

A Federated Learning-based project for predicting disease diagnoses using Artificial Intelligence

Soobia Saeed, Habibollah Haron, Izaz Riaz, Halah Khadija Shah, Muhammad Riaz, Mohsin Qadeer

^{1,2}Department of Computer Science, University Malaysia of Computer Science & Engineering (UNIMY)

³Kyber Medical College, Peshawar, Pakistan

⁴Istanbul Atlas University, Hamidiye, Turkey

⁵Department of Neurosurgery, University of Colorado | Denver Health Medical Center/Children's Hospital, Colorado, Usa

⁶Department of Neurosurgery, Ziauddin University, Karachi, Pakistan

Soobiasaeed1@gmail.com, habibollah@unimy.edu.my, izazriaz243@gmail.com, halahshah01@gmail.com,

Muhammad.Riaz2@dhha.org, Mohsin.qadeer@gmail.com

*Correspondence Author: soobiasaeed1@gmail.com

ABSTRACT

The healthcare sector has undergone an accelerated digital transformation which has resulted in the creation of large volumes of sensitive medical information and has simultaneously made it possible to use the power of AI to predict diseases and have serious data privacy, security, and compliance issues arise. Centralized healthcare systems, which are the norm, are finding it harder to secure their data and are against the possible legal and ethical battles posed by the strict regulations such as HIPAA and GDPR. This study presents an innovative solution of an AI-based disease prediction system that is meticulously designed to work with federated learning and a decentralized healthcare management framework. The main feature of the solution is that it allows different hospital networks to jointly train the same model without the need of sending the actual patient data among the hospitals thus, it is not only able to maintain the privacy of the data but also the power over its usage. The system combines federated learning with machine learning and deep learning-based models, which consist of, among others, Convolutional Neural Networks (CNNs) for medical imaging analysis, Support Vector Machines (SVMs) for structured clinical data, and Random Forest classifiers for multimodal prediction. Secure aggregation, encryption methods, and role-based access management are implemented to protect data authenticity and privacy during the course of the system operation. The application of such data sets that are freely accessible and stripped of personal identifiers is meant to create a representation of the real-world multi-institutional healthcare settings and this is also a way to demonstrate the proposed method's feasibility, scalability, and robustness. Decentralized federated learning models, according to the experiment outcomes, are capable of performing at the same level as centralized methods with the exception that they come with much lesser privacy risk. The authors of the paper emphasize the potential that the use of federated learning-based decentralized healthcare systems can have in the areas of early disease detection support, clinical decision making, and the deployment of responsible AI that is compatible with the ethical standards prevailing in the modern healthcare ecosystem.

Keywords: *Federated Learning with decentralized healthcare, disease prediction, and privacy-preserving AI.*

How to cite this article: Saeed S, Haron H, Riaz I, Shah HK, Riaz M, Qadeer M. A Federated Learning-based project for predicting disease diagnoses using Artificial Intelligence. *Int J Drug Deliv Technol.* 2026;16(10s): 975-984. DOI: 10.25258/ijddt.16.10s.113

1. INTRODUCTION

The very rapid digital transformation of healthcare systems resulted in the creation of very large amounts of sensitive clinical data via electronic health records (EHRs), medical imaging, wearable devices, and mobile health applications. On one side, this data explosion opens up vast opportunities for the use of artificial intelligence (AI) in disease prediction and drug personalization, while on the other hand, it raises quite serious issues of data privacy, security, interoperability and regulations that must be complied with. The centralized healthcare management systems, which have

been the standard, 'storing' all patient data in single repositories, have increasingly been exposed to a number of threats in terms of cyberattacks, data breaches, and ethical issues, particularly in contexts of the most stringent data protection regulations like HIPAA, GDPR, and various national personal data protection acts [1] Decentralized healthcare management systems have come up as a revolutionary and novel answer to the mentioned drawbacks by scattering data ownership and computation among various healthcare organizations and at the same time reducing the dependence on a single centralized authority. Federated learning (FL), among

A Federated Learning-based project for predicting disease diagnoses using Artificial Intelligence

decentralized AI technologies, has achieved the highest popularity as a privacy-preserving machine-learning framework that allows teaching the model collaboratively without transferring raw patient data from the local servers. Instead, encrypted model updates are shared and aggregated, which results in an enormous reduction of the likelihood of data leakage while the model's performance and universality are kept intact [2,3] This scenario is closely related to the core values of decentralized healthcare by allowing control over the data, independence of the institutions, and co-operation among them.

Recently, several studies have demonstrated the application of federated learning in different clinical scenarios such as cancer detection, cardiovascular risk assessment, and even prediction of infectious disease outcomes. The latter being a significant hint as it can support the healthcare sector with the provision of real-time, trustworthy, and scalable intelligence [4,5]. Moreover, the combination of federated learning with secure communication protocols and encrypted storage may become even more powerful when used along with role-based access control as it gives more power to decentralized healthcare infrastructures, which can now manage the constantly changing regulatory and ethical standards [6]. With this, the current research proposal reveals an AI-powered decentralized healthcare management system that makes use of federated learning to conduct disease predictions with privacy among multiple institutions. Mimicking actual hospital scenarios through the use of public, anonymized datasets, the proposed system demonstrates the strength of decentralized intelligence in the areas of early disease detection, reducing diagnostic delays, and empowering both patients and clinicians without jeopardizing data confidentiality. It is thus, this research work, which aligns AI innovation with decentralization and privacy-by-design principles, that adds to the growing body of studies advocating for secure, scalable, and ethically responsible healthcare management systems for the next generation of digital health ecosystems.

