„CAROL DAVILA" UNIVERSITY OF MEDICINE AND PHARMACY, BUCHAREST

DOCTORAL SCHOOL OF MEDICINE

LEVERAGING PREDICTIVE ANALYTICS TO IMPROVE PATIENT OUTCOMES

DATA SCIENCE METHODS AND APPLICATIONS IN HEALTHCARE

THESIS SUMMARY

Scientific Coordinator:

PROF. UNIV. DR. SEBASTIAN IONESCU

PhD Candidate:

ANDREEA-MĂDĂLINA ȘERBAN

-2025-

Contents

Nomenclature

AI	Artificial Intelligence
AVE	Average Variance Extracted
AUC	Area Under the (Receiver Operating Characteristic - ROC) Curve
CYBER	Cyber Insecurity
DSE	Digital Self-Efficacy
EHR	Electronic Health Records
mHealth	Mobile health
ML	Machine Learning
OOB	Out-of-bag
PEOU	Perceived Ease of Use
PGHD	Patient Generated Health Data
PU	Perceived Usefulness
RF	Random Forest
RISK	Perceived Risk
RITAM	Risk-Integrated Technology Acceptance Model
SHAP	SHapley Additive Explanations
TAM	Technology Acceptance Model

1. Introduction and Objectives

Despite ongoing advances in pediatric care, clinicians face persistent challenges in anticipating patient trajectories and identifying those at highest risk for adverse outcomes. The increasing availability of clinical data—from electronic health records (EHRs), patient registries, and patient-generated health data (PGHD) via mobile health (mHealth) tools—offers unprecedented opportunities to improve risk stratification, optimize treatment pathways, and enhance patient-centered care [1–3]. Predictive analytics, at the intersection of clinical medicine and data science, provides the methodological backbone for this transformative shift, from reactive risk stratification, based on late clinical signs, to proactive, personalized care.

Building on these foundations, the growing integration of predictive analytics into healthcare reflects a paradigm shift from data collection to meaningful clinical application. The true value of these tools lies not only in their predictive capacity but in their ability to support real-time clinical decisions, streamline workflows, and enhance continuity of care. In pediatrics, where timely interventions can dramatically alter outcomes, predictive models offer a bridge between complex datasets and actionable foresight [4,5]. Yet, to translate these advances into daily practice, their development must be underpinned by rigorous methodological standards that ensure validity, interpretability, and trust among clinicians and patients alike.

Each step in the predictive analytics process is critical for ensuring the intended impact of any intervention implemented in this domain [5,6]. Within this context, the present thesis aims to build a robust theoretical and practical framework to support the implementation of predictive analytics tools in healthcare. The objectives of this research can be grouped into three interrelated domains: data collection, technology adoption, and machine learning (ML) model development and validation.

First, regarding data collection, the thesis establishes principles for ensuring high-quality, well-preprocessed data, which serve as a foundation for ML training and validation. Specific attention is given to EHRs, surgical patient registries, and mHealth platforms for structured PGHD collection and outcome prediction in pediatric populations [7]. Second, in the domain of behavioral medicine and technology adoption, the research examines factors influencing the uptake of digital tools for predictive analytics. Stakeholder attitudes, including digital self-efficacy (DSE) and privacy concerns, are identified as critical determinants of acceptance [8]. The conceptual model is an original extended Risk-Integrated Technology Adoption Model (RITAM) applied to the healthcare context [9–13]. These insights inform strategies to enhance adoption, ensuring that collected data can meaningfully support predictive analytics. Finally, in ML modeling and predictive tool validation, the thesis demonstrates practical applications of predictive analytics. A Random Forest (RF) classifier is developed and validated for short-term mortality prediction in pediatric surgical patients, using a large-scale dataset. Interpretability tools, such as Partial Dependence Plots (PDPs) and SHapley Additive exPlanations (SHAP), enable clinicians to understand the influence of the individual predictors in the survival prediction model, thus ensuring transparency in decision-making [14–18].

