Duality

Case Study

Secure Federated Learning for Oncology Research

Case Study at the Dana Farber Cancer Institute in collaboration with Duality Technologies

Secure Federated Learning in Digital Pathology

The Oncology Research Lab at Dana Farber Cancer Institute (DFCI) is at the forefront of cancer classification research. Facing the challenge of analyzing large digital pathology data sets distributed across multiple institutions, Dana Farber needs an agile, flexible, and regulation-compliant computational framework for model training on large distributed datasets, without the risk of exposing sensitive data.

Challenges in Collaborative Research

To build high quality models, Dana Farber need access to a large set of pathology images, including outside of their organization. Accessing external large-scale oncology data is a complex and slow process, with significant privacy, security, and IP challenges. Pathology images are considered Protected Health Information (PHI), creating a barrier to sharing these between organizations, necessitating a secure method of collaborative analysis.

Accessing and utilizing large-scale oncology data from multiple medical centers is a complex and slow process with significant privacy, security, and IP control challenges.

The Solution - Secure Data Collaboration Platforms

Duality Technologies provides a secure AI collaboration platform with federated learning and a trusted execution environment that adheres to stringent security and privacy standards.

This framework enables collaborations for medical centers where data stays securely in their own environment. Using the federated training process, model training occurs locally, and only the weights are transmitted, while encrypted. These weights are aggregated on a collaboration management server.

To ensure maximum security, Duality employs a trusted execution environment for collaboration management and secure aggregation. This guarantees the weights are encrypted during transmission and remain inaccessible even while data is in use.

HEADQUARTERS

5 Marine View Plaza

Hoboken, NJ 07030

CONTACT US

support@dualitytech.com

Duality Technologies

dualitytech.com

Key Features

Privacy
Sensitive data never moves out of its secure location and is not shared between institutions.

Cost-Effective
Utilize cloud computing resources for collaboration management and confidential computing, minimizing additional costs to research budgets.

Security
Advanced data-in-use protection methods, including encryption and TEEs, prevent threats to the models and data

Flexible and Scalable
Supports various models and scales according to the collaborative needs - participants, models, and new images.

Model
Train any model using a federated architecture.

Data Pre-Processing
Pre-process locally and prepare the image data before sending it encrypted for model training.

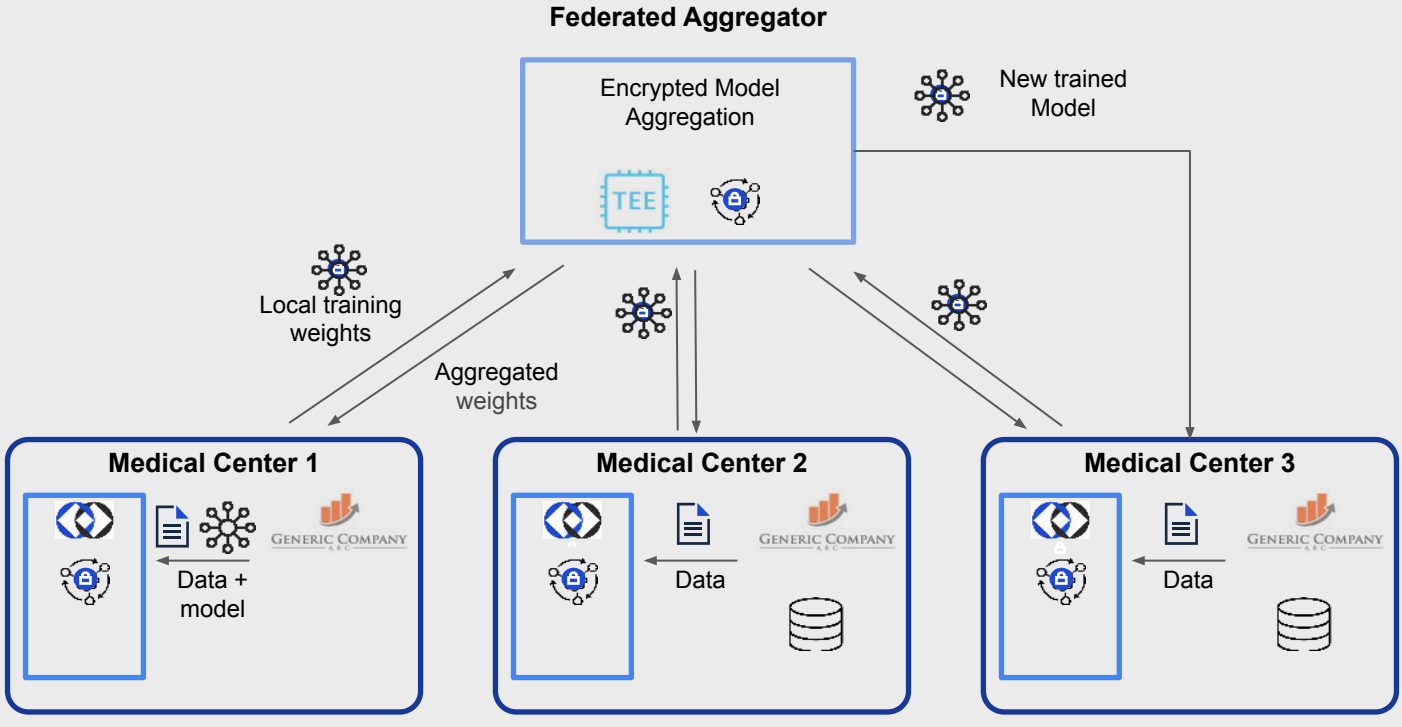
User-Friendly
The platform is designed for ease of use without requiring extensive technical knowledge from participants.

