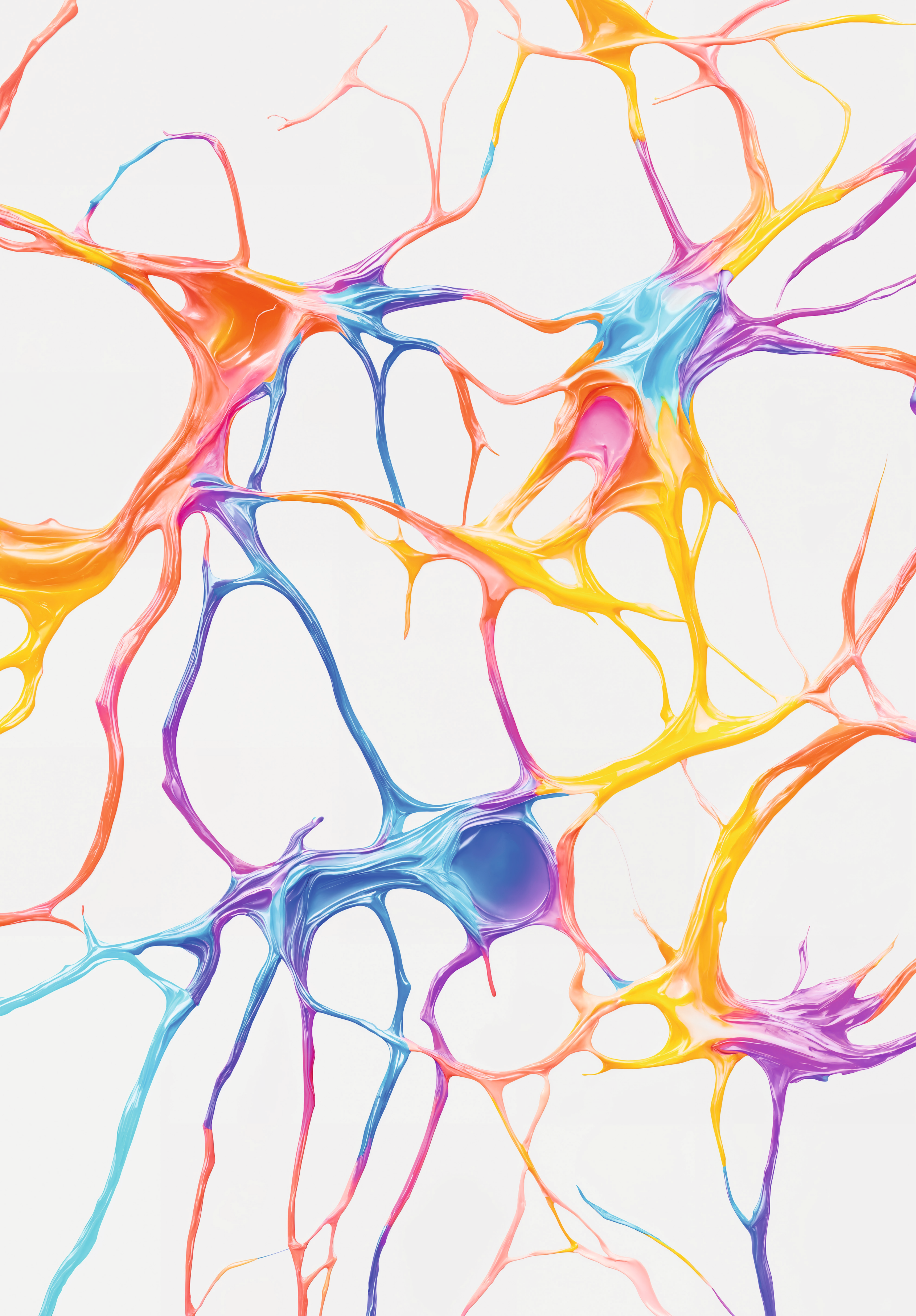
From code to clinic: theory and practice for artificial intelligence prediction algorithms

Hond, A.A.H. de



10

Summary written together with GPT-3 [1]

This thesis looks at Artificial Intelligence (AI) and its potential to revolutionize the healthcare sector. The first part of this thesis focuses on the responsible development and validation of AI-based clinical prediction algorithms, exploring the prime considerations in this process. The second part of this thesis addresses the opportunities for classical statistics and machine learning techniques for developing prediction algorithms as well as the performance, potential, and challenges of AI prediction algorithms for clinical practice.

