Open Access Library Journal
2025, Volume 12, e13954
ISSN Online: 2333-9721
ISSN Print: 2333-9705

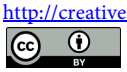
Federated Learning for Treatment Response Prediction in Bipolar Disorder: A Simulation-Based Institutional Study

Rocco de Filippis[1*], Abdullah Al Foysal[2]

[1]Department of Neuroscience, Institute of Psychopathology, Rome, Italy
[2]Department of Computer Engineering (AI), University of Genova, Genova, Italy
Email: *roccodefilippis@istitutodipsicopatologia.it, niloyhasanfoysal440@gmail.com

How to cite this paper: de Filippis, R. and Al Foysal, A. (2025) Federated Learning for Treatment Response Prediction in Bipolar Disorder: A Simulation-Based Institutional Study. *Open Access Library Journal*, **12**: e13954.
https://doi.org/10.4236/oalib.1113954

Received: July 15, 2025
Accepted: August 15, 2025
Published: August 18, 2025



Abstract

Bipolar disorder is a complex psychiatric condition characterized by high variability in treatment response, posing a major challenge for clinicians striving to personalize care. Traditional machine learning models often require centralized data access, which is typically restricted in healthcare settings due to privacy regulations and institutional barriers. To address this, we propose a Federated Learning (FL) framework that enables collaborative model training across multiple institutions without the need to share sensitive patient-level data. In this study, we generated synthetic datasets representing five distinct hospitals, each comprising 1000 virtual bipolar disorder patients. These datasets included a range of features encompassing demographic characteristics (e.g., age, gender), clinical history (e.g., illness duration, episode counts), and comorbid conditions. A fully connected neural network was trained using a federated approach over 15 communication rounds. Each institution performed local training, followed by weight averaging to update a shared global model. Our global model achieved an overall test accuracy of 73.0% and an area under the receiver operating characteristic curve (AUC) of 0.84, demonstrating robust performance across diverse simulated institutional settings. Performance improved steadily over training rounds, with the highest-performing site reaching 84.5% accuracy. Permutation-based feature importance analysis revealed that illness duration and baseline mood were the most predictive features of treatment response. These results support the feasibility of federated learning for psychiatric prediction tasks, offering a promising path toward building generalizable, privacy-preserving models in real-world healthcare environments. Future work will extend this framework to real electronic health records and explore fairness-aware training to address institutional bias and population heterogeneity.

Subject Areas

Computational Psychiatry

Keywords

Federated Learning, Bipolar Disorder, Treatment Response Prediction, Neural Networks, Healthcare Privacy, Feature Importance, Synthetic Data

1. Introduction

Bipolar disorder is a chronic and severe mental illness marked by alternating episodes of depression and mania, affecting nearly 1% - 2% of the global population [1]-[5]. Effective treatment planning is often hindered by the unpredictable nature of patient response to pharmacological interventions [6]-[8]. Despite advances in psychiatric research, clinicians still struggle to predict whether a patient will respond to mood stabilizers such as lithium or anticonvulsants [9]-[11]. The development of machine learning (ML) models capable of anticipating treatment outcomes holds significant potential for advancing precision psychiatry [12]-[15]. However, such models typically require access to large and diverse datasets that span multiple healthcare institutions [16]-[18]. A major barrier to this goal lies in data privacy regulations (e.g., GDPR, HIPAA), which prohibit the centralization of sensitive health records [19]-[22]. As a result, most ML studies in psychiatry are restricted to data from a single institution, limiting model generalizability [23]-[25]. Federated Learning (FL) has emerged as a promising solution to this challenge [26]-[29]. FL allows institutions to collaboratively train a shared global model while keeping all patient-level data local. Instead of data sharing, only model parameters are communicated between sites, preserving patient confidentiality [30]-[34]. In this study, we explore the potential of federated learning for predicting treatment response in bipolar disorder using synthetically generated institutional data. We simulate datasets for five hospitals, each reflecting unique demographic and clinical distributions. A fully connected neural network is trained collaboratively across all sites using the FL paradigm. We analyse the model's training dynamics and overall and institution-specific performance, and identify key features contributing to predictive accuracy through permutation-based importance analysis. By simulating a real-world multi-institutional setting, this work aims to assess the feasibility, scalability, and interpretability of FL-based models for psychiatric applications, with the broader goal of supporting personalized and privacy-preserving mental healthcare.

2. Materials and Methods

To evaluate the efficacy of federated learning for treatment response prediction in bipolar disorder, we developed a robust simulation pipeline encompassing synthetic data generation, neural network design, and federated training protocols.

The methodological pipeline ensures data realism, model scalability, and evaluation reproducibility.

