Finding Barrett’s oesophagus: is there a machine learning approach in our future?

Currently, when diagnosing Barrett’s oesophagus, screening is advocated by gastrointestinal societies for only a small number of patients who are at increased risk. However, this restricted screening approach has never been tested in terms of cost efficacy or improved outcomes and is based solely on expert opinion. Screening for Barrett’s oesophagus is currently only done via endoscopy.

In *The Lancet Digital Health*, Avi Rosenfeld and colleagues report the application of machine learning techniques to a cohort from the BEST2 case-control screening study, comprising 880 patients with Barrett’s oesophagus, including 40 with invasive oesophageal adenocarcinoma, and 419 controls. Using their machine learning approach, the authors analysed the questionnaires that were administered in the BEST2 study and clinical, demographic, and physical features to identify eight independent features that were risk factors for Barrett’s oesophagus including sex, age, smoking, waist circumference, frequency of abdominal pain, duration of acidic taste, and use of antireflux medication. They combine these features into a prediction panel and used a second case-control cohort, the BOOST study, to independently validate the panel of features. This cohort included 398 patients, of whom 198 had Barrett’s oesophagus including 23 with oesophageal adenocarcinoma, and 200 controls.

Application of machine learning to analyse problems dates back over five decades, when predefined algorithms and models were used to allow computers to assess large datasets. This technology, a subset of what is termed artificial intelligence, can now be applied to identify populations at risk of diseases, such as Rosenfeld and colleagues have done with Barrett’s oesophagus since large electronic databases are available for analysis. The goal of these exercises is to see if associations exist between, for example, Barrett’s oesophagus and any collected information in the database, and if these associations can be quantified into an algorithm that can help identify these patients through a simple assessment, either conducted by a health-care professional or via self-assessment. This type of supervised machine learning has been applied in cardiology to find cardiomyopathy and in a wide assortment of medical conditions.

Reassuringly, most of the features identified through this approach to predict those at risk of Barrett’s oesophagus match those identified in recent studies that also examined factors in multivariate regression analysis, although not involving machine learning. However, similarity in the risk factors identified is not surprising because one form of supervised machine learning involves logistic regression. Machine learning logistic regression approaches examine all of the dataset using logistic regression and attempt to use the data to build a predictive model, and then the model is optimised for the fewest datapoints needed to find patients with the relative condition without human supervision or intervention. The key difference between machine learning and non-machine learning is the lack of any human filtering to restrict the data elements used in the analysis. Thus, the similarities between machine learning and non-machine learning logistic regression are to be anticipated.

Machine learning can involve several different modes including the traditional supervised and unsupervised modes, and a semi-supervised mode in which there are some classified patients and some unclassified patients—eg, in the case of this study, a proportion of patients would be classified as having Barrett’s oesophagus, leaving the rest to be predicted by the machine. The authors chose to use a supervised method, which is commonly done in the initial development of an algorithm. In addition to these learning algorithms, different approaches exist that involve regression algorithms, such as ordinary least squares regression analysis and linear regression. Additionally, instance-based algorithms are used in the field, such as k-nearest neighbors, as well as are clustering algorithms, which are based on association rules such as the Apriori algorithm. Finally, artificial neural networks are currently generating great interest because of analytical tools that include back-propagation, stochastic gradient descent, and other interesting classes of pattern matching. These clustering algorithms all influence the results obtained, which means that different machine learning approaches
can easily arrive at different results even looking at the same datasets. All of these machine learning approaches have been used previously in diagnostic approaches; however, only the logistic regression method used by Rosenfeld and colleagues is probably the most suitable because of the relatively small size of the dataset (by machine learning standards).

Given the wide variety of algorithms potentially generated, it is an essential next step for Rosenfeld and colleagues to do validation tests of these findings. The first is construct validation—ie, whether the study actually targets the condition it is investigating, which is determined by the nature of the question being studied. The construct in this study targets key outcomes in Barrett’s oesophagus. The conditions of content validity (ie, whether the study content addresses the condition of interest) and face validity (ie, does the study, at face value, appear to address the condition being screened) also seem to be met. A less known important validation criterion is construct validity, which is whether the findings in this study can be found in other studies. Although surprising criteria can be identified via machine learning, whether these criteria can be identified in other databases is important to enhance their validity. Criterion validation has been achieved since most of the criteria in this study have been previously described.

To determine if this algorithm will be applicable to larger datasets, it will need to be validated prospectively to see if it can identify people with Barrett’s oesophagus in community practices. All the patients from the BEST2 screening studies were derived from specialty clinics before endoscopy. Therefore, data on this feature might not be collected in most primary care records and so might be difficult to apply broadly.

Machine learning will be used in the future to aid clinicians to determine if their patients will need to be screened for specific conditions. This study shows that steps are already being taken to look for ways to trigger further investigation in patients based on commonly available information. However, if the goal of the study is truly population-based screening, we believe that one key criterion for machine learning validation should be application in a general clinical database.

*Kenneth K Wang, Cadman Leggett
Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN 55905, USA
wang.kenneth@mayo.edu
We declare no competing interests.

Copyright © 2020 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license.

1 Parasa S, Vennalaganti S, Gaddam S, et al. Development and validation of a model to determine risk of progression of Barrett’s esophagus to neoplasia. Gastroenterology 2018; 154: 1282–89.
2 Kunzmann AT, Thrift AP, Cardwell CR, et al. Model for identifying individuals at risk for esophageal adenocarcinoma. Clin Gastroenterol Hepatol 2018; 16: 1229–36.
3 Ross-Innes CS, Debridar-Bechar D, O’Donovan M, et al. Evaluation of a minimally invasive cell sampling device coupled with assessment of trefoil factor 3 expression for diagnosing Barrett’s esophagus: a multi-center case-control study. PLoS Med 2015; 12: e1001780.
4 Rosenfeld A, Grahm DG, Jevons S, et al. Development and validation of a risk prediction model to diagnose Barrett’s oesophagus (MARK-BE): a case-control machine learning approach. Lancet Digital Health 2020; 2: e37–48.
5 Galler BA. The value of computers to medicine. JAMA 1960; 174: 2151–62.
6 Huda A, Shah SJ, Castano A, et al. A machine learning model for the systematic identification of wild-type transthyretin cardiomyopathy. J Card Fail 2019; 25 (suppl): S53–54.
7 Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Med 2019; 25: 65–69.
8 Cherif W. Optimization of K-NN algorithm by clustering and reliability coefficients: application to breast-cancer diagnosis. Procedia Comput Sci 2018; 127: 293–99.