



## Application of a machine learning algorithm for detection of atrial fibrillation in secondary care



Kevin G. Pollock<sup>a,\*</sup>, Sara Sekelj<sup>b</sup>, Ellie Johnston<sup>b</sup>, Belinda Sandler<sup>a</sup>, Nathan R. Hill<sup>a</sup>, Fu Siong Ng<sup>c,d</sup>, Sadia Khan<sup>c,d</sup>, Ayman Nassar<sup>a</sup>, Usman Farooqui<sup>a</sup>

<sup>a</sup>Bristol-Myers Squibb Pharmaceuticals Ltd, Uxbridge Business Park, Sanderson Road, Uxbridge, Middlesex UB8 1DH, UK

<sup>b</sup>Imperial College Health Partners, 30 Euston Square, London NW1 2FB, UK

<sup>c</sup>Chelsea & Westminster Hospital NHS Foundation Trust, 369 Fulham Road, London SW10 9NH, UK

<sup>d</sup>Imperial College London, London W12 0NN, UK

### ARTICLE INFO

#### Article history:

Received 27 October 2020

Accepted 2 November 2020

Available online 29 November 2020

#### Keywords:

Atrial fibrillation

Machine learning

Artificial intelligence

Diagnosis

### ABSTRACT

Atrial fibrillation (AF) is the most common sustained heart arrhythmia and significantly increases risk of stroke. Opportunistic AF testing in high-risk patients typically requires frequent electrocardiogram tests to capture the arrhythmia. Risk-prediction algorithms may help to more accurately identify people with undiagnosed AF and machine learning (ML) may aid in the diagnosis of AF. Here, we applied an AF-risk prediction algorithm to secondary care data linked to primary care data in the DISCOVER database in order to evaluate changes in model performance, and identify patients not previously detected in primary care. We identified an additional 5,444 patients who had an AF diagnosis only in secondary care during the data extraction period. 2,696 (49.5%) were accepted by the algorithm and the algorithm correctly assigned 2,637 (97.8%) patients to the AF cohort. Using a risk threshold of 7.4% in patients aged  $\geq 30$  years, algorithm sensitivity and specificity was 38% and 95%, respectively. Approximately 15% of AF patients assigned to the AF cohort by the algorithm had a secondary care diagnosis with no record of AF in primary care. These additional patients did not substantially alter algorithm performance. The additional detection of previously undiagnosed AF patients in secondary care highlights unexpected potential utility of this ML algorithm.

Crown Copyright © 2020 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### 1. Introduction

Atrial fibrillation (AF) is the most common sustained heart arrhythmia [1] and may increase the risk of thromboembolic stroke by five-fold [2]. Furthermore, patients with AF experience more severe strokes than those without AF [3]. The global incidence of AF has significantly increased over time and prevalence is estimated to be 3% in the UK population [4,5]. However, AF can be difficult to diagnose because it is often paroxysmal and/or asymptomatic or minimally symptomatic [5]. Currently, there is no formal screening programme for AF in the UK. Detection of undiagnosed AF must equally consider the associated patient burden, healthcare resource use, diagnostic sensitivity and specificity. Opportunistic testing for AF in high-risk patients (such as those with irregular pulse or aged  $\geq 65$  years as risk of AF increases with age) typically requires frequent electrocardiogram (ECG) tests to capture the arrhythmia [5]. Risk-prediction algorithms may help

to more accurately identify people with undiagnosed AF and machine learning (ML) may aid in the diagnosis of AF, through targeted screening modalities [6,7]. Machine learning is particularly useful for examining non-linear associations and complex interactions between variables without having to specify these relationships *a priori*. Investment in, development, and adoption of artificial intelligence across the NHS is at the forefront of the UK government's healthcare agenda [8].

As AF has a complex aetiology, models developed using ML methods may offer improved predictive performance compared with models built with classical statistical methods to estimate AF incidence. A recently published AF risk prediction algorithm, developed using routinely collected UK primary care data from the Clinical Practice Research Datalink (CPRD) was better able to identify patients at highest risk of AF compared with existing models [6]. Compared to the CHARGE-AF model, the AF-risk prediction algorithm was able to reduce the number of high-risk patients needed to be screened (NNS) to identify one case of AF by 31%, from 13 to 9. This was further validated in the DISCOVER primary care database (general or medical practitioner data) in North-West

\* Corresponding author.

