



Editorial

Research on the Application and Interpretability of Predictive Statistical Data Analysis Methods in Medicine

Pentti Nieminen

Medical Informatics and Data Analysis Research Group, University of Oulu, 90014 Oulu, Finland;
pentti.nieminen@oulu.fi

1. Introduction

Multivariable statistical analysis involves the dichotomy of modeling and predicting. Yet, in clinical medicine, epidemiology, and health care research, the main goal is to understand, rather than to predict. Physicians are more eager to understand the phenomenon of a severe disease and to learn how to help a patient to survive as long as possible, than to be able to predict the survival time. For epidemiologists, knowing how to reduce the prevalence or incidence of a disease in a population is more important than predicting who will develop a disease. For clinicians and subject-specific researchers, the point of looking at health data is often to intervene to change the expected outcomes.

Artificial intelligence (AI) methods, including machine learning (ML), deep learning (DL) and random forests, have come to show promise in the field of medicine, and these methods can develop effective diagnostic and predictive tools to identify various diseases [1]. However, prediction methods suffer from the “black box” problem: inputs are fed to the algorithm and an output emerges, but it is not entirely clear which variables were identified, or how they contributed to the final output [2]. Models do not provide an effect size index that is familiar to medical researchers and that helps them evaluate the effect of specific explanatory variables. In contrast, estimated classical multivariable regression models, although not always as powerful as ML, DL, or random forests, are easy to interpret. Due to the lack of transparency in how predictive algorithms work, it may become difficult to implement these models in daily medical research, and especially in clinical practice, where interpretable models are preferred [3,4]. This may ultimately precipitate a declining preference for machine learning models in medical research. Limited data on the actual effects on patient outcomes do not support evidence-based practice in medicine, which relies on the judicious use of the best current evidence to make health care decisions. Meta-analysis is the basic statistical tool in evidence-based medicine since it provides strong evidence by combining the effect sizes of a number of individual studies on a particular topic.

Predictive algorithms offer significant opportunities in terms of identifying and predicting treatment options and outcomes for patients. Therefore, it is important to explore how best to explain in detail to the physician the decisions made by an AI-based algorithm regarding the diagnosis, treatment, or prognosis of a disease. Numerous efforts are underway to build more explainable AI (XAI) models, as well as to develop methods that attempt to interpret predictive algorithms, for example, by visualizing the results [5–7]. One area of active innovation is the establishment of new methods, such as parallel models, where one is used for core computation and the other for interpretation [4].

BioMedInformatics has been an opportunity for the scientific community to present research on the application and complexity of data analysis methods, and to provide insight into new challenges in biostatistics, biomedicine, epidemiology, health sciences, dentistry, and clinical medicine. Special Issues of *BioMedInformatics* have brought together information on established statistical and data analysis methods in these fields. One area of innovative articles is the synthesis and presentation of where and how AI results are



Citation: Nieminen, P. Research on the Application and Interpretability of Predictive Statistical Data Analysis Methods in Medicine. *BioMedInformatics* **2024**, *4*, 321–325. <https://doi.org/10.3390/biomedinformatics4010018>

Received: 25 January 2024
Accepted: 26 January 2024
Published: 30 January 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

presented to medical researchers, in many cases to aid in interpretability. As the demand for more explainable machine learning models with interpretable predictions grows, so does the need for methods that can help to achieve these goals. In this editorial, I highlight some important studies published in *BioMedInformatics* that have provided in-depth identification, analysis, and comparison of machine learning interpretability methods, which I believe will be prominent in the coming years.

2. Interpretable Complex Predictive Algorithms

Artificial intelligence technology has already been applied to modern biology and biomedicine; however, the lack of interpretability in predictive models can undermine confidence in these models, especially in healthcare. A paper authored by Konstantina Athanasopoulou and her co-authors [3] describes the basic principles of predictive approaches. The authors highlight the role of AI-based methods in various biological research areas and explore the impact of AI on everyday clinical practice and healthcare systems. The development of new computational tools has opened up new vistas for both biological and medical sciences. Although advances in AI will lead to widespread investment across the biomedical industry and their impact will be far-reaching, AI will augment, but not replace, human expertise. The authors also discuss challenges and future directions for predictive models—these include a lack of interpretability, but also dependence on the quality of the datasets used to train each model, as well as several emerging social and legal issues.