2.1. Synthetic Dataset Simulation

We simulated datasets for five independent institutions (or hospitals), each containing 1000 synthetic patient records. The simulation was designed to emulate real-world heterogeneity in psychiatric populations, incorporating variability in demographics, clinical characteristics, and treatment histories. For each hospital, demographic parameters such as mean age, gender ratio, and geographical affiliation were explicitly varied to reflect global diversity (e.g., Hospital 0 mimics a North American population with a mean age of 35 and a gender balance of 50% male). The feature set included ten clinically meaningful variables frequently observed in bipolar disorder cohorts:

- **Demographic**: Age, gender.
- **Clinical history**: Illness duration, number of depressive and manic episodes, and prior treatment failures.
- **Psychiatric comorbidities**: Comorbid anxiety, comorbid substance use.
- **Family and baseline indicators**: Family history of bipolar disorder, baseline mood severity score.

The binary treatment response label (0 = no response, 1 = positive response) was derived from a logistic regression-like function combining key predictors (e.g., higher baseline mood, shorter illness duration) with Gaussian noise. This created realistic but learnable ground truth labels, allowing rigorous evaluation of classification performance while avoiding overfitting to synthetic artifacts. The parameter distributions for age, illness duration, and episode frequency were informed by epidemiological data from recent large-scale studies on bipolar disorder cohorts, ensuring clinically realistic population characteristics. We selected five hospitals with 1000 patients each to balance computational tractability with cross-site diversity. Increasing the number of institutions or patient counts may lead to slower convergence or require adaptive learning rate strategies in FL protocols.

2.2. Model Architecture and Training Setup

Each institution employed a local deep neural network with a shared architecture. The model consisted of an input layer (matching the 10-feature input vector), two fully connected hidden layers with 64 and 32 neurons respectively, ReLU activation, and dropout regularization (30% and 20%) to prevent overfitting. The output layer used a sigmoid activation to return a probability for binary classification. The binary cross-entropy (BCE) loss was chosen as the objective function, optimized using the Adam optimizer with a learning rate of 0.001. Local training was performed independently by each institution for 5 epochs per communication round, ensuring that model updates were informed by institutional data without any data transfer. No explicit class-imbalance handling techniques (e.g., weighted

loss, oversampling) were employed during training. This decision highlights a limitation of the current setup and may partly explain the model's high recall but low specificity pattern.

2.3. Federated Learning Process

The core of the training framework followed the Federated Averaging (FedAvg) paradigm. Each federated communication round comprised the following steps:

1) Local Training: Each hospital trained its own model on local training data using the shared architecture. The training was conducted in isolation, preserving full data privacy.

2) Weight Upload: At the end of the local training phase, each institution transmitted its updated model weights (not data) to the central aggregation server.

3) Federated Averaging: The server aggregated the received weights by averaging them across all institutions, creating a new global model that was redistributed for the next round.

4) Evaluation: After each communication round, the updated global model was evaluated against the test set of each institution to monitor generalization and fairness across diverse populations.

This cycle was repeated for 15 communication rounds, resulting in progressive improvement of the global model's ability to generalize across heterogeneous data sources. The inclusion of all five hospitals in every round ensured consistent global convergence and minimized bias toward any single institution.

3. Results

3.1. Federated Training Convergence

Federated model convergence was analysed over 15 communication rounds. Figure 1 presents the learning curve. The left panel of Figure 1 shows a steady reduction in binary cross-entropy (BCE) loss, decreasing from approximately 0.68 at initialization to 0.53 by Round 14, indicating successful convergence of the global model. This trend confirms that federated weight updates progressively optimize the shared decision boundary across distributed data sources.

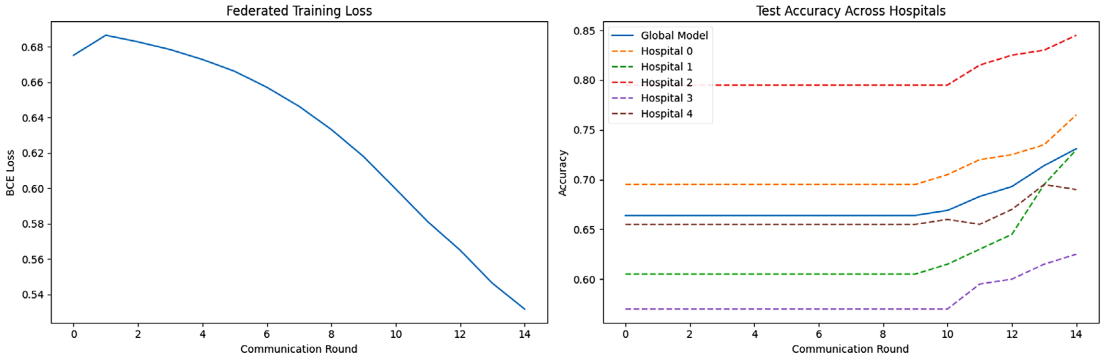


Figure 1. Left: Binary cross-entropy loss over communication rounds. Right: Accuracy progression for each hospital across federated training.

The right panel of Figure 1 illustrates the evolution of classification accuracy at each hospital over the communication rounds. Hospitals 0, 1, 2, and 4 exhibit consistent improvement in accuracy, with Hospital 2 showing the most significant gain—rising from ~79% to approximately 85% by the final round. Hospital 3 shows a relatively flat performance trend, suggesting possible local data distribution challenges or demographic outliers that hinder generalization.