2. LITERATURE REVIEW

2.1 Decentralization in Modern Healthcare Systems

The rapid transformation of healthcare through digital technologies has resulted in an extraordinary volume of medical data, among which are electronic health records (EHRs), diagnostic imaging (CT, MRI), laboratory reports, and real-time patient monitoring data. Traditionally, healthcare information systems have heavily relied on centralized architectures that consolidate patient data from various healthcare institutions into one repository. Although this method facilitates analytics and interoperability, it also brings along major problems with respect to data privacy, security risks, regulatory compliance, and centralized

points of failure. The radical changes that have been critically necessary are the ones that have brought decentralized healthcare management systems on the scene as a very promising and progressive option, thereby opening up distributed data governance that still permits collaborative intelligence across the medical sectors. In such systems, data is kept with local entities such as hospitals or clinics, and only their insights, encrypted parameters, or model updates get shared. This model is compliant with the prevailing international regulatory frameworks like HIPAA, GDPR, and PDPA that emphasize the need for data minimization and consumer sovereignty [7].

2.2 Federated Learning as a Core Enabler of Decentralized Healthcare

Federated Learning (FL) has solidified its standing as a critical technology in the middleware of decentralized health care management. The privacy-preserving feature of FL makes it different from conventional machine learning as it allows the collaboration of training models but without the transfer of raw patient data to the central server, thereby still maintaining the privacy and autonomy of the institutions. In a way, each participating location or node trains a local model on its dataset and only sends the coordinates of the model parameters (or encrypted model updates) to a central server for merging [8,9]. Current researches validate the potential of FL in decentralized disease prediction systems especially for delicate medical tasks like cancer detection, cardiovascular risk assessment, and infectious disease prognosis. In the same trend, Dayan et al. (2021) applied FL over different hospitals in order to predict the clinical outcomes of COVID-19, and they got performance that was on the same level as centralized models but with privacy being significantly fortified. Likewise, Kapila and Saleti (2024) reported the use of FL-based feature fusion methods to bring about an increase in generalization in heart disease prediction. Besides, the attached project report further supports the relevance of FL in the decentralized healthcare field by suggesting an AI-assisted disease prediction system where the hospitals sharing the collaboration won't have to reveal the records of their patients in order to improve the predictive performance. This directly correlates with the conference theme in that FL is the one that operationalizes decentralization while still offering clinical utility[10].

2.3 Security and Privacy Preservation in Decentralized Healthcare Systems

Though decentralization diminishes the risks related to centralized data storage, it brings along new security issues such as model inversion attacks, poisoning attacks, and inference threats. A detailed review is presented by Li et al. (2023), discussing the security of FL in healthcare which is dependent on differential privacy,

A Federated Learning-based project for predicting disease diagnoses using Artificial Intelligence

secure multiparty computation, and homomorphic encryption, as major protective measures[3]. Decentralized healthcare management systems widely adopt encryption protocols like TLS and AES-256 for data protection during transmission and storage. The document illustrates the use of end-to-end encryption, role-based access control (RBAC), and audit logging as measures to confirm accountability and compliance with regulations in a decentralized architecture. Additionally, Shen et al. (2023) proved that the integration of FL with homomorphic encryption could result in an accuracy of diagnosis higher than 95% whereas keeping communication overhead low, thus making such systems suitable for real-time e-healthcare applications. However, despite the above-mentioned developments, privacy and performance issues still exist, particularly with heterogeneous and non-IID medical data across decentralized nodes[7,11-14].

2.4 Decentralized Disease Prediction and Clinical Decision Support

The use of AI-driven disease prediction and clinical decision support in decentralized healthcare management systems is becoming more widespread as it helps to speed up diagnosis and enhance patient care. Deep learning models based on CNNs have been trained on decentralized medical imaging data and have shown to be very effective in tumor, heart, and lung disease detection (Amritanjali and Gupta, 2025). The system reviewed in the document submitted illustrates that decentralized AI models can deliver preliminary diagnostic help not just to doctors but also to patients, thus narrowing the time span between medical tests and consultations with doctors. Such an approach is usually needed in health care systems that are suffering from a shortage of doctors or have long waiting times, as is the case in many developing and middle-income countries[11,15-18]. On top of that, decentralized decision support systems give power to the patients by offering transparent and interpretable predictions while making sure that final clinical decisions are still under professional supervision.

2.5 Integration with Emerging Decentralized Healthcare Ecosystems

Literature to date incorporates decentralized healthcare management via blockchain, edge computing, and the healthcare metaverse in addition to FL. Kugan et al. (2023) claim that FL is the underpinning of decentralized healthcare ecosystems allowing for safe participation in immersion, IoT-enabled and spread out geographically environments. In a like manner, Zhang et al. (2023) put forward a classification of FL-based medical applications indicating its suitability with edge devices and decentralized infrastructures. These trends show that decentralized healthcare management solutions are not stand-alone but rather move together with the general

transition to patient-centered, trust-based digital health ecosystems[15-19-21].

