

Rezumat română

MACHINE LEARNING ÎN AUTOMATIZAREA PROCESELOR MEDICALE

Cuprins

1. INTRODUCERE	1
2. ASPECTE ALE INTELIGENȚEI ARTIFICIALE ÎN ÎNGRIJIRILE DE SĂNĂTATE	4
2.1. GENERALITĂȚI	4
2.1.1. Conceptul de Inteligență artificială	4
2.1.2. Repere istorice în evoluția Inteligenței artificiale	4
2.2. SUBDOMENIILE INTELIGENȚEI ARTIFICIALE	5
2.2.1. Sisteme expert	6
2.2.2. Automatizarea proceselor robotizate	6
2.2.3. Procesarea limbajului natural	7
2.2.4. Învățare automată	7
2.2.4.1. Învățare supravegheată	7
2.2.4.2. Învățare nesupravegheată	8
2.2.4.3. Învățare prin consolidare	10
2.2.4.4. Învățare profundă	11
2.2.5. Viziune computerizată	16
2.3. MACHINE LEARNING ȘI DEEP LEARNING ÎN MEDICINĂ	19
2.3.1. Exemple ale utilizării Machine Learning în medicină	19
2.3.2. Exemple ale utilizării Deep Learning în medicină	19
2.3.3. Cerințele de tranziție de la Machine Learning la Deep Learning	20
2.3.4. De la date clinice la big data	21
2.3.4.1. Date din imagini medicale	21
2.3.4.2. Date de la Fișa electronică sau Dosarul electronic al pacientului	23
2.3.4.3. Big data în medicina personalizată	23
2.3.4.4. Predicția structurii și funcției proteinelor	24
2.3.4.5. Descoperirea și dezvoltarea medicamentelor	24
2.4. MODALITĂȚI PRIN CARE ÎNVĂȚAREA AUTOMATĂ POATE AJUTA MUNCA CLINICIENILOR	24
2.4.1. Prognoză	24
2.4.2. Diagnostic	25
2.4.3. Tratament	25
2.4.4. Fluxul de lucru al clinicianului	25
2.4.5. Extinderea disponibilității expertizei clinice	25
2.5. INTELIGENȚA ARTIFICIALĂ GENERATIVĂ ÎN MEDICINĂ	25
2.5.1. Suport administrativ clinic	26
2.5.2. Suport pentru decizii clinice	26
2.5.3. Implicarea pacientului	26
2.5.4. Generarea de date sintetice	27
2.5.5. Educație profesională	27
2.5.6. Tehnologii Chatbot	27
2.6. ASPECTE GENERALE ALE AUTOMATIZĂRII PROCESELOR ROBOTIZATE	29
2.6.1. Tipuri de Automatizare a Proceselor Robotizate	29
2.6.2. Automatizarea Proceselor Robotizate în sarcinile repetitive din front office	30
2.6.3. Automatizarea Proceselor Robotizate în sarcinile repetitive din back-office	30
2.6.4. Avantajele Automatizării Proceselor Robotizate	30
2.6.5. Dezavantajele Automatizării Proceselor Robotizate	31
2.7. SISTEME DE ASISTENȚĂ PENTRU DECIZII CLINICE	31
2.7.1. Funcții ale sistemelor de asistență pentru decizii clinice	32
2.7.1.1. Siguranța pacientului	32

