Standardized use of PNGs/JPEGs for AI-Based Detection of Thyroid Eye Disease via Federated Learning

Michael Lau*, Vishwanath Prathikanti*, Angela McCarthy, Ye Tian, Christopher Nielsen, Sina Gholami, Andrea Kossler, Eric Brown, Minhaj Alam, Lora Dagi Glass, Kaveri A. Thakoor
Columbia University

**denotes equal contribution*

Background

Thyroid eye disease (TED), is an autoimmune disorder caused by the same antibody as autoimmune thyroid disorder, causing inflammation, swelling, excess and scarring[1]. Diagnosis of TED is classically based on clinical appearance, and radiologic imaging can be used for confirmation and/or preparing for surgery. Early diagnosis via Artificial Intelligence (AI) can help clinicians quickly identify TED patients and provide timely intervention by applying deep learning to facial images of patients. However, it is often hard to obtain a large and diverse dataset of facial images for training deep learning models. At the same time, facial images are highly sensitive and cannot be freely shared between institutions. In this work, we address this challenge through collaborative training between institutions with federated training (FL). FL enables collaboration across different institutions while preserving data privacy.

Data standardization is particularly relevant for FL since it streamlines processing of data across multiple institutions, reducing the potential for inhomogeneity and manual errors arising from variations in data collection methods, formats, and quality that can threaten the development of robust AI models. The Observational Health Data Sciences and Informatics (ODSHI) initiative is working to address this issue by promoting the use of the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) to standardize data across institutions. Park and colleagues[2] developed robust image feature tables for the use of DICOM metadata, but it remains untested. Part of our goal was to test the tables for PNG metadata as well and understand how it would be used practically. With the development of new medical technology such as portable OCT devices, it is necessary to have robust data standardization for image file types outside of DICOM, such as PNGs and JPEGs.

Our work represents the first effort to integrate ophthalmic imaging into the OMOP CDM and address the sensitivity of eye images, tackling problems uniquely difficult for ophthalmology. To test the tables and their use for PNGs and JPEGs, we leveraged the OMOP CDM and deep learning to predict Thyroid eye disease (TED) in a federated learning setting. This paper will discuss the results of both the image standardization table and federated learning AI results.

Methods

Our FL model takes facial images as input, and predicts TED vs. control. We combine FL with masked autoencoder (MAE)[2] pretraining to enhance model representations and robustness of our model. There are two client institutions, Columbia and Stanford, each with 291 and 200 images, respectively, with different data distributions. Federated Averaging is used as the main FL algorithm; both clients train for 10 epochs in each round, and we perform 10 rounds of FL training in total. ResNet-18 and ViT-B were used as image encoders for MAE and SimCLR pre-training, respectively.

We extended the OMOP image_feature and image_occurrence tables to store relevant metadata extracted from PNG and JPEG files such as the image occurrence date, path, and features. In an effort to remain accessible to investigators at multiple sites with limited SQL knowledge, our data model was simplified from the full OMOP CDM. However, the relationships between the image extension tables was designed to mirror tables in the full CDM closely. Each image was linked to clinical features in the OMOP CDM via person_id and visit_occurrence_id, enabling multimodal analyses. To simulate a federated environment, we deployed our data model at two institutions and used a federated learning framework (FLARE via Rhino Cloud Platform).

Data quality checks were implemented to assess metadata completeness, and a set of example queries were used to validate the practical utility of the modified schema.

Results

We divide our experiment into two stages: local and FL. From our local experiments we found that the same methods applied to both sites yielded different results, with highest AUC across methods being 88.35% for the Columbia dataset and 97.17% for the Stanford dataset with only one encoder. The difference in model performance shows that datasets have different difficulties, with the Columbia dataset having more diverse control conditions, further motivating the need of FL. Due to this reason, we focused on personalization in FL, fine tuning FL models to better fit the needs of each institution while benefiting from collaborative training. MAE combined with FL yielded the best results with 89.26% AUC and 98.70% AUC for Columbia and Stanford, respectively (Figure 1).

