniu [3]. Folosind informația destul de generală din [21] acest sistem a fost replicat în limbajul R și testată performanța lui pe setul de date din studiul curent. Au fost obținute rezultate ce corespund celor raportate de autori. În continuare, aceste rezultate au servit drept jaloane în crearea și evaluarea sistemului propriu. În Tabelul 4.1 este prezentată metrica comparativă a performanței acestor două sisteme.

**Tabelul 4.1. Metrica comparativă a două sisteme de prezicere a sepsisului (InSight vs sistemul creat în cadrul cercetării curente)**

Metrica	Sistemul InSight	Sistemul creat
<b>AUC</b>	0.914 (95% Î: 0.902 – 0.926)	0.929 (95% Î: 0.919 – 0.939)
<b>Acuratețe diagnostică</b>	0.925	0.940
<b>Sensibilitate</b>	0.888	0.916
<b>Specificitate</b>	0.941	0.942

*Notă: AUC – area under the curve (suprafața de sub curba ROC).*

Performanța acestor două sisteme este destul de apropiată. O sensibilitate mai înaltă în cazul sistemului creat ar indica la un sistem mai echilibrat, care puțin mai bine detectează cazurile cu sepsis, decât sistemul InSight. Sistemul creat diferă de InSight după datele inițiale folosite pentru învățarea automată, metodele de procesare a datelor și formatul datelor finale, care în ultimă instanță sunt livrate pentru învățare și validare. Spre exemplu, datele finale în cazul sistemului creat reprezintă vectori cu lungimea 14 (șiruri numerice alcătuite din 14 numere/valori) comparativ cu 30, în cazul InSight, adică o reducere a dimensiunilor datelor mai mult de 2 ori. Aspectul datelor finale folosite de sistemul creat este ilustrat în tabelul 4.2.

**Tabelul 4.2. Formatul final al datelor transmise algoritmului de învățare automată**

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

*Notă: Acestea sunt datele finale care descriu 5 pacienți: 2 cu sepsis și 3 – cu alte patologii. “CAT”- este eticheta fiecărui subset/pacient (1 – sepsis, 0 – alte patologii); “BDM-MR/BDM-CIRC” – reprezintă valorile BDM pentru grupurile de date respirator-metabolice și cele care se referă la sistemul circulator; “V1 – V12” reprezintă dinamica celor 6 parametri fiziologici (FCC, SpO<sub>2</sub>, Temp, PAS, PAD, Resp) pe o perioadă de 3 ore.*

Datele finale organizate în formatul descris au fost în ultimă instanță livrate unui model în baza algoritmului GBM (Gradient Boosting Machine - algoritm în baza arborilor decizionali), care a efectuat predicția prin clasificare (sepsis vs non-sepsis). Hiperparametrii optimi (rata învățării, numărul de arbori, profunzimea, nr. de ramificații) au fost identificați inițial prin învățare automată automatizată (AutoML) utilizând platforma H2O. Acești hiperparametri au fost apoi justificați în contextul datelor prezente printr-o căutare grilă grosieră cu validare multiplă (10 folds cross-validation).

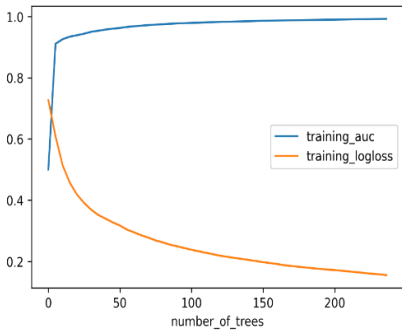


Figura 4.3. Curba diminuării pierderii la antrenare (logloss – de culoare orange) – pe axa Y. Pe axa X – numărul de arbori. Algoritm GBM.