Overall, our research integrates data quality, user-centered technology adoption, and advanced ML methodologies to provide a comprehensive framework for predictive analytics in healthcare. By combining theoretical rigor with applied studies, it offers actionable insights to improve patient outcomes, support clinical decision-making, and guide the responsible implementation of predictive tools in pediatric healthcare settings.

2. Research Methodology

This thesis employs a multidisciplinary approach integrating clinical pediatrics, data science, and behavioral research to develop and validate predictive statistical tools for modeling patient outcomes. The overarching methodological framework combines retrospective outcome analysis, behavioral modeling of technology adoption, and advanced ML to bridge the gap between data generation and clinical application.

The retrospective audit studies described in Section 5.1 exemplify how conventional outcome assessment can evolve through predictive analytics. By systematically capturing detailed surgical and patient-level data, these studies demonstrate the potential of structured clinical information to inform predictive models and guide evidence-based decisions in pediatric care. They highlight how real-world hospital data, when rigorously standardized, can provide meaningful insights into postoperative risk and outcomes—laying the foundation for data-driven clinical decision support.

In Section 5.2, the thesis advances this approach through a scoping study of surgical patient registries, employing the PRISMA framework to identify and categorize the methodological challenges of developing high-quality registry-based datasets. This approach also guided the selection of exemplary registries that illustrate optimal strategies for ensuring reliability, completeness, and interoperability—essential traits for their integration into predictive analytics pipelines [19]. These findings emphasize that the clinical value of AI models depends fundamentally on the quality and consistency of the data on which they are built.

The next component focuses on PGHD through the design and presentation of KIDoc, an original mHealth platform created to support longitudinal monitoring in children with chronic diseases. Complementing the high-fidelity data from EHRs and registries, PGHD provides continuous, real-time insight into patients' daily experiences, treatment adherence, and evolving symptoms. This integration of clinical and self-reported data enhances temporal resolution and enables a more nuanced understanding of patient trajectories—critical for predictive modeling in pediatrics, where early deviations often precede acute deterioration.

Chapter 6 extends the methodological scope into behavioral modeling, addressing the persistent challenge of digital health adoption. Using an original extended RITAM, the thesis identifies the key determinants of patients' and caregivers' intention to use mHealth solutions, as well as the relative weight of each determinant in decision-making. Survey data were analyzed with Partial Least Squares Structural Equation Modeling (PLS-SEM), a method chosen for its flexibility with small, heterogeneous samples and its predictive orientation. PLS-SEM operates through a two-stage algorithm: the measurement model, which estimates latent constructs such as DSE, PEOU, and privacy concerns; and the structural model, which assesses causal relationships among them using path analysis. This analytical approach emphasizes prediction over confirmation, making it particularly suitable for studying behavioral intentions toward digital health adoption. All analyses in this section of the thesis were performed using WarpPLS 8.0 software.

Chapter 7 represents an original contribution to data science applications in pediatric surgery. It focuses on developing and validating an RF algorithm for short-term mortality prediction in pediatric patients with gastrointestinal surgical conditions. Through the use of a large-scale dataset, we ensured both heterogeneity and statistical robustness. Data preprocessing was a critical step: missing values were imputed using RF-based methods (missForest and rfImpute), preserving inter-feature dependencies; categorical and continuous variables were appropriately encoded and normalized; and new composite indicators, such as comorbidity indices and perioperative risk scores, were engineered to enhance clinical interpretability. Given the class imbalance between survivors and non-survivors, a hybrid resampling strategy and weighted classification were applied to ensure model robustness. Model training and validation were performed using a stratified train-test split to maintain outcome prevalence. Hyperparameters—including the number of trees, maximum depth, and node size—were tuned using grid search and out-of-bag (OOB) error estimation. Model performance was assessed through a comprehensive suite of metrics: accuracy, sensitivity, specificity, F1-score, precision-recall curves, and the area under the receiver operating characteristic curve (AUC). This ensured both statistical rigor and clinical interpretability of the results.

Recognizing that interpretability is critical for clinical acceptance, model-agnostic explainability techniques were applied. PDPs illustrated the marginal effect of key predictors—such as gestational age, preoperative status, and time to intervention—on survival probabilities, while SHAP quantified global and patient-specific feature importance. These visualization techniques bridge the gap between algorithmic prediction and clinical reasoning, allowing clinicians to understand, audit, and trust model outputs.

Overall, this thesis employs a methodologically integrated framework that spans from high-quality data collection and behavioral adoption modeling to advanced machine learning and interpretability methods. By aligning technical innovation with clinical applicability, it provides a scalable model for developing transparent, reliable, and clinically relevant predictive analytics tools in pediatric healthcare. This comprehensive approach ensures that predictive models are not only statistically sound but also usable, trustworthy, and impactful at the bedside.

3. Results Summary

Building upon the multidisciplinary methodological framework, the findings presented in the thesis demonstrate how predictive analytics can be effectively operationalized across different layers of pediatric healthcare. The studies are organized to reflect the sequential logic of research: from evaluating data sources and quality, to understanding user adoption of predictive tools, and finally to implementing and validating ML models for clinical outcome prediction.

The first chapter of the original contribution section of the thesis analyzes the various sources of clinical data that underpin predictive analytics algorithms. The audit studies and surgical registries demonstrate that well-structured clinical data can reliably inform predictive modeling. Retrospective audits identify key patient- and procedure-level risk factors, while surgical registries—enhanced through standardized protocols and centralized data management—ensure high-quality, multi-source datasets suitable for predictive applications. The original KIDoc platform further illustrates how these predictive insights can be extended to children with chronic diseases, providing real-time, longitudinal monitoring and enabling outcome forecasts that support personalized, anticipatory care. Collectively, these results highlight the foundational role of high-quality clinical data in predictive analytics and their potential to transform decision-making, improve risk stratification, and advance preventive, patient-centered healthcare.

Chapter 6 builds on this foundation by addressing the behavioral and technological dimensions that determine whether predictive analytics tools can be successfully adopted in practice. The study validates an extended RITAM, offering empirical insights into how users perceive and engage with mobile health technologies. The measurement model demonstrated excellent reliability and validity, with composite reliability ranging from 0.901 to 0.991, Cronbach's alpha exceeding 0.83, and average variance extracted (AVE) above 0.69 for all constructs—confirming strong internal consistency and convergent validity. Discriminant validity was supported by correlation analyses, heterotrait-monotrait ratios below 0.9, and square roots of AVEs exceeding inter-construct correlations, while a standardized root mean squared residual value of 0.063 indicated excellent model fit [20].

Structural model analyses confirmed that DSE strongly predicts PEOU ($f^2 = 0.486$, $p < 0.001$) and that cybersecurity concerns (CYBER) strongly predict perceived risk (RISK) ($f^2 = 0.448$, $p < 0.001$), with age emerging as an additional significant factor. The model further supported key hypothesized pathways: PEOU positively influences perceived usefulness (PU) ($\beta = 0.693$) and intention to use (INT) ($\beta = 0.293$); PU negatively predicts RISK ($\beta = -0.717$) and positively predicts INT ($\beta = 0.342$); and RISK negatively affects INT ($\beta = -0.321$). Goodness-of-fit indices (Tenenhaus index = 0.703–0.817) and R^2 values (PEOU 68.7%, RISK 53.2%, INT 72.9%) indicate substantial explanatory power. Diagnostic analyses using Warp2 and Warp3 confirmed the hypothesized causal directions, with no evidence supporting reversal between PU and RISK [20]. Overall, these findings validate the RITAM framework in a healthcare context, emphasizing that usability, perceived benefit, and risk perception are key behavioral determinants of technology adoption.

Chapter 7 translates these theoretical and behavioral insights into applied data science, demonstrating the practical value of ML approaches for patient outcome prediction. The RF analysis included 3,559 pediatric patients with 3,680 congenital gastrointestinal conditions, with surgical intervention as the primary treatment for 89.6% of cases. Typical for a clinical dataset, it exhibited class imbalance—survivors substantially outnumbering non-survivors. This was mitigated through a hybrid resampling strategy, ensuring balanced training and robust model performance. The final model achieved 89.3% accuracy, 89.98% sensitivity for non-survivors, 84.1% specificity for survivors, an F1-score of 0.746, and an AUC of 0.941, significantly outperforming the no-information baseline. Model performance remained consistent across congenital malformation subtypes and income strata [21].

