The Journal of Laryngology & Otology

cambridge.org/jlo

Patient classification of two-week wait referrals for suspected head and neck cancer: a machine learning approach

J W Moor[1], V Paleri[2] and J Edwards[3]

[1]ENT Department, Leeds Teaching Hospitals NHS Trust, [2]ENT Department, Royal Marsden Hospital, London and [3]Temporal Computing Ltd, Department of Computer Science, Newcastle upon Tyne, UK

Main Article

Mr J W Moor takes responsibility for the integrity of the content of the paper

Cite this article: Moor JW, Paleri V, Edwards J. Patient classification of two-week wait referrals for suspected head and neck cancer: a machine learning approach. *J Laryngol Otol* 2019;**133**:875–878. https://doi.org/10.1017/S0022215119001634

Accepted: 17 May 2019
First published online: 2 September 2019

Key words:
Machine Learning; Head And Neck Neoplasms; Diagnostic Techniques And Procedures

Author for correspondence:
Mr James Moor,
ENT Dept, Leeds General Infirmary,
Leeds LS1 3EX, UK
E-mail: jamesmoor@nhs.net

Abstract

Background. Machine learning algorithms could potentially be used to classify patients referred on the two-week wait pathway for suspected head and neck cancer. Patients could be classified into 'predicted cancer' or 'predicted non-cancer' groups.
Methods. A variety of machine learning algorithms were assessed using the clinical data of 5082 patients. These patients had previously been referred via the two-week wait pathway for suspected head and neck cancer to two separate tertiary referral centres in the UK. Outcomes from machine learning classification were analysed in comparison to known clinical diagnoses.
Results. Variational logistic regression was the most clinically useful technique of those chosen to perform the analysis and patient classification; the proportion of patients correctly classified as having 'non-cancer' was 25.8 per cent, with a false negative rate of 1 out of 1000.
Conclusion. Machine learning algorithms can accurately and effectively classify patients referred with suspected head and neck cancer symptoms.

Introduction

Machine learning techniques for the analysis of clinical data have yet to become the norm. The term 'machine learning' relates to a variety of relatively novel statistical techniques used for the analysis of large datasets.[1] It is said to have evolved from artificial intelligence, where algorithms provide computers with the ability to learn without being explicitly programmed. Machine learning techniques are the domain of the computer scientist; this is in contrast to standard statistical techniques, which are regularly applied to data by healthcare professionals and biostatisticians to demonstrate clinical or statistical significance. The benefit of using machine learning techniques lies with their ability to offer predictions relating to data that may be accurately applied from the test population to a non-test population.[2]

Head and neck cancer is an umbrella term relating to a heterogeneous group of tumours arising from the mucosa of the upper aerodigestive tract, paranasal sinuses and salivary glands. The incidence is generally low when compared to other common cancers. Age-standardised incidence rates in the UK have been reported as approximately 3 per 100 000 population for oral cavity and larynx cancer, and 0.6 per 100 000 population for salivary gland malignancy.[3] Referrals for suspected cancer are rising in the UK, with over 1.5 million urgent general practitioner referrals for suspected cancer (at any site) made in 2015, an increase of 50 per cent in the last four years.[4] The rapid referral system will continue to be a top priority for NHS England owing to lower cancer outcomes compared to Europe.[4]

For patients presenting with suspected cancer in primary care, the current recommendations are referral via the two-week wait pathway for investigation and management[5] as appropriate. The published literature in this area has consistently demonstrated that the rate of head and neck cancer diagnoses from all suspected two-week wait referrals is low, ranging from 5 to 15 per cent; a recent systematic review identified a pooled detection rate of 11.1 per cent, from 6 studies involving 1809 patients.[6]

The provision of the two-week wait service within secondary care is consultant led; current guidance states that patients are seen by clinicians who are core members of the head and neck multidisciplinary team, or by trainees who are directly supervised by core members.[5]

This study investigated whether machine learning could be part of the process that leads to efficiencies within this setting. There are differences in the strengths of machine learning compared to more conventional statistical analysis. A comparison summary of these strengths includes 'learning' (as opposed to 'fitting', used with more conventional statistical techniques), 'generalisation' versus 'test set performance', and 'supervised learning' versus 'regression or classification'.[7] We used these parameters to classify patients (with symptom and demographic data) into different diagnostic groups, albeit very broad ones: cancer and non-cancer.

