Federated Learning Model Application to Gastric Cancer Diagnosis

Project Proposal - Group 7.2

Warisara Siriponpaiboon, David Bain, James Murray, Sinh Thanh Nguyen, Taekjin Jeong, Emmanuel Niko Sindayen

Gastric cancer (GC) remains a formidable global health challenge, ranking as the fifth most common cancer worldwide in 2022. Research indicates that early detection significantly improves patient outcomes, with survival rates exceeding 90% in early-stage GC cases. This underscores the urgent need for diagnostic systems that enable faster and more accurate detection, thus facilitating timely medical interventions.

Recent advances in data science and machine learning present promising opportunities to enhance diagnostic capabilities in the medical field. Deep learning models have demonstrated significant potential in improving diagnostic accuracy by employing artificial intelligence (AI) algorithms to analyse medical images for disease detection. However, despite these advancements, the clinical implementation of AI-based diagnostic tools requires further validation through large-scale datasets. This process, however, is often impeded by concerns regarding patient privacy and data security which blocks the sharing of live training data by medical institutions.

To address these challenges, the team proposes the development of a federated learning model to support faster and more accurate GC diagnosis using histopathological images. This approach allows each participating institution to train the model locally on its respective data while sharing only the model weights for centralized aggregation. By eliminating the need for direct data sharing, this framework ensures the preservation of patient privacy while enabling collaborative model training across different medical clinics.

Our study will evaluate the framework by comparing the diagnostic performance of the baseline local algorithms with their outcomes after undergoing the federated learning process. Through this project, the team aims to contribute to the ongoing discourse on the implementation of federated learning models as a viable solution to the data privacy and availability challenges hindering the adoption of AI algorithms in the medical field, specifically in diagnostic imaging.

Group 7.2 Members:

- 25099704 Sinh Thanh Nguyen
- 91082596 David Bain
- 25099654 Taekjin Jeong
- 25014616 Warisara Siriponpaiboon
- 24888796 Emmanuel Niko Sindayen
- 13879046 James Murray

Contents

List of Figures

List of Tables

1 Introduction

Gastric cancer (GC) ranks fifth for incidence of all cancers and fourth for mortality with over 1 million new cases and 768,000 deaths worldwide in 2020. These numbers are predicted to increase by 2040 with 1.77 million new cases and 1.27 million deaths predicted1 worldwide, making it a significant global burden on health systems. Most GCs generally display no apparent symptoms in their early stages, aside from those shared with more common gastrointestinal issues such as ulcers and gastritis (Sheller et al. 2020) Therefore consultation and subsequent diagnosis are often delayed, heavily affecting the patients' chances as a late diagnosis often carries poor observed survival rates to as low as 30%

The current gold standard of GC detection is through histopathology screening of a biopsy or surgical specimen using a microscope to identify the cancerous features which is conventionally done by pathologists through a thorough manual screening of the tissue biopsies. However, visual analysis of tissue biopsies by pathologists is extremely laborious, time-consuming, and subjective. Conclusions drawn by one pathologist can be different from another. The correct analysis of histopathology is highly dependent upon the expertise and experience of the clinician, which makes manual histopathological analysis prone to possible misdetection and misdiagnosis due to human errors. Coupled with the current shortage of pathologists, long backlogs in the processing of patient cases are becoming common which consequently increases the likelihood of delayed cancer detection (Yong et al. 2023).

Deep learning offers promising solutions to these unique challenges through the development of automation and standardisation in histopathological analysis. By integrating this technology into diagnostic workflows, specific stages of the current analysis process can be enhanced, potentially reducing the overall rate of misdiagnosis and expediting turnaround times, ultimately alleviating the workload of pathologists while ensuring more consistent and reliable outcomes (Yong et al. 2023).

Currently, several studies [yan@dengArtificialIntelligenceApplications2022] have already attempted to develop deep learning models that can assist pathologists in diagnosing gastric cancer, aiming to improve accuracy and reduce human error. However, these models face challenges due to limited training datasets that inhibit the validation and generalisation of these models on unseen data.