2.6 Research Gaps and Future Directions

The decentralized healthcare management system has gaps despite the promising advancements made. First, the issue of different types and amounts of data in healthcare institutions negatively impacts model convergence and fairness. Another major concern is that there are no standardized benchmarks and evaluation frameworks in decentralized healthcare AI. Up to now, the systems have had very few large-scale real-world deployments; most of them have undergone validation through either simulation or public datasets. To be accepted, the system comes up with a solution as a scalable and privacy-preserving prototype for decentralized healthcare analytics. Research to be done in the future includes hybrid FL-blockchain architectures, explainable AI for decentralized clinical decision-making, and cross-border regulatory harmonization. In the end, decentralized healthcare management systems imply a transition to privacy-preserving collaboration intelligence rather than data-centric centralization. Federated learning combined with strong security procedures and AI-powered analytics enables healthcare institutions to simultaneously create disease prediction and decision support systems while keeping patient data secret. The literature review and the ongoing project confirm that decentralized systems are not only technically feasible but also the future of the ethical, scalable, and trustworthy digital healthcare.

3. METHODOLOGY

The use of a hybrid research approach consisting of design-science and experimentation has enabled the creation and evaluation of a decentralized healthcare management system that employs federated learning (FL) to maintain privacy in disease prediction. The research is structured with respect to system analysis, design, and implementation workflows as documented under the project but in a way that is still compliant with the current digital healthcare environments' main principles of decentralization, data sovereignty, and secure collaboration.

3.1 Research Design and Approach

The approach taken was a hybrid one, involving system design, testing federated machine learning, and evaluating performance, all together ensuring that the system was both robust from a technical standpoint and suitable for practical use. The investigation proceeds through four successive steps: creating a model of the decentralized system, developing a federated learning-based disease prediction model, implementing privacy, security, and governance measures, and performing a thorough evaluation of model and system performance.

A Federated Learning-based project for predicting disease diagnoses using Artificial Intelligence

The main advantage of such an approach is that clinical intelligence is jointly developed amongst the scattered healthcare organizations keeping the sensitive patient data at their own location, thus directly supporting the goals of decentralized healthcare management systems.

3.2 System Architecture for Decentralized Healthcare

The decentralized healthcare system is a system that not only permits joining of all healthcare players but also does it with the maximum privacy and at the same time, each party retains its power. The architecture is built on the concept of distributed intelligence and secure coordination to infer diseases without revealing granular medical data.

3.2.1 Overall Architecture Design

The proposed system utilizes a client-edge-aggregator structure whereby patient data is always stored in hospitals, clinics, or on personal devices. The local nodes are responsible for the patients' data storage and processing while the edge servers are conducting the training of local models with institution-specific datasets. A federated aggregator coordinates the collection and integration of encrypted model updates to develop a global model. There is no passing of raw medical records beyond their local settings, thereby ensuring data locality, regulatory compliance, and institutional sovereignty, which are the fundamental needs for decentralized healthcare ecosystems as shown in Figure 1.



Figure.1: Privacy Preserving and Decentralized Training

3.2.2 Functional Components

A wide range of integrated functional modules forms the backbone of the system architecture that altogether enables decentralized operation. The mobile healthcare interface, built with React Native, ensures the effortless and quick communication of medical reports between patients and doctors, and the display of prediction outcomes is very friendly to the user. The federated learning engine, which is developed using either PySyft or TensorFlow Federated, helps in the decentralized training of models on distributed datasets without compromising data privacy. Backend services set up with Django, are responsible for user authentication that is secure and also for communication that is encrypted between system components and based on the user roles. Decentralized data storage using MongoDB is also employed to store metadata, system logs, and anonymized outputs, thereby ensuring that sensitive medical records are distributed rather than centralized while still allowing for scalability and auditability.

3.3 Data Sources and Decentralized Data Handling

The decentralized healthcare management system is built to operate within a strict ethical and regulatory framework but also permits the exchange of information between different organizations. To achieve this, the system adopts a data management strategy that treats privacy, data sovereignty, and regulatory compliance as the major concerns, thereby ensuring that sensitive healthcare data is neither stored nor communicated during the entire process of model development and deployment.

3.3.1 Data Sources

To comply with the ethical and legal standards, the system has to depend on a combination of publicly available medical datasets, anonymized clinical records, and synthetic data. Public datasets from sources like Kaggle, NIH, and MONAI are used to showcase real-life clinical scenarios without any patient data being identified. Furthermore, synthetic datasets are also used in the mix to evaluate system scalability and robustness with the increased data load. The medical imaging data that has been anonymized, which includes CT and MRI scans, as well as structured health reports, is being used to assist in disease prediction tasks. All these datasets, together, create a multi-institutional healthcare environment that can be used for collaborative model training while at the same time completely eliminating the risk of disclosing patient identities or breaching their privacy.

