ORIGINAL ARTICLE

Role of Big Data and Machine Learning in Diagnostic Decision Support in Radiology

Tanveer Syeda-Mahmood, PhD

Abstract

The field of diagnostic decision support in radiology is undergoing rapid transformation with the availability of large amounts of patient data and the development of new artificial intelligence methods of machine learning such as deep learning. They hold the promise of providing imaging specialists with tools for improving the accuracy and efficiency of diagnosis and treatment. In this article, we will describe the growth of this field for radiology and outline general trends highlighting progress in the field of diagnostic decision support from the early days of rule-based expert systems to cognitive assistants of the modern era.

Key Words: Diagnostic decision support, artificial intelligence, deep learning, machine learning, cognitive assistants, medical image analysis, knowledge and reasoning

J Am Coll Radiol 2018;15:569-576.

INTRODUCTION

The field of artificial intelligence (AI) is making a strong comeback literally in all its senses. From the sense of vision with self-driving cars [1] to the sense of taste through AI-generated recipes [2], many new applications are emerging that expand the role of AI. This revolution has largely been spurred by big data, that is, large amounts of data now easily available in many fields either on the web, collected by smart devices, or in the case of health care, through large-scale health records becoming electronically available. In health care, the role of AI is particularly felt in nearly all fields, from drug discovery where drug candidates are being found faster through the use of machine learning techniques on big data [3] to consumer health where wearable devices are collecting large amounts of data to enable better monitoring and prediction through use of machine learning techniques [4].

With medical imaging now being analyzed through AI and deep learning techniques, the role of big data and machine learning has taken on an added significance for radiologists [5-7]. The largest impact of machine learning on big data is being felt in the field of diagnostic decision support, which has been undergoing a dramatic transformation since the early days of AI and rule-based expert systems [42,43]. In this article, we will describe the growth of this field for radiology and outline general trends. Specifically, we describe the growth of the field from rule-based expert systems, through computer-aided diagnosis systems and big data-driven decision support systems, to data and knowledge-driven systems in their current form as cognitive assistants. We use specific examples from our own research in this field to illustrate the evolved thinking in clinical decision support.

RULE-BASED EXPERT SYSTEMS—EARLY APPLICATIONS

Early applications of AI in radiology were in rule-based expert systems for decision support [8,9]. The rules would form associations of specific conditions and symptoms with relevant tests to order [10], with differential diagnosis or recommended treatments including drugs [9]. The AI technology used in these cases was rule-based inference and reasoning using several knowledge representation methods including semantic networks [11,12], which have survived in the form of knowledge graphs in the unified medical language system currently [13]. For diagnosis in particular, these systems did not really scale because the rules were either incomplete for a specialty or did not completely apply to a patient to trigger in appropriate systems. The most common use of clinical decision

IBM Almaden Research Center, San Jose, California.

Corresponding author and reprints: Tanveer Syeda-Mahmood, PhD, IBM Almaden Research Center, 650 Harry Road, San Jose, CA 95120; e-mail: stf@us.ibm.com.

This work was funded by IBM Research Internal Funding. The author has no conflicts of interest related to the material discussed in this article.

1546-1440/18/$36.00 ■ https://doi.org/10.1016/j.jacr.2018.01.028

support is in the form of alerts in the electronic medical record system such as for drug dosage [14], based on a priori defined rules and not necessarily customized for the current patient. Later rule-based systems point to guidelines based on matches to a patient's conditions [15,18]. The rules in these systems are derived from guidelines composed by experts such as American College of Cardiology guidelines [15] into conveniently organized pages such as those in UptoDate (Wolters Kluwer, Waltham, Massachusetts) [16], Medscape (WebMD, New York) [17], StatDx (Elsevier, Philadelphia) [63], DiagnosisPro [19], Dynamed (Ebsco Health Ipswich, Massachusetts) [20], Pepid (Chicago, Illinois) [21], among others. The knowledge offered through these browser-based technologies is primarily meant for visual examination by specialists. The mapping of a patient's condition to these guidelines is provided through simple search techniques. Thus, most of the rule-based expert systems for clinical decision support in use in clinical practice are based on fixed a priori developed rules and using simple search or rule-based inference techniques to pull up the relevant information for diagnosis, treatment, or outcome with input provided to such systems in structured textual or numeric data form.