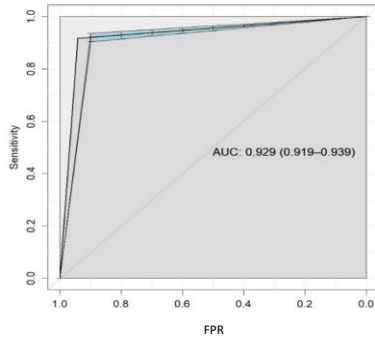


Figura 4.4. Curba ROC la testare pe setul-test (A): pe axa X – rata fals positive (FPR), pe axa Y- rata adevărat pozitivă (TPR).

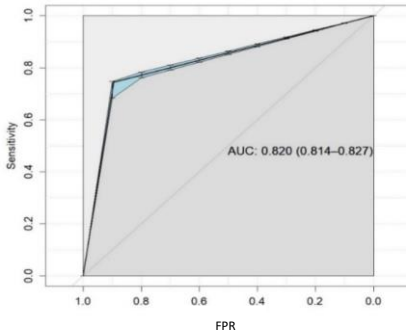


Figura 4.5. Curba Roc la testare pe setul test (B) a unui model antrenat pe date din setul A fără coloanele cu toate valorile lipsă.

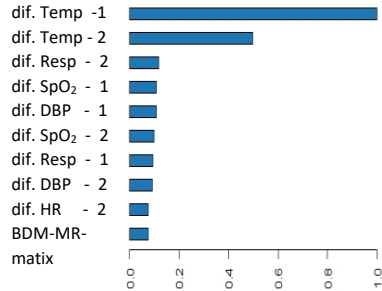


Figura 4.6. Importanța (pondera) variabilelor determinată în timpul învățării și cross-validării. Axa X – scara importanței (max = 1,0)

Conform terminologiei statistice tradiționale sistemul creat este un model de predicție multivariabil (cu 14 predictorii) în baza clasificării binomiale. Conform principiilor PROBAST (2019) [10] – este un model de predicție prognostică (unde în calitate de factori de prognozare servesc 6 parametri fiziologici și dinamica lor care generează 14 caracteristici), iar conform ghidului TRIPOD (2015) [9] – cercetarea actuală este de “Tip 3” ce presupune dezvoltarea unui model predictiv cu utilizarea unui set de date și evaluarea performanței lui utilizând date separate, ultima fiind numită și “validare externă”.

Sistemul a fost creat conform principiilor și urmând pașii descriși anterior. S-a efectuat validarea lui (10 folds cross-validation) și testarea pe setul-test. Performanța sistemului creat este ilustrată în figurile 4.3 – 4.6 și tabelul ce urmează:

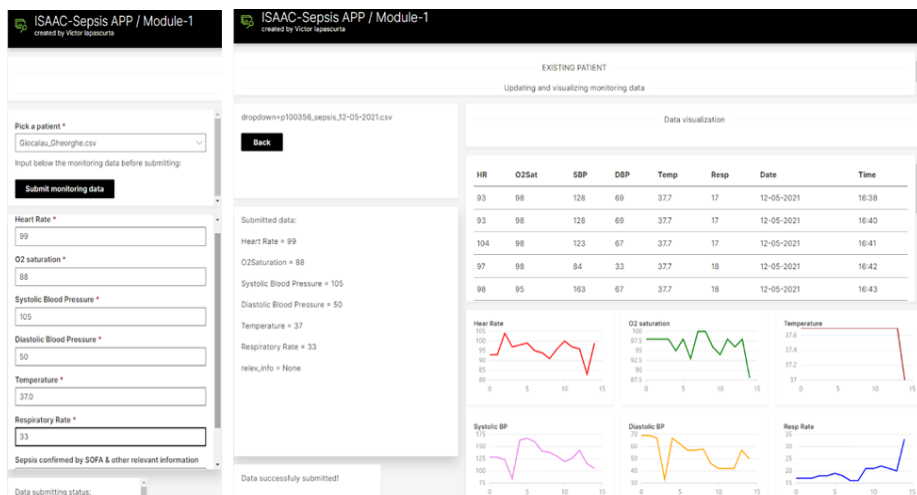
**Tabelul. 4.3. Sumar al performanței sistemului în baza algoritmului GBM**