The paper by Thomas Krause and his co-authors [8] provides a comprehensive overview of ML methods in metagenomics. They clearly illustrate the different methods within different datasets. The challenges faced by machine learning in biomedical research are also discussed in the hope of further improving the explainability and reproducibility of outcomes. Both determining the need for explainability and managing the possible trade-offs between explainability and accuracy are challenges when implementing new clinical solutions using ML.

The most popular explanatory technique for AI models is the feature importance approach [9]; however, there are several different approaches to measuring feature importance. Feature importance is a step of building a machine learning model that involves calculating the scores for all the input variables in a model in order to determine the importance of each variable in the decision-making process. The higher the score for a variable, the more influence it has on the model predicting a particular response variable. Like a correlation matrix, feature importance allows one to understand the relationship between the features and the target variable, but it also helps one understand which features are irrelevant to the model. By calculating scores for each feature, it is possible to determine which features contribute the most to the predictive power of the model.

Machine learning models can be highly accurate in their predictions, but there are still some major problems associated with their implementation. Yiqiao Yin and Yash Bingi [10] conducted a study using machine learning models to classify fetal health. These models provided no help in identifying a problem with a fetus if it was deemed pathological, and obstetricians will not be able to properly treat their patients if they do not know why the fetus is in danger. Another problem is that patients may not trust a machine to give them a diagnosis, especially when there is no way for them to see how it arrived at a given classification. In order to better explain the model, the authors proposed a new feature metric and a feature alternative for explanations of black box models (FAB) was created. This metric analyzes the importance of features through the unique technique of removing individual features and rechecking the model's accuracy. Overall, this technique allows physicians and medical professionals to classify fetal health with high accuracy and to also determine which features were the most influential in the process.

A detailed understanding of the mathematical details of a predictive algorithm may be possible for experts in statistics or computer science; however, when it comes to the fate of human beings, this “developer’s explanation” is often not deemed sufficient. In

their work, Jörn Lötch and his coworkers [11] explored the concept of explainable AI as a solution to this problem. Explainable AI is an emerging line of research that helps the user or developer of machine learning models understand why the models behave the way they do. Their report highlights the need for comprehensibility of AI-based biomedical decisions. A truly explainable AI system is one that draws its conclusions based on a model that is thoroughly understood and accepted in depth by a human expert in the field in which the XAI is deployed. This understanding and acceptance of the AI inference model must be such that the expert is ultimately willing to take legal responsibility for the AI's decisions. XAI systems must be able to explain each decision and its derivations in a way that can be understandable and comprehensible to the medical expert on the ground.

In medicine, machine learning (ML) methods are used to analyze magnetic resonance imaging (MRI) scans in order to support medical staff in their decision-making. It is often the case that the algorithms used do not explain their internal decision-making process. As a result, it is difficult to validate or interpret the results. The paper by Matthias Eder and his colleagues [12] is an interesting practical example of how to overcome this problem by using methods of explainable AI (XAI). The authors present an application of visual explanations to interpret the decision of an ML algorithm in the case of predicting the survival rate of brain tumor patients based on their MRI scans.

Visual inspection is considered to be an easy and quick way to recognize associations while analyzing complex multivariable data. In their paper, Milot Gashi and co-authors [13] investigated the visualization techniques on glioma datasets to provide an intuitive explanation of AI results. They show how to support the understanding of the results of a black-box model in glioma classification to find novel biomarkers. Visual analytics can also increase user acceptance and the adaptation of AI models in medical research.

Recently Alfred Ultsch and his colleagues [7] presented a novel XAI method called algorithmic population descriptions, which is able to classify cases based on subpopulations in high-dimensional data. A visualization method allows human experts to understand the reasoning used by the AI system.

3. Multivariable Data Can Be Analyzed in Several Ways

There is a clear trade-off between the performance of a machine learning model and its ability to produce explainable and interpretable predictions. On the one hand, there are the black-box models, which include machine learning and deep learning. On the other hand, there are standard regression models (linear, logistic, Cox, and negative binomial regressions) and decision tree-based models that easily produce explainable results. Although the latter models are easier to interpret, they often fail to achieve state-of-the-art predictive performance compared to the former. However, one systematic review showed that, for the most common machine learning models used for data analysis (i.e., classification trees, random forests, artificial neural networks, and support vector machines), there was no evidence of superior performance when compared to logistic regression for clinical prediction modeling [14]. When evaluating the use of machine learning models, it may be useful to estimate standard regression models on the same data on which predictions were made using the machine learning or random forest methods. Having these parallel models provides another axis of evaluation in terms of implementation cost, interpretability, and relative performance. Medical and public health researchers are familiar with basic biostatistics and know how to interpret traditional regression models, which makes it easy to communicate the results of regression models to researchers in medicine and related fields.