3.3.2 Local Data Processing

In a decentralized setting, every node that becomes a part of the network carries out all data processing operations on the datasets stored in its location independently. These data processing operations are referred to as data preprocessing actions. They consist of activities such as normalization, resizing, and encoding which are done to maintain standardization across the different data

A Federated Learning-based project for predicting disease diagnoses using Artificial Intelligence

sources. Afterwards, the features are extracted locally, and the data is further prepared for the machine learning by local model training, which uses only the institution-specific datasets. With this method, it is guaranteed that the original medical data will never be transferred to any other institution, and the data rights will always be respected. This also goes hand in hand with the decentralized health care governance models which advocate institutional autonomy and privacy preservation as their main tenets.

3.4 Federated Learning Methodology

The federated learning approach is at the center of the system's methodological choice, which facilitates the decentralized training of models in different healthcare organizations. By using this technique, the institutions can train their models together without compromising their data security and having to set up a central data storage, thus the privacy risks involved in the classic ML process are minimized.

3.4.1 Federated Training Strategy

A horizontal federated learning strategy is adopted in which the institutions involved have comparable feature spaces but distinct patient populations and data distributions. The workflow for federated training commences with global model initialization, which is then distributed to all the participating nodes. Each node conducts local training on its private dataset and computes updates to the model. These updates are securely sent to a central aggregator in encrypted form, where they are combined using the Federated Averaging (FedAvg) algorithm to refresh the global model. The refreshed model is then sent back to the participating nodes who can thus benefit from the iterative and privacy preserving improvement of the model.

3.4.2 Disease Prediction Models

Different machine learning models are assessed to ensure reliable and accurate disease forecasting in a decentralized context. Medical imaging data is further processed through Convolutional Neural Networks (CNNs), since these networks are good at getting the spatial features out of the CT and MRI scans. Support Vector Machines (SVMs) are compared with the health metrics based on the structure, thus giving the classification tasks with limited features good performance. Moreover, Random Forest classifiers are used for multimodal prediction by combining both imaging and structured data. The models are chosen because of their reliability, interpretability, and compatibility with the federated and decentralized healthcare systems' deployment as shown in Figure2.

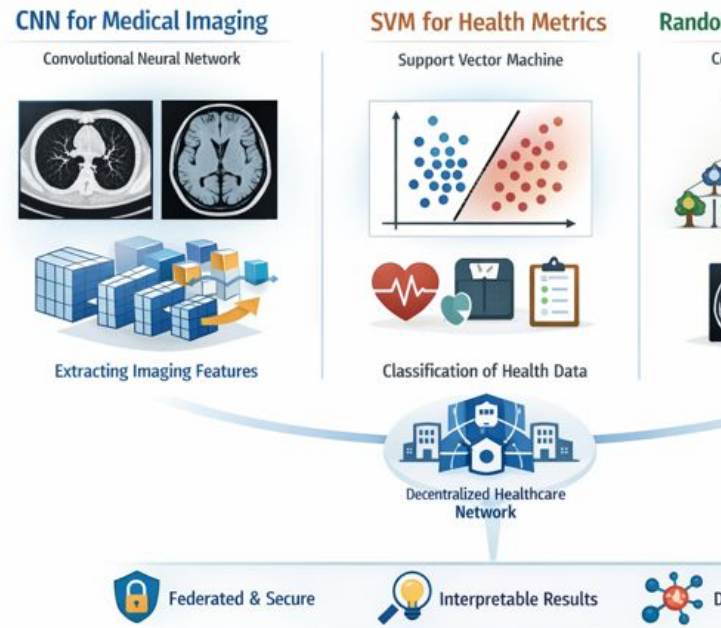


Figure.2: AI based disease prediction model

3.5 Security, privacy, and governance

Security, privacy, and governance are integral and interrelated elements of the proposed decentralized healthcare management system. The whole process from data processing to model aggregation and result dissemination is covered by the methodology which has considered both technical and administrative measures of varying degrees of toughness; thus, ensuring that sensitive medical data will be protected during the entire lifecycle of the system.

3.5.1 Privacy Preservation

The privacy preservation process is mainly through federated learning, which guarantees the non-sharing and non-transfer of raw patient data outside the local environment. All the stored data is protected with AES-256 encryption, while the communication of data between nodes and aggregators is done through Transport Layer Security (TLS) protocols, which offer high-level security. The secure aggregation method is used to protect the model update exchanges from inference attacks by preventing individual data contributions from being reconstructed or reverse-engineered through the aggregation of model parameters.

3.5.2 Access Control and Auditability

A Role-Based Access Control (RBAC) framework is instituted to manage access to the system and to allot rights to patients, physicians, and administrators. Access for each role is strictly confined to its tasks, thereby

A Federated Learning-based project for predicting disease diagnoses using Artificial Intelligence

minimizing the risk of data being accessed without authorization. Moreover, comprehensive audit trails are kept to record all system interactions, including data access, model updates, and administrative actions. These logs offer transparency, time, and accountability, which are the essential attributes of trustable and well-governed decentralized healthcare systems.

3.6 System Implementation Environment

The scheme for the system execution milieu is to back a decentralized healthcare management framework and at the same time push for scaling, security, and interoperability. The application's front end is constructed via React Native, which gives the possibility for a mobile interface that is cross-platform and, thus, accessible and easy to use for both patients and medical staff. The backend services are constructed using Django (Python), that is a powerful and secure framework for dealing with authentication, implementing role-based access control, and integrating with workflows of federated learning. MongoDB serves as the main database because of its capability to handle both structured and unstructured healthcare data, like medical reports and metadata, effortlessly. The functionality of federated learning is realized by utilizing PySyft and TensorFlow Federated, which allows the training of models in a decentralized manner without transferring raw patient data. For data transmission and storage, Transport Layer Security (TLS) and AES-256 encryption are used to ensure security. The whole system is put up at a simulation environment based on the cloud that is capable of testing decentralized learning scenarios in a scalable manner without actually depending on the infrastructure of a real hospital.

