A New Framework for Interoperability and Reproducibility of Predictive Models

The advancement of data-driven approaches in biomedical applications is fuelled by increased computational power and access to diverse data sources like electronic health records (EHRs), omics, imaging, and mobile health data. Artificial intelligence and machine learning (AI/ML) methods hold promise in healthcare, offering novel predictive capabilities for screening, diagnosis, and treatment improvement. Government agencies in regions like the European Union, Canada, and the United States are issuing guidance on the deployment of predictive models. Despite this, challenges persist around reproducibility, stemming from the lack of standard approaches in model development, incomplete reporting in publications, and obstacles posed by privacy-protected datasets and closed-source codebases. A recent study published in the Journal of Biomedical Informatics aimed to close this gap and offer a structured approach towards interoperability and reproducibility of predictive models.

Addressing the challenge of reproducibility in AI/ML models for healthcare

The effective use of AI/ML models in clinical settings necessitates a thorough understanding of their contextual usage and capabilities. The absence of detailed and transparent reporting makes it challenging for the scientific and healthcare communities to objectively assess the strengths and weaknesses of predictive model studies. Additionally, the inability to reproduce AI/ML models hampers comparisons across competing models, highlighting the need for standardized methods in the development, analysis, and evaluation of predictive models to enhance reproducibility in healthcare. Al Rahrooh's team addressed the lack of standard methodologies in developing and deploying machine learning (ML) models for biomedical research and healthcare. Existing tools for model replication lack a unifying blueprint, making it challenging to reproduce predictive ML models. The study introduces the Automated Metadata Pipeline (AMP), which extends the Predictive Model Markup Language (PMML) to automate the completion of an ML-based checklist, ensuring model elements are assessed for interoperability and reproducibility. The AMP is demonstrated on multiple test cases with various ML algorithms and health-related datasets, offering a framework to enhance compliance and reproducibility in predictive model research by ensuring all necessary information is available for model sharing and comparison.

Enhancing biomedical model dissemination and reproducibility with PREMIERE

The paper underscores the potential acceleration of biomedical discoveries by leveraging data and machine learning (ML) models. Despite the substantial number of predictive models published annually, their translation into practical applications faces obstacles due to issues surrounding scientific reproducibility. To address this, the authors introduce PREMIERE, a platform designed to support the dissemination of PMML-extended models while adhering to FAIR principles, thus ensuring their findability, accessibility, interoperability, and reusability. While acknowledging existing ML frameworks like MLFlow and Neptune, which primarily focus on model lifecycle management, the paper identifies a gap in the comprehensive documentation of data and processing steps crucial for biomedical predictive models. PREMIERE stands out by emphasizing not only model hosting but also reproducibility and documentation, thereby catering to the specific needs of the biomedical field.

Automating reproducibility assessment with the AMP framework

The Automated Metadata Pipeline (AMP) is a key component of PREMIERE, designed to automate the completion and evaluation of comprehensive ML model checklists. AMP captures crucial information regarding data preprocessing, model training, and evaluation, streamlining the process of assessing model reproducibility and facilitating external validation. The paper delves into the distinction between replicability and reproducibility, stressing the significance of detailed documentation and external validation in evaluating predictive model utility. Looking ahead, the authors outline plans to extend AMP's support for deep learning models and enhance compatibility with other model representations like ONNX. These efforts aim to bolster AMP's versatility and adaptability within the dynamic landscape of machine learning techniques. By providing detailed documentation and facilitating external validation, PREMIERE and AMP aim to address the challenges surrounding the reproducibility of biomedical predictive models, thereby advancing their utility in healthcare and research domains.

The Automated Metadata Pipeline (AMP) presented in this study serves as an early-stage proof of concept for the PREMIERE platform, aimed at assisting authors and model developers in reproducing and utilizing published predictive ML models. AMP functions by converting a predictive

ML model into an extended PMML file, which is then automatically evaluated against a checklist to ensure reproducibility. The study demonstrates that PMML holds promise as a universal standard for ML interoperability and can be expanded to include the necessary information for supporting reproducibility. AMP integrates workflows and pipelines into a single framework, allowing users to assess the completion of reproducibility criteria. Through this tool, more comprehensive checklists can be adopted to fully document a predictive model's design, testing, and usage. Ultimately, as part of the PREMIERE platform, AMP aims to foster a community and standard approaches for infrastructure supporting the reproducibility, comparison, and enhancement of biomedical AI/ML models, ensuring they are FAIR and scientifically reproducible.

Source: [Journal of Biomedical Informatics](#)

Image Credit: [iStock](#)

Published on : Mon, 18 Mar 2024