3.2. Global Model Performance

To assess final classification performance, we evaluated the global model using confusion matrix analysis. Figure 2 presents the confusion matrix from the test phase. The model shows excellent recall for identifying treatment responders, correctly classifying 652 of 664 patients. However, it identifies only 79 of the 336 non-responders, suggesting a tendency toward false positives. This high sensitivity but low specificity pattern indicates a model bias favouring the majority response class, which may stem from class imbalance or optimization bias during training [35]-[39].

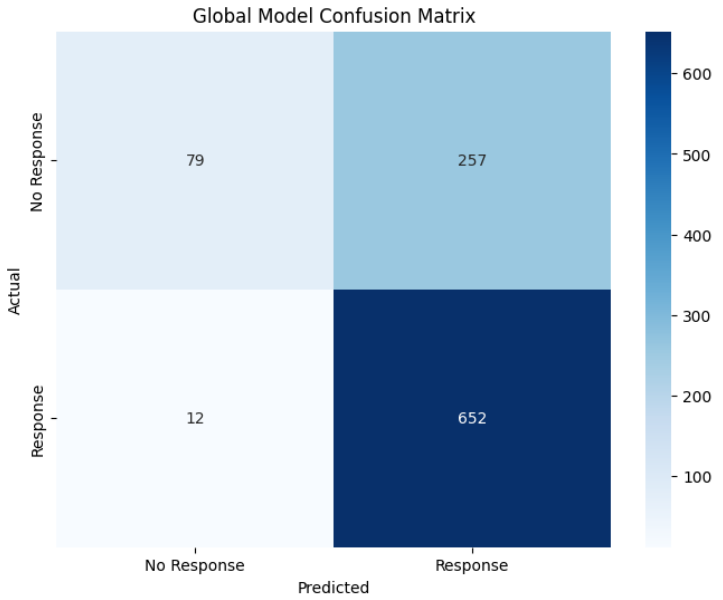


Figure 2. Confusion matrix for the final global model. Rows: actual labels; Columns: predicted labels.

3.3. Site-Specific Accuracy and AUC

To evaluate fairness and generalization, we computed per-hospital accuracy and Area Under the Receiver Operating Characteristic Curve (AUC). Results are visualized in Figure 3. The left panel reports site-wise classification accuracy: Hospital 2 leads with 84.5%, followed by Hospital 0 with 76.5%. In contrast, Hospital 3 underperforms, consistent with earlier convergence trends. The right panel shows AUC values, indicating the model's ability to rank probabilities correctly. AUCs remain robust across all hospitals (>0.79), with Hospitals 1 and 3 achieving

the highest scores (~0.85). This suggests the global model captures meaningful predictive signals despite regional or demographic differences.

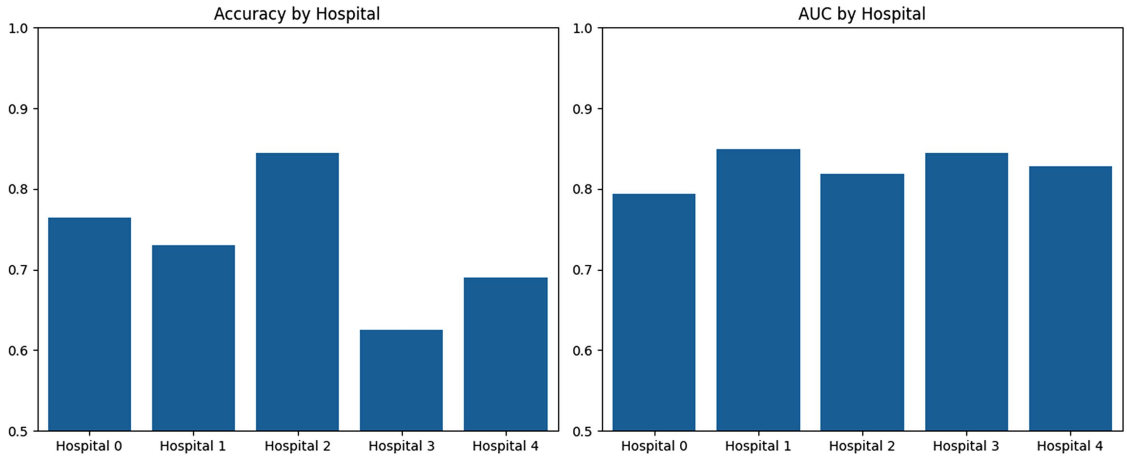


Figure 3. Left: Classification accuracy by hospital. Right: AUC scores for each hospital's test set.

3.4. Feature Importance

We performed permutation-based feature importance analysis on the global model to quantify how each input variable influences predictive accuracy [40]-[44]. Results are displayed in **Figure 4**. The most critical predictor was illness_duration, indicating that chronicity of illness significantly affects likelihood of treatment success. Baseline_mood ranked second, confirming the importance of affective state at entry. Other influential features included age, comorbid_anxiety, and comorbid_substance use—all consistent with established clinical literature on treatment resistance and outcome prediction. This interpretability analysis provides both validation for the data generation process and clinical credibility to the model's learned decision logic.