3.7 Evaluation metrics and validation

Both the prediction accuracy of the machine learning models and the operational efficiency of the decentralized system are evaluated and validated under this strategy.

3.7.1 Model performance metrics

The classification performance metrics of accuracy, precision, recall, F1-score, and ROC-AUC are standard ones for evaluating disease prediction models. They are the ones measuring the correctness, robustness, and reliability of the predictions made by the federated learning models. The system monitors performance diagnostics among the decentralized nodes which distribute data across multiple institutions to ensure that, even then, there is consistency and reliability in the performance of the system.

3.7.2 System-Level Evaluation

The decentralized healthcare platform is assessed at the system level under the following criteria: communication

overhead, model convergence rate, prediction latency, and scalability with varying node participation. Communication overhead gauges the efficiency of sending encrypted model updates from local nodes to the aggregator. Model convergence rate measures the speed at which the federated models come to a halt during training. Prediction latency determines how fast the system will react in real-time or near-real-time clinical situations, and the analysis of scalability looks at whether the system can still be high-performing when more medical facilities or users are participating.

3.8 Ethical and Regulatory Compliance

The suggested technique meets the requirements of the data protection and ethics standards in the healthcare sector, which include the principles of HIPAA, the requirements of GDPR related to data minimization and data sovereignty, and the guidelines that are recognized for ethical AI in clinical decision support systems. There will be no compromise on privacy as only encrypted model updates will be sent out while sensitive medical data is kept in local environments. The system has been made clear as just a support tool for decision-making, offering AI-assisted insights to both doctors and patients while also restricting itself from taking over the medical professional's judgment or clinical diagnosis.

This approach makes it possible to implement decentralized healthcare management by getting rid of centralized data storage for medical records, therefore risks concerning privacy and security are substantially lessened. It empowers the organizations to share their knowledge across, without having to expose the data directly, which implies that the collaboratively developed model can be trained while the data remains owned by the source. Moreover, the new way of working makes it possible to develop and use healthcare AI in a way that protects privacy and is still scalable, which means that the system can expand across institutions and populations while still the principles of decentralization, security, and ethical healthcare are respected as shown in Figure 3.

A Federated Learning-based project for predicting disease diagnoses using Artificial Intelligence

