

Machine Learning-based Asthma Exacerbation Prediction Model from Routinely Collected Primary Care Data

Arif Budiarto¹, Andrew Wilson², Aziz Sheikh¹, Syed Ahmar Shah¹
¹University of Edinburgh
²University of East Anglia

✉ arif.budiarto@ed.ac.uk



THE UNIVERSITY of EDINBURGH

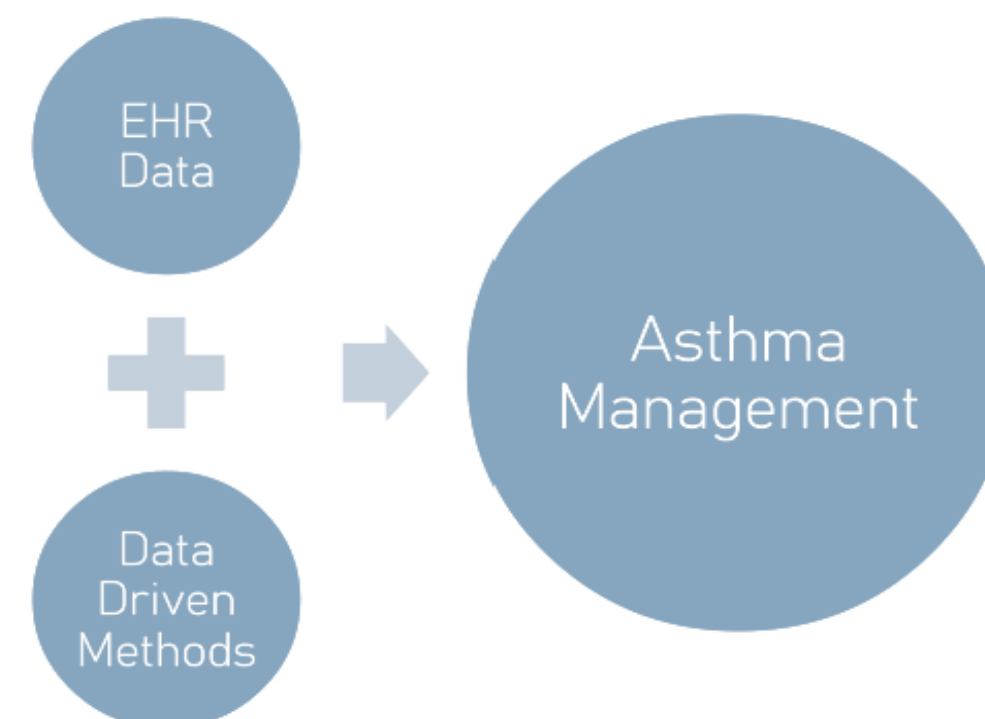
usher institute

BACKGROUND

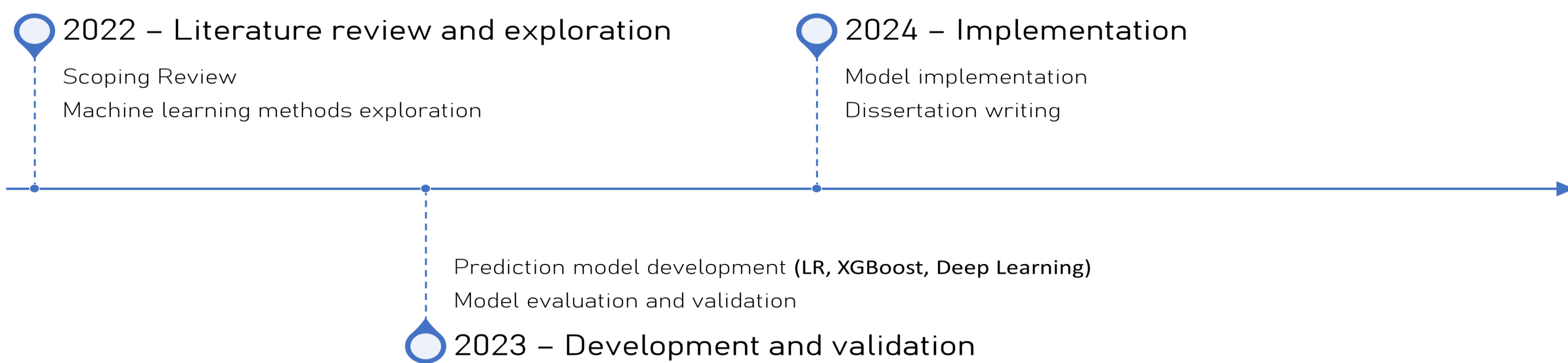
- **Asthma attacks** can potentially be life-threatening, but prompt action facilitated by timely prediction of deterioration may help
- Studies have been conducted in developing asthma attack prediction tools from **electronic health records (EHRs)**
- **Logistic regression (LR)** remains the most popular method because of its simplicity and interpretability
- Previous evidence in other domains suggests that more **advanced machine learning (ML) methods** can outperform LR
- Only limited previous studies focused on exploring this opportunity in asthma attack prognosis

AIM

“Investigate, evaluate, develop, and implement machine learning methods for developing an asthma attack predictive tool that could be used in primary care to help manage patients with asthma.”