One possible solution being presented is Federated Learning, a collaborative learning model that enables algorithms to learn from de-centralised data distributed across various institutions. With this process, classified data is never transferred beyond the safety of institutional firewalls and only the model specifics (e.g., parameters or gradients) are transferred for aggregation (Lu et al. 2020).

This project seeks to investigate the potential of federated learning in deep-learning classifier models for gastric histopathology. The team aims to develop a model that leverages the federated learning framework, enabling models to be trained collaboratively on distributed gastric histopathology datasets without compromising data privacy. The success of this project will demonstrate the efficacy of federated learning in enhancing the training and generalisation of gastric cancer diagnostic models on unseen data, utilising contributions from external institutions. Furthermore, this project will provide evidence that federated learning can be effectively applied to other domains that will benefit to the training of artificial intelligence models on sensitive, decentralised datasets.

2 Related Literature

The application of deep learning techniques in medical imaging has garnered significant attention due to their ability to detect and classify abnormalities with high accuracy. Numerous studies have demonstrated the efficacy of various convolutional neural network (CNN) techniques in diagnosing diseases, including gastric cancer (GC), often surpassing the performance of human diagnosticians. For example, Hirasawa et al. (2018) utilised a Single Shot MultiBox Detector CNN diagnostic system, which successfully detected 98.6% of lesions with a diameter of at least 6 mm. The missed cases were identified as specific edge cases, emphasising the system's overall robustness. Similarly, Ikenoyama et al. (2021) employed an Inception-v3 CNN, which detected early gastric cancer (EGC) cases more rapidly than endoscopists, and Tang et al. (2020) applied a Darknet-53 CNN, which not only outperformed endoscopists in diagnostic accuracy but also significantly enhanced clinicians' diagnostic capabilities.

In the context of this proposal, several studies have explored the application of deep learning models trained on histopathological images to support the diagnostic process. Yoshida et al. (2018) investigated the use of e-Pathologist, the first automated image analysis software, which demonstrated promising potential for aiding gastric cancer diagnosis after further enhancements to its screening capabilities. Building on this, Song et al. (2020) developed an AI system that assisted pathologists by analysing gastric histopathology images, flagging slides for re-examination or additional testing. The study's results indicated that the model not only functioned as a pre-analytical tool that can help prioritize cases but also as a second opinion system, particularly in challenging cases.

A systematic review by Klang et al. (2023) on the application of deep learning to GC detection further reinforced the success of AI systems in improving diagnostic accuracy and segmentation. However, the review identified key challenges in ensuring robust model performance across diverse populations and imaging conditions, which require larger and more varied datasets. Issues around the validation and standardization of AI models also emerged as critical barriers to clinical adoption. The authors noted, however, that the limitations of the reviewed studies may have compounded these challenges, especially on the fact that majority of the papers reviewed were conducted in single-centre settings.

In response to these data availability and privacy concerns, Deng et al. (2022) proposed federated learning (FL) as a solution. FL allows multiple institutions to collaborate on model training without sharing sensitive patient data, addressing concerns related to data privacy and security while promoting model validation and standardization. Sheller et al. (2020) compared FL to other collaborative learning approaches, such as Institutional Incremental Learning (IIL) and Cyclic Institutional Incremental Learning (CIIL), finding that FL yielded the best results in terms of rate of model improvement over the set learning method epoch and the development of better-performing models on average.

An in-depth application of FL was explored by Feng et al. (2024), who developed a customized FL model for identifying high-risk patients with postoperative gastric cancer recurrence. The model reportedly enhanced local hospital models and improved the detection local algorithm's performance when applied to a separate public lung cancer dataset. Nevertheless, the peer review of the study has raised methodological concerns, including the identification of regions of interest (ROI), sample sizes, and the privacy protection of generated data.

Other FL projects have also shown promise in improving diagnostic imaging and, by extension, patient outcomes. Pati et al. (2022) successfully employed FL in the diagnosis of common and fatal brain tumours, demonstrating positive contributions to the final consensus model's performance. Similarly, Almufareh et al. (2023) noted significant improvements in breast cancer diagnosis through FL, which consolidated data from multiple healthcare institutions without compromising patient privacy. Lu et al. (2020) utilized federated learning to effectively develop deep learning models that use whole-slide histopathological images opening the possibility for other institutions to contribute their data and train models that generalize better on unseen data.

Despite the advantages, FL is not without limitations. It remains susceptible to security vulnerabilities, including backdoors, data poisoning, membership inference, generative adversarial network (GAN)-based attacks, and differential privacy breaches (Hasan 2023). Moreover, even the transfer of model weights could expose information related to the training datasets through reverse engineering (Sheller et al. 2020). Bias propagation is another concern, as it may lead to unequal treatment outcomes for different patient groups (Chang and Shokri 2023). Nonetheless, FL still represents a transformative approach to AI in healthcare by facilitating the development of more accurate diagnostic models without direct interaction with confidential data. Specific interventions can also be implemented to mitigate these concerns such as the usage of differential privacy to further mask real data.

Regarding the dataset to be used (GasHisSDB) (Hu, Li, and Li, n.d.), the developers of the dataset sourced the images from Longhua Hospital Shanghai University of Traditional Chinese Medicine, which were then augmented by biomedical researchers from Northeastern University and experienced pathologists from Liaoning Cancer Hospital and Institute. To assess the useability of the images, the developers performed classification through classical machine learning methods (Random Forest, linear Support Vector Machine) and, deep learning (VGG16, Resnet50 and ViT), demonstrating that classification models performed competently on the database. Other deep-learning studies were also completed using this dataset by Yong et al. (2023) and Khayatian et al. (2024) which further solidified the viability of the dataset for training gastric image classifiers.

3 Project Problems

3.1 Project Aims and Objectives

The objective of this project is to develop a predictive tool for faster intervention and treatment of potential gastric cancer diagnoses. The project will focus on the histological analysis of gastrointestinal tissues, leveraging advanced image classification techniques and Federated Learning to ensure data privacy. The emphasis on Federated Learning enables collaboration between multiple institutions without sharing sensitive patient data, offering potential applications beyond gastric cancer diagnosis to other clinical contexts. The tool aims to aid in early detection, improving treatment outcomes while maintaining compliance with data privacy regulations.

3.2 Project Questions

- Can a single image classification model trained on a small dataset (single clinic) outperform the same model developed across multiple datasets (multiple clinics) using Federated Learning to ensure data privacy?
- Can Federated Learning enhance the predictive performance of models trained in different clinical settings?

3.3 Project Scope

Model Development: Implement a federated learning model in PyTorch to detect gastric cancer from medical images.

Data Preprocessing: Build pipelines to clean, normalise, and augment imaging data across nodes for consistent input quality.

Performance Evaluation: Validate the model using accuracy, precision, recall, F1-score, and AUC-ROC metrics across diverse datasets.

Model Deployment: Deploy the model on distributed systems (e.g., Docker, AWS) to enable collaborative training without data sharing.

Documentation and Reporting: Document model architecture, implementation, and results, and deliver a final report summarizing research and outcomes.

4 Methodologies

The methods to solve the problem will mainly focus on Federated Learning. To do this, the team will start with a pre-trained model and leverage transfer learning to apply it to this context. Referencing other studies, the team plan to test and use one of the following common deep learning architectures: VGG16, VGG19, ResNet50, or EfficientNet v2. Specifically, the team will "unfreeze" some layers of the model, ensuring that the model can be retrained for the planned binary classification task. The resulting model will serve as the baseline for future experiments. To begin the process, the training will be done on a small subset of the data to simulate how the model performs in a single clinical setting.

The team will then build a Federated Learning framework that will allow the simulation of training across multiple clinics without sharing raw data, effectively ensuring data privacy. The approach would be model-centric and cross-silo, meaning that the global model will be trained collaboratively across multiple clinics (each with their own assigned local data) to improve the global model. The goal is to see how well the federated model performs compared to the baseline trained on the simulated single clinic's data.