<b>Metrică performanță</b>	<b>Cross-validare</b>	<b>Setul test-1/ validare internă</b>	<b>Setul test-2/ validare externă</b>
<i>Sensibilitate/ Recall/True Positive Rate (TPR)</i>	0.885 (0.016)	0.916 (0.898 – 0.932)	0.713 (0.701 – 0.725)
<i>Specifitate/ True Negative Rate (TNR)</i>	0.951 (0.009)	0.942 (0.931 – 0.952)	0.913 (0.908 – 0.918)
<i>Acuratețe diagnostică</i>	0.928 (0.006)	0.940 (0.931 – 0.948)	0.902 (0.897 – 0.906)
<i>Valoarea predictivă pozitivă/ Positive predictive value (PPV)</i>	0.915 (0.91-0.92)	0.605 (0.561 – 0.646)	0.332 (0.320 – 0.344)
<i>Valoare predictivă negativă/ Negative predictive value (NPV)</i>	0.931 (0.93-0.93)	0.992 (0.987 – 0.993)	0.981 (0.981 – 0.982)
<i>Rata fals pozitivă/ False Positive Rate (FPR)</i>	0.048	0.058	0.087
<i>Rata fals negativă/ False Negative Rate (FNR)</i>	0.121	0.084	0.287
<i>Raport de probabilitate pozitiv/ Positive likelihood ratio (LR+)</i>	18.31 (17.1-19.6)	15.85 (13.26 – 18.94)	8.20 (7.76 – 8.66)
<i>Raport de probabilitate negativ/ Negative likelihood ratio (LR-)</i>	0.13 (0.12-0.13)	0.09 (0.07 – 0.11)	0.31 (0.30 – 0.33)
<i>Raportul șanselor/ Diagnostic odds ratio (DOR)</i>	140.85	176.11	26.45
<i>Analiza ROC (AUC)</i>	0.970 (0.004)	0.929 (95% ÎI: 0.919 – 0.939)	0.813 (95% ÎI: 0.807 - 0,820)
<i>Analiza PR (prAUC)</i>	0.959 (0.004)	0.864	0.647

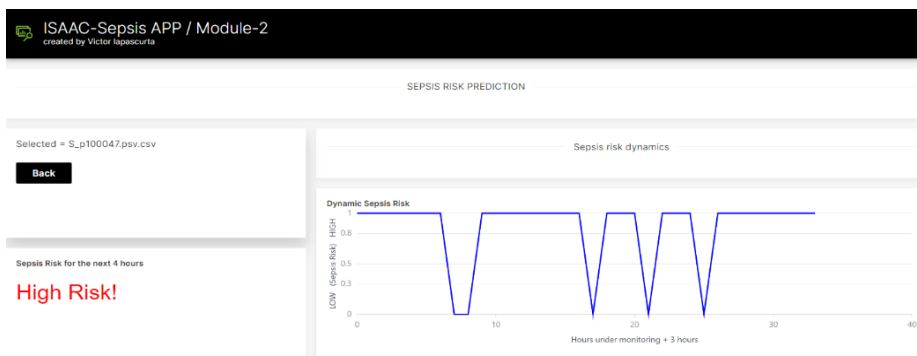
### 4.3. Incorporarea sistemului creat într-o aplicație pentru utilizare clinică.

Pentru utilizarea clinică a sistemului, el a fost realizat ca o aplicație software [28], care este comod de utilizat de către medicul ATI. Figurile 4.8 și 4.9 ilustrează interfața grafică a acestei aplicații.

S-a tins spre comoditatea aplicației pentru utilizator cu posibilitatea de a introduce datele la interval de o oră (care pot fi la necesitate descărcate ca fișier “.csv”), iar în cazul unor date acumulate anterior – e posibilă importarea datelor cu vizualizarea lor.