The presented research lays the groundwork for open

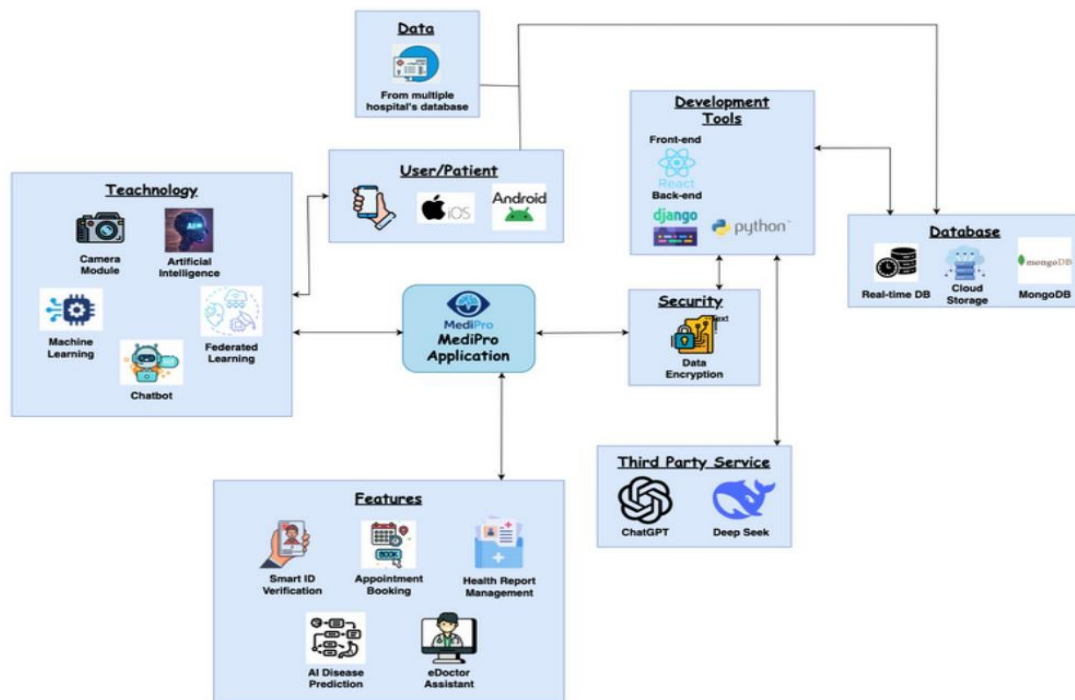


Figure.3: System design architecture

4. SYSTEM DESIGN

The structure of the AI-powered disease prediction system for privacy-preserving healthcare is clearly around a division of roles and responsibilities to ensure efficient development and seamless integration of decentralized machine learning features. The federated learning infrastructure using PySyft and TensorFlow Federated is created and maintained by Shaun Liew. Data preprocessing, model updating, and monitoring model accuracy are among the tasks he performs to maintain a consistent performance across the decentralized nodes. Lim Kean Yee is involved in the design of the overall system architecture, building its infrastructure, integrating third-party API services such as ChatGPT or DeepSeek to create AI-based medical reports, and developing and maintaining CNN-based image classification models. This research is aimed at improving the user experience and the designs of the interface, the interplay of the group leader and the user being discussed with the group leader to create a front-end that offers intuitive and accessible interaction for the healthcare professionals and the patients. Lee Seonghyeok is responsible for the back-end infrastructure, working together with Kean Yee and Shaun on database management, protecting data storage with the combination of MongoDB, TLS and AES-256 encryption, and implementing access control mechanisms to shield sensitive health information.

communication and team coordination throughout the different stages of task execution that involve the mixing of front-end, back-end, and machine learning parts, for example, integration of features, system testing, and deployment. Federated learning, a privacy-preserving AI technology is the backbone of system architecture. In federated learning, the model is trained at different hospital nodes without moving the raw patient data to the central server. The mobile app is built using the React Native technology so that it can be used on both iOS and Android. Wireframes and layouts are developed for smooth and user-friendly navigation. The app has different categories of users: patients, medical staff, and admins, with access control rights limited to their roles. The database architecture is capable of managing both data types, i.e., structured and unstructured, and is designed for quick access as well as safe data transfer. The AI part utilizes the CNN models for prediction in disease, giving preliminary diagnoses and report generation through third-party APIs, whereas the federated learning method allows for keeping patient data in the hospital, thereby guaranteeing adherence to data protection laws. This configuration of the system encourages the healthcare field to work together but in a decentralized way, all the while providing the benefits of user-friendliness, security, privacy, and accurate AI-supported medical suggestions.

5. EXPERIMENTAL WORK

The research methodology applied for the establishment of the AI-based system for disease prediction using

A Federated Learning-based project for predicting disease diagnoses using Artificial Intelligence

federated learning for privacy-preserving healthcare was a well-structured and collaborative approach which was followed by the whole team. The project was started with a mobile center design strategy, taking advantage of Flutter and Firebase for the front and back end and database management respectively, because of their user-friendliness and cross-platform compatibility. Federated learning was carried out by using PySyft and TensorFlow Federated so that the private patient data did not leave the local devices thus keeping privacy intact and observing data protection laws. Public datasets like the ones that can be found on Kaggle were used during the Capstone 1 phase of simulating multi-hospital data collaboration with the agreement that real hospital data would be considered in future phases with official approvals. The roles of each team member were unambiguously indicated, where the front-end developers concentrated on making a user-friendly interface with the help of wireframes, mid-fidelity prototypes, and role-based access management while backend developers took care of the federated learning setup, trained CNN-based disease prediction models and also made sure that the data was securely stored using TLS and AES-256 encryption on MongoDB.

The first stage of the experimental work was the creation of the dataset, further processed, and integrated into one CSV. The federated learning configuration permitted local model updates for the various hospital departments, which were, among others, dermatology, orthopedics, neurology, and internal medicine. The team reused the already available codes on GitHub and Kaggle, for instance, the bleeding detection notebooks, and tailored them to the project's particular datasets. AI disease diagnostic modules were put together with the mobile application, thus giving the users the option to upload medical reports or pictures, ask for AI-assisted diagnoses, and receive prescribed actions. Model accuracy, simulation hospital node effectiveness, and secure data handling integrity analyzing audit log verification and encryption testing were the criteria for the system's evaluation. Continuous feedback from the supervisor was part of the development process, which meant that the UI features were sharpened, federated learning workflows assured, and AI predictions were made in conformity with ethical norms.

6. DISCUSSION

The MediPro system, which is an AI-based disease prediction model utilizing federated learning, seems to significantly help in solving issues related to distributed healthcare management. The application is tailored to various types of users such as patients, volunteer and professional doctors, and administrative personnel. The architecture based on roles allows users to communicate with the application in accordance with their functions while preserving privacy and security of data. Volunteer doctors are the ones that provide free consultations,

whereas paid ones are for professional healthcare providers, hence both access and service sustainability are being supported. The user interface (UI) was created as a provisional prototype to assist the initial implementation of coding and give a low to mid-fidelity representation of the system operation. This method permits design enhancements to be made iteratively depending on the feedback from the supervisors and the considerations of the users' experiences.

Dataset management and preprocessing are unavoidable parts of MediPro's decentralized structure. Preprocessing was applied to combined datasets in order to unify parameters and thus publicly accessible datasets, like those from Kaggle and MONAI, were used to simulate multi-hospital collaboration which led to the effective training of AI models without the need for patient data to be centralized, thus complying with the privacy-preserving requirements. The group of researchers scrutinized and raised highlights of the literature from multiple sources in order to strengthen the literature review and make sure that existing AI-based healthcare solutions were covered comprehensively, thus pinpointing the project's opportunity to fill a gap.