Feature importance analyses identified a set of clinically intuitive predictors of 30-day post-intervention mortality, including the presence of complications, duration of postoperative antibiotic therapy, ventilation or parenteral nutrition requirements, ASA score, patient weight at admission, and adherence to surgical safety protocols, as well as sociodemographic factors such as continent and income level. PDPs highlighted nonlinear associations—such as between gestational age and survival probabilities—while SHAP analyses provided both global and patient-specific interpretability [21]. These findings confirm the RF model's capacity to generate accurate and transparent predictions, supporting individualized risk assessment and data-driven clinical decision-making in pediatric surgery.

Taken together, the results across all three components—data quality, technology adoption, and ML—illustrate a coherent and clinically grounded framework for predictive analytics in healthcare. Together, they demonstrate how integrating diverse data sources, human factors, and explainable AI models can transform predictive insights into actionable strategies for improving pediatric patient outcomes.

4. Conclusions and Personal Contributions

This thesis establishes predictive analytics as a pivotal tool for advancing healthcare by transforming complex, multi-source data into actionable clinical insights. Through a combination of theoretical frameworks, methodological innovation, and empirical studies, it addresses critical gaps in the translation of AI into practice: inconsistent data collection, limited understanding of technology adoption, and challenges in validating predictive models in real-world contexts.

A major conclusion is that structured, high-quality datasets—including EHRs, surgical registries, and mHealth platforms—can reliably support predictive modeling when coupled with rigorous quality assurance measures. The research demonstrates that retrospective audits and carefully curated registries provide actionable insights into patient risk factors, outcomes, and procedural patterns, forming a foundation for accurate, clinically relevant predictive models.

On technology adoption, the thesis adapts and extends the RITAM framework, showing quantitatively that behavioral intention to use digital health solutions is driven by perceived usefulness and ease of use, while perceived risk—mediated by cyber-insecurity and digital self-efficacy—remains a key barrier. These findings offer a validated framework for anticipating the uptake of patient-generated health data in predictive analytics.

From a predictive modeling standpoint, the development of an RF algorithm to forecast short-term mortality in pediatric surgical patients demonstrates high accuracy and clinical relevance, with key predictors identified such as postoperative complications, nutrition and ventilation requirements, ASA score, and adherence to safety protocols. This illustrates the practical value of predictive analytics in supporting risk stratification and informed decision-making. By leveraging explainable AI techniques like PDPs and SHAP, the study ensured that predictive models remain understandable, auditable, and relevant for clinical practice.

Personal contributions of this thesis span conceptual, methodological, technical, and applied domains. Conceptually, it bridges theoretical understanding with real-world implementation, identifies adoption barriers, and integrates retrospective outcome analysis with predictive modeling. Methodologically, it applies PLS-SEM to analyze determinants of technology adoption. Technically, it implements and validates a robust ML pipeline tailored to healthcare datasets, including preprocessing strategies for missing data and class imbalance. Practically, it provides actionable guidance for clinicians and policymakers on integrating predictive analytics into clinical workflows, demonstrating both feasibility and impact.

Viewed comprehensively, this thesis is both timely and highly relevant, as predictive analytics has emerged as a cornerstone of AI in medicine, transforming complex clinical data into actionable insights. By enabling proactive, personalized, and data-driven interventions, the work contributes not only to scientific knowledge but also to practical improvements in patient care, bridging the gap between AI research and clinical practice.

Selected References

[1] Structured vs Unstructured Data in Healthcare. TrueNorth ITG 2022. https://www.truenorthitg.com/healthcare-unstructured-data/ (accessed January 30, 2025).

[2] Grosman-Rimon L, Li DHY, Collins BE, Wegier P. Can we improve healthcare with centralized management systems, supported by information technology, predictive analytics, and real-time data?: A review. Medicine (Baltimore) 2023;102:e35769. https://doi.org/10.1097/MD.0000000000035769.