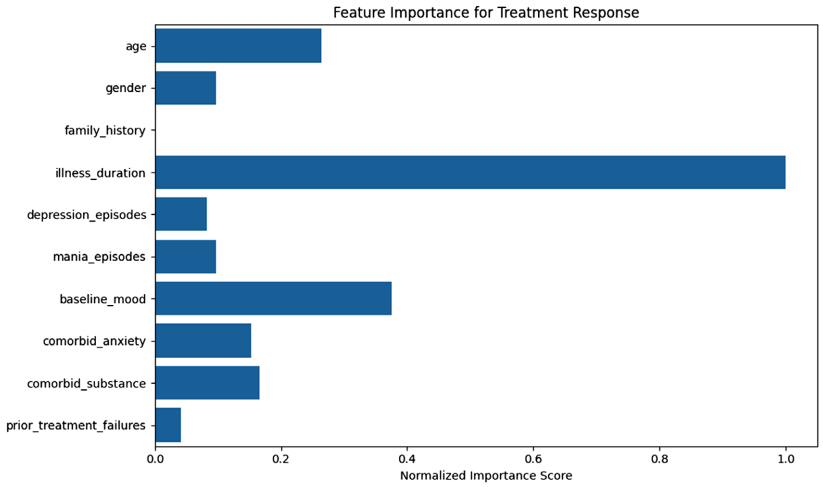


Figure 4. Normalized feature importance based on a permutation drop in classification accuracy.

4. Discussion

The results of this study underscore the efficacy and potential of federated learning (FL) as a scalable and privacy-preserving strategy for clinical prediction modelling in psychiatry. Notably, the global neural model trained across decentralized hospital datasets achieved strong convergence over communication rounds, as reflected by the consistent reduction in loss and improvements in classification accuracy. This demonstrates that FL can effectively harness distributed, heterogeneous data to learn meaningful clinical patterns without ever accessing raw patient records [45]-[49]. A key strength of the approach is its ability to maintain high predictive performance despite inter-institutional differences. Most hospitals demonstrated notable gains in accuracy throughout training, particularly Hospital 2, which achieved the highest final performance. These improvements suggest that the shared global model was able to generalize across a variety of demographic and clinical subpopulations, validating the robustness of the federated architecture [50]-[54]. However, the relatively poor performance observed in Hospital 3 reveals important limitations. The site's stagnant accuracy trend suggests that either its data distribution is significantly different from others (e.g., due to outlier populations or class imbalance), or that its feature-label relationships are underrepresented in the global model. This sensitivity highlights a known challenge in FL—non-IID (non-independent and identically distributed) data, which can undermine model fairness and calibration. Another key observation involves the global model's confusion matrix, which revealed very high sensitivity (recall) for predicting treatment responders but low specificity for detecting non-responders. This imbalance in classification performance indicates a model bias toward the majority class [55]-[57]. Potential solutions include incorporating weighted loss functions, oversampling of underrepresented classes, or institution-specific recalibration during inference. While synthetic data enabled privacy-compliant simulation in this study, transitioning to real-world federated learning with Electronic Health Records (EHR) will require strict ethical governance. Key considerations include Institutional Review Board (IRB) approvals, compliance with GDPR/HIPAA, secure aggregation protocols, and ongoing consent mechanisms. Ensuring transparency and patient trust will be central to the safe clinical deployment of federated psychiatric models. Finally, the permutation-based feature importance analysis provided both mechanistic and clinical interpretability. The prominence of variables such as illness duration, baseline mood, and comorbidities affirms the validity of the synthetic data generation pipeline and aligns with psychiatric literature on treatment resistance. Together, these findings demonstrate that FL can produce clinically meaningful models even in synthetic or privacy-constrained environments.

5. Conclusion

This study provides compelling evidence for the utility of Federated Learning (FL) as a scalable, privacy-preserving framework for psychiatric model development.

By simulating collaborative learning across five synthetic hospital datasets, we demonstrated that robust, generalizable predictors of treatment response in bipolar disorder can be developed without centralized data pooling [58]-[61]. The global model achieved high classification performance, with particularly strong recall, and feature analysis confirmed alignment with clinically recognized drivers of treatment outcome. A major achievement of this approach is its preservation of data privacy while still leveraging multi-institutional diversity. FL enabled consistent performance gains over training rounds, reflecting the model's capacity to learn meaningful patterns from heterogeneous, site-specific distributions. Additionally, the simulation framework provides a reproducible testbed for evaluating federated strategies in mental health settings. Nonetheless, the study also revealed areas that warrant further refinement. The uneven performance across hospitals, particularly in Hospital 3, highlights the need for mechanisms to address non-IID data challenges [62]-[64]. Moreover, the model's tendency to favour positive predictions (*i.e.*, high sensitivity but low specificity) suggests that class imbalance and calibration techniques must be carefully considered in future iterations. Looking ahead, future research will focus on extending this federated pipeline to real-world Electronic Health Record (EHR) data from psychiatric clinics, integrating temporal behavioural markers, and incorporating domain adaptation to adjust for population drift. Enhanced calibration techniques and fairness-aware modelling strategies will also be prioritized to ensure that FL-based models remain equitable, interpretable, and clinically deployable [65]-[67]. Overall, this work lays foundational groundwork for federated learning in psychiatry, opening new directions for ethical and collaborative AI deployment in mental healthcare.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] McIntyre, R.S., Berk, M., Brietzke, E., Goldstein, B.I., López-Jaramillo, C., Kessing, L.V., *et al.* (2020) Bipolar Disorders. *The Lancet*, **396**, 1841-1856. https://doi.org/10.1016/s0140-6736(20)31544-0