2.7.1.2.	Managementul clinic	32
2.7.1.3.	Controlul costurilor	32
2.7.1.4	Funcții administrative	32
2.7.1.5.	Suport pentru diagnosticare	33
2.7.1.6.	Suport decizional orientat către pacient	34
2.7.2.	Probleme ale folosirii sistemelor de asistență pentru decizii clinice	34
2.7.2.1.	Fluxuri de lucru fragmentate	34
2.7.2.2.	Alertele nepotrivite	34
2.7.2.3.	Impactul asupra abilităților utilizatorilor	35
2.7.2.4.	Sistemul de asistență pentru decizii clinice poate depinde de cunoștințele de informatică	35
2.7.2.5.	Întreținerea sistemului și a conținutului	35
2.7.2.6.	Impactul operațional al calității slabe a datelor și al conținutului incorect	35
2.7.2.7.	Lipsa transportabilității (portabilității) și a interoperabilității	36
2.7.2.8.	Provocări financiare	36
2.7.3.	Sisteme de asistență pentru deciziile clinice în asistența medicală primară	36
2.8.	DISPOZITIVE INTELIGENTE PURTABILE CA INSTRUMENTE DE DIAGNOSTICARE DIGITALĂ	37
2.9.	OBSERVAȚII	39
3.	CADRUL CONCEPTUAL AL CERCETĂRII EFECTUATE	41
3.1.	MANAGEMENTUL INCERTITUDINII MEDICALE ÎN MEDICINA MODERNĂ	41
3.2.	NECESITATEA ALGORITMILOR MEDICALI	42
3.3.	OPTIMIZAREA MUNCII MEDICULUI	43
3.4.	OBSERVAȚII	44
4.	METODOLOGIA ȘI REZULTATELE CERCETĂRII	46
4.1.	CONSIDERENTE PRELIMINARE	46
4.1.1.	Seturile de date folosite în algoritmii de Machine Learning și Deep Learning	48
4.1.2.	Considerente legate de algoritmii de Machine Learning folosiți	49
4.1.2.1.	K-Nearest Neighbors	49
4.1.2.2.	Decision Tree	50
4.1.2.3.	Support Vector Machine	51
4.1.2.4.	Logistic Regression	51
4.1.2.5.	Random Forest	52
4.1.2.6.	Regresie liniară	53
4.1.2.7.	Regresie polinomială	53
4.1.2.8.	Gradient Boosting	54
4.1.2.9.	Extreme Gradient Boosting	55
4.1.3.	Observații și discuții	56
4.2.	INTEGRAREA AUTOMATIZĂRII PROCESELOR ROBOTIZATE ȘI MACHINE LEARNING ÎN PROCESELE MEDICALE	56
4.3.	APLICABILITATEA MACHINE LEARNING ÎN PREDICȚIA PREDIABETULUI	58
4.3.1.	Considerații despre prediabet	58
4.3.2.	Metodologie	59
4.3.2.1.	K Nearest Neighbors	60
4.3.2.2.	Decision Tree	60
4.3.2.3.	Support Vector Machine	61
4.3.2.4.	Logistic Regression	61
4.3.2.5.	Random Forest	61
4.3.3.	Rezultate	62
4.3.3.1.	K Nearest Neighbors	62
4.3.3.2.	Decision Tree	62
4.3.3.3.	Support Vector Machine	63

4.3.3.4. Logistic Regression	64
4.3.3.5. Random Forest	64
4.3.4. Observații și discuții	66
4.4. PREDICȚIA PREDIABETULUI FOLOSIND O REȚEA NEURONALĂ CU MAI MULTE STRATURI	68
4.4.1. Metodologie	68
4.4.2. Rezultate	68
4.4.3. Observații și discuții	70
4.5. APLICABILITATEA UNOR ALGORITMI DE ÎNVĂȚARE AUTOMATĂ ÎN PREDICȚIA DIABETULUI ZAHARAT DE TIP 2	70
4.5.1. Introducere	70
4.5.2. Repere din literatura de specialitate referitoare la predicția diabetului zaharat de tip 2 cu ajutorul algoritmilor de inteligență artificială	72
4.5.2.1. Modele care folosesc PIDD ca set de date	72
4.5.2.2. Modele care folosesc seturi de date diferite de PIDD	77
4.5.2.3. Modele care folosesc alte abordări	80
4.5.3. Metodologie	83
4.5.4. Rezultate	84
4.5.4.1. Decision Tree	84
4.5.4.2. Logistic Regression	86
4.5.4.3. Support Vector Machine	87
4.5.4.4. K Nearest Neighbors	87
4.5.4.5. Random Forest	89
4.5.4.6. XGBoost	91
4.5.5. Observații și discuții	92
4.6. APLICABILITATEA UNOR ALGORITMI DE ÎNVĂȚARE AUTOMATĂ SUPRAVEGHEATĂ ÎN STADIUL TERMINAL AL BOLII HEPATICE	94
4.6.1. MELD - Model for End-stage Liver Disease	94
4.6.2. Câteva repere din literatură privind folosirea algoritmilor de inteligență artificială în stadiul terminal al bolii de ficat	95
4.6.3. Metodologie	96
4.6.4. Rezultate	96
4.6.4.1. Decision Tree	96
4.6.4.2. K-Nearest Neighbors	99
4.6.4.3. Logistic Regression	100
4.6.5. Observații și discuții	101
4.7. SOLUȚIE PENTRU AUTOMATIZAREA ACTULUI MEDICAL ÎNTR-UN CABINET DE MEDICINĂ DE FAMILIE	102
4.7.1. Extragerea datelor din buletine medicale și generarea de alerte medicale	103
4.7.1.1. Introducere	103
4.7.1.2. Scopul soluției informatice propuse	104
4.7.1.3. Metodologie	104
4.7.1.4. Materiale	107
4.7.1.5. Rezultate	109
4.7.1.6. Observații și discuții	116
4.7.2. Harta colorimetrica individuala pentru monitorizarea alcoolicii cronici	118
4.7.2.1. Introducere	118
4.7.2.2. Materiale și metode	118
4.7.2.3. Rezultate	118
4.7.2.4. Observații și discuții	120
4.7.3. Impactul modificării greutateii corporale asupra parametrilor metabolici	120
4.7.3.1. Introducere	120