COMPUTER-AIDED DIAGNOSIS SYSTEMS

The computer-aided diagnosis (CAD) systems were developed as a specific field of AI that used input from images to reach conclusions about potential anomalies or offer differential diagnosis [22-24]. They coupled data-driven feature extraction methods from image processing, computer vision, and medical image analysis with inferences rules to reason about regions in images containing potential anomalies. Most of these stopped short of giving a diagnosis (computer aided diagnosis [CADx]) [25,26] and instead simply point to potential anomalies and allow semi-automatic calculation of measurements (computer aided detection [CADe]) [22,27]. Due to the large number of false-positives generated and because they do not offer differential diagnosis, many of these systems are therefore used as second readers in radiology to ensure that an anomaly is not accidentally missed. Whether they are CADe or CADx systems, rule-based inference principles of AI are still employed, and the deduction is made on the basis of a priori rules built into the system and applied to a single patient's data. More recent CAD systems use machine learning to do a feature-based classification, and still newer methods have used deep learning as well [27]. Figure 1 illustrates the evolution of CADe systems that identified potential anomalies in a single patient's data based on image analysis, feature extraction, and classification. Thus, although initial CAD systems were rule-based, all the newer systems use some form of machine learning to classify candidate regions as normal or abnormal after sufficient training images are provided, so that they are also now beginning to exploit big data.

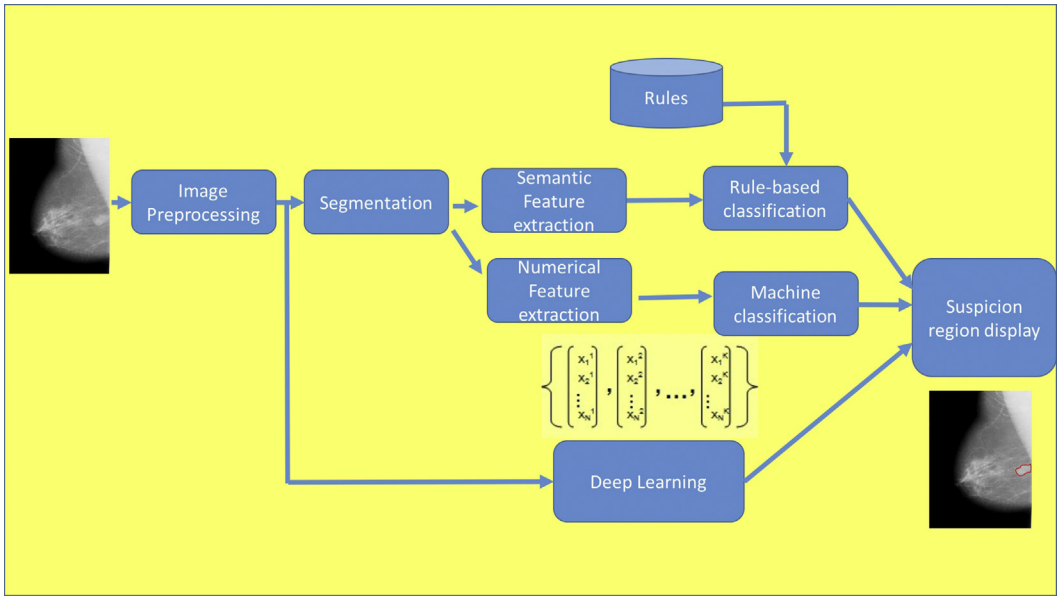

Fig 1. Illustration of computer-aided diagnosis (CAD) systems over three generations of classifiers. The first generation was based on rules and used semantic and qualitative features. The second generation used statistical machine learning with handcrafted features. The most recent CAD systems directly use deep learning to do simultaneous feature extraction and classification.