To assess the practical utility of the schema, we simulated the process of building cohorts using the data collected. Figure 2 shows the integration of the image extension tables into a simple data model. Cohorts built on various demographic data such as race, ethnicity, age and gender allowed us to understand generalizability based on parameters from different sites. Furthermore, by connecting diagnosis data from non-TED patients, we are able to understand how different diagnoses may affect parameters. The image extension tables afforded convenient storage of metadata and accessibility for retrieval.

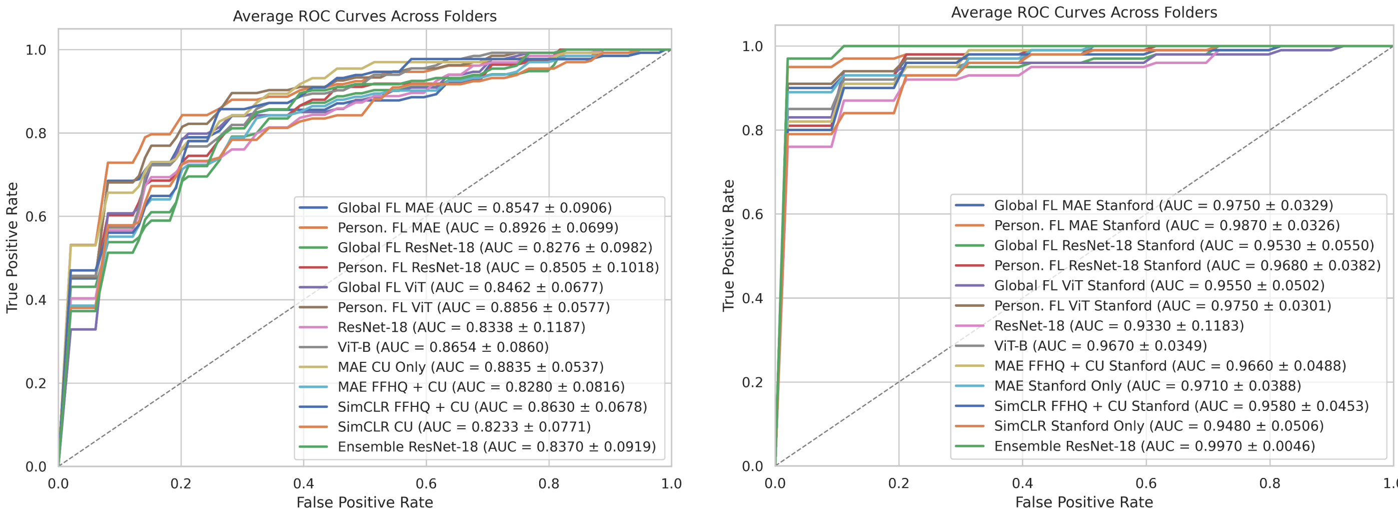

Figure 1: AUC Curves of Local and FL models on both clients sites**. Left:** Results on Columbia Dataset. Personalized FL MAE has the highest AUC, and ensemble models do not improve local performance. **Right**: Results on Stanford Dataset. Personalized FL MAE also achieves highest performance for a single model.