[3] Van Calster B, Wynants L, Timmerman D, Steyerberg EW, Collins GS. Predictive analytics in health care: how can we know it works? J Am Med Inform Assoc 2019;26:1651–4. https://doi.org/10.1093/jamia/ocz130.

[4] Jayatilake SMDAC, Ganegoda GU. Involvement of Machine Learning Tools in Healthcare Decision Making. J Healthc Eng 2021;2021:6679512. https://doi.org/10.1155/2021/6679512.

[5] Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. JAMA 2018;319:1317–8. https://doi.org/10.1001/jama.2017.18391.

[6] Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. IEEE J Biomed Health Inform 2018;22:1589–604. https://doi.org/10.1109/JBHI.2017.2767063.

[7] Lee TC, Shah NU, Haack A, Baxter SL. Clinical Implementation of Predictive Models Embedded within Electronic Health Record Systems: A Systematic Review. Informatics (MDPI) 2020;7:25. https://doi.org/10.3390/informatics7030025.

[8] Ulfert-Blank A-S, Schmidt I. Assessing digital self-efficacy: Review and scale development. Computers and Education 2022;191. https://doi.org/10.1016/j.compedu.2022.104626.

[9] Davis FD. A technology acceptance model for empirically testing new end-user information systems : theory and results, Thesis 1985.

[10] Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989;13:319. https://doi.org/10.2307/249008.

[11] Park C-K, Kim H-J, Kim Y-S. A study of factors enhancing smart grid consumer engagement. Energy Policy 2014;72:211–8. https://doi.org/10.1016/j.enpol.2014.03.017.

[12] Shapira S, Cauchard JR. Integrating drones in response to public health emergencies: A combined framework to explore technology acceptance. Front Public Health 2022;10:1019626. https://doi.org/10.3389/fpubh.2022.1019626.

[13] Alaiad A, Zhou L. The determinants of home healthcare robots adoption: an empirical investigation. Int J Med Inform 2014;83:825–40. https://doi.org/10.1016/j.ijmedinf.2014.07.003.

[14] Imrie F, Davis R, van der Schaar M. Multiple stakeholders drive diverse interpretability requirements for machine learning in healthcare. Nat Mach Intell 2023;5:824–9. https://doi.org/10.1038/s42256-023-00698-2.

[15] Lundberg SM, Erion GG, Lee S-I. Consistent Individualized Feature Attribution for Tree Ensembles 2019. https://doi.org/10.48550/arXiv.1802.03888.

[16] Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, vol. 30, Curran Associates, Inc.; 2017.

[17] Ribeiro MT, Singh S, Guestrin C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco California USA: ACM; 2016, p. 1135–44. https://doi.org/10.1145/2939672.2939778.
[18] Zhang Y, Song K, Sun Y, Tan S, Udell M. “Why Should You Trust My Explanation?” Understanding Uncertainty in LIME Explanations 2019. https://doi.org/10.48550/arXiv.1904.12991.
[19] Serban AM, Ionescu NS. Surgical patient registries: scoping study of challenges and solutions. J Public Health Policy 2023;44:523–34. https://doi.org/10.1057/s41271-023-00442-5.
[20] Serban AM, Druică E. Mobile Technology Adoption in Healthcare—A Behavioral Understanding of Chronic Patients’ Perspective. Clinics and Practice 2025;15:181. https://doi.org/10.3390/clinpract15100181.
[21] Serban AM. Short-term mortality prediction in children with gastrointestinal congenital anomalies using a random forest classifier. Pediatr Res 2025. https://doi.org/10.1038/s41390-025-04378-2.