**Figura 4.7. Interfața grafică a aplicației pentru prezicerea timpurie a sepsisului: introducerea datelor și vizualizarea lor în formă de tabel și grafică**



**Figura. 4.8. Interfața grafică a aplicației pentru prezicerea timpurie a sepsisului: obținerea rezultatului predicției și vizualizarea dinamicii riscului**

Această aplicație poate fi utilizată în felul următor: (1) se introduc datele ce descriu cazul/pacientul (cei 6 parametri) – pot fi introduse orar sau sub formă de tabel (când sunt date pentru mai multe intervale orare); (2) datele pot fi vizualizate grafic în formă de serii

temporale. Odată cu acumularea a 3 ferestre orare automat se determină riscul de sepsis la pacientul vizat (“Risc înalt” vs “Risc neînsemnat”); (3) pe măsura avansării timpului după fiecare oră starea pacientului (în privința riscului sepsisului) este reevaluată cu afișarea gradului de risc pentru fiecare evaluare.

#### 4.4. Utilizarea aplicației pentru precizare continuă a riscului sepsisului

Pentru evaluarea performanței sistemului creat în privința urmării dinamicii riscului sepsisului în timp, prin eșantionare aleatorie au fost selectați câte 10 pacienți din setul test cu sepsis (confirmat prin metode tradiționale) și din setul cu alte patologii. Pentru uniformitate pentru fiecare caz selectat s-a extras o perioadă de observație de 11 ore. În cazul pacienților cu sepsis această perioadă include 8 ore până la confirmarea clinica a sepsisului și 3 ore – după. Pentru pacienții non-septici s-au extras 11 ore fără valori lipsă a parametrilor de interes. Rezultatul aplicării sistemului este ilustrat de Tab. 4.4.

Tabelul 4.4. **Precizarea riscului sepsisului în dinamică**

Pacient ID	ora 1	ora 2	ora 3	ora 4	ora 5	ora 6	ora 7	ora 8	ora 9
Pacienți cu 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
Pacienți fără sepsis									
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

*Notă: Rezultatul predicției: 1 – sepsis, 0 – non-sepsis. Sepsisul la septici - diagnosticat prin metode tradiționale la ora a 7-a*

## 5. EXPLICABILITATEA MODELELOR IA. DISCUȚII ȘI DIRECȚII DE CERCETARE DE VIITOR

### 5.1. Sumar al lucrului de cercetare efectuat. Discuția rezultatelor. Limitări

Anterior performanța modelului creat în acest studiu a fost comparată cu performanța sistemului InSight (vezi Tab. 4.1). O sursă credibilă pentru comparare poate fi “Ghidul internațional pentru managementul sepsisului și al socului septic” din noiembrie, 2021 [29], unde în premieră este menționat potențialul rol al învățării automate, care “poate îmbunătăți performanța instrumentelor de screening”, în baza unei meta-analize a 42623 de pacienți din șapte studii pentru previziunea sepsisului spitalicesc, care raportează o arie de sub curba ROC de 0,89 (95% ÎI, 0,86-0,92); sensibilitate - 81%; (95% ÎI, 80-81) și specificitate - 72% (95% ÎI, 72-72). Performanța sistemului creat în studiul curent este superioară după fiecare din metricele de performanță enumerate.

Una din principalele limitări ale sistemului de predicție creat în cadrul cercetării curente este legată de un subiect care la moment este în dezbatere. O meta-analiză recentă [30] a indicat că majoritatea pacienților cu COVID-19 grav bolnavi (78%) au îndeplinit criteriile Sepsis 3.0 pentru sepsis/șoc septic cu sindrom de detresă respiratorie acută (ARDS), ca fiind cea mai frecventă disfuncție de organ (88%). Datele în baza cărora a fost creat sistemul includ numai cazuri de sepsis bacterian.