E-mail address: [kevin.pollock@bms.com](mailto:kevin.pollock@bms.com) (K.G. Pollock).

**Table 1**  
Algorithm performance in WSIC dataset with the CPRD risk threshold of 7.4% and 5.5% after addition of patients with AF diagnosis in secondary care only.

Risk threshold 7.4%	AF		Risk threshold 5.5%	AF	
	Yes	No		Yes	No
<b>Patients aged <math>\geq 30</math> years</b>					
Algorithm			Algorithm		
Yes	8737	27,700	Yes	11,558	45,090
No	14,355	558,659	No	11,534	541,269
Sensitivity	38%		Sensitivity	50%	
Specificity	95%		Specificity	95%	
PPV	24%		PPV	20%	
NPV	97%		NPV	98%	
1/PPV (NNS)	4		1/PPV (NNS)	5	
<b>Patients aged <math>\geq 65</math> years</b>					
Algorithm			Algorithm		
Yes	7797	23,169	Yes	10,252	37,143
No	8182	82,298	No	5727	68,324
Sensitivity	49%		Sensitivity	64%	
Specificity	78%		Specificity	65%	
PPV	25%		PPV	22%	
NPV	91%		NPV	92%	
1/PPV (NNS)	4		1/PPV (NNS)	5	

London (NWL) encompassing a population of 2.5 million [9], with a sensitivity of 75% and specificity of 99.1%. Here, we employed a machine learning algorithm that had been developed and validated in primary care datasets, with the aim of assessing its performance and utility when applied to secondary care linked data (hospital data). This machine learning model leveraged secondary care data to detect undiagnosed AF with reasonable performance.

## 2. Methods

A retrospective cohort study was undertaken using coded secondary care data from the Whole Systems Integrated Care (WSIC) dataset, which is one of Europe's largest patient-level datasets, containing data from approximately 2.5 million patients across NWL. Study data were obtained through the DISCOVER secure environment, which was developed by Imperial College Health Partners, the Academic Health Science Network for NWL. Favourable ethical opinion was secured in October 2018 to use the Discover Research Platform for research purposes for a period of five years. Local Research and Development Department approval was obtained from the NWL Data Research Access Group on 18th October 2018. Patient consent was not required because the study was retrospective study using anonymised data.

The development, application and validation of the machine learning algorithm in primary care systems has been described in detail elsewhere [5,9]. De-identified secondary care hospital data for all patients were extracted via DISCOVER for the period 01 January 2001 to 31 December 2016. In order to assess the model's predictive ability to distinguish between patients at high and relatively lower risk of AF, the first step was to generate a threshold for the risk of AF among patients not flagged by the algorithm for screening. These thresholds were derived from baseline risk factors (age, previous cardiovascular disease, antihypertensive medication usage) and additional time-varying predictors (proximity of cardiovascular events, body mass index [levels and changes], pulse pressure, and frequency of blood pressure measurements). The risk-prediction model did not look at ECG data as it was initially intended to identify patients at risk of AF in primary care, before they developed disease that might result in consequent ECGs. Based on the per-patient risk scores returned by the algorithm in the original primary care dataset (CPRD) [6], a risk threshold of 7.4% was set for AF versus not AF. Thus, all patients with risk scores  $\geq 7.4\%$  were categorised as being at high risk of developing

AF, and those with a risk scores  $< 7.4\%$  were categorised as being at low risk.

The predictive performance of the model was then assessed using the following metrics: sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and number needed to screen (NNS). All analyses were performed using Microsoft Excel 2013, Stata version 15.0 and R version 3.6.0.

## 3. Results

We identified an additional 5,444 patients who had an AF diagnosis only in secondary care during the data extraction period. These patients were not diagnosed with AF in primary care. 2,696 (49.5%) were accepted by the algorithm and 2,748 (50.5%) were excluded. Among the 2,696 patients included, the algorithm correctly assigned 2,637 (97.8%) patients to the AF cohort and incorrectly assigned 59 patients (2.2%) to the non-AF cohort. These additional patients did not substantially alter algorithm performance (Table 1). Using a risk threshold of 7.4% in patients aged  $\geq 30$  years, algorithm sensitivity and specificity was 38% and 95%, respectively. Altering the risk threshold to 5.5% increased the sensitivity to 50% with specificity remaining at 95%.

