Chapter 8
Systematic Literature Review:
XAI and Clinical Decision Support

Thomas Connolly
https://orcid.org/0000-0002-4276-7301
DS Partnership, UK

Mario Soflano
https://orcid.org/0000-0003-0758-1509
Glasgow Caledonian University, UK

Petros Papadopoulos
https://orcid.org/0000-0002-8110-7576
University of Strathclyde, UK

ABSTRACT

Machine learning (ML) applications hold significant promise for innovation within healthcare; however, their full potential has not yet been realised, with limited reports of their clinical and cost benefits in clinical practice. This is due to complex clinical, ethical, and legal questions arising from the lack of understanding about how some ML models operate and come to make decisions. eXplainable AI (XAI) is an approach to help address this problem and make ML models understandable. This chapter reports on a systematic literature review investigating the use of XAI in healthcare within the last six years. Three research questions identified as issues in the literature were examined around how bias was dealt with, which XAI techniques were used, and how the applications were evaluated. Findings show that other than class imbalance and missing values, no other types of bias were accounted for in the shortlisted papers. There were no evaluations of the explainability outputs with clinicians and none of the shortlisted papers used an interventional study or RCT.

DOI: 10.4018/978-1-6684-5092-5.ch008

INTRODUCTION

A Clinical Decision Support System (CDSS) is a technology and a system architecture that uses medical knowledge with clinical data to provide customised advice for an individual patient's care. CDSSs are designed for a variety of purposes such as diagnosis, treatment response prediction, treatment recommendation, prognosis, and the prioritisation of patient care according to their level of risk. CDSSs can help to improve patients' safety, quality of care, and healthcare efficiency, as well as reducing the costs of healthcare. They can improve patient safety not only by reducing medical errors but also through reminders for medications or other medical events for patients or clinicians. Additionally, CDSSs can be useful in low-resource settings where the number of medical institutions, equipment, and qualified clinicians is limited.

CDSSs fall under two main categories with regards to how they capture medical knowledge that have different approaches on how to infer clinical knowledge: knowledge-based and non-knowledge-based (Kim et al., 2014; Velickovski et al., 2014). Knowledge-based systems, also known as expert systems or rule-based systems, use an extensive clinical knowledge base containing subject-specific knowledge taken from the medical literature, or from one or more experts. Non-knowledge-based systems, also known as case-based systems, use machine learning (ML) to look for patterns in clinical assessment data to suggest a diagnostic probability. They are described as non-knowledge based, as the system does not actually have any knowledge prior to a ML algorithm process taking place.

A typical ML model first learns the knowledge from the data it is exposed to and then applies this knowledge to provide predictions about emerging (future) data. Supervised ML is when the program is trained on one or more datasets. Unsupervised ML is when the software is provided with data but must discover patterns and relationships in that data.

The ability of ML for characterising complex health data structures has been successfully demonstrated. However, even though complex classification patterns can be identified, these patterns are only partially useful because often clinicians do not obtain knowledge of their internal workings and, therefore, the meaning and relevance of the results cannot be explained. Given that clinical decisions need to be justifiable for transparency and accountability and decisions need to be fully explained to patients, this poses a significant issue for the use of ML in CDSSs. eXplainable AI (XAI) has emerged to address this problem.

This chapter provides a systematic review of the use of XAI in healthcare, specifically looking at how XAI has been applied and what types of ML models are being used. The next section discusses some of the key concepts in XAI and the following section presents the research questions being addressed in this chapter and outlines the methodology used. An analysis of the papers found in the systematic literature review are discussed and the three research questions answered. Some discussion and conclusions are presented at the end.

BACKGROUND

Chapter 2 provided an introduction to eXplainable AI (XAI). Explainable AI aims to explain the way that AI systems work. At a high-level, two types of models can be distinguished: