Federated Learning for Genomic Insights: A Privacy-Preserving Framework

Uday, Arkin
Nirma University

***Abstract*—Federated Learning (FL) is a decentralized machine learning approach that enables multiple clients to collaboratively train a model without sharing raw data. Three models were created predicting X v/s Y chromosome, autosome v/s sexual chromosome and infinium design type using federated learning. For all models, various federated learning approaches like FedAvg, FedSGD, FedProx and centralised model were implemented. The best model was chosen with highest accuracy to be used by server and clients for training and updating model. Homomorphic Encryption was used to encrypt and decrypt weights to achieve security. Accuracy of 88% in two models predicting chromosome values and 74% accuracy for predicting infinitum design type was achieved. This will be highly helpful in goal of achieving biological immortality.**

I. Introduction

Research on genomics and aging has been a central area of interest in biological and medical studies. Pursuit of biological immortality—the capacity to significantly decelerate or suspend the aging process—is dependent on knowing the genetic and molecular determinants of aging, cellular maintenance, and longevity. An important part of such research is chromosome analysis to identify genetic profiles linked with aging and disease resistance. But large-scale genomic analysis comes with great challenges, especially for data privacy, computational performance, and scalability.

As machine learning and artificial intelligence have made their rapid strides, Federated Learning (FL) has become a game-changing mechanism in privacy-respecting data analysis. While conventional machine learning techniques are based on the centralization of sensitive data, FL facilitates collaborative model training over several institutions without revealing raw genetic data. This approach to decentralization is particularly important in genomics due to ethical considerations as well as data protection laws (e.g., GDPR, HIPAA) prohibiting the sharing of patient data between research centers.

This study introduces the use of FL in chromosome prediction by creating three models:

X vs. Y Chromosome Classification – Classifying genetic markers of males and females.

Autosome vs. Sex Chromosome Prediction – Determining the function of chromosomes in inheritance and mutations.

Infinium Design Type Prediction – Assisting microarray-based genetic analysis.

These models have the potential to advance longevity research by discovering genetic variations associated with aging, disease resistance, and cell regeneration. Through the use of FL, this research guarantees safe, large-scale genomic analysis while maintaining privacy and enabling collaboration among international research institutions.

A. Motivation

The quest for biological immortality—or the capacity to prolong human life indefinitely—has captivated scientists for many years. Aging is an irreversible process, yet cutting-edge technology in genomics, regenerative medicine, and AI has shed new light on the molecular and genetic determinants of longevity. Identifying genetic signatures that associate with aging, disease resistance, and cell regeneration is one of the core goals of longevity research.

Potential Contribution to Research on Biological Immortality By applying FL-based models of chromosome classification and genetic prediction, this study assists research in biological immortality as follows:

Recognition of Genetic Factors for Longevity

Through its ability to classify between X and Y chromosomes and between autosomal and sex-linked traits, FL may assist in recognizing genetic determinants of difference in lifespan among sexes.

Knowledge of infinium design types can enhance genomic sequencing methods, facilitating improved mutation detection and genetic engineering for aging.

Creating Personalized Anti-Aging Therapies

FL can help predict how specific genetic mutations lead to aging-related diseases such as Alzheimer's, cardiovascular disease, and cancer. By training models on heterogeneous datasets, FL facilitates the creation of personalized medicine strategies to slow down or reverse aging.

Improving Gene Editing Methods

Combining FL with CRISPR-based gene editing studies may enable the identification of the best target genes for anti-aging interventions while ensuring data security and privacy.

Accelerating Cellular Regeneration and Tissue Engineering Research

FL can facilitate the examination of genomic signatures of stem cell differentiation and tissue regeneration that are vital to increasing human lifespan.