Chapter 2 provides an overview of actionable guidelines and quality criteria for the development, evaluation, and implementation cycle of Artificial Intelligence prediction algorithms. It covers topics such as data quality, external performance evaluation, and impact assessment. This chapter concludes that more research is needed on the latter three phases of the AI development cycle (software development, clinical impact assessment, and implementation).

Chapter 3 discusses considerations for measuring and validating the predictive performance of algorithms. First, the differences in discrimination metrics between the fields of statistics and machine learning are addressed. We call for one common approach to increase cross-discipline collaboration. Secondly, we show that labels (e.g., 'good', 'bad') for specific AUC values are largely arbitrary and that further measures, such as Net Benefit, should be reported to evaluate the value of an algorithm for decision-making. Lastly, we provide an overview of four types of generalizability associated with clinical predictive algorithms, and encourage validation practices that are aligned with the intended use of the algorithm.

In the second part of this thesis, several clinical use cases are discussed. In **chapter 4**, we investigate whether machine learning algorithms could predict hospital admission of emergency department patients more accurately than logistic regression. Results show that machine learning algorithms have an excellent but similar predictive performance as the logistic regression algorithm. These results imply that machine learning has little benefit for predicting hospital admission over conventional algorithms.

Chapter 5 compares the performance of two machine learning algorithms, a statistical algorithm, and a simple asthma action plan rule for predicting severe asthma exacerbations in patients with mild-to-moderate chronic asthma based on home monitoring data. Results show that the logistic regression algorithm has higher discriminative performance than the XGBoost algorithm and the simple clinical rule, but still produces many false positives at high levels of sensitivity. The study concludes that machine learning algorithms may not outperform classical regression prediction algorithms in predicting short-term asthma exacerbations based on home monitoring data.

Chapter 6 compares the predictive performance of machine learning and regression competing risks algorithms in 11 datasets of patients undergoing arthroplasty surgery. The results show that the machine learning and regression algorithms perform similarly. The findings highlight the importance of not over-relying on machine learning methods and the need to collect more relevant predictors for the benefit of individual patients.

Chapter 7 describes the development of a prediction algorithm for early identification of patients with cancer at risk of depression at the start of chemo-or radiotherapy treatment. The best performing algorithm (LASSO logistic regression with structured data) has reasonable discrimination and calibration performance. Structured text is more relevant to predicting depression risk than unstructured text in this specific clinical workflow. The results imply that this algorithm has potential as an early triage tool to guide psycho-oncologist resources, but further validation is necessary.

In **chapter 8**, we assess the external validity of a certified machine learning algorithm predicting readmission or death within 7 days after ICU discharge. Results show external validity cannot be taken for granted and further retraining is necessary to improve algorithm performance at the new site.

In conclusion, cross-discipline collaboration, exchangeability of knowledge and results, and validation of AI for healthcare practice are essential for realizing the potential of AI in healthcare.

Too long, didn't read:

This thesis looks at Artificial Intelligence (AI) and its potential to revolutionize the healthcare sector. It focuses on the responsible development and validation of AI-based clinical prediction algorithms, exploring the prime considerations in this process as well as the performance of AI for several healthcare use cases and the potential of prediction algorithms for healthcare practice.

REFERENCE

1. Ouyang, L., J. Wu, X. Jiang, et al., *Training language models to follow instructions with human feedback.* arXiv preprint arXiv:2203.02155, 2022.