4.7.3.2. Scopul studiului	121
4.7.3.3. Metodologie	121
4.7.3.4. Rezultate	126
4.7.3.5. Observații și discuții	134
5. CONCLUZII FINALE	138
Postscriptum	142
Bibliografie	143
Lista acronimelor folosite frecvent	155
Lista tabelelor	156
Lista figurilor	157

Cuvinte cheie: Machine Learning, Automatizare, Diagnostic medical, Predicție riscuri medicale, Monitorizare

Teza explorează integrarea metodelor de automatizare și a algoritmilor de Machine Learning în domeniul sănătății, cu scopul de a îmbunătăți procesul de diagnosticare, managementul riscurilor medicale și eficientizarea fluxurilor de lucru. Lucrarea propune o soluție software hibridă care combină două direcții esențiale: predicția riscurilor pentru diverse afecțiuni și automatizarea activităților administrative și medicale.

O componentă poate extrage automat datele din buletinele de analize ale pacienților și le salvează într-o bază de date locală. Aceste date sunt apoi analizate cu ajutorul unor algoritmi de inteligență artificială și clasici, oferind sprijin în diagnosticarea și prognosticul afecțiunilor precum diabetul zaharat de tip 2, bolile cardiovasculare, hepatice și renale.

Soluția software permite generarea automată de grafice care reflectă evoluția parametrilor biologici ai unui pacient în timp, precum și o hartă colorimetrică ce vizualizează variațiile stării de sănătate bazate pe parametri specifici. Această reprezentare vizuală are un impact educativ și motivațional important, în special în cazul pacienților cu boli cronice sau stil de viață nesănătos (ex: alcoolism), încurajându-i să adere la tratamente și să adopte comportamente sănătoase.

Soluția propusă contribuie, de asemenea, la personalizarea actului medical: oferă estimări privind impactul modificărilor stilului de viață (ex: scădere în greutate) asupra valorilor biologice și permite ajustarea recomandărilor medicale, nutriționale și terapeutice în mod individualizat.

Beneficiile majore ale aplicației se regăsesc în reducerea erorilor umane, prin automatizarea deciziilor bazate pe date precise, creșterea eficienței, economisind timp pentru cadrele medicale prin digitalizarea proceselor, răspuns rapid, prin monitorizare continuă și generare automată de alerte la modificări semnificative în starea pacientului.

Originalitatea cercetării constă în integrarea completă a predicției clinice cu automatizarea proceselor medicale. Soluția software nu doar stochează date, ci oferă și analize predictive și recomandări automate, textuale și de tip grafic, inclusiv prin utilizarea hărților colorimetrice ca instrument vizual inovator și aplicabil pe diverse categorii de pacienți.

În concluzie, teza aduce o contribuție semnificativă prin dezvoltarea unui sistem integrat, predictiv și automatizat, ce are potențialul de a transforma pozitiv modul în care este oferită îngrijirea medicală, atât din perspectiva specialiștilor din domeniu, cât și a pacienților.

