Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230098

Enabling Clinical Trials of Artificial Intelligence: Infrastructure for Heart Failure Predictions

Arian RANJBAR[a,1], Jesper RAVN[a], Erland RONNINGEN[a] and Ole HANSETH[b]
[a] *Medical Technology and E-health, Akershus University Hospital, Norway*
[b] *Department of Informatics, University of Oslo, Norway*

Abstract. The last decade has seen a large increase in artificial intelligence research within healthcare. However, relatively few attempts of clinical trials have been made for such configurations. One of the main challenges arise in the extensive infrastructure necessary, both for development, but particularly to run prospective studies. In this paper, infrastructural requirements are first presented, together with constraints due to underlying production systems. Then, an architectural solution is presented, with the aim of both enabling clinical trials and streamline model development. Specifically, the suggested design is intended for research of heart failure prediction from ECG, but is generalizable to projects using similar data protocols and installed base.

Keywords. Artificial Intelligence, Clinical trials, Infrastructure, Decision support

1. Introduction

The last decade has seen a large increase in research of artificial intelligence (AI)-enabled algorithms within many industries, including health care. Particularly for specialties relying on extensive data during diagnostics, such as radiology and pathology. Initial development, by these modelling techniques, often use performance evaluations limited to retrospective data in laboratory settings, [1]. To realize the potential, and incorporate the models into production, additional clinical testing is necessary; ensuring real world performance, effectiveness and safety. However, there are currently relatively few attempts of such studies and most rely on small populations or non-randomized tests, [2].

Recently the development of electrocardiogram (ECG) classification using machine learning reached desirable performance for various tasks on retrospective data, see e.g. [3]. Similarly, such ECG classification models need further clinical testing before deployment, [4]. Nevertheless, clinical trials require comprehensive preparatory work; both with regards to study design but especially in enabling infrastructure, connecting parts of the underlying health care system. Particularly, allowing prediction models to run in real-time and clinicians to interact with inference results. Then, how can we design an infrastructure that supports efficient development cycles of AI in health care, including clinical trials?

[1] Corresponding Author: Arian Ranjbar, Akershus Universitetssykehus HF, 1478 Lørenskog, Norway; E-mail: arian.ranjbar@ahus.no. This research was funded by Nasjonalforeningen for folkehelsen, and parts of the infrastructure was sponsored by Novartis Norway AS.

This paper aim to first present the infrastructural challenges concerning clinical trials of AI-enabled algorithms, including the precedent stages of model development, within an established healthcare production system. Then, present an infrastructure not only enabling clinical trials but also supporting faster development; by integrating data mining, training and prospective studies into a single platform. Specifically, the presented design is intended for development of heart failure prediction models using ECG, but is generalizable to systems relying on the same design principles.

2. Background

AI-enabled algorithms in healthcare have traditionally been built based on machine learning algorithms, such as neural networks, [5]. Most often, retrospective data is used for both training and evaluation, presuming the recorded data is representative for the total inference population. However, the model may not necessarily be able to extrapolate or generalize outside of the training domain. To prevent deployment of models with limitations in real-world performance, either due to discrepancies in training and inference population, overfitting or other model deficiencies; further prospective studies are carried out for validation, [1].

Prospective studies are generally divided into performance validations of the model or clinical trials, i.e. effectiveness evaluation of the system as e.g. decision support. In the former, after reaching sufficient accuracy on retrospective testing data, the model is queried for inference on prospective data. Such processes can be executed in the background without interfering with the regular production, since involvement of clinicians are unnecessary. Monitoring processes can also be implemented to detect performance drifts, either due to temporal changes or erroneous assumptions of the inference domain. After extensive testing, reaching sufficient performance on both retro- and prospective data, the latter type of study can be implemented. Clinical trials add an additional layer by including a feedback loop between clinicians and the model, to try out the intended interaction. This type of study may also provide metrics for usefulness, efficiency, quality, ease of learning, response time etc., in a real-world setting.

The key enabler to all developing stages is an efficient infrastructure. Model development requires access to relevant databases with the possibility of linking information to build datasets, in the process of data mining. Validation requires data streaming of prospective data and real-time inference capabilities. Finally, the clinical study needs writing access to production systems, without risking interruption of regular activity and patient safety. All requirements need to be delivered while also being compliant with data and privacy regulations, [6].