Probabilistic classification is better served by a machine learning approach, as this offers better predictive capacity than statistical analysis. This is very important when considering 'out of sample' predictive performance, which refers to the accuracy of diagnostic classification for a new patient who is not part of the dataset from which the analysis, and algorithm development, has been undertaken. This, specifically, is regarded as the realm of machine learning.[8]

Materials and methods

An existing dataset of 5082 patients with suspected head and neck cancer, referred via the two-week wait system to 2 tertiary hospitals in England, was used for the analysis.[9] These data were collected in a mixed prospective and retrospective fashion. Data fields included basic demographics, presenting symptoms and final diagnosis. Data were subjected to various machine learning analytical processes that are regarded as 'best in class' classifiers.

A standard 10-fold cross-validation framework was applied to ensure the integrity of results.[10] This means that a rotating 10 per cent of the data were held back from analysis for all of the machine learning classifiers and used as a test set; thus, only 90 per cent of the data were subjected to a training cycle. The test set was then used to test each machine learning classifier, and each machine learning classifier was tested 10 times, each time with a test set comprising different patients.

Mean and standard deviation values were produced for patients classified into either 'predicted cancer' or 'predicted non-cancer' groups. The test sets were then compared with known outcomes for patients (actual cancer and actual non-cancer). Thus, four separate outcome groups were produced: predicted non-cancer, actual non-cancer (true negatives); predicted non-cancer, actual cancer (false negatives); predicted cancer, actual cancer (true positives); and predicted cancer, actual non-cancer (false positives).

The data analysis was implemented using Scikit-learn, a leading open source web-based collection of machine learning tools.[11] For each machine learning classifier, the classification of predicted cancer was made if the calculated probability was more than 0.5, consistent with a 'winner takes all' approach. Two modifications to the standard logistic regression approach were made, as described below.

Regarding the first modification, for variational logistic regression, the threshold value to assert whether cancer was true was modified to: p (cancer) > 0.08. This was based on empirical data from previously published area under the receiver operator characteristic curve analysis.[9] This was considered a cautious approach to classifying patients, as they would be classified as having a cancer diagnosis without strong suspicion based on their data. Hence, the patients classified as 'predicted cancer' would intentionally have a high false positive rate, to include all patients who did have cancer (all the true positives) as well as a large number who did not (false positives). However, it also acted to decrease the risk of false negative classifications (i.e. classifying an 'actual cancer' patient as 'predicted non-cancer').

In a second modification of logistic regression, the importance of an 'actual cancer' diagnosis during the analytical part of the machine learning process was biased empirically by a factor of 100. This reflected both the imbalanced nature of the data (large numbers of patients with benign diagnoses, and relatively few cancer diagnoses), and the clinical importance of a cancer diagnosis over a non-cancer diagnosis. Thus, the technique erred towards classifying patients as 'predicted cancer'.

The following machine learning techniques were assessed: logistic regression (a standard linear classifier); K nearest neighbour (a similarity-based classifier that utilises the proximity of a new example to examples from the original data); support vector machine (data are transformed into a domain more conducive for linear analysis, and support vectors (which represent the edge of the decision boundary) are established and then used for subsequent classification); decision tree classifier (an algorithm that generates a tree of clause statements which classifies the data in a piecemeal fashion); random forest (an ensemble of decision trees that produce a more moderated classification as a result of the averaging of decisions from multiple separate classifiers (this is thought to give lower out-of-sample error and require less 'tuning' for a specific dataset)); AdaBoost (weak classifiers are boosted and combined to strengthen their overall predictive capacity (an ensemble decision)); naive Bayes (a histogram method of analysis that relies on a simple application of conditional probability); linear discriminant analysis (a traditional technique similar to and now superseded by linear support vector machine); weighted logistic regression (an algorithmic modification that amplifies the logistic regressions to reflect the importance of specific class classification); and variational logistic regression (a Bayesian logistic regression approach that is more moderate in its prediction where indicative data are sparse).