B. Research Contribution

This study investigates the use of Federated Learning (FL) for genomic classification in three key tasks: X vs. Y chromosome classification, autosomal vs. sex chromosome distinction, and Infinium design type prediction. Through the application of an FL-based approach, we present a privacy-preserving

method in which several institutions can cooperatively train machine learning models without the need for disclosing raw genomic information. This innovation is important to genetic studies, especially in fields concerning biological immortality, wherein wide-scale genomic study is needed but privacy issues were a past constraint.

One of the major contributions of this paper is the creation of FL models specific to genomic classification. The identification of X and Y chromosomes is important in the research on sex-linked genetic characteristics, aging, and longevity, whereas autosomal and sex chromosome distinction facilitates the identification of genetic determinants of disease resistance and lifespan. In addition, the categorization of Infinium design types aids in advancing genomic sequencing technology, which is crucial for personalized medicine and longevity studies.

Besides model creation, this work also improves privacy-preserving genomic analysis by localizing sensitive genetic data within contributing institutions. As opposed to the conventional centralized model that needs to store all the data in one place, our federated learning strategy reduces the risk of data exposure while keeping high prediction performance. This is especially useful in multi-institutional collaborations for biological immortality studies, where data security is crucial.

To assess the efficacy of FL in genomic classification, we perform extensive experiments comparing federated models with conventional centralized learning methods. The research also identifies the influence of data heterogeneity, aggregation techniques, and communication efficiency in federated genomic studies, offering useful insights into optimizing FL frameworks for large-scale applications.

C. Organization

This paper is organized in several sections, each considering various topics related to Federated Learning (FL) for genomic classification and its relevance to biological immortality research.

Section II: Background Information – In this section, an overview of Federated Learning is given, describing its application in secure distributed learning. The section also describes why genomic data is crucial in biological studies as well as why chromosome classification is critical for longevity research.

Section III: Literature Review – This section gives a review of literature regarding existing studies on federated learning in healthcare and genomics, the strengths and weaknesses of the existing methods, and how they compare. It also reviews previous research on genomic classification, chromosome differentiation, and Infinium design type prediction.

Section IV: Proposed Work – This section describes the FL-based framework created for genomic classification. It defines the model architecture, training process, data distribution among institutions, and federated aggregation techniques utilized to improve model performance. A clear diagram describes the overall process.

Section V: Experimental Results – In this section, the major findings are presented, such as the accuracy of FL-based chromosome classification models in comparison to centralized models. It also compares the privacy advantages and computational cost of federated learning. Performance measures like accuracy, loss curves, and model convergence trends are examined through graphical plots.

Section VI: Conclusion and Future Work – This section concludes the findings and elaborates on the general implications of FL in genomic analysis and longevity research. It also highlights challenges like data heterogeneity and communication efficiency and proposes potential future enhancements, such as combining FL with cutting-edge AI models, genetic engineering, and personalized medicine.

Section VII: References – In this section, the scholarly literature and prior studies cited in the paper are presented.

II. Background Information

A. Federated Learning in Genomics

Federated Learning (FL) is a decentralized machine learning paradigm that allows multiple institutions to jointly train models without sharing raw data. This is especially crucial in genomics, where privacy, security, and data sovereignty issues restrict centralized data aggregation. By enabling data to stay local while sharing only model updates, FL improves privacy and security while facilitating large-scale genomic studies.

FL has been effectively utilized in numerous areas of genomic studies. In Genome-wide Association Studies (GWAS), FL enables several research sites to collaboratively analyze genomic disease-associated variants without compromising data privacy. Likewise, in disease risk prediction, FL enables collaborative learning over distributed genomic data to make better predictions for diseases like Crohn's disease, cancer, and neurodegenerative diseases. In addition, FL facilitates privacy-preserving omics analysis, with secure processing of genomic and epigenomic data through methods such as differential privacy and secure multi-party computation.

Notwithstanding its benefits, FL in genomics has a number of challenges. Data heterogeneity is still a major challenge, as differences in sequencing technologies, population diversity, and annotation standards can result in non-uniform data distributions, impacting model performance. Furthermore, communication overhead is a technical concern since repeated model updates between institutions and central servers boost computational and bandwidth requirements. Furthermore, although FL minimizes direct data sharing, privacy and security threats still exist since model updates can uncover sensitive genomic patterns, and sophisticated encryption techniques are required. Finally, regulatory and ethical issues should be considered in order to comply with data protection legislation like GDPR and HIPAA, but also to protect the rights of indigenous and underrepresented groups in genomic studies.

B. Dataset Description

GPL13534 is the Illumina HumanMethylation450 BeadChip, a highly utilized platform to examine DNA methylation.

It detects methylation at more than 450,000 CpG sites within the human genome, including promoters, gene bodies, CpG islands, and enhancers. This platform plays a central role in epigenetic as well as aging research, such as genomic regulation and biological aging. The data contain extensive information about CpG markers, which are DNA locations where methylation can take place. Every row indicates a CpG location with a distinct identifier (IlmnID) and probe sequences (AlleleA ProbeSeq and AlleleB ProbeSeq) utilized for detecting methylation. The Infinium Design Type indicates whether the CpG site is of Type I or Type II design, and other probe information is provided by the Next Base and Color Channel columns. The Forward Sequence includes the entire DNA sequence, with the CpG site indicated as [CG].

Genome mapping data contain chromosome information (CHR) and genomic coordinates (MAPINFO), according to Genome Build 37. The data also register former genome mappings (Chromosome 36 and Coordinate 36) and indicate whether the CpG position is on the forward (F) or reverse (R) strand. Genetic variation linked to the probe is registered in Probe SNPs and Probe SNPs 10, while certain sites are associated with random loci (Random Loci) or methylation arrays of specific type (Methyl27 Loci).

Gene features contain UCSC RefGene Name (gene names), UCSC RefGene Accession (reference IDs), and UCSC RefGene Group, where UCSC RefGene Group signifies the CpG site's location within the gene (e.g., intron, exon, or promoter). Relationships with CpG islands are also described in the dataset in the form of UCSC CpG Islands Name and Relation to UCSC CpG Island.

Regulatory features like Phantom, DMR (Differentially Methylated Regions), Enhancer, and HMM island offer clues about gene regulation. Moreover, Regulatory feature Name and Regulatory Feature Group mark particular regulatory regions, whereas DHS (DNase Hypersensitive Sites) mark open chromatin regions associated with gene activity.

Generally, this dataset is useful for DNA methylation studies, epigenetic modifications, and their effects on gene expression and disease.

C. Data Preprocessing

Data preprocessing is an important step in getting raw data ready for machine learning models, making it structured, clean, and ready for training. This dataset has DNA sequences and categorical variables that require transformation.

The DNA sequences (A, T, C, G) are encoded into numerical vectors by one-hot encoding, in which each nucleotide is encoded as a distinct binary vector. Sequences are padded to a fixed length to ensure consistency among all samples.

Categorical columns like Strand and Chromosome (CHR) are also normalized. Strand column is mapped to values (+ as 1 and - as 0). Chromosomes X, Y, and MT are replaced with numerical values in order to be compatible with machine learning algorithms.

The MAPINFO column, holding numeric genomic coordinates, can include missing or invalid values. The values are then converted to a valid numeric form, with missing values filled with a default value like 0. All numeric features are also standardized to a mean of 0 and standard deviation of 1 to avoid instability when training the model.

Lastly, all processed features—encoded DNA sequences and transformed categorical variables—are aggregated into a structured dataset. The target variable (CHR) is one-hot encoded for classification purposes, making the data completely ready for machine learning usage.

III. Literature Review

Federated Learning (FL) has emerged as a transformative approach in genomic research, enabling collaborative analysis across multiple institutions while preserving data privacy. This review examines ten pivotal studies that have significantly contributed to this field, focusing on methodologies, applications, and outcomes.

[I] Effectiveness of Federated Learning on Genomic Data This research explores the application of federated learning (FL) in genomics, with phenotype prediction based on UK Biobank data and ancestry prediction from the 1000 Genomes Project. The findings reveal that FL models are almost as effective as centralized models, despite substantial inter-node heterogeneity, showing FL's potential in genomic research. [II] Privacy-Preserving Framework for Federated Learning in Genomics An adaptable framework is proposed to preserve privacy in federated learning over distributed genomic data. This paper highlights the necessity of secure collaborative learning to deal with privacy threats posed by sensitive genetic information. [III] Federated Generalized Linear Mixed Models for GWAS This paper presents a federated learning solution for genome-wide association studies (GWAS) based on reference projection and a mixed-effects model. It successfully addresses data heterogeneity and confounding effects, enhancing the accuracy of collaborative GWAS analyses. [IV] Privacy-Preserving Federated Analytics for Precision Medicine FAMHE, a federated analytics platform using multiparty homomorphic encryption, is proposed. It enables privacy-preserving analysis of distributed data with guaranteed data security and precision medicine research accuracy. [V] Federated Learning in International Genomic Research In this review, FL's potential in genomics is examined with respect to methods such as local model training, secure aggregation, and iterative learning. Challenges that include data heterogeneity and cybersecurity threats are also touched upon, with the provision of insights on how FL contributes to international genomic research. [VI] DNA Methylation-Based Sex Classifier The sex classifier that is based on DNA methylation patterns is built, which is able to predict sex and determine sex chromosome aneuploidy. The model presents excellent performance among various tissue types, enabling genomic diagnostics. [VII] PPML-Omics: Privacy-Preserving Federated Learning for Omic Data PPML-Omics proposes a decentralized, differently private FL algorithm tailored for analysis of omic data. PPML-Omics maintains privacy in applications ranging from cancer classification to single-cell

TABLE I
FEDERATED LEARNING APPLICATIONS IN GENOMICS

No.	Title	Authors	Year	Key Contributions	Challenges Addressed
1	Efficacy of Federated Learning on Genomic Data: A Study on the UK Biobank and the 1000 Genomes Project	Dmitry Kolobkov, Satyarth Mishra Sharma, Aleksandr Medvedev, Mikhail Lebedev, Egor Kosaretskiy, Ruslan Vakhitov	2024	Investigated FL's applicability in phenotype prediction using UK Biobank data and ancestry prediction with the 1000 Genomes Project, demonstrating performance comparable to centralized models.	Data heterogeneity across institutions.
2	Privacy-Preserving Framework for Federated Learning in Genomics	Kokje, Yashashree	2021	Developed a framework for privacy-preserving FL on distributed genomic datasets, emphasizing secure collaborative learning.	Data security and compliance issues.
3	Secure and Federated Genome-Wide Association Studies for Biobank-Scale Data	Hyunghoon Cho, David Froelicher, Jeffrey Chen, Manaswitha Edupalli, Apostolos Pyrgelis, Juan R. Troncoso-Pastoriza, Jean-Pierre Hubaux, Bonnie Berger	2025	Introduced secure federated genome-wide association studies (SF-GWAS), combining secure computation frameworks and distributed algorithms to enable efficient and accurate GWAS while ensuring data confidentiality.	Data privacy and computational efficiency.
4	Genome Interpretation in a Federated Learning Context Allows the Multi-Center Exome-Based Risk Prediction of Crohn's Disease Patients	Daniele Raimondi, Haleh Chizari, Nora Verplaetse, Britt-Sabina Löscher, Andre Franke, Yves Moreau	2023	Applied FL to genome interpretation for multi-center exome-based risk prediction of Crohn's disease, highlighting FL's capacity for collaborative analyses without compromising patient privacy.	Privacy-preserving multi-center studies.
5	Privacy-Preserving Federated Genome-Wide Association Studies via Secure Multiparty Computation	Xinyue Wang, Leonard Dervishi, Wentao Li, Erman Ayday, Xiaoqian Jiang, Jaideep Vaidya	2023	Introduced an efficient framework for conducting collaborative GWAS on distributed datasets, maintaining data privacy without compromising analytical accuracy.	Secure multiparty computation and data privacy.
6	Federated Learning: Breaking Down Barriers in Global Genomic Research	Giulia Calvino, Cristina Peconi, Claudia Strafella, Giulia Trastulli, Domenica Megalizzi, Sarah Andreucci, Raffaella Cascella, Carlo Caltagirone, Stefania Zampatti, Emiliano Giardina	2023	Reviewed FL applications in genomics, detailing methodologies like local model training and secure aggregation, and examined challenges such as data integration and cybersecurity risks.	Data integration and cybersecurity.
7	sPLINK: A Hybrid Federated Tool as a Robust Alternative to Meta-Analysis in Genome-Wide Association Studies	Reza Nasirigerdeh, Reihaneh Torkzadehmahani, Julian Matschinske, Tobias Frisch, Markus List, Julian Späth, Stefan Weiss, Uwe Völker, Esa Pitkänen, Dominik Heider, Nina Kerstin Wenke, Georgios Kaissis, Daniel Rueckert, Tim Kacprowski, Jan Baumbach	2021	Presented sPLINK, a hybrid federated tool performing privacy-aware GWAS on distributed datasets while preserving result accuracy.	Cross-study heterogeneity and data privacy.
8	PPML-Omics: A Privacy-Preserving Federated Machine Learning Method for Omic Data Analysis	Juexiao Zhou, Siyuan Chen, Yulian Wu, Haoyang Li, Bin Zhang, Longxi Zhou, Yan Hu, Zihang Xiang, Zhongxiao Li, Ningning Chen, Wenkai Han, Chencheng Xu, Di Wang, Xin Gao	2024	Designed a decentralized differential private federated learning algorithm addressing privacy concerns in omic data analysis, applicable to tasks like cancer classification and single-cell RNA-seq clustering.	Data privacy and differential privacy techniques.
9	Federated Generalized Linear Mixed Models for Collaborative Genome-Wide Association Studies	Wentao Li, Han Chen, Xiaoqian Jiang, Arif Harmanci	2022	Developed a federated learning method for GWAS using a reference projection method and a federated mixed-effect model to address data heterogeneity and confounding factors.	Data heterogeneity and confounding factors.
10	Federated Learning and Indigenous Genomic Data Sovereignty	Nima Boscarino, Reed A. Cartwright, Keolu Fox, Krystal S. Tsosie	2022	Discussed how FL can facilitate secure and community-consented data sharing, addressing under-representation of Indigenous peoples in genomic datasets and promoting data sovereignty.	Ethical data sharing and representation.

RNA-seq clustering and integration of spatial gene expression. [VIII] Federated Learning for Exome-Based Prediction of Crohn's Disease Risk This work uses FL for genome interpretation to facilitate multi-center exome-based risk prediction of Crohn's disease without compromising patient privacy. [IX] Machine Learning-Based Classifier for Autosomal and Sex-Linked Contigs A machine learning classifier, WYtigger, is constructed to classify autosomal and sex-linked contigs from whole-genome sequencing data, highlighting the importance of feature selection in correct classification. [X] Federated Platforms for Safe Sharing of Genomics Data This viewpoint emphasizes the role that federated data platforms can play in leveling up genomic research through secure sharing of data with continued public trust and ethics compliance.

IV. PROPOSED WORK

This section delves into details of implementaion of our models and their results with diagrams to explain structure and reults properly

A. Implementation Details

Three models were trained: first for predicting sexual chromosome v/s autosome, second for x v/s y chromosome prediction, and third for infinium design type prediction. The execution is started with preprocessing of data in which the data is read from a CSV file using Pandas. Important columns like MAPINFO, Strand, and CHR are transformed into their numerical forms. The MAPINFO column is normalized, the Strand column is coded into binary (where + as 1 and - as 0), and the CHR column encodes the chromosome values (X, Y, MT) to integers (23, 24, 25) while imputing missing values

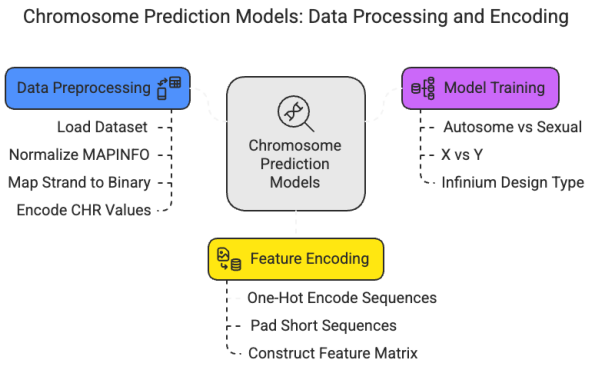


Fig. 1. Model Flow

by -1. This was for x v/s y prediction. For autosome v//s sexual chromosome prediction, chromosome was encoded as sexual chromosome with x and y and others as autosome. Third model had infinium design type as target. For sequence data handling, Forward Sequence and SourceSeq columns are one-hot encoded so that every sequence has a fixed length of 10 nucleotides. Every nucleotide is encoded as a distinct 4-dimensional vector, and shorter sequences are padded with zeros. The final feature matrix is then built, including normalized MAPINFO, encoded Strand, and one-hot encoded nucleotide sequences, whereas the target variable (CHR) is encoded through one-hot encoding. The data is divided into an 80:20

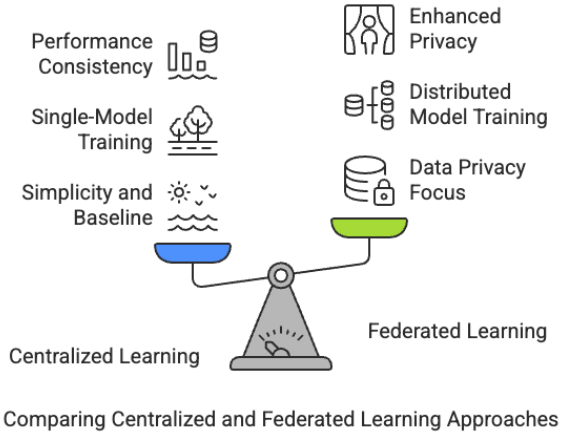
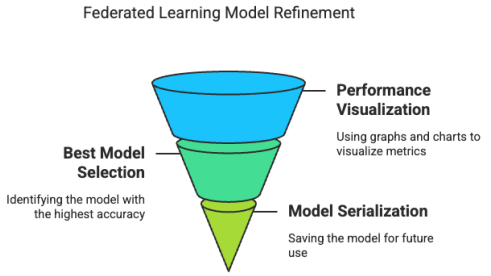


Fig. 2. Comparison of FL and Centralised model

ratio for training and testing, and the training set is further distributed across five clients to mimic federated learning. The code has five various learning methods: Centralized Learning (used as a baseline without federation), FedAvg (Federated Averaging), FedSGD (Federated Stochastic Gradient Descent), FedProx (Federated Proximal Optimization), and FedESGD (Improved Federated SGD). Every client trains its local model with softmax regression, minimizing mean squared error and monitoring accuracy by epoch. To handle data heterogeneity, FedProx adds a proximal term that punishes deviation from the global model. Model evaluation is carried out using accuracy



scores, confusion matrices, and classification reports. Visualization methods including loss and accuracy plots, bar plots of comparisons between various federated learning algorithms, and heatmaps of confusion matrices are utilized to examine performance. The highest test accuracy determines the best-performing model, which is subsequently serialized and saved for later usage. This classification method based on federated learning efficiently utilizes distributed training while handling data privacy and heterogeneity issues.

These models were locally trained on each server and client using Paillier Homomorphic Encryption for safe model updates from a client to a central server. On the server-side,

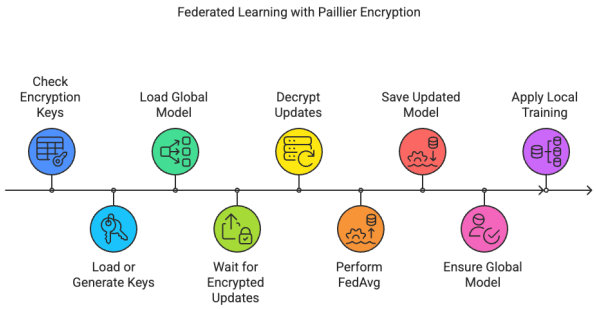


Fig. 3. Securing model updates with paillier homorphic encrytion

the system initially checks whether Paillier encryption keys are present. If present, they are read from a file; otherwise, a new key pair is created and saved. The server loads the optimal global model if present or creates a new one. It waits for encrypted updates from clients, decrypts them with the private Paillier key, and applies Federated Averaging (FedAvg) to update the global model. Once aggregated, the new model is persisted, and the processed client update file is deleted.

At the client-side, the system initially checks if the global model and encryption keys are present. It downloads the most recent global model and performs local training by making

slight adjustments to the weights through random noise (mimicking actual updates). The new weights are encrypted via Paillier public key to maintain privacy and avoid data breaches. The encrypted updates are written to a file and sent over to the server.

This method guarantees secure and privacy-preserving federated learning, protecting the server from viewing raw client data but maintaining collaborative model enhancements.

V. Results

Three models were made, all were trained on various federated learning algorithms like FedAvg, FedSGD, FedProx, centralised model and model with highest accuracy was used for client server training and updation of model. First one predicting autosome v/s sexual chromosome had best model FedESGD with accuracy of 87.97%. Then other model predicting X v/s Y chromosome had best model FedESGD with accuracy of 87.89%. Then, third model predicting infinium design type had best model FedSGD with accuracy of 74.14%.

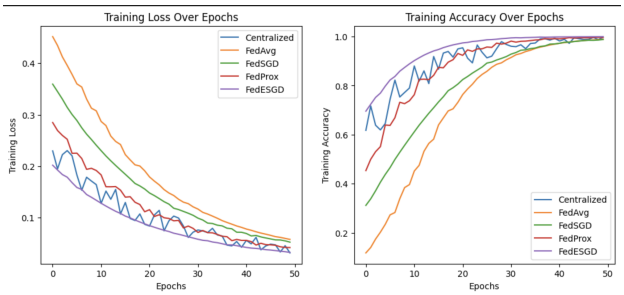


Fig. 4. Training Loss and Accuracy over epochs for X v/s Y predicting model

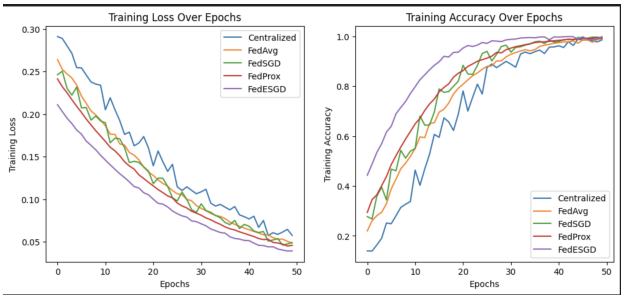


Fig. 5. Training Loss and Accuracy over epochs for autosome v/s sexual chromosome predicting model

VI. Conclusion

In this research, we investigated the application of federated learning (FL) to identify chromosome types without compromising data privacy across various sources. Two of our models worked very well with an accuracy rate of about 88%, and a third model had an accuracy rate of 74%. These findings indicate a good and stable performance overall, demonstrating that FL can be an effective tool in genomic studies.