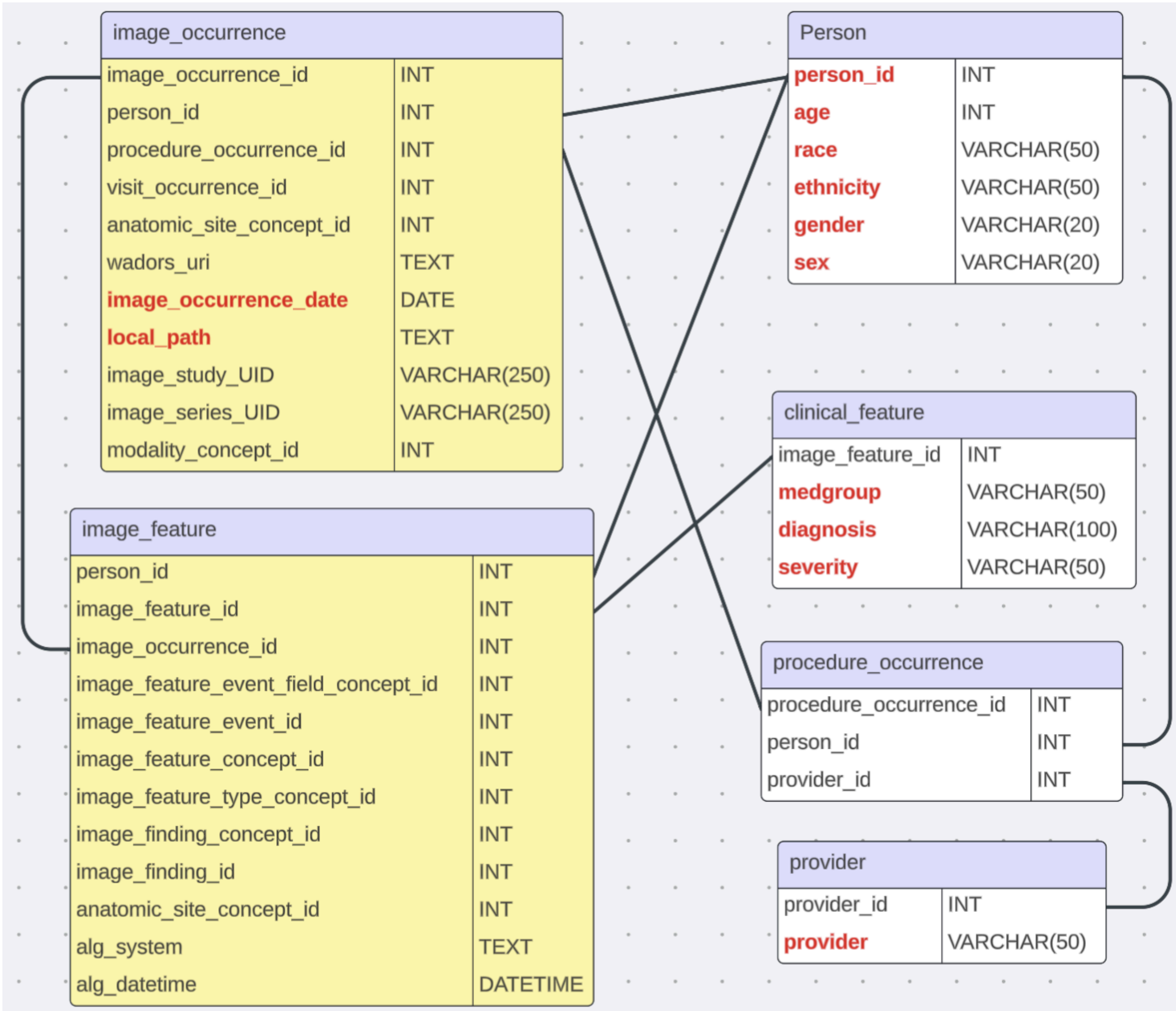

Figure 2: Integration of image extension tables in the data model. Constructed image extension tables (highlighted yellow) and simplified tables representing collected data on TED/Control images. Red text represents manually collected/imputed data. Black text represents metadata or generated values.

Conclusion

We developed and evaluated a framework for FL in TED, integrated self-supervised pretraining to improve model representation and robustness. The future vision for further improving not only TED detection but also models for detection of other diseases would require multimodal data, where data standardization would play a key role in enabling cross-institution collaboration. To that end, standardized image data table integration for multiple file types is essential.

References

1. N. F. Xavier, D. T. Lucena, and A. A. V. Cruz, “Monoclonal antibodies for the treatment of graves orbitopathy: Precision medicine?” Ophthal. Plast. Reconstr. Surg., vol. 39, no. 4, pp. 307–315, Jan. 2023.
2. Park WY, Jeon K, Schmidt TS, Kondylakis H, Alkasab T, Dewey BE, You SC, Nagy P. “Development of Medical Imaging Data Standardization for Imaging-Based Observational Research: OMOP Common Data Model Extension.” J Imaging Inform Med. 2024 Apr;37(2):899-908.
3. K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 16000–16009.