The performance of each machine learning classifier is presented using confusion matrix scores. This is a standard machine learning approach that provides all combinations of results classified into actual and predicted categories.

Results

There were 5082 patients in the dataset; 367 patients were excluded because of missing data. Of the remaining 4715 patients, 397 received a cancer diagnosis (8.4 per cent).

The performance results for each machine learning classifier are presented in Table 1. Data are presented as percentages of patients ± standard deviation within test sets correctly classified as 'cancer' or 'non-cancer'.

The technique that proved the most clinically useful for accurate classification was variational logistic regression. This technique correctly classified 25.8 ± 8.9 per cent of patients as not having cancer (true negatives), with a false negative rate of only 0.1 ± 0.16 per cent. The true positive rate was 7.7 ± 1.4 per cent, and the false positive rate was 66.4 ± 8.9 per cent.

Discussion

This study investigated machine learning approaches for the classification of patients referred on the two-week wait pathway for suspected head and neck cancer. The most clinically useful technique of the 11 machine learning approaches assessed was variational logistic regression. This technique demonstrated that approximately one-quarter of patients referred could accurately be classified as not having cancer and receive a non-cancer diagnosis, whilst the risk of a patient being classified as having non-cancer but ultimately receiving a cancer diagnosis was only 1 in 1000. We estimate that this false negative rate is probably broadly equivalent to clinical practice, although this conjecture is purely anecdotal (derived

https://doi.org/10.1017/S0022215119001634 Published online by Cambridge University Press

Table 1. Comparison of different machine learning techniques used for classifying patients

Machine learning technique	True negative	False negative	False positive	True positive
Logistic regression	91.8 ± 2.0	7.6 ± 1	0.3 ± 0.5	0.2 ± 0.4
K nearest neighbour (k = 5)	92.2 ± 1.4	7.8 ± 1.4	0.0 ± 0.0	0.0 ± 0.0
Support vector machine	91.8 ± 1.3	7.6.1 ± 1.3	0.5 ± 0.4	0.1 ± 0.2
Decision tree classifier	92.2 ± 1.4	7.8 ± 1.4	0.0 ± 0.0	0.0 ± 0.0
Random forest	92.0 ± 1.4	7.6 ± 1.3	0.2 ± 0.2	0.2 ± 0.2
AdaBoost	84.7 ± 8.3	6.3 ± 2.1	7.5 ± 8.2	1.6 ± 1.6
Naive Bayes	91.7 ± 1.4	7.5 ± 1.3	0.5 ± 0.4	0.3 ± 0.2
Linear discriminant analysis	73.5 ± 25.2	5.4 ± 2.7	18.6 ± 25.5	2.4 ± 2.4
Weighted logistic regression	28.6 ± 8.4	0.2 ± 0.2	63.6 ± 8.2	7.6 ± 1.4
Variational logistic regression	25.8 ± 8.9	0.1 ± 0.16	66.4 ± 8.9	7.7 ± 1.4

Data represent means ± standard deviations

from the clinical authors) and the acquisition of such data has not been undertaken. There are no published data available in the literature to confirm this assertion; however, it is accepted that diagnostic delay, of which 'professional delay' (a false negative cancer diagnosis would be included in this category) may be implicated. In a systematic review published in 2012, a false negative rate diagnosis rate for head and neck cancer in secondary care was not evaluated.[12]

We have presented the results of the analysis in terms of percentages of patients classified according to a particular diagnostic group and standard deviation values, to ensure relevance to a clinical audience; we have assumed no prior knowledge of machine learning and have presented the most clinically relevant statistics (false negative and true negative decision rates) as percentages. These statistics clearly illuminate the underlying implications of the study: namely, the proportion of referrals that could potentially be removed from further investigation and the probability of a patient with a potential diagnosis of cancer not receiving any further investigation.