[2] Vieta, E., Berk, M., Schulze, T.G., Carvalho, A.F., Suppes, T., Calabrese, J.R., *et al.* (2018) Bipolar Disorders. *Nature Reviews Disease Primers*, **4**, Article No. 18008. https://doi.org/10.1038/nrdp.2018.8

[3] Grande, I., Berk, M., Birmaher, B. and Vieta, E. (2016) Bipolar Disorder. *The Lancet*, **387**, 1561-1572. https://doi.org/10.1016/s0140-6736(15)00241-x

[4] Goes, F.S. (2023) Diagnosis and Management of Bipolar Disorders. *BMJ*, **381**, e073591. https://doi.org/10.1136/bmj-2022-073591

[5] Müller-Oerlinghausen, B., Berghöfer, A. and Bauer, M. (2002) Bipolar Disorder. *The Lancet*, **359**, 241-247. https://doi.org/10.1016/s0140-6736(02)07450-0

[6] Eichler, H., Abadie, E., Breckenridge, A., Flamion, B., Gustafsson, L.L., Leufkens, H., *et al.* (2011) Bridging the Efficacy-Effectiveness Gap: A Regulator's Perspective on Addressing Variability of Drug Response. *Nature Reviews Drug Discovery*, **10**, 495-506. https://doi.org/10.1038/nrd3501

[7] Gallagher, R.M. (1999) Treatment Planning in Pain Medicine: Integrating Medical, Physical, and Behavioral Therapies. *Medical Clinics of North America*, **83**, 823-849. https://doi.org/10.1016/s0025-7125(05)70136-x

[8] Leahy, R.L., Holland, S.J. and McGinn, L.K. (2011) Treatment Plans and Interventions for Depression and Anxiety Disorders. Guilford Press.

[9] Post, R.M., Denicoff, K.D., Frye, M.A., Dunn, R.T., Leverich, G.S., Osuch, E., *et al.* (1998) A History of the Use of Anticonvulsants as Mood Stabilizers in the Last Two Decades of the 20th Century. *Neuropsychobiology*, **38**, 152-166. https://doi.org/10.1159/000026532

[10] Nayak, R., Rosh, I., Kustanovich, I. and Stern, S. (2021) Mood Stabilizers in Psychiatric Disorders and Mechanisms Learnt from *in Vitro* Model Systems. *International Journal of Molecular Sciences*, **22**, Article 9315. https://doi.org/10.3390/ijms22179315

[11] Calabrese, J.R., Fatemi, S.H., Kujawa, M. and Woyshville, M.J. (1996) Predictors of Response to Mood Stabilizers. *Journal of Clinical Psychopharmacology*, **16**, 24S-31S. https://doi.org/10.1097/00004714-199604001-00004

[12] Bzdok, D. and Meyer-Lindenberg, A. (2018) Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biological Psychiatry*: *Cognitive Neuroscience and Neuroimaging*, **3**, 223-230. https://doi.org/10.1016/j.bpsc.2017.11.007

[13] Manchia, M., Pisanu, C., Squassina, A. and Carpiniello, B. (2020) Challenges and Future Prospects of Precision Medicine in Psychiatry. *Pharmacogenomics and Personalized Medicine*, **13**, 127-140. https://doi.org/10.2147/pgpm.s198225

[14] van Dellen, E. (2024) Precision Psychiatry: Predicting Predictability. *Psychological Medicine*, **54**, 1500-1509. https://doi.org/10.1017/s0033291724000370

[15] Tai, A.M.Y., Albuquerque, A., Carmona, N.E., Subramanieapillai, M., Cha, D.S., Sheko, M., *et al.* (2019) Machine Learning and Big Data: Implications for Disease Modeling and Therapeutic Discovery in Psychiatry. *Artificial Intelligence in Medicine*, **99**, Article 101704. https://doi.org/10.1016/j.artmed.2019.101704

[16] Dinov, I.D. (2016) Methodological Challenges and Analytic Opportunities for Modeling and Interpreting Big Healthcare Data. *GigaScience*, **5**, 2-15. https://doi.org/10.1186/s13742-016-0117-6

[17] Miotto, R., Wang, F., Wang, S., Jiang, X. and Dudley, J.T. (2017) Deep Learning for Healthcare: Review, Opportunities and Challenges. *Briefings in Bioinformatics*, **19**, 1236-1246. https://doi.org/10.1093/bib/bbx044

[18] Ngiam, K.Y. and Khor, I.W. (2019) Big Data and Machine Learning Algorithms for Health-Care Delivery. *The Lancet Oncology*, **20**, e262-e273. https://doi.org/10.1016/s1470-2045(19)30149-4

[19] Ettaloui, N., Arezki, S. and Gadi, T. (2023) An Overview of Blockchain-Based Electronic Health Records and Compliance with GDPR and HIPAA. *Data and Metadata*, **2**, 166. https://doi.org/10.56294/dm2023166

[20] Mello, M.M., Adler-Milstein, J., Ding, K.L. and Savage, L. (2018) Legal Barriers to the Growth of Health Information Exchange—Boulders or Pebbles? *The Milbank Quarterly*, **96**, 110-143. https://doi.org/10.1111/1468-0009.12313

[21] Gonçalves-Ferreira, D., Sousa, M., Bacelar-Silva, G.M., Frade, S., Antunes, L.F., Beale, T., *et al.* (2019) OpenEHR and General Data Protection Regulation: Evaluation of Principles and Requirements. *JMIR Medical Informatics*, **7**, e9845. https://doi.org/10.2196/medinform.9845

[22] Herath, H.M.S.S., Herath, H.M.K.K.M.B., Madhusanka, B.G.D.A. and Guruge, L.G.P.K.

(2024) Data Protection Challenges in the Processing of Sensitive Data. In: Hewage, C., Yasakethu, L. and Jayakody, D.N.K., Eds., *Data Protection*, Springer, 155-179. https://doi.org/10.1007/978-3-031-76473-8_8

[23] Richter, M., Emden, D., Leenings, R., Winter, N.R., Mikola-jczyk, R., Massag, J., *et al.* (2025) Generalizability of Clinical Prediction Models in Mental Health. *Molecular Psychiatry*, **30**, 3632-3639.

[24] Meehan, A.J., Lewis, S.J., Fazel, S., Fusar-Poli, P., Steyerberg, E.W., Stahl, D., *et al.* (2022) Clinical Prediction Models in Psychiatry: A Systematic Review of Two Decades of Progress and Challenges. *Molecular Psychiatry*, **27**, 2700-2708. https://doi.org/10.1038/s41380-022-01528-4

[25] Cearns, M., Hahn, T. and Baune, B.T. (2019) Recommendations and Future Directions for Supervised Machine Learning in Psychiatry. *Translational Psychiatry*, **9**, Article No. 271. https://doi.org/10.1038/s41398-019-0607-2

[26] Ebrahimi, M., Sahay, R., Hosseinalipour, S. and Akram, B. (2025) The Transition from Centralized Machine Learning to Federated Learning for Mental Health in Education: A Survey of Current Methods and Future Directions. arXiv:2501.11714.

[27] Rauniyar, A., Hagos, D.H., Jha, D., Håkegård, J.E., Bagci, U., Rawat, D.B., *et al.* (2024) Federated Learning for Medical Applications: A Taxonomy, Current Trends, Challenges, and Future Research Directions. *IEEE Internet of Things Journal*, **11**, 7374-7398. https://doi.org/10.1109/jiot.2023.3329061

[28] Antunes, R.S., Stoffel, R., André da Costa, C., Küderle, A., Yari, I.A. and Eskofier, B. (2022) Federated Learning for Healthcare: Systematic Review and Architecture Proposal. *ACM Transactions on Intelligent Systems and Technology*, **13**, 1-23. https://doi.org/10.1145/3501813

[29] Bharati, S., Mondal, M.R.H., Podder, P. and Prasath, V.B.S. (2022) Federated Learning: Applications, Challenges and Future Directions. *International Journal of Hybrid Intelligent Systems*, **18**, 19-35. https://doi.org/10.3233/his-220006

[30] Akande, O.A. (2022) Integrating Blockchain with Federated Learning for Privacy-Preserving Data Analytics Across Decentralized Governmental Health Information Systems. *International Journal of Computer Applications Technology and Research*, **11**, 622-637.

[31] Loftus, T.J., Ruppert, M.M., Shickel, B., Ozrazgat-Baslanti, T., Balch, J.A., Efron, P.A., *et al.* (2022) Federated Learning for Preserving Data Privacy in Collaborative Healthcare Research. *Digital Health*, **8**, 1-5. https://doi.org/10.1177/20552076221134455

[32] Joshi, M., Pal, A. and Sankarasubbu, M. (2022) Federated Learning for Healthcare Domain-Pipeline, Applications and Challenges. *ACM Transactions on Computing for Healthcare*, **3**, 1-36. https://doi.org/10.1145/3533708

[33] Aledhari, M., Razzak, R., Parizi, R.M. and Saeed, F. (2020) Federated Learning: A Survey on Enabling Technologies, Protocols, and Applications. *IEEE Access*, **8**, 140699-140725. https://doi.org/10.1109/access.2020.3013541