4.1 Data Collection and Data Analysis

The dataset for this project will be the GasHisSDB dataset. This dataset consists of 245,196 image patches from six hundred gastric cancer whole-slide images (2048x2048 pixels). Normal images contain no cancerous regions, with cells showing little to no atypia, regular single-layer arrangement, and minimal mitosis. These characteristics make them easily identifiable as normal under a microscope, allowing whole images to be directly cropped for dataset creation. Abnormal images feature gastric cancer, typically ulcer-shaped, infiltrating from the mucosal to muscle and serosal layers. Cancer cells may appear in nests, glands, or irregular arrangements, with unclear boundaries when infiltrating stroma. Cropped images focus on regions with at least 50% cancerous areas. The H&E staining method, the gold standard in histology, stains nuclei purplish-blue and cytoplasm pink, making cellular structures easily distinguishable. Normal images display more pink and white areas, while abnormal ones have more disorganised purplish-blue regions (Hu et al. 2022).

4.2 Data Preprocessing

Images will be resized to 224x224 pixels and divided into four subsets to simulate different clinical settings. Since the images are limited to two colour channels, three approaches are available for conversion to three channels:

1. Repeating one of the existing channels.

2. Converting to grayscale by averaging the existing channels.
3. Stacking three grayscale channels.

Normalisation in this context means converting the pixel values of the image (which range from 0 to 255) into a range between 0 and 1 by dividing each value by 255. After that, the pixel values are adjusted using known parameters for existing models, mean [0.485, 0.456, 0.406] and standard deviation [0.229, 0.224, 0.225] for each channel. This adjustment ensures that the input data matches the distribution that the pre-trained model was originally trained on. Furthermore, the data will be divided into four different batches to simulate four distinct clinics, each representing a client in the federated learning setup. The batches will then be split into training, validation, and testing sets in a 60:20:20 ratio.

4.3 Predictive Modelling and Evaluation

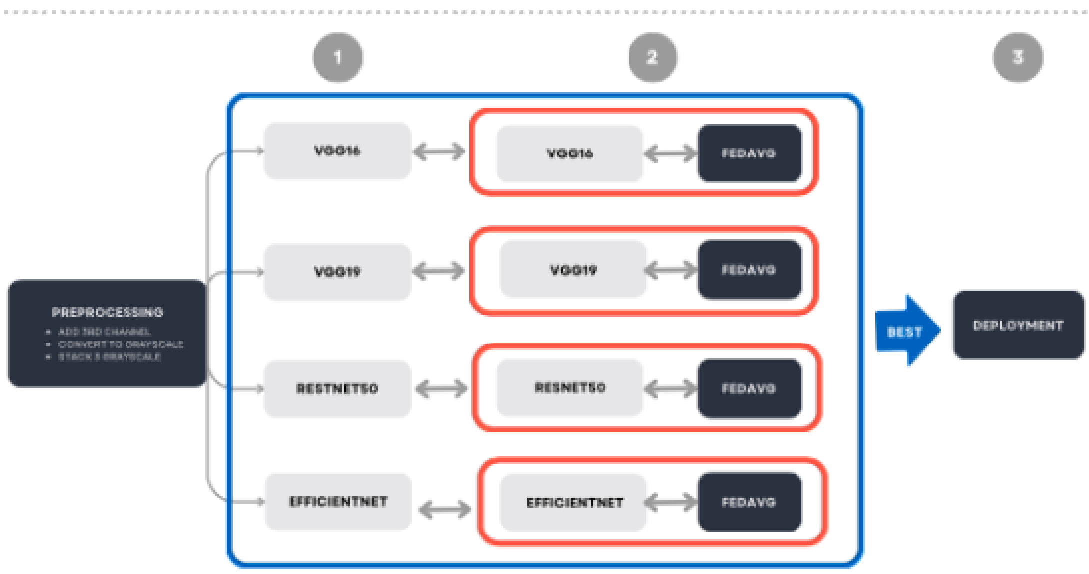

Figure 1: Model Evaluation Method

The evaluation will be conducted in two stages:

1. **Training the Predictive Model (Pre-trained Models).** Four pre-trained models (VGG19, VGG16, ResNet50, EfficientNet) will be fine-tuned for binary classification, and trained on centralised data. The performances will be evaluated using accuracy, precision, recall, and F1-score to assess their suitability for the task.
2. **Applying Federated Learning (FedAvg) to Pre-trained Models.** After evaluating the standalone pre-trained models, Federated Averaging (FedAvg) will be applied. The global model, initialised

with ImageNet weights, will be distributed to clients. Each client fine-tunes the model on local data, and updated model parameters are aggregated by the server. This process will be repeated for multiple communication rounds. The final performance of these federated models will be compared to the standalone pre-trained models to determine if Federated Learning improves their performance.

3. **Deployment of the Best Performing Model.** Once the evaluation is complete, the best performing model, whether a standalone pre-trained model or a federated model, will be selected for deployment in production.

5 Resources

5.1 Materials

- **PyTorch:** Frameworks for building and training the federated learning model.
- **OpenMined / PySyft:** Tools for enabling federated learning and privacy-preserving machine learning.
- **Poetry:** Python dependency management and environment setup.
- **AWS S3:** For storing data such as gastric cancer image datasets.
- **Google Colab:** For enabling deep learning with GPU standards.
- **GitHub:** For version control and collaboration.
- **Slack:** For commucation and collaboration.
- **Microsoft Teams:** For project management, sharing dataset and online meeting.

5.2 Roles and Responsibilities

Table 1: Roles and Responsibilities

Name	Roles & Responsibilities
David Bain	Develop Client model, implement integrated model, handle version control (GitHub), report writing
Emmanuel Niko Sindayen	Direct the research aspects of the federated learning model, and develop federated learning model, report writing
James Murray	Develop Client model, implement integrated model, manage all written reports
inh Thanh Nguyen	ead models development (Build and train the federated learning model using PyTorch), report writing
Taekjin Jeong	Develop Client model, Project management, support research aspects, report writing
Warisara Siriponpaiboon	Direct the research aspects of the federated learning model, and develop federated learning mode, coordinate communication. Report writing

6 Expected Outcomes

6.1 Materials

- **Federated Learning Model:** A fully trained federated learning model capable of diagnosing gastric cancer from medical imaging data. This model will be distributed across multiple nodes, allowing for decentralised data processing while maintaining patient privacy.
- **Data Preprocessing Pipeline:** A robust pipeline for cleaning, normalising, and augmenting the medical imaging data to ensure high-quality input for the federated learning model.
- **Thesis:** The Thesis will provide the entire research and development process of the project. It will cover the theoretical background, methodology, experimental results, and discussions. The thesis will serve as a key academic deliverable and contribution to the field of medical AI and federated learning.

7 Milestone - Schedule

A project plan was developed, and the Gantt chart below highlights the key timelines associated with the project's critical elements. While this seems linear in its application, there is a significant amount of concurrent activity in preparation for each task. There are also numerous sub-tasks associated with each of the milestones outlined below.