We briefly described (above) the current provision of two-week wait services for suspected head and neck cancer patients, and this raises several issues relating to provision of this service. Firstly, referral rates from primary care may be too high. However, this statement only holds merit on purely numerical terms; given the low age-standardised incidence rates of new head and neck cancer diagnoses compared to other common tumour sites (e.g. bowel, breast), primary care physicians may have a low threshold for referral, to avoid the low but definite risk of a missed cancer diagnosis.

Secondly, we could state that the provision of two-week wait clinics are an inefficient use of consultant time. Given that the provision of consultant-led out-patient clinics are an expensive resource, it stands to reason that cheaper models of delivering the two-week wait system could be implemented, to help relieve financial pressures. This is with the proviso that any variation from the current pathway is provided by appropriately trained staff, with safeguards and quality assurance processes embedded, and subject to serial quality assurance and/or audit. The logical extension of this position may be the development of a model where the only patients who ever see a core multidisciplinary team consultant are those with a cancer diagnosis, pre-cancerous conditions or complex benign pathologies. We practice within a finite National Health Service (NHS) financial framework, and all healthcare professionals are regularly and repeatedly asked to identify areas where potential efficient savings can be made.

There are strong counter-arguments to this view. At a patient-centric level (although when is the delivery of healthcare not patient-centric?), the delivery of a non-cancer diagnosis to a worried patient from a clinician who is an expert in treating cancer patients is extremely powerful, and may represent the most efficient way of reassuring the majority of patients who do not have cancer, and thus need no further investigations or interventions. Facilitating the next generation of consultant surgeons to learn this skill, and implement it within their clinical workload, is an acknowledged expectation of all supervisors and trainers.

- Two-week wait referrals to head and neck services include a large majority of patients with non-cancer diagnoses
- The identification of high-risk individuals could lead to prioritisation of some investigations and would therefore be advantageous
- Machine learning analysis of symptom data and demographics can accurately classify patients, at an early stage, into cancer and non-cancer groups
- This can be achieved with a low false negative rate, avoiding inaccurate non-cancer diagnosis classifications

The utility of adopting machine learning based patient classification in the suspected head and neck cancer setting relates to the proportion of patients accurately classified as true negatives; these are the patients who could potentially be managed in a different way to the current patient pathway. In our analysis, this represented approximately one-quarter of all patients referred on the two-week wait pathway. We do not advocate that these patients be removed from the pathway; that would be a decision for individual NHS Trusts to explore and develop. Nevertheless, this represents a potential opportunity for the development of alternative pathways within secondary care, in which non-cancerous diagnoses can be filtered out early in the pathway. Similarly, when considering the low diagnostic rate of two-week wait referrals from primary care, the opportunity exists to reduce the overall referral rate. Both scenarios can easily be seen to represent potential efficiency savings for the NHS as a whole, especially in the context of approximately 100 000 two-week wait referrals annually for suspected head and neck cancer in the UK.[4,5]

https://doi.org/10.1017/S0022215119001634 Published online by Cambridge University Press

The use of risk stratifiers is not new for a variety of healthcare domains. Existing clinical tools include QRisk, QCancer and Cancer Decision Support.[13–15] However, the statistical approach utilised in these existing tools would be regarded as conventional, as opposed to the machine learning approach we have undertaken. There is a general paucity of literature relating to the use of machine learning within clinical medicine and surgery.[16–18] We consider this to be the first presentation of data relating to the machine learning classification of suspected head and neck cancer patients referred from primary to secondary care.