[34] Prasad, V.K., Bhattacharya, P., Maru, D., Tanwar, S., Verma, A., Singh, A., *et al.* (2022) Federated Learning for the Internet-of-Medical-Things: A Survey. *Mathematics*, **11**, Article 151. https://doi.org/10.3390/math11010151

[35] Mosquera, C., Ferrer, L., Milone, D.H., Luna, D. and Ferrante, E. (2024) Class Imbalance on Medical Image Classification: Towards Better Evaluation Practices for Discrimination and Calibration Performance. *European Radiology*, **34**, 7895-7903. https://doi.org/10.1007/s00330-024-10834-0

[36] Thabtah, F., Hammoud, S., Kamalov, F. and Gonsalves, A. (2020) Data imbalance in

classification: Experimental Evaluation. *Information Sciences*, **513**, 429-441. https://doi.org/10.1016/j.ins.2019.11.004

[37] Parker, B.J., Günter, S. and Bedo, J. (2007) Stratification Bias in Low Signal Microarray Studies. *BMC Bioinformatics*, **8**, Article No. 326. https://doi.org/10.1186/1471-2105-8-326

[38] Araf, I., Idri, A. and Chairi, I. (2024) Cost-Sensitive Learning for Imbalanced Medical Data: A Review. *Artificial Intelligence Review*, **57**, Article No. 80. https://doi.org/10.1007/s10462-023-10652-8

[39] Fernández, A., García, S., Galar, M., Prati, R.C., Krawczyk, B. and Herrera, F. (2018) Learning from Imbalanced Data Sets. Vol. 10, Springer. https://doi.org/10.1007/978-3-319-98074-4

[40] Mi, X., Zou, B., Zou, F. and Hu, J. (2021) Permutation-based Identification of Important Biomarkers for Complex Diseases via Machine Learning Models. *Nature Communications*, **12**, Article No. 3008. https://doi.org/10.1038/s41467-021-22756-2

[41] Gómez-Ramírez, J., Ávila-Villanueva, M. and Fernández-Blázquez, M.Á. (2020) Selecting the Most Important Self-Assessed Features for Predicting Conversion to Mild Cognitive Impairment with Random Forest and Permutation-Based Methods. *Scientific Reports*, **10**, Article No. 20630. https://doi.org/10.1038/s41598-020-77296-4

[42] Biswas, S., Grundlingh, N., Boardman, J., White, J. and Le, L. (2025) A Target Permutation Test for Statistical Significance of Feature Importance in Differentiable Models. *Electronics*, **14**, Article 571. https://doi.org/10.3390/electronics14030571

[43] Liu, S.Y., David, L.C., Susana, G.-R., James, S.M. and Charles, M.P. (2025) Crafted Experiments to Evaluate Feature Selection Methods for Single-Cell RNA-Seq Data. *NAR Genomics and Bioinformatics*, **7**, lqaf023. https://doi.org/10.1093/nargab/lqaf023

[44] Espinosa, R., Sánchez, G., Palma, J. and Jiménez, F. (2025) Permutation-Based Multi-Objective Evolutionary Feature Selection for High-Dimensional Data. arXiv:2501.14310.

[45] Ali, S., Ahsan, M., Tasnim, L., Afrin, S., Biswas, K., Hossain, M., Ahmed, M., Hashan, R., Islam, K. and Raman, S. (2024) Federated Learning in Healthcare: Model Misconducts, Security, Challenges, Applications, and Future Research Directions—A Systematic Review. arXiv:2405.13832.

[46] Sachin D.N, Annappa, B, Hegde, S., Abhijit, C.S. and Ambesange, S. (2024) Fedcure: A Heterogeneity-Aware Personalized Federated Learning Framework for Intelligent Healthcare Applications in IoMT Environments. *IEEE Access*, **12**, 15867-15883. https://doi.org/10.1109/access.2024.3357514

[47] Abbas, S.R., Abbas, Z., Zahir, A. and Lee, S.W. (2024) Federated Learning in Smart Healthcare: A Comprehensive Review on Privacy, Security, and Predictive Analytics with IoT Integration. *Healthcare*, **12**, Article 2587. https://doi.org/10.3390/healthcare12242587

[48] Dasaradharami Reddy, K. and Gadekallu, T.R. (2023) A Comprehensive Survey on Federated Learning Techniques for Healthcare Informatics. *Computational Intelligence and Neuroscience*, **2023**, Article ID: 8393990. https://doi.org/10.1155/2023/8393990

[49] Fathima, A.S., Basha, S.M., Ahmed, S.T., Mathivanan, S.K., Rajendran, S., Mallik, S., *et al.* (2023) Federated Learning Based Futuristic Biomedical Big-Data Analysis and Standardization. *PLOS ONE*, **18**, e0291631. https://doi.org/10.1371/journal.pone.0291631

[50] Zulueta, N.M. (2024) Development of Federated Learning Models for Improved Ge-

netic Variant Assessment in a Multi-Site Clinical Setting. PhD Diss., Université Paris Cité.