Alte limitări previzibile la această etapă pot fi legate de diferența dintre datele din studiul curent și cele din secția ATI unde sistemul ar putea fi utilizat în viitor, cauzată de un spectru diferit de bolnavi septici, tehnicile diferite de monitorizare și tratament. O potențială soluție în acest caz ar fi crearea unui sistem nou, folosind experiența acumulată.

### 5.2. Aspecte de explicabilitate a modelelor IA

Explicabilitatea modelelor IA are o importanță majoră pentru înțelegerea cum modelul funcționează, ceea ce influențează adoptarea lui de comunitatea medicală. Acest domeniu poartă denumirea XAI (eXplainable Artificial Intelligence) și cu toate ca este la început de cale se dezvoltă rapid. Un exemplu de abordare este estimarea importanței variabilelor pentru predicție, aceasta fiind prezentat anterior (Fig. 4.6). Cu țel similar se utilizează metodele PD (Partial Dependence), ICE (Individual Conditional Expectations), LIME (Local Interpretable Model-agnostic Explanations), etc., care sunt descrise în textul de bază al tezei.

O abordare mai complexă, care include elemente din metodele descrise este metoda SHAP (SHapley Additive exPlanations). Diagrama SHAP pentru un modelul GBM din studiul curent este prezentată în Fig. 5.1. Ea îmbină importanța caracteristicilor (poziția caracteristicii pe axa Y; cu cât mai sus – cu atât mai important) cu efectele caracteristicilor (impactul asupra predicției - pe axa X; spre dreapta de linia verticală “0” – în favoarea sepsisului, iar spre stânga – în defavoarea sepsisului și cu cât mai departe de “0” – cu atât

impactul este mai mare). Culoarea reprezintă valoarea normalizată a caracteristicii de la mică la mare. Astfel, în cazul temperaturii, o diferență de mărime cât mai mare (spre exemplu, când temperatura corpului crește esențial și rapid) se traduce printr-un risc de sepsis jos, fapt cunoscut (și aparent înțeles de la sine) de către clinician. Mergând de la contrariu, dinamica modestă a temperaturii ar putea denota situația când forțele de apărare ale organismului pacientului septic sunt epuizate, nefiind în stare să asigure o reacție adecvată (la agentul patogen), exprimată în acest caz prin reacția piretică “atenuată”.