RESEARCH PLAN



RESEARCH PROGRESS

SCOPING REVIEW METHOD

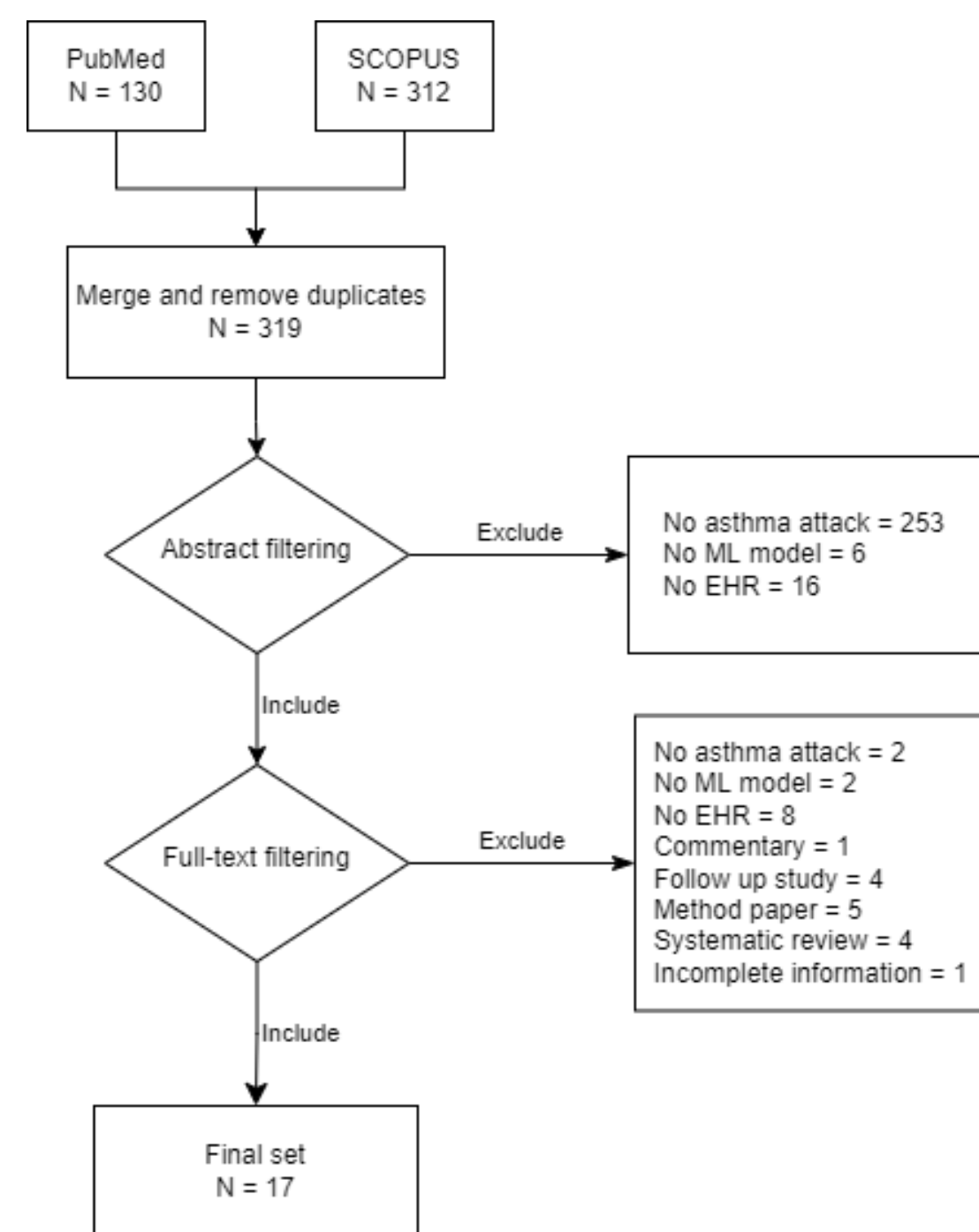


Fig 1. PRISMA Diagram



Fig 2. Word Cloud from Selected Abstracts

SCOPING REVIEW RESULT

- **59%** of the studies used routinely collected data from **primary care practices**.
- **More than 70%** of the studies used extreme imbalanced data, but only **six** of them employed **imbalanced data handling method**.
- **LR** was used as the baseline method in **12 out of 17** studies.
- **Gradient Boosting Methods (including XGBoost)** was the best method in **half of the studies**.
- **AUC** is the most used performance metric to evaluate the models.
- Only **three studies** included **Deep Learning** method in their analysis pipeline.
- **Almost three-quarters** of the studies employed a **model explanation method** to determine the important predictors based on the best model.
- **None** of the studies followed standard **reporting guidelines** and have been **deployed** in routine practices.

MODEL DEVELOPMENT PLAN

- Clinical Practice Research Datalink (**CPRD**) and Optimum Patient Care Research Database (**OPCRD**) will be the main datasets.
- Define **operational definition** of exacerbation in primary care **through stakeholder engagement sessions**.
- Define a **robust sample selection** method from database.
- Define the most **clinically useful time horizon** to assess someone's risk of exacerbation (as the model outcome).
- Explore and compare **the various of imbalance data handling methods**
- ML methods to be developed and compared: **LR (as the baseline), SVM, RF, XGBoost, LSTM with attention mechanism**.
- Explore the use of **temporal data representation techniques**.
- Follow The Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis (**TRIPOD**) and Developmental and Exploratory Clinical Investigations of Decision support systems driven by Artificial Intelligence (**DECIDE-AI**) as reporting guidelines.



THE UNIVERSITY of EDINBURGH

University of East Anglia



Asthma UK Centre for Applied Research



ASTHMA LUNG UK