Learning to Collaborate via Explainable AI in Medical Education

A Data Management Plan created using DMPonline

Creator: Manxi Lin

Affiliation: Danmarks Tekniske Universitet / Technical University of Denmark

Template: National research center DIREC

ORCID iD: 0000-0003-3399-8682

Project abstract:

Every minute, over 150 newborns bring hope for the future to their families. Infants' health is always a vital concern of parents and doctors. Evidence shows that abnormal fetal growth in both industrialized and developing countries is one of the leading causes of perinatal morbidity and mortality. Ultrasound scanning is widely employed in prenatal examination. It provides a safe and convenient way to detect early fetal abnormalities, due to its low cost, real-time capability, and absence of harmful radiation. During the examination, clinicians manipulate the ultrasound probe to achieve certain standardized scan planes. However, obtaining high-quality ultrasound images is very challenging and depends on operator experience. The risk of diagnostic failure increases when ultrasound examinations are of insufficient quality. This is emphasized by the high variance due to different operators, devices, and acoustic windows. As a result, even experienced experts take a long time to find and assess standard scan planes. For those inexperienced young clinicians, working with minimal expert supervision and support increases the risk of medical errors. In the Western World, medical errors are only exceeded by cancer and heart diseases in the number of fatalities caused. About one in ten diagnoses is estimated to be wrong, resulting in inadequate and even harmful care. Recently, many works have explored the possibility of processing ultrasound data with deep neural networks (DNNs). Trained DNNs are faster and far more reproducible than humans, giving potential for standardized quality of care and more efficient use of clinicians' time. However, in these existing works, DNNs are rarely designed as a collaborator for the healthcare professionals, but rather as a mechanical substitute for part of a diagnostic workflow - researchers only develop models to beat state-of-the-art on narrow performance parameters. As a result, clinicians do not always perceive these solutions as helpful in solving their clinical tasks, as they only solve part of the problem sufficiently well and face the risk of overfitting. To address these problems, we need not only to provide interpretability in the form of explainable models -- but we also need to provide models whose explanations are easy to understand and utilize during the clinician's workflow. Put simply, we need to provide good explanations. This project aims to address these problems: - We will develop explainable models based on deep learning, which recognize standard planes. - We will develop a system that provides optimized, continual feedback on improving image quality. This system will be trained to optimize the quality of explanations as well. - We will provide an additional feedback parameter to the clinician, that detects whether the image is out of distribution, or whether demographics put the patient at risk of suboptimal algorithmic performance.

Last modified: 03-05-2022

creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Learning to Collaborate via Explainable AI in Medical Education

Data Collection

This project already has access to almost all ultrasound scannings of pregnant women in Denmark in the period from 2009 to 2018, giving more than 30 million ultrasound scans as well as follow-up on the pregnancy, birth weight, and time of birth. Structural and chromosomal anomalies errors found during pregnancy are recorded and we have documented pregnancy outcomes (abortion, live birth, birth outcomes, neonatal outcomes). The data are collected by the clinicians in the hospital, by picking frames from the recorded videos from ultrasound scanners.

The main form of the scans is DICOM images.

These are in the process of being transferred to and organized at DTU Compute. Since manual annotation is expensive, there are only around 6.000 annotations in second-trimester scanning and 2.000 annotations in third-trimester scanning. We are getting more scans and annotations from the clinicians. The annotations are formatted in JSON files. Other information about the pregnancy is stored in the MySQL database. The data are versioned by the batch index, each batch includes around 500 images.

In this project, the use of existing data is under the data access agreement. The collection of new data will be compliant with GDPR and DTU ethics. As the data involves the personal information of pregnant women, we consider the protection of privacy when processing the data. This is under the data access agreement (databehandleraftale, G2019-34-ANYM-gravid).

Data Storage

The data is totally stored on the DTU server. Only registered users with an authentication code can access the server. The backup is done by server administrators. Users on the server can not access the Internet, except DTU Gitlab for version control. We manage the data under the data access agreement (databehandleraftale, G2019-34-ANYM-gravid) to ensure the security of sensitive data.

Documentation

All the image paths, annotation paths, date of data collection, and device number are stored in a MySQL database. Excepting the paths, all the other metadata is recorded by the ultrasound scanner and our collaborators in clinicians. The metadata is generated and stored in a way under the data access agreement(databehandleraftale, G2019-34-ANYM-gravid).

The MySQL file has a headline, describing the data. There is a pdf file documenting the data. Both of these are understandable for secondary users. The reproducibility is evaluated by all the users on the server.

Data Sharing

We have no plan to share the data with anyone not affiliated with our project at this stage. If we are going to share the data at some stage in the future. the data will be shared in line with the data access agreement (databehandleraftale, G2019-34-ANYM-gravid).

We have no plan to share the data at this stage. If we are going to share the data at some stage in the future. the data will be shared in line with the data access agreement (databehandleraftale, G2019-34-ANYM-gravid).

Long-term Preservation

The criteria used to select data to be archived will be in line with the discussion with our collaborators from the clinician. They will be archived in a safe way according to the data access agreement (databehandleraftale, G2019-34-ANYM-gravid).