[51] Nasajpour, M., Pouriyeh, S., Parizi, R.M., Han, M., Mosaiyebzadeh, F., Liu, L., *et al.* (2025) Federated Learning in Smart Healthcare: A Survey of Applications, Challenges, and Future Directions. *Electronics*, **14**, Article 1750. https://doi.org/10.3390/electronics14091750

[52] Sadilek, A., Liu, L., Nguyen, D., Kamruzzaman, M., Serghiou, S., Rader, B., *et al.* (2021) Privacy-First Health Research with Federated Learning. *npj Digital Medicine*, **4**, Article No. 132. https://doi.org/10.1038/s41746-021-00489-2

[53] Li, S., Cai, T. and Duan, R. (2023) Targeting Underrepresented Populations in Precision Medicine: A Federated Transfer Learning Approach. *The Annals of Applied Statistics*, **17**, Article No. 2970. https://doi.org/10.1214/23-aoas1747

[54] Gholami, S., Jannat, F., Thompson, A.C., Ong, S.S.Y., Lim, J.I., Leng, T., *et al.* (2025) Distributed Training of Foundation Models for Ophthalmic Diagnosis. *Communications Engineering*, **4**, Article No. 6. https://doi.org/10.1038/s44172-025-00341-5

[55] Kang, S. (2020) Model Validation Failure in Class Imbalance Problems. *Expert Systems with Applications*, **146**, Article 113190. https://doi.org/10.1016/j.eswa.2020.113190

[56] Johnson, J.M. and Khoshgoftaar, T.M. (2019) Survey on Deep Learning with Class Imbalance. *Journal of Big Data*, **6**, Article No. 27. https://doi.org/10.1186/s40537-019-0192-5

[57] Nguyen, G.H., Bouzerdoum, A. and Phung, S.L. (2009) Learning Pattern Classification Tasks with Imbalanced Data Sets. *Pattern Recognition*, **10**, 1322-1328.

[58] Rebouças, D.B., Barreto, P.A.P.M., Noronha, L.T., Roza, T.H. and Passos, I.C. (2025) Machine Learning Techniques in Bipolar Disorder. In: Passos, I.C., Berk, M. and Kapczinski, F., Eds., *Bipolar Disorder*, Springer, 815-835. https://doi.org/10.1007/978-3-031-85519-1_40

[59] Walsh, C.G., Ripperger, M.A., Hu, Y., Sheu, Y., Lee, H., Wilimitis, D., *et al.* (2024) Development and Multi-Site External Validation of a Generalizable Risk Prediction Model for Bipolar Disorder. *Translational Psychiatry*, **14**, Article No. 58. https://doi.org/10.1038/s41398-023-02720-y

[60] Bişkin, O.T., Candemir, C. and Selver, M.A. (2025) Detection of Bipolar Disorder and Schizophrenia Employing Bayesian-Optimized Grad-Cam-Driven Deep Learning. *Applied Sciences*, **15**, Article 1717. https://doi.org/10.3390/app15041717

[61] Ford, T., Stewart, R. and Downs, J. (2020) Surveillance, Case Registers, and Big Data. In: Jayati Das-Munshi, *et al.*, Eds., *Practical Psychiatric Epidemiology*, Oxford University Press, 219-C13.P111. https://doi.org/10.1093/med/9780198735564.003.0013

[62] Lu, Z., Pan, H., Dai, Y., Si, X. and Zhang, Y. (2024) Federated Learning with Non-IID Data: A Survey. *IEEE Internet of Things Journal*, **11**, 19188-19209. https://doi.org/10.1109/jiot.2024.3376548

[63] Nandan, U., Sai, C., Sai, N. and Viswanadapalli, A. (2024) Enhanced Brain Tumor Detection and Privacy Preserving Using Federated Learning. *International Journal of Scientific Research in Science and Technology*, **11**, 131-144. https://doi.org/10.32628/IJSRST24116166

[64] Hsieh, K., Phanishayee, A., Mutlu, O. and Gibbons, P. (2020) The Non-IID Data Quagmire of Decentralized Machine Learning. In: Daumé, H. and Singh, A., Eds., *International Conference on Machine Learning*, PMLR, 4387-4398.

[65] Benmalek, M. and Seddiki, A. (2024) Bias in Federated Learning: Factors, Effects,

Mitigations, and Open Issues. *Ingénierie des Systèmes d Information*, **29**, 2137-2160. https://doi.org/10.18280/isi.290605

[66] Gadekallu, T.R., Dev, K., Khowaja, S.A., Wang, W., Feng, H., Fang, K., Pandya, S. and Wang, W. (2025) Framework, Standards, Applications and Best practices of Responsible AI: A Comprehensive Survey. arXiv:2504.13979.

[67] Thirupathi, L., Anusha, G., Reddy Boya, T. and Prasad Cheerla, H.C. (2025) Federated Learning and Personalized Medicine. In: Reddy C, K.K. and Nag, A., Eds., *Federated Learning for Neural Disorders in Healthcare* 6.0, CRC Press, 248-278. https://doi.org/10.1201/9781003591085-9