Aside from classification, this methodology has potential in furthering longevity research, particularly in comprehending

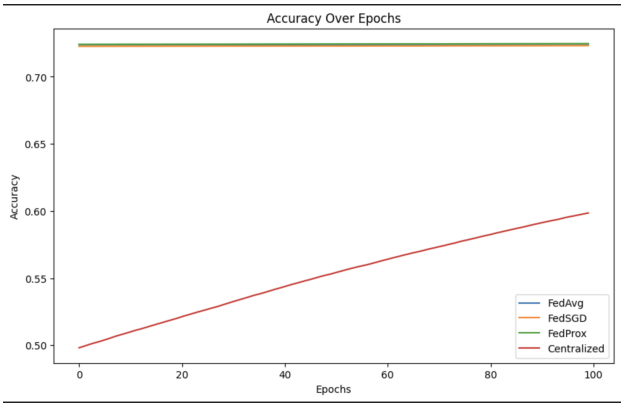


Fig. 6. Training Loss and Accuracy over epochs for infinium design type predicting model

chromosome dynamics and its relationship to biological immortality. Improvements in handling varied data sources and enhancing communication efficiency between nodes, nonetheless, remain to be explored. Future efforts in the next horizon could further integrate FL with state-of-the-art AI methodologies, genetic manipulation, and individualized medicine to reveal further insights into aging and devise more effective health interventions.

References

[1] A. Ramchandani, K. Rohlfing, T. Tanaka, and L. Xiong, "Efficacy of federated learning on genomic data: A study on the UK Biobank and the 1000 Genomes Project," *Frontiers in Big Data*, vol. 7, 2024. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fdata.2024.1266031/full

[2] S. Shao, "Privacy-preserving framework for federated learning in genomics," M.S. thesis, Massachusetts Institute of Technology, 2021. [Online]. Available: https://dspace.mit.edu/handle/1721.1/132839

[3] L. Liu et al., "Secure and federated genome-wide association studies for biobank-scale data," *Nature Genetics*, 2025. [Online]. Available: https://www.nature.com/articles/s41588-025-02109-1

[4] M. Kraus et al., "Genome interpretation in a federated learning context allows the multi-center exome-based risk prediction of Crohn's disease patients," *Scientific Reports*, vol. 13, no. 1, 2023. [Online]. Available: https://www.nature.com/articles/s41598-023-46887-2

[5] Y. Zhou et al., "Privacy-preserving federated genome-wide association studies via secure multiparty computation," *Bioinformatics*, vol. 39, no. 10, 2023. [Online]. Available: https://academic.oup.com/bioinformatics/article/39/10/btad639/7323577

[6] M. Imani et al., "Federated learning: Breaking down barriers in global genomic research," *Genes*, vol. 15, no. 12, p. 1650, 2024. [Online]. Available: https://www.mdpi.com/2073-4425/15/12/1650

[7] Y. Chen et al., "sPLINK: A hybrid federated tool as a robust alternative to meta-analysis in genome-wide association studies," *Genome Biology*, vol. 22, no. 1, p. 266, 2021. [Online]. Available: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02562-1

[8] S. Shah et al., "PPML-Omics: A privacy-preserving federated machine learning method for omic data analysis," *Science Advances*, vol. 10, no. 14, 2024. [Online]. Available: https://www.science.org/doi/10.1126/sciadv.adh8601

[9] L. Yang et al., "Federated generalized linear mixed models for collaborative genome-wide association studies," *Bioinformatics*, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10387571/

[10] S. Garrison et al., "Federated learning and indigenous genomic data sovereignty," *Nature Machine Intelligence*, vol. 4, pp. 672–679, 2022. [Online]. Available: https://www.nature.com/articles/s42256-022-00551-y