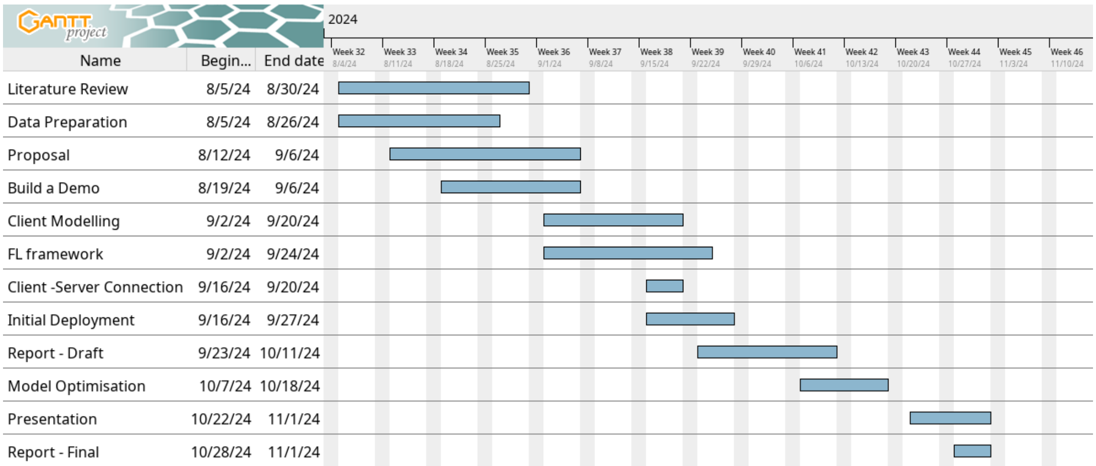

Figure 2: Project Schedule - Gantt Chart

8 Conclusion

The project proposal outlined above highlights state of the art methods that could be applied in a medical context. While this has been done by others the opportunity to test and subsequently demonstrate some of these techniques in a Data sensitive context should not be underestimated, it should however, be embraced. These state of the art methods have the potential to assist in overcoming some of the significant challenges of applying machine learning in the Data sensitive context of the medical domain. The potential for machine learning to contribute to better patient care is significant. Federated Learning is one method that could enable a significant acceleration and uptake machine learning in the medical field, assisting physicians and ultimately improving patient outcomes.

References

Almufareh, Maram Fahaad, Noshina Tariq, Mamoona Humayun, and Bushra Almas. 2023. "A Federated Learning Approach to Breast Cancer Prediction in a Collaborative Learning Framework." *Healthcare* 11 (24): 3185. https://doi.org/10.3390/healthcare11243185.

Chang, Hongyan, and Reza Shokri. 2023. "Bias Propagation in Federated Learning." arXiv. https://doi.org/10.48550/arXiv.2309.02160.

Deng, Yang, Hang-Yu Qin, Yan-Yan Zhou, Hong-Hong Liu, Yong Jiang, Jian-Ping Liu, and Ji Bao. 2022. "Artificial Intelligence Applications in Pathological Diagnosis of Gastric Cancer." *Heliyon* 8 (12): e12431. https://doi.org/10.1016/j.heliyon.2022.e12431.

Feng, Bao, Jiangfeng Shi, Liebin Huang, Zhiqi Yang, Shi-Ting Feng, Jianpeng Li, Qinxian Chen, et al. 2024. "Robustly Federated Learning Model for Identifying High-Risk Patients with Postoperative Gastric Cancer Recurrence." *Nature Communications* 15 (January): 742. https://doi.org/10.1038/s41467-024-44946-4.

Hasan, Jahid. 2023. "Security and Privacy Issues of Federated Learning." arXiv. https://doi.org/10.48550/arXiv.2307.12181.

Hirasawa, Toshiaki, Kazuharu Aoyama, Tetsuya Tanimoto, Soichiro Ishihara, Satoki Shichijo, Tsuyoshi Ozawa, Tatsuya Ohnishi, et al. 2018. "Application of Artificial Intelligence Using a Convolutional Neural Network for Detecting Gastric Cancer in Endoscopic Images." *Gastric Cancer* 21 (4): 653–60. https://doi.org/10.1007/s10120-018-0793-2.

Hu, Weiming, Chen Li, and Xiaoyan Li. n.d. "GasHisSDB: A New Gastric Histopathology Image Dataset for Computer Aided Diagnosis of Gastric Cancer."

Hu, Weiming, Chen Li, Xiaoyan Li, Md Mamunur Rahaman, Jiquan Ma, Yong Zhang, Haoyuan Chen, et al. 2022. "GasHisSDB: A New Gastric Histopathology Image Dataset for Computer Aided Diagnosis of Gastric Cancer." *Computers in Biology and Medicine* 142 (March): 105207. https://doi.org/10.1016/j.compbiomed.2021.105207.