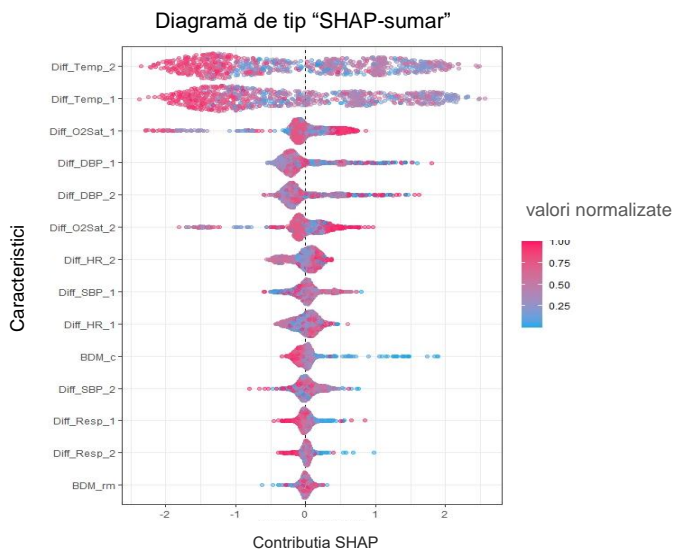


Figura 5.1. Diagrama SHAP-sumar pentru modelul GBM

### 5.3. Transferabilitatea sistemelor IA. Direcții de cercetare de viitor

Un aspect separat discutat în literatură [3, 21] îl constituie transferabilitatea SÎA, adică folosirea unui sistem creat în baza datelor dintr-un spital în alt spital. De obicei în acest caz performanța este mai mică. În cadrul studiului curent sistemul creat în baza setului “A” a fost testat pe date din setul “B” cu o performanță de 0.813 (95% ÎI: 0.807 - 0,820, după AUC ROC). Soluții pentru rezolvarea problemei cu performanța în acest caz sunt propuse cel puțin două: (a) crearea unui model nou în baza datelor noi, folosind cunoștințele acumulate la elaborarea modelelor similare în alte studii (adică replicarea unui studiu deja realizat pe date noi); (b) învățarea prin transfer, când un model deja creat este suplimentar antrenat pe date noi.



Eventualele direcții de cercetare pe viitor ar include: (a) calibrarea sistemului de precizie a sepsisului în condiții clinice în Republica Moldova; (b) îmbunătățirea performanței sistemului prin învățarea prin transfer în baza datelor locale; (c) studierea interacțiunii specialistului ATI cu sistemul în condiții clinice

## CONCLUZII GENERALE

1. Fenomenul pătrunderii tehnologiilor IA în ATI este unul relativ nou, însă de proporții și cu perspective greu de prezis. Aspectele activității medicului anesteziolog-reanimatolog care pot beneficia de aceste tehnologii sunt numeroase, pornind de la diagnosticul/monitorizarea/predicția diferitor fațete ale stării critice. Conform literaturii de specialitate, cea mai frecvent utilizată la etapa actuală tehnologie de IA este învățarea automată supravegheată, care se regăsește în peste 90% de lucrări.
2. Algoritmii elaborați și utilizați în cercetarea curentă pentru restabilirea datelor biomedicale continui cu valori lipsă, parvenite de la pacienți cu sepsis, furnizează rezultate plauzibile, fapt confirmat de performanța înaltă a sistemului de predicție. În baza rezultatelor obținute (modificarea neînsemnată a performanței sistemului la testarea lui pe date-restabilite cu algoritmul vs cazuri complete) se poate afirma că dacă și există părtinire introdusă de algoritm, ea este minoră.
3. Metoda de reprezentare și procesare a datelor în formă de serii temporale a parametrilor fiziologici la pacienții septici, cu ajutorul metricii de complexitate algoritmică Kolmogorov - Chaitin, care a fost în premieră utilizată cu așa scop în cercetarea curentă, este o metodă reușită de prelucrare a datelor pentru învățarea automată, fapt demonstrat de rezultatele finale obținute.
4. Cei șase parametri fiziologici, reprezentați prin datele - valori în timp ale lor, au fost suficienți pentru crearea unui sistem de predicție performant, utilizând metodele de procesare descrise mai sus. De importanță este și aceea că parametrii respectivi sunt de rutină monitorizați în ATI, inclusiv și în Republica Moldova.
5. Performanță înaltă (mai mare de 92% vs 89% în ultimul Ghid pentru managementul sepsisului și șocului septic, 2021) a sistemului de predicție creat face rațională cercetarea de mai departe cu calibrarea sistemului în UTI țintă din Moldova și posibilitatea ulterioară de utilizare în practica clinică pentru predicția timpurie a unui risc iminent de dezvoltare a sepsisului cu inițierea în timp util a tratamentului (în primul rând aspectele care se referă la infecție). Aceasta poate avea un impact pozitiv asupra rezultatelor tratamentului, concomitent asigurând suport decizional pentru anesteziolog-reanimatolog.