The system can predict diseases based on images because of CNNs (Convolutional Neural Networks) used for processing imaging data which include CT and MRI scans of the brain, skin cancer, and orthopedic injuries conditions. Federated learning enables the local updates of the model at the nodes participating, so the raw data is not leaving the hospital that originally processed it, which is a good way of minimizing privacy risk. Connection to third-party APIs like DeepSeek and ChatGPT gives interpretability, as it the predictions are explained in natural language. It thus makes the difference between complicated medical analysis and the patient's understanding smaller. The dependency on third-party APIs, however, could lead to risks such as unavailability of services or downtimes; therefore, the system design has already taken such events into account through the use of fallback mechanisms and dummy data for demonstration purposes.

The project scope was redefined to limit it to one or two main disease categories, in accordance with the comments of the supervisor and examiner, in order to secure practical feasibility within the Capstone 1 timeline. The phasing of the project ensures that the model's training, testing, and validation can be done in a proper way but may also extend the system for future enhancements that will allow it to support several diseases and incorporate hospital datasets from the real world. The whole project brought to the fore various ethical aspects, especially with respect to patient data handling, validation of AI predictions with health professionals and government cooperation. The system prototype demonstrates the functioning of a federated portal as a way of modeling inter-hospital collaboration,

A Federated Learning-based project for predicting disease diagnoses using Artificial Intelligence

thus making it easier for the future integration of the entire country while still following the privacy and ethical guidelines very strictly.

Among the foremost issues raised by the risk analysis were those related to privacy of data, accuracy of predictions, required computation resources, user confidence and difficulties in usability. Data privacy is guaranteed through encryption protocols and the use of federated learning, which means that only a small amount of sensitive patient information is revealed. The accuracy of AI predictions is guaranteed by regular model retraining cycles and offering disclaimers to ensure that professional consultation is sought after for the end-users. The problem of high computational demands was circumvented by picking the most lightweight frameworks like Django and Flutter that provide a good compromise between performance and development efficiency. User trust and usability issues, especially regarding the non-technical user, were handled via intuitive UI design, output explanations, and feedback initiation through iterations.

At last, the initiative is in sync with the larger opportunities in the digital healthcare sector of Malaysia. It can be supported by government or NGO, it can be expanded to more diagnostic areas, and it can be linked with the national health systems as well. On the other hand, changes in regulations, bias in models, and competition among applications were some of the threats that were mentioned which pointed to the necessity of regular observation, update of the system, and keeping ethical standards. To sum up, MediPro has shown a way that is both feasible and morally sound to manage healthcare through a decentralized system combining federated learning, AI-driven interpretability, and secure, role-based system design.

7. CONCLUSION AND FUTURE WORK

The projection of the AI-powered disease prediction system, which is the outcome of the project, proves the effectiveness and the main benefits of the use of federated learning in a decentralized healthcare setting. The system, under such conditions, allows hospitals to come together and provide a common training ground for the machine learning models while preventing the sharing of any sensitive patient data directly. In this way, privacy along with preservation, data security, and compliance with ethical guidelines are all ensured. The use of a CNN-based disease prediction model together with a user-friendly mobile application is a strong and unambiguous sign of data locality remaining during the whole process of giving accurate diagnostic insights to both healthcare providers and patients. Furthermore, the system provides strong security measures including data encryption, role-based access control, and audit logging which together create an impregnable fortress and an absolute traceability for all health records. The project showcases

the potential of such decentralized platforms to draw medical services to be concentrated and to encourage government initiatives like the collaboration with organizations such as MUFON or the Malaysian Ministry of Health for auditing and coordinated healthcare delivery.

The prediction of several diseases in various medical fields by utilizing both hospital datasets already in use and third-party AI tools such as DeepSeek or ChatGPT for enriching the reports is a part of future stages of the system development. Moreover, among the improvements may be the introduction of differential privacy mechanisms, multi-class disease classification, and blockchain-based auditability that would help to further fortify the security and trustworthiness of decentralized healthcare systems. The real-world implementation with hospitals and governments working together, in addition to the professional validation by doctors, will make it easier for a large-scale adoption across the country which will eventually lead to the establishment of an efficient, privacy-protecting, and centralized healthcare management system that would be applicable in Selangor and other regions.

ACKNOWLEDGMENTS

The authors wish to sincerely thank the project supervisor for his/her continuous guidance, constructive feedback, and valuable insights during the entire research process. Acknowledgment is particularly given to all team members for their input in system design, federated learning implementation, backend and frontend development, and experimental validation. The authors are grateful for the publicly available datasets and open-source frameworks like PySyft, TensorFlow Federated, Kaggle, NIH, and MONAI, which directly contributed to the experimental evaluation of the proposed system. Last but not least, the authors acknowledge the support and encouragement from their institutions, without which this research would not have been possible.