To ensure the accurate analysis of biomedical datasets, it is critical to recognize and address the limitations of both regression analysis and predictive machine learning methods. Jörn Lötch and Alfred Ultsch [15] propose a mixture of experts on machine learning models, which may include regression models, to handle classification problems in biomedical data. They note that it is highly recommended to avoid relying solely on a single analysis method

when studying biomedical datasets and, by incorporating additional analytical approaches, researchers can obtain more reliable and interpretable results.

Different classification approaches have specific advantages and shortcomings. In their paper, Emilija Strelcenia and Simant Prakoonwit [16] present a detailed review and comparison of the application of six popular machine learning models in the field of breast cancer diagnosis. Their study combined classifiers with feature selection for breast cancer diagnosis and they note that the integration of multiple algorithms could improve the accuracy and reliability of breast cancer detection. In addition, further transparency in variable selection methods could help to identify the most relevant features for breast cancer diagnosis.

4. Statistical Synthesis of Results from a Series of Studies

Meta-analysis is a very common and important statistical technique in medical research. Systematic reviews and meta-analyses are used in many disciplines to evaluate previous research studies to draw conclusions about a particular medical topic. Reviews of original articles and research syntheses expand our knowledge by combining and comparing original studies. A major problem in analyzing, evaluating, and summarizing the reported results of studies on the association between explanatory variables and an outcome variable is that the results are analyzed and reported in different ways [17]. When using a systematic literature review with a meta-analytic approach to learn from pooled studies, we are dependent on the research methods and reporting of the underlying studies. The measure used to represent the study results in a meta-analysis is called an effect size statistic. If the reviewed research articles based on complex predictive algorithms do not report any effect sizes, it becomes difficult for the meta-analyst to combine the results. A topic for future research is how to combine results from multivariable regression models and predictive algorithms. This issue also touches on the problem of reproducibility in biomedicine.

Scoping reviews are considered a valid approach when systematic reviews and meta-analyses are unable to meet the necessary objectives or the needs of knowledge users. A scoping review conducted by Alexandre Hudson and coworkers [18] focused on the use of neural networks in the context of psychotherapeutic approaches. From the eight studies reviewed, three main uses were identified: prediction of therapeutic outcomes, content analysis, and automated categorization of psychotherapeutic interactions. Uses of neural networks were identified with limited evidence of their effects and the authors found that most studies achieved, at best, a low quality of evidence.

Conventional meta-analysis requires a lot of human effort, is labor-intensive, and requires expert knowledge. Stella C. Chriatopoulou [19] presents an overview of automated meta-analysis tools. The goal is to provide a system that automates as much of the meta-analysis process as possible to reduce the time required to conduct a meta-analysis without reducing expert confidence in methodological and scientific rigor. However, although important steps have been taken to date, there is currently no application that can fully replace the human effort involved in conducting a systematic review to draw conclusions from published studies.