Approximately 15% of AF patients assigned to the AF cohort by the algorithm had only a secondary care diagnosis, with no READ codes relating to AF in their linked primary care record.

## 4. Discussion

Granular data collected from secondary care may represent an under-utilized method for improving the diagnosis of many health conditions [10]. Here, we employed a machine learning algorithm that had been developed and validated in primary care datasets, with the aim of assessing its performance and utility when applied to secondary care linked data. This machine-learning model leveraged secondary care data to detect undiagnosed AF with reasonable performance. However, the low prevalence of the disease resulted in a low positive predictive value, and for clinically meaningful sensitivity thresholds to be actionable, confirmatory testing with high specificity (e.g. electrocardiogram) would be required following model detection. However, the additional detection of AF in previously undiagnosed patients in secondary care highlights unexpected potential utility of this machine learning algorithm, with a number needed to screen (NNS) of 5. Further studies are required to externally validate algorithm performance in other sec-

ondary care datasets and explore the feasibility of embedding algorithms into the perioperative electronic health record for real-world use by clinicians.

### Summary declarations of interest

KGP, BS, NRH, AN and UF are employees of Bristol-Myers Squibb Pharmaceuticals Ltd. SS and EJ are employees of Imperial College Health Partners, which received funding from BMS to undertake this study. FSN acknowledges funding from the British Heart Foundation (RG/16/3/32175) and the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre (BRC).

### Authors' contributions

BS, NRH, and UF contributed to the conception or design of the work. All authors contributed to the acquisition, formal analysis, or interpretation of data for the work. KP drafted the manuscript. All authors critically revised the manuscript, gave final approval and agree to be accountable for all aspects of work ensuring integrity and accuracy.

### Role of the funding source

This work was supported by Bristol-Myers Squibb Pharmaceuticals Ltd who provided funding for data collection and data analysis.

### Data sharing statement

The datasets analysed in this study are available in the DISCOVER database and available on request from Imperial College

Health Partners at <https://www.registerfordiscover.org.uk/researchers/how-to-access-discover>.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### References

- [1] Atrial Fibrillation Association Anticoagulation Europe (UK), The AF Report - Atrial Fibrillation: Preventing A Stroke Crisis, 2012.
- [2] P.A. Wolf, R. Abbott, W. Kannel, Atrial fibrillation as an independent risk factor for stroke: the Framingham Study, *Stroke* 22 (1991) 983–988.
- [3] H.J. Lin, P.A. Wolf, M. Kelly-Hayes, et al., Stroke severity in atrial fibrillation. The framingham study, *Stroke* 27 (1996) 1760–1764.
- [4] S.S. Chugh, R. Havmoeller, K. Narayanan, et al., Worldwide epidemiology of atrial fibrillation: a global burden of disease 2010 study, *Circulation* 129 (2014) 837–847.
- [5] National Institute for Health and Care Excellence. Clinical guideline [CG180], Atrial fibrillation: management, 2014.
- [6] N.R. Hill, D. Ayoubkhani, P. McEwan, et al., Predicting atrial fibrillation in primary care using machine learning, *PLoS ONE* 14 (2019) e0224582.
- [7] J. Oster, J.C. Hopewell, K. Zibera, et al., Identification of patients with atrial fibrillation: a big data exploratory analysis of the UK Biobank, *Physiol. Meas.* 41 (2020), <https://doi.org/10.1088/1361-6579/ab6f9a>.
- [8] Department of Health, NHS Long Term Plan, 2019.
- [9] S. Sekelj, B. Sandler, E. Johnston, et al., Detecting undiagnosed atrial fibrillation in UK primary care: validation of a machine learning prediction algorithm in a retrospective cohort study, *Eur. J. Prev. Cardiol.* (2020), Aug 13:2047487320942338. doi: 10.1177/2047487320942338. Epub ahead of print.
- [10] M.R. Mathis, M.C. Engoren, H. Joo, et al., Early detection of heart failure with reduced ejection fraction using perioperative data among noncardiac surgical patients: a machine-learning approach, *Anesth. Analg.* 130 (2020) 1188–1200, <https://doi.org/10.1213/ANE.0000000000004630>.