REFERENCE

1. Dayan, I., Roth, H.R., Zhong, A., Harouni, A., Gentili, A., Abidin, A.Z., Liu, A., Costa, A.B., Wood, B.J., Tsai, C.S. and Wang, C.H., 2021. Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature medicine*, 27(10), pp.1735-1743.
2. Kapila, R. and Saleti, S., 2025. Federated learning-based disease prediction: A fusion approach with feature selection and extraction. *Biomedical Signal Processing and Control*, 100, p.106961. <https://doi.org/10.1016/j.bspc.2024.106961>
3. Li, H., Li, C., Wang, J., Yang, A., Ma, Z., Zhang, Z. and Hua, D., 2023. Review on security of federated learning and its application in healthcare. *Future Generation Computer Systems*, 144, pp.271-290.
4. Pati, S., Kumar, S., Varma, A., Edwards, B., Lu, C., Qu, L., Wang, J.J., Lakshminarayanan, A., Wang,

A Federated Learning-based project for predicting disease diagnoses using Artificial Intelligence

- S.H., Sheller, M.J. and Chang, K., 2024. Privacy preservation for federated learning in health care. *Patterns*, 5(7).
<https://doi.org/10.1016/j.patter.2024.100974>
5. Rauniyar, A., Hagos, D.H., Jha, D., Håkegård, J.E., Bagci, U., Rawat, D.B. and Vlassov, V., 2023. Federated learning for medical applications: A taxonomy, current trends, challenges, and future research directions. *IEEE Internet of Things Journal*, 11(5), pp.7374-7398.
6. Sharma, R., Miller, J., Iyer, P. and James, C., 2024. Federated Learning in Healthcare: Privacy-Preserving AI for Secure Medical Data Analysis. *ResearchGate. Viitattu*, 25, p.2025.
7. Shen, G., Fu, Z., Gui, Y., Susilo, W. and Zhang, M., 2023. Efficient and privacy-preserving online diagnosis scheme based on federated learning in e-healthcare system. *Information Sciences*, 647, p.119261.
<https://doi.org/10.1016/j.ins.2023.119261>
8. Amritanjali and Gupta, R., 2025, January. Federated Learning for Privacy Preserving Intelligent Healthcare Application to Breast Cancer Detection. In *Proceedings of the 26th International Conference on Distributed Computing and Networking* (pp. 302-306).
<https://doi.org/10.1145/3700838.3703679>
9. Sharma, R., Miller, J., Iyer, P. and James, C., 2024. Federated Learning in Healthcare: Privacy-Preserving AI for Secure Medical Data Analysis. *ResearchGate. Viitattu*, 25, p.2025.
10. Najib, T., Wasi, W., Muntasir, F. and Ahmed, N., 2024. Federated Learning in Healthcare: Preserving Privacy, Unleashing Potential.
<https://doi.org/10.13140/RG.2.2.24248.97280>
11. Z. Dougeri, Z. and Fasoulas, J., 2002, September. Stable grasping control under gravity by dual robotic fingers with soft rolling contacts. In *IEEE/RSJ International Conference on Intelligent Robots and Systems* (Vol. 2, pp. 1681-1686). IEEE.
<https://doi.org/10.1109/irids.2002.1043997>
12. Heath, M., Porter, T.H. and Silvera, G., 2022. Hospital characteristics associated with HIPAA breaches. *International Journal of Healthcare Management*, 15(2), pp.171-180.
<https://doi.org/10.1080/20479700.2020.1870349>
13. Kugan, S., Islam, M.Q.U. and Kashef, R., 2023, May. Decentralized Federated Deep Learning Image Recognition Models. In *2023 4th International Conference on Artificial Intelligence, Robotics and Control (AIRC)* (pp. 1-8). IEEE.
14. [14] Zhang, B., Shi, H. and Wang, H., 2023. Machine learning and AI in cancer prognosis, prediction, and treatment selection: a critical approach. *Journal of multidisciplinary healthcare*, pp.1779-1791.
<https://doi.org/10.2147/jmdh.s410301>
15. Manzoor, S.I., Jain, S., Singh, Y. and Singh, H., 2023. Federated learning based privacy ensured sensor communication in IoT networks: a taxonomy, threats and attacks. *Ieee Access*, 11, pp.42248-42275.
16. Venkateshalu, S.G. and Deshpande, S.L., 2023. Spatial-Spectral Sparse Optimized CNN for Hyper Spectral Image Classification. *International Journal of Intelligent Engineering & Systems*, 16(6).
17. Zhang B, Chen Y, Yao R, Xiong S, Xiong S, Lu X. SSPNet: Spatial-Spectral Perception Network for Mineral Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*. 2025 Oct 6.
18. Sun H, Xu J, Meng F, Cheng M, Cao Q. Spectral-spatial convolutional hybrid Transformer for hyperspectral image classification (January 2025). *IEEE Access*. 2025 Mar 28.
19. Banerjee, A., Swain, S., Rout, M., & Bandyopadhyay, M. (2025). Composite spectral spatial pixel CNN for land-use hyperspectral image classification with hybrid activation function. *Multimedia Tools and Applications*, 84(12), 10527-10550.
20. Zhang J, Qu H, Jia J, Li Y, Jiang B, Chen X, Peng J. Multi-scale Spatial-Spectral CNN-Transformer Network for Hyperspectral Image Super-Resolution. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2025 Apr 30.
21. Sah J, Ghosh DK, Chauhan U. A Review on Hyper Spectral Image Classification using CNN and Deep Learning Techniques. In *2025 International Conference on Intelligent and Secure Engineering Solutions (CISES)* 2025 Aug 11 (pp. 988-993). IEEE.