Designing an architecture that satisfies the requirements, is heavily dependent on the underlying IT infrastructure. The installed base is commonly classified as either a centralized system, via a platform, or distributed system, consisting of IT silos. The former usually hosts a common core with the distinct modules running as applications on top of it, transforming the infrastructural- to a software problem. In this scenario, clinical trials can be executed as additional applications with permission to access relevant data sources. However, most health care systems today run the latter, distributed system; with modules hosted within their own IT silos, often using different standards, [7]. Although not inherently wrong, IT silos generally provide more flexible solutions; the interconnection gets increasingly challenging in this scenario, relying on both hardware and software, [8].

3. Method

An architectural design of a platform for AI development, including clinical trials, is suggested based on the general requirements discussed in the previous section. The platform is specifically intended to be used for heart failure prediction models using ECG data on a distributed installed base. The explicit requirements for the system are:

- Linking retrospective ECG data from a picture archiving and communication system (PACS), [9], with medical records and external data (national medical registers), to do data mining in a safe contained environment.
- Connect external compute infrastructure for model training and inference.
- Enable real time data streaming of incoming ECG data.
- Enable write capabilities into the PACS.
- Prevent the system from causing interruption of regular production and comply with regulation.

4. Infrastructure

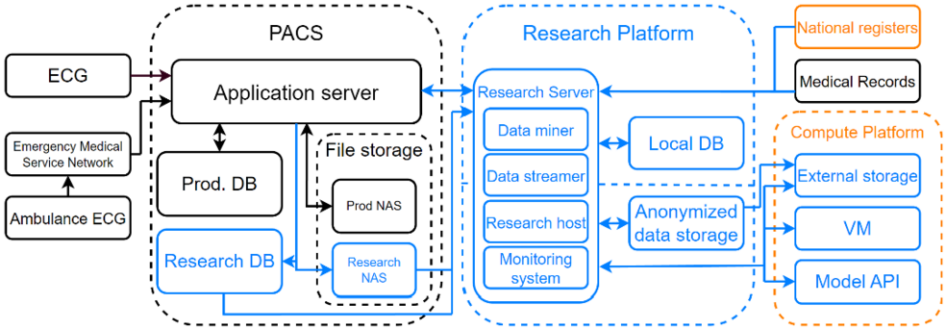

Figure 1. Schematic overview of the suggested infrastructure, arrows indicate data flow. Components of the installed base are represented by black, externally hosted resources are indicated by orange and suggested infrastructure by blue.

The main part of the suggested infrastructure consists of a research platform, serving as an interconnection of the required subsystem hosted within their representative IT silos, see Fig. 1. The PACS handles incoming ECG data, either from the clinics or from ambulances over an emergency medical service network. The network uses the digital imaging and communication in medicine (DICOM)-standard, [10], and data transfer is done via a main PACS application server. Each case is stored within a production database with the respective ECG waveform data in a dedicated file storage. To prevent interruption of the regular production pipeline, a clone of the database and ECG storage is installed. The clones are not only connected to the research platform, but also recipient of external information relayed through the application platform using DICOM-SR. The latter closes the feedback-loop, again, without potentially interfering with production.

The research platform mainly acts as a gateway, running data services such as data mining, anonymization and compute communication. These services can be hosted at two distinct layers with different permissions, where the first layer handles sensitive data and the second layer only has access to anonymized data and non-sensitive information. Therefore, ECG data from the PACS enters the platform at the first layer, together with medical records and externally provided sensitive data. This allows for data mining,

generating training datasets of the recorded retrospective data, necessary for model development. As datasets are generated, indexing is stored in a local database at the first layer before anonymization and exportation to the second layer. At the second layer the platform has access to external compute infrastructure allowing for model training and inference. Finally, as previously mentioned, the research server may communicate directly with the PACS application server, allowing feedback in terms of inference results. This channel also allows the PACS to directly send inference requests to the research platform without going through the PACS database.

5. Implementation

The implementation was done at Akershus University Hospital using ComPACS as PACS, LifeNET as provider for the emergency medical service network and DIPS for medical records. The research platform was hosted using two servers with mirroring for redundancy and real-time patching capabilities. Each service could be hosted in a contained virtual machine with the required permissions. Finally, cloud computing was provided by Google Cloud Platform, connected through the corresponding developer kit, only having access to data storage containing pseudonymized data.