## RECOMANDĂRI PRACTICE

1. Atunci când cercetarea are de a face cu date lipsă și în special în cazul datelor reprezentate prin serii temporale, care în ATI sunt ubicuitare, utilizarea algoritmului de reconstrucție-restabilire a datelor ar putea facilita lucrul cercetătorului și contribui la amplificarea performanței lui.
2. Complexitatea algoritmică ca o posibilă metrică a proceselor ce derulează în subiectul de studiu și reprezentarea datelor obținute cu ajutorul ei (spre ex. cu ajutorul metodei de decompoziție în blocuri – BDM) are un potențial încă puțin explorat. De aceea experimentarea cu aceste instrumente chiar și în studii în afara domeniului IA este binevenită. Ea ar putea reuși să suplimentă și chiar să întrecă metodele statistice tradiționale.
3. În pofida faptului că procesul de raționament al clinicianului probabil diferă esențial de logica computațională, iar performanțele de calcul nu pot fi comparate cu cele ale unui calculator, atragerea atenției asupra dinamicii factorilor cu cea mai mare valoare predictivă determinați de model (ex. dinamica subtilă a temperaturii, a saturației sângelui periferic cu oxigen – SpO<sub>2</sub>, a presiunii arteriale diastolice, etc) ar putea fortifica abilitățile de prognozare (subconștiente) ale specialistului uman în cadrul managementului clinic al pacienților cu sepsis. În acest context importante sunt și aspectele legate de explicabilitatea modelelor.
4. La elaborarea unor sisteme de suport decizional similare, pentru utilizare practică în ATI este recomandabil de folosit ghidul TRIPOD și principiile PROBAST, care de fapt sunt aplicabile și pentru studii ce nu țin de IA, de rând cu raționamentul clinic și bunele practici ale medicinei.
5. Numărul de tehnologii și dispozitive în baza învățării automate și IA va crește și în acest context este necesară o atitudine corectă a practicianului față de aceste dispozitive. Una reușită pare a fi acceptarea lor în calitate de asistenți, și nicidecum - rivali, cu atât mai mult ca cel care ia decizia finală și o realizează, cel puțin la moment și în viitorul previzibil, rămâne să fie specialistul uman.

## BIBLIOGRAFIE

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. Finkelsztejn, 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*. [citată: 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]. [citată: 24.12.2022]. Disponibil: [www.algorithmicdynamics.net/software.html](http://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]. [citată 24.12.2022]. Disponibil: <https://towardsdatascience.com/workflow-of-a-machine-learning-project-ec1dba419b94>
28. **Iapăscurtă, V.**, Belfi, 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.

## LISTA LUCRĂRILOR ȘTIINȚIFICE PUBLICATE LA TEMA TEZEI

- **Articole în reviste științifice peste hotare**
  - ✓ **articole în reviste ISI, SCOPUS și alte baze de date internaționale**
  - 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.,** Belii, 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
- **Articole în reviste științifice naționale acreditate**
  - ✓ **articole în reviste de categoria B**
  - 4. **Iapăscurtă, V.** Managementul pacientului în stare critică actualități și perspective. In: *Curierul Medical*, Chișinău, 1/1997, pp.15-18. ISSN 1875-0666
  - 5. Belii, 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
- **Articole în lucrările conferințelor științifice:**
  - ✓ **internaționale desfășurate în Republica 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 12<sup>th</sup> International Conference on Electronics, Communications and Computing, October 20-21, 2022, Chișinău*, pp. 1-5.

- **Rezumate/abstracte/teze în lucrările conferințelor științifice naționale și internaționale**

9. **Iapăscurtă V.**, Ghereg V., Revițchi V. Tehnologii informaționale în terapia intensivă. In: *Materialele conferinței științifice anuale a colaboratorilor și studenților, Universitatea de Stat de Medicină și Farmacie, Chișinău*, 1995, pp. 92-92.
10. Sevastianov E., Ghereg V., **Iapăscurtă V.** Automatizarea procesului de diagnostic diferențial cu utilizarea indicelui leucocitar de intoxicație la pacienții cu procese toxico-septice. In: *Materialele conferinței științifice anuale a colaboratorilor și studenților, 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 conferinței științifice anuale a colaboratorilor și studenților, Universitatea de Stat de Medicină și Farmacie, Chișinău*, 1997, pp. 223-223.
14. **Iapăscurtă V.**, Belii 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 și sănătate: calitate, excelență și performanță, 19-21 octombrie, 2022, Chișinău. MJHS, 29(3), 2022, pp. 250-250. ISSN 2345-1467.
16. Monastîrșchi S., **Iapăscurtă V.** Modele sistemice dinamice pentru anestezia clinică (pe exemplul propofolului). *Conferința științifică anuală a USMF „N. Testemițanu”*. Cercetarea în biomedicină și sănătate: calitate, excelență și performanță, 19-21 octombrie, 2022, Chișinău. MJHS, 29(3), 2022, pp. 314-314. ISSN 2345-1467.

