Protocol: A nationwide deep learning pipeline to predict stroke and COVID-19 death in atrial fibrillation

Sub-project of CCU004 COVID and Cardiovascular Disease Risk Prediction – Aim 2 Derive, validate and undertake public health modelling of contemporary CVD risk prediction models

Amendments

October 20th 2021: base version

November 10th 2021: update v1

- Minor adjustments to background and methods wording – broader feedback on discussion topics will be incorporated into manuscript for publication
- Updated experiment runs from 2 to 3

November 30th 2021: update v2

- Updated stroke prediction outcome to ischaemic stroke only
- Updated sampling approach and scenarios reported on
- More detailed explanations of final specifications for methodology will be incorporated into manuscript for publication

January 24th 2022: update v3

- Added interpretability analyses (calibration curve and shapley additive explanations (SHAP) values) for most performant model

Background

Recent advances in artificial intelligence can provide the basis for more personalised medical recommendations[1]. In particular, advances in modelling large sequences of text using deep learning (DL)[2, 3] has opened up the possibility of using long term patient trajectories held in electronic health records [4, 5]. The performance of these DL models on the selected disease prediction tasks is impressive [4,5] but there is limited comparison to existing clinical prediction tools or exploration of how they could practically improve clinical practice.

Anticoagulant prescribing decisions in atrial fibrillation (AF) offer a use case where the benchmark stroke risk prediction tool (CHA_2DS_2-VASc) used regularly in clinical practice could be meaningfully improved. AF is a disturbance of heart rhythm affecting 37.5 million people globally[6] and significantly increases stroke risk[7]. Anticoagulants reduce the risk of stroke[8] and are recommended for people with AF and a high risk of stroke, broadly defined as a CHA_2DS_2-VASc >=2 based on the

National Institute for Health and Care Excellence (NICE) threshold[9,10]. However, NICE's own evidence review highlights the need for improved stroke risk assessment[11] and shows that whilst CHA_2DS_2-VASc is good for identifying people potentially at risk of stroke (high sensitivity) it is poor at identifying people who may not have a stroke (low specificity)[12]. Whilst evidence is limited, the predictive performance of CHA_2DS_2-VASc appears even worse at assessing the risk of first stroke for people with AF, with discriminatory statistics (e.g. c-index, area under the receiver operating characteristics curve (AUC) below 0.60 [13].

COVID-19 has presented another risk factor for people with AF, who are at increased risk of poor outcomes if they become infected[14]. Recent research from our group and others [15, 16] has observed that pre-existing use of antithrombotics, particularly anticoagulants, is associated with lower odds of people with AF dying from COVID-19. A prediction model that could also identify which people with AF were at greatest risk of COVID-19 death and could further inform anticoagulant prescribing decisions.

This study, therefore, aims to develop and test a DL model that uses an individual's medical history (represented as a sequence of codes) to predict first ischaemic stroke in people with AF, and as a secondary outcome, COVID-19 death. Results will be directly compared against more conventional machine learning (ML) methods and CHA_2DS_2-VASc to support translation to clinical practice. As the first DL model to be developed using the nationwide linked electronic health record (EHR) data in the NHS Digital Trusted Research Environment (TRE) for England, this study will also provide a framework for other researchers and clinical use cases.

Methods

Study population

Individuals will be eligible for the sample cohorts if they have five or more recorded events across GDPPR and HES, were >= 18 years old and alive on January 1st 2020, have available sex, ethnicity and GP practice location data (based on most recent, available data across primary care (GDPPR), secondary care (HES APC) and death registrations (ONS)) and have a diagnosis of AF (coded in GDPPR). People with AF who also had a stroke diagnosis will be included only if their stroke occurred after the date of their first AF diagnosis.

To predict COVID-19 death for people with AF, the inclusion criteria of a COVID-19 event will be defined as any of a positive test (polymerase chain reaction or lateral flow), a coded diagnosis in primary or secondary care or a COVID-19 diagnosis on a death certificate.

The COVID-19 death outcome included people with a COVID-19 diagnosis on their death certificate in any position, a registered death within 28 days of their first recorded COVID-19 event or a discharge destination denoting death after a COVID-19 hospitalisation (see Thygesen et al (https://github.com/BHFDSC/CCU013_01_ENG-COVID-19_event_phenotyping/tree/main/phenotypes) for further details and phenotyping algorithms).

Follow-up for both stroke and COVID-19 death will be conducted up to May 1st 2021.

Prediction task

The primary task will be to predict the binary outcome of first ischaemic stroke in people with AF.
Two types of inputs will be used, sequential medical histories and static covariates. Sequential medical histories are an ordered list of unique biomedical codes assembled from an individual's medical history in GDPPR and HES APC (e.g. [SNOMED-CT code 1, SNOMED-CT code 2, ICD-10 code 1, SNOMED-CT code 3, SNOMED-CT code n]) up to the date of AF diagnosis.
Static covariates are variables that represent information for each individual at a single point in time (e.g. age at AF diagnosis, ethnicity, sex).

The same approach will be applied to predicting COVID-19 death with sequential medical histories collated up to the first COVID-19 event date, age recorded at January 1st 2020 and index of multiple deprivation decile and most recent BMI added as static covariates.

Model architecture

Transformer and LSTM network architectures will be selected for the DL models due to their suitability for sequence modelling [3,17,18] with logistic regression, random forest and XGboost as benchmark ML methods and CHA_2DS_2-VASc>=2 as the baseline

For the ML methods (logistic regression, random forest and XGboost) sequential medical histories will be represented as one hot encoded variables for each unique code in the cohort sample. Static inputs will then be represented as covariates in their continuous or categorical form.

