Workstream timelines.xlsx

Recommendations for Conducting and Reporting Studies on AI-Based Clinical Predictive Models (AICPMs) Trained on Imaging Datasets in the NHS

Introduction

AI-based clinical predictive models (AICPMs), particularly those trained on imaging datasets (e.g., radiology, MRI, CT scans), offer significant potential for improving diagnostic or prognostic accuracy and patient outcomes (Dai, L., Zhou, M. & Liu, H. 2023)

In the NHS, where imaging datasets especially radiology images are abundant, developing rigorous standards for the conduct and reporting of studies to evidence these models trained on image datasets is crucial to ensure their validity, reliability, and utility in clinical practice. Organisations like NICE is place to support developers by providing a clear guidance of 'what good looks like'.

This study outlines best practices for the proper conduct and reporting of studies involving clinical predictive models trained on imaging data in the NHS.

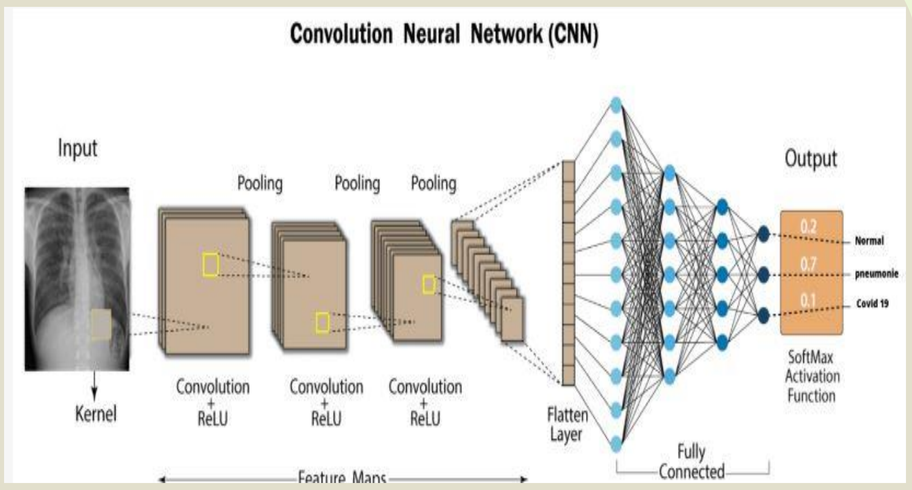
Study Design and Dataset Preparation

- The clinical prediction tasks the model is intended to address, imaging modalities used (e.g., MRI, CT scans), and the clinical outcomes of interest should be clearly defined, along with the specific healthcare setting where it will be implemented.
- Prior to developing AI-CPM for healthcare study, appropriate ethical approval must be obtained, including patient consent where necessary, particularly when identifiable patient data is used. Developers should ensure datasets are anonymized and comply with ethical standards and regulations (e.g., GDPR).
- The use of datasets that are representative of the target population, including diversity in age, gender, ethnicity, and disease prevalence is paramount in developing AI-CPMs that are generalizable and to explore aspects of model fairness.
- In terms of data sample size, it is a good practice to use the entire dataset to make the results more accurate and reproducible .

Park, S. H., et al. (2018)

Implementation of AI-Based CPMs in Healthcare

- The most used AI models for imaging data are convolutional neural networks (CNNs) for their superior ability in image classification and images important features identification as well as their ability to capture spatial hierarchies in images, but the choice of best AI-CPMs should be justified based on the specific clinical task.
- The complexity of most AI-CPMs for imaging dataset analysis, requires explainability tools (Grad-CAM, LIME or SHAP) to provide insights into how the model arrives at its predictions.



(Zakaria et al., 2022)

Model Development and Evaluation

- Selection of clinically relevant features and expert validation will ensure model's predictions are based on meaningful clinical data.
- The CPMs development involves splitting dataset into training, validation, and test sets using robust methods such as k-fold cross-validation or bootstrapping.
- The model evaluation is done by checking accuracy, sensitivity, specificity, and area under the curve (AUC). Heatmaps or saliency maps, which highlight areas of the image that contributed most to the decision need to be analysed.
- Internal validation needed for model performance while external validation involves assessing model generalizability across diverse patient populations and clinical settings.

(Dai, L., Zhou, M. & Liu, H. 2023)

Clinical Relevance, Regulatory Compliance and Integration in NHS

- Approval from relevant regulatory bodies, such as the MHRA is necessary, before deploying the model in clinical practice.
- The developed AICPMs should integrate seamlessly with existing clinical workflows and electronic health record (EHR) systems with a clear instruction on how to interpret and use the model's outputs in clinical practice.
- Prospective studies or clinical trials must be conducted to validate the model's effectiveness in real-world clinical settings. Pilot studies in real-world NHS settings can also be conducted to assess the model's performance and gather user feedback.
- Reporting standards such as TRIPOD-AI, including detailed descriptions of model performance, limitations, and potential biases, should be followed by any AICPMs developer. (BMJ 2024;385:e078378)

By
Shereefdeen Adebayo
HDRUK Intern @ NICE : https://www.linkedin.com/in/shereefdeenadebayo/