Ikenoyama, Yohei, Toshiaki Hirasawa, Mitsuaki Ishioka, Ken Namikawa, Shoichi Yoshimizu, Yusuke Horiuchi, Akiyoshi Ishiyama, et al. 2021. "Detecting Early Gastric Cancer: Comparison Between the Diagnostic Ability of Convolutional Neural Networks and Endoscopists." *Digestive Endoscopy* 33 (1): 141–50. https://doi.org/10.1111/den.13688.

Khayatian, Danial, Alireza Maleki, Hamid Nasiri, and Morteza Dorrigiv. 2024. "Histopathology Image Analysis for Gastric Cancer Detection: A Hybrid Deep Learning and Catboost Approach." *Multimedia Tools and Applications*, August. https://doi.org/10.1007/s11042-024-19816-2.

Klang, Eyal, Ali Sourosh, Girish N. Nadkarni, Kassem Sharif, and Adi Lahat. 2023. "Deep Learning and Gastric Cancer: Systematic Review of AI-Assisted Endoscopy." *Diagnostics* 13 (24): 3613. https://doi.org/10.3390/diagnostics13243613.

Lu, Ming Y., Dehan Kong, Jana Lipkova, Richard J. Chen, Rajendra Singh, Drew F. K. Williamson, Tiffany Y. Chen, and Faisal Mahmood. 2020. "Federated Learning for Computational Pathology on Gigapixel

Whole Slide Images." arXiv. https://doi.org/10.48550/arXiv.2009.10190.

Pati, Sarthak, Ujjwal Baid, Brandon Edwards, Micah Sheller, Shih-Han Wang, G. Anthony Reina, Patrick Foley, et al. 2022. "Federated Learning Enables Big Data for Rare Cancer Boundary Detection." *Nature Communications* 13 (1): 7346. https://doi.org/10.1038/s41467-022-33407-5.

Sheller, Micah J., Brandon Edwards, G. Anthony Reina, Jason Martin, Sarthak Pati, Aikaterini Kotrotsou, Mikhail Milchenko, et al. 2020. "Federated Learning in Medicine: Facilitating Multi-Institutional Collaborations Without Sharing Patient Data." *Scientific Reports* 10 (1): 12598. https://doi.org/10.1038/s41598-020-69250-1.

Song, Zhigang, Shuangmei Zou, Weixun Zhou, Yong Huang, Liwei Shao, Jing Yuan, Xiangnan Gou, et al. 2020. "Clinically Applicable Histopathological Diagnosis System for Gastric Cancer Detection Using Deep Learning." *Nature Communications* 11 (1): 4294. https://doi.org/10.1038/s41467-020-18147-8.

Tang, Dehua, Lei Wang, Tingsheng Ling, Ying Lv, Muhan Ni, Qiang Zhan, Yiwei Fu, et al. 2020. "Development and Validation of a Real-Time Artificial Intelligence-Assisted System for Detecting Early Gastric Cancer: A Multicentre Retrospective Diagnostic Study." *EBioMedicine* 62 (December): 103146. https://doi.org/10.1016/j.ebiom.2020.103146.

Yong, Ming Ping, Yan Chai Hum, Khin Wee Lai, Ying Loong Lee, Choon-Hian Goh, Wun-She Yap, and Yee Kai Tee. 2023. "Histopathological Gastric Cancer Detection on GasHisSDB Dataset Using Deep Ensemble Learning." *Diagnostics* 13 (10): 1793. https://doi.org/10.3390/diagnostics13101793.

Yoshida, Hiroshi, Taichi Shimazu, Tomoharu Kiyuna, Atsushi Marugame, Yoshiko Yamashita, Eric Cosatto, Hirokazu Taniguchi, Shigeki Sekine, and Atsushi Ochiai. 2018. "Automated Histological Classification of Whole-Slide Images of Gastric Biopsy Specimens." *Gastric Cancer* 21 (2): 249–57. https://doi.org/10.1007/s10120-017-0731-8.