Rezumat engleză

MACHINE LEARNING IN THE AUTOMATION OF MEDICAL PROCESSES

Table of Contents

1.INTRODUCTION – 1
2.ASPECTS OF ARTIFICIAL INTELLIGENCE IN HEALTHCARE – 4
2.1. General Considerations – 4
2.1.1. The Concept of Artificial Intelligence – 4
2.1.2. Historical Milestones in the Evolution of AI – 4
2.2. Subfields of Artificial Intelligence – 5
2.2.1. Expert Systems – 6
2.2.2. Robotic Process Automation – 6
2.2.3. Natural Language Processing – 7
2.2.4. Machine Learning – 7
2.2.4.1. Supervised Learning – 7
2.2.4.2. Unsupervised Learning – 8
2.2.4.3. Reinforcement Learning – 10
2.2.4.4. Deep Learning – 11
2.2.5. Computer Vision – 16
2.3. Machine Learning and Deep Learning in Medicine – 19
2.3.1. Examples of Machine Learning Applications in Medicine – 19
2.3.2. Examples of Deep Learning Applications in Medicine – 19
2.3.3. Transition Requirements from Machine Learning to Deep Learning – 20
2.3.4. From Clinical Data to Big Data – 21
2.3.4.1. Medical Imaging Data – 21
2.3.4.2. Data from Electronic Health Records – 23
2.3.4.3. Big Data in Personalized Medicine – 23
2.3.4.4. Prediction of Protein Structure and Function – 24
2.3.4.5. Drug Discovery and Development – 24
2.4. Ways in Which Machine Learning Can Support Clinicians' Work – 24
2.4.1. Prognosis – 24
2.4.2. Diagnosis – 25
2.4.3. Treatment – 25
2.4.4. Clinician Workflow – 25
2.4.5. Expanding Access to Clinical Expertise – 25
2.5. Generative AI in Medicine – 25
2.5.1. Administrative Support – 26
2.5.2. Clinical Decision Support – 26
2.5.3. Patient Engagement – 26
2.5.4. Synthetic Data Generation – 27
2.5.5. Professional Education – 27
2.5.6. Chatbot Technologies – 27
2.6. General Aspects of Robotic Process Automation (RPA) – 29
2.6.1. Types of RPA – 29
2.6.2. RPA in Repetitive Front-Office Tasks – 30
2.6.3. RPA in Repetitive Back-Office Tasks – 30
2.6.4. Advantages of RPA – 30
2.6.5. Disadvantages of RPA – 31
2.7. Clinical Decision Support Systems (CDSS) – 31
2.7.1. Functions of CDSS – 32

2.7.1.1. Patient Safety – 32	32
2.7.1.2. Clinical Management – 32	32
2.7.1.3. Cost Control – 32	32
2.7.1.4. Administrative Functions – 32	32
2.7.1.5. Diagnostic Support – 33	33
2.7.1.6. Patient-Centered Decision Support – 34	34
2.7.2. Issues with CDSS Usage – 34	34
2.7.2.1. Fragmented Workflows – 34	34
2.7.2.2. Inappropriate Alerts – 34	34
2.7.2.3. Impact on Users' Skills – 35	35
2.7.2.4. Dependence on IT Knowledge – 35	35
2.7.2.5. System and Content Maintenance – 35	35
2.7.2.6. Poor Data Quality and Incorrect Content – 35	35
2.7.2.7. Lack of Portability and Interoperability – 36	36
2.7.2.8. Financial Challenges – 36	36
2.7.3. CDSS in Primary Healthcare – 36	36
2.8. Smart Wearable Devices as Digital Diagnostic Tools – 37	37
2.9. Observations – 39	39
3.CONCEPTUAL FRAMEWORK OF THE RESEARCH – 41	41
3.1. Managing Medical Uncertainty in Modern Medicine – 41	41
3.2. The Need for Medical Algorithms – 42	42
3.3. Optimizing the Physician's Work – 43	43
3.4. Observations – 44	44
4.RESEARCH METHODOLOGY AND RESULTS – 46	46
4.1. Preliminary Considerations – 46	46
4.1.1. Datasets Used in ML and DL Algorithms – 48	48
4.1.2. Considerations Regarding the ML Algorithms Used – 49	49
4.1.2.1. K-Nearest Neighbors – 49	49
4.1.2.2. Decision Tree – 50	50
4.1.2.3. Support Vector Machine – 51	51
4.1.2.4. Logistic Regression – 51	51
4.1.2.5. Random Forest – 52	52
4.1.2.6. Linear Regression – 53	53
4.1.2.7. Polynomial Regression – 53	53
4.1.2.8. Gradient Boosting – 54	54
4.1.2.9. Extreme Gradient Boosting – 55	55
4.1.3. Observations and Discussions – 56	56
4.2. Integrating RPA and ML into Medical Processes – 56	56
4.3. Applicability of ML in Prediabetes Prediction – 58	58
4.3.1. Considerations on Prediabetes – 58	58
4.3.2. Methodology – 59	59
4.3.2.1. K-Nearest Neighbors – 60	60
4.3.2.2. Decision Tree – 60	60
4.3.2.3. Support Vector Machine – 61	61
4.3.2.4. Logistic Regression – 61	61
4.3.2.5. Random Forest – 61	61
4.3.3. Results – 62	62
4.3.3.1. K-Nearest Neighbors – 62	62
4.3.3.2. Decision Tree – 62	62
4.3.3.3. Support Vector Machine – 63	63
4.3.3.4. Logistic Regression – 64	64
4.3.3.5. Random Forest – 64	64