Compliant
An IRB template is provided to expedite approvals for participating organizations.

Reproducible
The ability to re-train the model and reproduce results.

Deployed Anywhere
The platform supports deployment on any cloud and on-prem.

Federated Learning with Trusted Execution Environment



Technical Implementation and Results

The federated learning model was implemented with two participating collaborators, each contributing 100 whole slide images which are high-resolution digital scans of entire pathology slides. The images, each approximately 1GB in size, are segmented into smaller images, each being passed through a pre-trained CNN for feature extraction.

The resulting vectors are then used to train a state-of-the-art cancer classification model named CLAM. CLAM, using an architecture of self-attention and SVM, is trained to provide a slide-level diagnosis of the entire initial image.

The model was tested under three different scenarios:

- **Centralized Training**: The whole dataset consolidated and processed at a single location.
- **Federated Training**: Data processed at all locations locally, with training results aggregated centrally.
- **Single Site Training**: Train on a single site data.

	Test Accuracy	Test AUC
Single Site	0.91	0.936
Federated	**0.95**	**0.960**
Centralized	0.99	0.994

These results, which reflect an average of 5 running iterations, demonstrate that federated learning accuracy and AUC results for the federated CLAM achieve high accuracy compared to running the CLAM model on a centralized dataset, while providing the advantages of distributed collaboration, including data security and integrity.

Once the model has been built, the participants can deploy and incorporate it as part of their MLOP workflow for model serving.

Duality

Case Study

Secure Federated Learning for Oncology Research

Case Study at the Dana Farber Cancer Institute in collaboration with Duality Technologies

Conclusion

The secure federated learning study by the Dana Farber Cancer Institute in collaboration with Duality Technologies showcases significant operational benefits for medical research, including agility, flexibility, and privacy-protection for sensitive data for all participants in federated collaborations.

This approach demonstrates full control over sensitive medical data by the data custodians, eliminating the need for data transfer, as well as many additional benefits, proving to be a pivotal solution for multi-institutional collaborative oncology research.

Get Started
Email: support@dualitytech.com

Recognized across Public and Private Sectors

HEADQUARTERS
5 Marine View Plaza
Hoboken, NJ 07030

CONTACT US
support@dualitytech.com
Duality Technologies

dualitytech.com