- **Certificate de inovator**

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 Institutul 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 Institutul 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 artificială, 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 inteligență artificială și, în special, a învățării automate, ca una din tehnologiile de bază ale IA la etapa actuală, utilizate în anestezioterapie 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 ambigue. Metodele propuse pentru reconstrucția și 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 sepsis prediction 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. Выявленные 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.

## GLOSAR DE TERMENI TEHNICI

**Algoritm** - suită de raționamente sau operații având ca scop rezolvarea unor probleme.

**Algoritm de învățare automată** - programul folosit pentru a învăța un model de învățare automată din date.

**Caracteristici** - intrările utilizate pentru predicție sau clasificare. O caracteristică este o coloană din setul de date. Se presupune că ele sunt interpretabile, ceea ce înseamnă că este ușor de înțeles ce înseamnă. Dar dacă este greu de înțeles caracteristicile de intrare, este și mai greu de înțeles ce face modelul.

**GBM** (*eng.* Gradient boosting machine) – algoritm de învățare automată în baza arborilor decizionali cu utilizarea metodei de stimulare a coborârii (diminuării) gradientului.

**Inteligență artificială** - știința dezvoltării sistemelor informatice care pot îndeplini sarcini care necesită în mod normal inteligență umană.

**Învățare automată (ÎA)** - set de metode care permit computerelor să învețe din date pentru a face și îmbunătăți predicții.

**Învățare automată automatizată** – set de metode prin care computerul aplică pe datele livrate mai mulți algoritmi ÎA și determină automat cele mai performante modele de învățare automată.

**Model de învățare automată** - programul învățat care mapează intrările la predicții.

**Model “cutie neagră”** – (*eng.* “Black Box” Model) - sistem care nu își dezvăluie mecanismele interne. În învățarea automată, „cutia neagră” descrie modele care nu pot fi înțelese analizând parametrii lor.

**Model “cutie de sticlă”** - (*eng.* “Glass Box” Model) - model interpretabil.

**Predicția** - ceea ce modelul de învățare automată „ghicește” care ar trebui să fie valoarea țintă pe baza caracteristicilor date.

**Rețele Neuronale Artificiale** – (*eng.* Artificial Neural Networks, ANN) - algoritm de învățare automată, care utilizează similaritatea cu funcționarea neuronilor crierului.

**Set de date** - tabel cu datele din care mașina învață. Setul de date conține caracteristicile și ținta de prezis. Când este folosit pentru a induce un model, setul de date se numește *date de antrenament*.

**IAPĂSCURTĂ Victor**

**PREZICEREA TIMPURIE A SEPSISULUI CU AJUTORUL  
UNEI APLICAȚII PROPRII ELABORATE ÎN BAZA ÎNVĂȚĂRII AUTOMATE  
(INTELIGENȚĂ ARTIFICIALĂ)**

**321.19 – ANESTEZIOLOGIE ȘI TERAPIE INTENSIVĂ**

Rezumatul tezei de doctor în științe medicale

---

Aprobat spre tipar 15.06.2023  
Hârtie ofset, tipar digital  
Coli de tipar: 2,0

Formatul A5  
Tiraj 30 ex.  
Comanda nr. 21

---

Tiparit la "SRL Sirius"  
Chisinau, str. A. Lăpușneanu 2, tel: 022 23 23 52