4.3.4. Observations and Discussions –	66
4.4. Prediabetes Prediction Using a Multilayer Neural Network –	68
4.4.1. Methodology –	68
4.4.2. Results –	68
4.4.3. Observations and Discussions –	70
4.5. Applicability of ML Algorithms in Type 2 Diabetes Prediction –	70
4.5.1. Introduction –	70
4.5.2. Literature Overview on AI-Based Type 2 Diabetes Prediction –	72
4.5.2.1. Models Using PIDD Dataset –	72
4.5.2.2. Models Using Other Datasets –	77
4.5.2.3. Models Using Alternative Approaches –	80
4.5.3. Methodology –	83
4.5.4. Results –	84
4.5.4.1. Decision Tree –	84
4.5.4.2. Logistic Regression –	86
4.5.4.3. Support Vector Machine –	87
4.5.4.4. K-Nearest Neighbors –	87
4.5.4.5. Random Forest –	89
4.5.4.6. XGBoost –	91
4.5.5. Observations and Discussions –	92
4.6. Applicability of Supervised ML Algorithms in End-Stage Liver Disease –	94
4.6.1. MELD – Model for End-stage Liver Disease –	94
4.6.2. Literature Highlights on AI in End-stage Liver Disease –	95
4.6.3. Methodology –	96
4.6.4. Results –	96
4.6.4.1. Decision Tree –	96
4.6.4.2. K-Nearest Neighbors –	99
4.6.4.3. Logistic Regression –	100
4.6.5. Observations and Discussions –	101
4.7. Medical Automation Solution for a Family Medicine Practice –	102
4.7.1. Data Extraction from Medical Reports and Medical Alerts Generation –	103
4.7.1.1. Introduction –	103
4.7.1.2. Objective of the Proposed Software Solution –	104
4.7.1.3. Methodology –	104
4.7.1.4. Materials –	107
4.7.1.5. Results –	109
4.7.1.6. Observations and Discussions –	116
4.7.2. Individual Colorimetric Map for Monitoring Chronic Alcoholics –	118
4.7.2.1. Introduction –	118
4.7.2.2. Materials and Methods –	118
4.7.2.3. Results –	118
4.7.2.4. Observations and Discussions –	120
4.7.3. Impact of Weight Change on Metabolic Parameters –	120
4.7.3.1. Introduction –	120
4.7.3.2. Study Objective –	121
4.7.3.3. Methodology –	121
4.7.3.4. Results –	126
4.7.3.5. Observations and Discussions –	134
5.FINAL CONCLUSIONS –	138
Postscriptum –	142
Bibliography –	143
List of Frequently Used Acronyms –	155

List of Tables – 156

List of Figures – 157

Keywords: Machine Learning, Automation, Medical Diagnosis, Medical Risk Prediction, Monitoring

The thesis explores the integration of automation methods and Machine Learning algorithms in the healthcare field, with the goal of improving the diagnostic process, managing medical risks, and optimizing workflow efficiency. The work proposes a hybrid software solution that combines two essential directions: risk prediction for various medical conditions and the automation of both administrative and clinical activities.

One component can automatically extract data from patients' blood test reports and store it in a local database. This data is then analyzed using both classical and artificial intelligence algorithms, providing support in diagnosing and forecasting conditions such as type 2 diabetes, cardiovascular, liver, and kidney diseases.

The software solution allows for the automatic generation of graphs that reflect the evolution of a patient's biological parameters over time, as well as a colorimetric map that visualizes health status variations based on specific parameters. This visual representation has a significant educational and motivational impact, especially for patients with chronic illnesses or unhealthy lifestyles (e.g., alcoholism), encouraging them to adhere to treatments and adopt healthier behaviors.

The proposed solution also contributes to the personalization of medical care: it offers estimates of how lifestyle changes (e.g., weight loss) impact biological values and allows for individualized adjustment of medical, nutritional, and therapeutic recommendations.

The major benefits of the application include reducing human errors through automated decision-making based on accurate data, increasing efficiency by saving time for medical staff through process digitization, and ensuring a fast response by continuously monitoring patient status and automatically generating alerts when significant changes are detected.

The originality of the research lies in the complete integration of clinical prediction with the automation of medical processes. The software solution not only stores data but also provides predictive analyses and automatic recommendations, both textual and graphical, including the use of colorimetric maps as an innovative visual tool applicable to various categories of patients.

In conclusion, the thesis makes a significant contribution through the development of an integrated, predictive, and automated system, with the potential to positively transform the way medical care is delivered, both from the perspective of healthcare professionals and that of patients.