6. Discussion

The presented infrastructure not only enables clinical trials, but streamlines the process from model development to production deployment. Data mining is done directly on the research platform, which is particularly useful if data inclusion needs to be iterated. Prospective studies can run as background processes of the production systems, with real-time monitoring from the research platform, without relying on clinicians relaying relevant cases. Finally, clinical studies are executed in the same environment as development, which is consistent with regular production. This also simplifies the process for clinicians accustomed to the production systems.

6.1. Data regulation

Although the platform has access to the production systems, the data available for processing is strictly limited by data regulation. This is an additional benefit of running clones of the production databases, serving as an intermediate data layer. However, in order to use medical records, both for retrospective and prospective data, relevant permits and ethical assessments needs approval. In particular, due to the consent paradigm, consent either needs to be collected from patients already in the registers, or new data needs to be collected from consenting patients. Such regulations are often adapted to traditional medical research, conforming to data minimization principles, unlike machine learning relying on extensive data, [11].

On this note, Norway recently legislated exemption from the consent paradigm when collection of consent or new data is unfeasible; and the data usage is considered to be of significant benefit and unharmful, in ethical reviews (Helseperonelloven Kap.5 §29). In projects approved for exemption, data mining is significantly easier since the targeted data is directly cloned into the separate storage.

6.2. Limitations

This study investigates infrastructure for clinical trials and efficient AI development cycles, particularly aimed to be used for heart failure prediction modelling. It is also constrained to an installed base using distributed IT silo systems. The suggested platform is thereby tailored for this specific use case. However, these constraints and installed base are commonly used within healthcare, [7]. The solution is also generalizable to projects using similar data setups e.g. data from an archiving system and medical records.

Performing clinical trials of AI-enabled models require a human machine interface (HMI) to provide feedback to clinicians. Detailed examination of HMIs is left out of this study, where PACS is used as communication channel. Although having the benefit of being a regular production system, the effectiveness as mediator is not investigated.

7. Conclusion

This paper investigates efficient designs of infrastructure for AI development in healthcare, enabling clinical trials. Such solutions are heavily dependent on the installed base, but in general needs data mining capabilities from relevant sources, access to compute power, real-time data streaming of prospective data and writing permissions in production systems; all while not risking interference with regular production and complying with data regulation. The suggested solution is intended for development of heart failure prediction models using ECG data, but is generalizable to projects using similar setups. By deploying the platform, the full development cycle is streamlined, from model training to clinical trials. However, further investigations are needed for effective HMI design.

References

[1] Magrabi F, Ammenwerth E, McNair JB, De Keizer NF, Hyppönen H, Nykänen P, et al. Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. Yearbook of medical informatics. 2019;28(01):128-34.
[2] Nagendran M, Chen Y, Lovejoy CA, Gordon AC, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. bmj. 2020;368.
[3] Ribeiro AH, Ribeiro MH, Paixao GM, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature communications. 2020;11(1):1-9.
[4] Attia ZI, Harmon DM, Behr ER, Friedman PA. Application of artificial intelligence to the electrocardiogram. European heart journal. 2021;42(46):4717-30.
[5] Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nature biomedical engineering. 2018;2(10):719-31.
[6] Rumbold JMM, Pierscionek B. The effect of the general data protection regulation on medical research. Journal of medical Internet research. 2017;19(2):e7108.
[7] Bygstad B, Hanseth O. Transforming digital infrastructures through platformization. Association for Information Systems; 2019.
[8] Aanestad M, Grisot M, Hanseth O, Vassilakopoulou P. Information infrastructures and the challenge of the installed base. Information infrastructures within European health care. 2017:25-33.
[9] Huang H, Taira RK. Infrastructure design of a picture archiving and communication system. American Journal of Roentgenology. 1992;158(4):743-9.
[10] Mildenberger P, Eichelberg M, Martin E. Introduction to the DICOM standard. European radiology. 2002;12(4):920-7.
[11] Ranjbar A, Skolt K, Aakenes Vik KT, et al. Fairness in Artificial Intelligence: Regulatory Sandbox Evaluation of Bias Prevention for ECG Classification. Proceedings of Medical Informatics Europe; 2023.