When analyzing epidemiological and clinical data, demonstrating the use of predictive algorithms often does not answer the real research question for medical researchers: how to help patients with their disease. Synthesizing effect sizes from multiple studies provides evidence about the importance of risk factors or the effectiveness of different treatment methods. Extensive future studies are needed to develop combinable effect sizes from predictive models that will help the medical and public health community to adopt and accept the results of predictive analytics. Unsustainable promises and unfulfilled expectations should be avoided in the context of machine learning methods and predictive algorithms. Combining results with simpler classical approaches can often provide elegant and sufficient answers to important questions and can convince clinicians of the results.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Rowe, M. An Introduction to Machine Learning for Clinicians. *Acad. Med.* **2019**, *94*, 1433–1436. [[CrossRef](#)] [[PubMed](#)]
2. Petch, J.; Di, S.; Nelson, W. Opening the Black Box: The Promise and Limitations of Explainable Machine Learning in Cardiology. *Can. J. Cardiol.* **2022**, *38*, 204–213. [[CrossRef](#)] [[PubMed](#)]
3. Athanasopoulou, K.; Daneva, G.N.; Adamopoulos, P.G.; Scorilas, A. Artificial Intelligence: The Milestone in Modern Biomedical Research. *BioMedInformatics* **2022**, *2*, 727–744. [[CrossRef](#)]
4. Matheny, M.; Israni, S.T.; Ahmed, M.; Whicher, D. (Eds.) *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*; National Academy of Medicine: Washington, DC, USA, 2019; ISBN 2013206534.
5. Carvalho, D.V.; Pereira, E.M.; Cardoso, J.S. Machine Learning Interpretability: A Survey on Methods and Metrics. *Electronics* **2019**, *8*, 832. [[CrossRef](#)]
6. Vilone, G.; Longo, L. Notions of Explainability and Evaluation Approaches for Explainable Artificial Intelligence. *Inf. Fusion* **2021**, *76*, 89–106. [[CrossRef](#)]
7. Ultsch, A.; Hoffmann, J.; Röhnert, M.A.; von Bonin, M.; Oelschlägel, U.; Brendel, C.; Thrun, M.C. An Explainable AI System for the Diagnosis of High Dimensional Biomedical Data. *BioMedInformatics* **2024**, *4*, 197–218. [[CrossRef](#)]
8. Krause, T.; Wassan, J.T.; Mc Kevitt, P.; Wang, H.; Zheng, H.; Hemmje, M. Analyzing Large Microbiome Datasets Using Machine Learning and Big Data. *BioMedInformatics* **2021**, *1*, 138–165. [[CrossRef](#)]
9. Saarela, M.; Jauhiainen, S. Comparison of Feature Importance Measures as Explanations for Classification Models. *SN Appl. Sci.* **2021**, *3*, 272. [[CrossRef](#)]
10. Yin, Y.; Bingi, Y. Using Machine Learning to Classify Human Fetal Health and Analyze Feature Importance. *BioMedInformatics* **2023**, *3*, 280–298. [[CrossRef](#)]
11. Lötsch, J.; Kringel, D.; Ultsch, A. Explainable Artificial Intelligence (XAI) in Biomedicine: Making AI Decisions Trustworthy for Physicians and Patients. *BioMedInformatics* **2022**, *2*, 1–17. [[CrossRef](#)]
12. Eder, M.; Moser, E.; Holzinger, A.; Jean-Quartier, C.; Jeanquartier, F. Interpretable Machine Learning with Brain Image and Survival Data. *BioMedInformatics* **2022**, *2*, 492–510. [[CrossRef](#)]
13. Gashi, M.; Vuković, M.; Jekic, N.; Thalmann, S.; Holzinger, A.; Jean-Quartier, C.; Jeanquartier, F. State-of-the-Art Explainability Methods with Focus on Visual Analytics Showcased by Glioma Classification. *BioMedInformatics* **2022**, *2*, 139–158. [[CrossRef](#)]
14. Christodoulou, E.; Ma, J.; Collins, G.S.; Steyerberg, E.W.; Verbakel, J.Y.; Van Calster, B. A Systematic Review Shows No Performance Benefit of Machine Learning over Logistic Regression for Clinical Prediction Models. *J. Clin. Epidemiol.* **2019**, *110*, 12–22. [[CrossRef](#)] [[PubMed](#)]
15. Lötsch, J.; Ultsch, A. Pitfalls of Using Multinomial Regression Analysis to Identify Class-Structure-Relevant Variables in Biomedical Data Sets: Why a Mixture of Experts (MOE) Approach Is Better. *BioMedInformatics* **2023**, *3*, 869–884. [[CrossRef](#)]
16. Strelcenia, E.; Prakoonwit, S. Effective Feature Engineering and Classification of Breast Cancer Diagnosis: A Comparative Study. *BioMedInformatics* **2023**, *3*, 616–631. [[CrossRef](#)]
17. Nieminen, P. Application of Standardized Regression Coefficient in Meta-Analysis. *BioMedInformatics* **2022**, *2*, 434–458. [[CrossRef](#)]
18. Hudon, A.; Aird, M.; La Haye-Caty, N. Deciphering the Mosaic of Therapeutic Potential: A Scoping Review of Neural Network Applications in Psychotherapy Enhancements. *BioMedInformatics* **2023**, *3*, 1101–1111. [[CrossRef](#)]
19. Christopoulou, S.C. Towards Automated Meta-Analysis of Clinical Trials: An Overview. *BioMedInformatics* **2023**, *3*, 115–140. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.