Scientific Papers Published

First Author

- **Serban AM**, Ionescu NS. Surgical patient registries: scoping study of challenges and solutions. J Public Health Policy. 2023 Dec;44(4):523-534. doi: 10.1057/s41271-023-00442-5. Epub 2023 Sep 19. PMID: 37726394 (**Impact Factor 2.7**)
- **Serban AM**, Druica E. Mobile Technology Adoption in Healthcare – A Behavioral Understanding of Chronic Patients' Perspective. *Clinics and Practice*. 2025; 15(10):181. https://doi.org/10.3390/clinpract15100181 (**Impact Factor 2.2**)
- **Serban AM.** Short-term Mortality Prediction in Children with Gastrointestinal Congenital Anomalies using a Random Forest Classifier. Pediatr Res. 2025 Sep 15. doi: 10.1038/s41390-025-04378-2. PMID: 40954231 (**Impact Factor 3.7**)

Scientific Papers Published

International Collaborations

- **National Institute for Health and Care Research Global Health Research Unit on Global Surgery.** Reducing the environmental impact of surgery on a global scale: systematic review and co-prioritization with healthcare workers in 132 countries. Br J Surg. 2023 Jun 12;110(7):804-817. doi: 10.1093/bjs/znad092. Erratum in: Br J Surg. 2023 Nov 9;110(12):1907. doi: 10.1093/bjs/znad317. PMID: 37079880; PMCID: PMC10364528 (**Impact Factor 8.6**).
- NIHR Global Health Unit on Global Surgery; **COVIDSurg Collaborative.** Elective surgery system strengthening: development, measurement, and validation of the surgical preparedness index across 1632 hospitals in 119 countries. Lancet. 2022 Nov 5;400(10363):1607-1617. doi: 10.1016/S0140-6736(22)01846-3. Epub 2022 Oct 31. PMID: 36328042; PMCID: PMC9621702 (**Impact Factor 202.7**).
- **COVIDSurg Collaborative**; GlobalSurg Collaborative. SARS-CoV-2 infection and venous thromboembolism after surgery: an international prospective cohort study. Anaesthesia. 2022 Jan;77(1):28-39. doi: 10.1111/anae.15563. Epub 2021 Aug 24. PMID: 34428858; PMCID: PMC8652887 (**Impact Factor 12.9**).
- **Global PaedSurg Research Collaboration**. Mortality from gastrointestinal congenital anomalies at 264 hospitals in 74 low-income, middle-income, and high-income countries: a multicentre, international, prospective cohort study. Lancet. 2021 Jul 24;398(10297):325-339. doi: 10.1016/S0140-6736(21)00767-4. Epub 2021 Jul 13. PMID: 34270932; PMCID: PMC8314066 (**Impact Factor 202.7**).
- **COVIDSurg Collaborative**, GlobalSurg Collaborative. SARS-CoV-2 vaccination modelling for safe surgery to save lives: data from an international prospective cohort study. Br J Surg. 2021 Sep 27;108(9):1056-1063. doi: 10.1093/bjs/znab101. PMID: 33761533; PMCID: PMC7995808 (**Impact Factor 11.12**).
- **COVIDSurg Collaborative**; GlobalSurg Collaborative. Timing of surgery following SARS-CoV-2 infection: an international prospective cohort study. Anaesthesia. 2021 Jun;76(6):748-758. doi: 10.1111/anae.15458. Epub 2021 Mar 9. PMID: 33690889; PMCID: PMC8206995 (**Impact Factor 12.89**).

Conference Presentations

- **Serban AM,** Ionescu NS. **National Disease Registries in Romania- status and perspectives.** AOSR (Scientist Academy Romania) National Scientific Conference 2022.
- Ionescu NS, **Serban AM**, Tirlea S. **Personalized medicine in urinary tract dilatation.** AOSR (Scientist Academy Romania) National Scientific Conference 2022.

Awards for Research on the Thesis Topic

- **Young Transatlantic Innovation Fellowship 2022** – US Department of State – Research developed in collaboration with the University of Pennsylvania – Advanced Statistical Analyses of Oral Insulin Therapy in a Canine Model – prediction of diabetes therapy success.
- **Innovators for Children Award 2021 Impact Hub**– for the KIDoc project, a platform for data collection and analysis dedicated to children with chronic diseases – identification of at-risk patients and mortality prediction.
- **Young Investigator Award** – Carol Davila University of Medicine and Pharmacy International Congress 2019 - **Șerban AM**, Spătaru R, Isac G, Ionescu NS. Prognostic factors of short-term mortality for patients with congenital diaphragmatic hernia.