We acknowledge some weaknesses in this study. The data used in this analysis were collected from two head and neck cancer centres in the UK, Newcastle upon Tyne (the majority) and Birmingham, and therefore are less subject to regional variation. However, application of the findings to the UK population, or internationally, cannot be assumed. We also recognise that whilst the whole dataset may be considered relatively large, only 397 patients received a cancer diagnosis, which is not a large number of patients against which to make prospective comparisons in a non-theoretical setting.

Conclusion

Machine learning classifiers can be used to accurately classify patients referred on the two-week wait pathway for suspected head and neck cancer. Variational linear regression demonstrated the best balance of accurate true negative classification against a low false negative classification.

Acknowledgement. This study was funded by DotForge Health and Data, Leeds, UK.

Competing interests. None declared

References

1 MacKay DJC. *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press, 2003
2 Clifton DA. *Machine Learning for Healthcare Technologies*. London: Institution of Technology and Engineering, 2016
3 Oxford Cancer Intelligence Unit. Profile of Head and Neck Cancers in England: Incidence, Mortality and Survival. In: http://www.ncin.org.uk/cancer_type_and_topic_specific_work/cancer_type_specific_work/head_and_neck_cancers/head_and_neck_cancer_hub/resources [5 October 2016]
4 NHS England. Achieving World Class Cancer Outcomes: Taking the strategy forward. In: https://www.england.nhs.uk/wp-content/uploads/2016/05/cancer-strategy.pdf [5 October 2016]
5 NHS England Interim Management and Support. Delivering Cancer Waiting Times: A Good Practical Guide. In: https://www.england.nhs.uk/wp-content/uploads/2015/03/delivering-cancer-wait-times.pdf [5 October 2016]
6 Drinnan M, Paleri V, Kumar R, Mehanna H. Efficacy of the two week wait referral system for head and neck cancer: a systematic review. *Ann R Coll Surg Engl* 2012;**94**:101–5
7 Statistics vs. Machine Learning, fight! In: http://brenocon.com/blog/2008/12/statistics-vs-machine-learning-fight/ [17 November 2016]
8 Breiman L. Statistical modeling: the two cultures. *Statistical Science* 2001;**16**:199–215
9 Tikka T, Pracy P, Paleri V. Refining the head and neck cancer referral guidelines: a two-centre analysis of 4715 referrals. *Clin Otolaryngol* 2016;**41**:66–75
10 Webb AR, Copsey AD. *Statistical Pattern Recognition*. Chichester: Wiley, 2011
11 Scikit-learn. In: http://scikit-learn.org/stable/index.html [17 November 2016]
12 Seone J, Takkouche B, Varela-Centelles P, Tomas I, Seoane-Romero JM. Impact of delay in diagnosis on survival to head and neck carcinomas: a systematic review with meta-analysis. *Clin Otolaryngol* 2012;**37**:99–106
13 Hippersley-Cox J, Coupland C, Robson J, Brindle P. Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database. *BMJ* 2010;**341**:c6624
14 QCancer. In: https://qcancer.org [17 November 2016]
15 Cancer Decision Support (CDS) tool. In: http://www.macmillan.org.uk/about-us/health-professionals/programmes-and-services/prevention-early-diagnosis-programme/cancer-decision-support-tool.html [17 November 2016]
16 Taylor RA, Pare JR, Venkatesh AK, Mowafi H, Melnick ER, Fleischman W *et al*. Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data-driven, machine learning approach. *Acad Emerg Med* 2016;**23**:269–78
17 Gultepe E, Green JP, Nguyen H, Adams J, Albertson T, Tagkopoulos I. From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system. *J Am Med Inform Assoc* 2014;**21**:315–25
18 Klann JG, Anand V, Downs SM. Patient-tailored prioritization for a pediatric care decision support system through machine learning. *J Am Med Inform Assoc* 2013;**20**:e267–74

https://doi.org/10.1017/S0022215119001634 Published online by Cambridge University Press