The DL models (transformer and LSTM) will require a more sophisticated input representation and architecture. A vocabulary of each unique code from the sequential medical histories of the cohort sample will be assembled and used to create a trainable set of vector embeddings for each code.

The hyperparameters of the DL architecture (e.g. number and types of layers, learning rate) will then be developed through optimisation on the training and validation data.

The same architecture will then be used for all experiments on the test data. It is anticipated that deeper layered models and hyperparameter optimisation using methods such as grid search will not be possible due to a lack of GPUs and memory capacity on the TRE.

Model sampling, training and evaluation

The entire eligible study population will be randomly split 80:20 into training and test datasets. Prevalence of first ischaemic stroke after AF is low (~2%) which means the target class (stroke) is likely to be highly imbalanced in the training data. To address this for the training data, we will create a rebalanced sample by selecting all stroke cases and randomly selecting (with replacement) controls at a ratio of 1 control to 1 stroke case. This ratio will be selected after initial experimentation for training on other ratios (1-to-3 and population prevalence) with the best performing ratio reported on. The testing dataset will be kept at the population prevalence. The same approach will be adopted for COVID- 19 death which has moderate prevalence (23.2%) in people with AF.

Random sub- samples of 10,000 people will then be selected from the rebalanced training and testing datasets. The training sub-sample will be split into model training data (n=8000) and validation data (n=2000), with the model with the highest AUC on the validation data selected for testing. To assess the reliability of model predictions, three versions of each training and test sub-sample will be created with averages and confidence intervals reported in results.
The maximum length of medical codes included for each individual will also be limited to reduce computational requirements. Models will be trained and tested with a limit of 100 medical codes which included all codes for 99% of the AF first stroke cohort and >75% of AF+COVID-19 cohort.

Training parameters for the DL models will be kept consistent (e.g. number of epochs, batch size) across LSTM and transformer architectures.

Model performance will be assessed using accuracy, AUC, sensitivity, specificity and precision with AUC reported as the primary metric and CHA2DS2-VASc>=2 used as the baseline.

Models will also be developed, analysed and reported on for sex, age and ethnicity sub-groups.

Interpretability analyses in the form of a calibration curve and shapley additive explanations (SHAP) values analysis will be conducted on the most performant model.

References

[1] Jiang et al, Artificial intelligence in healthcare: past, present and future - https://svn.bmj.com/content/2/4/230.abstract
[2] Vaswani, A. et al Attention Is All You Need - https://arxiv.org/abs/1706.03762
[3] Devlin, J. et al BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding - https://www.aclweb.org/anthology/N19-1423/
[4] Li, Y et al BEHRT: Transformer for Electronic Health Records - https://www.nature.com/articles/s41598-020-62922-y
[5] Rasmy, L et al Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction - https://www.nature.com/articles/s41746-021-00455-y
[6] Lippi G, Sanchis-Gomar F, Cervellin G. Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge. Int J Stroke 2021;16:217–21. doi:10.1177/1747493019897870
[7] Asplund K. High prevalence of atrial fibrillation among patients with ischemic stroke. Stroke 2014;45:2599–605. doi:10.1161/STROKEAHA.114.006070
[8] Campbell Cowan J, Wu J, Hall M, et al. A 10 year study of hospitalized atrial fibrillation-related stroke in England and its association with uptake of oral anticoagulation. Eur Heart J 2018;39:2975–83. doi:10.1093/eurheartj/ehy411
[9] NICE. Recommendations | Atrial fibrillation: diagnosis and management | Guidance | NICE. Published Online First: 2021.https://www.nice.org.uk/guidance/ng196/chapter/Recommendations#stroke-prevention (accessed 30 Jun 2021).
[10] Proietti M, Lane DA, Boriani G, et al. Stroke Prevention, Evaluation of Bleeding Risk, and Anticoagulant Treatment Management in Atrial Fibrillation Contemporary International Guidelines. Can J Cardiol 2019;35:619–33. doi:10.1016/J.CJCA.2019.02.009
[11] NICE guidance 2021, Atrial fibrillation: diagnosis and management: Recommendations for research - https://www.nice.org.uk/guidance/ng196/chapter/Recommendations-for-research
[12] NICE guidance 2021, Atrial fibrillation: diagnosis and management Evidence reviews C and D: Tools to predict stroke in people with atrial fibrillation - https://www.nice.org.uk/guidance/ng196/evidence/c-and-d-tools-to-predict-stroke-in-people-with-atrial-fibrillation-pdf-9081923440
[13] Sullivan, J et al Combining Clinical and Polygenic Risk Improves Stroke Prediction Among Individuals With Atrial Fibrillation -

https://www.ahajournals.org/doi/10.1161/CIRCGEN.120.003168?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200pubmed
[14] Yang H, Liang X, Xu J, et al. Meta-Analysis of Atrial Fibrillation in Patients With COVID-19. Am. J. Cardiol. 2021;144:152–6. doi:10.1016/j.amjcard.2021.01.010
[15] Handy, A. et al Evaluation of antithrombotic use and COVID-19 outcomes in a nationwide atrial fibrillation cohort - https://www.medrxiv.org/content/10.1101/2021.09.03.21263023v1
[16] The OpenSAFELY Collaborative / Wong, A. et al Association between oral anticoagulants and COVID-19 related outcomes: two cohort studies - https://www.medrxiv.org/content/10.1101/2021.04.30.21256119v1
[17] Olah, C “Understanding LSTM networks” - https://colah.github.io/posts/2015-08-Understanding-LSTMs/
[18] Pham, T., Tran, T., Phung, D. & Venkatesh, S. DeepCare: A Deep Dynamic Memory Model for Predictive Medicine. In Pacific-Asia Conference on Knowledge Discovery and Data Mining, 30–41 (Springer, 2016)