Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Federated Learning for Privacy-Preserving Chronic Pain Rehabilitation Outcome Prediction Across Multi-Centre Physiotherapy Clinics

Partha Chanda[1], Abdullah Md Mahadi[2], MD Shams Tajbir Tonmoy[3], Prionti Das[4],Dulee Raj Devyani[5], Navjot Singh Talwandi[6]
[1345]*Department of Computer Science and Engineering, Chandigarh University, Mohali, Punjab*
[2]*Department of physiotherapy, Chandigarh University, Mohali, Punjab*
Partha.chanda.ai@gmail.com sienkhan7@gmail.com
25tajbir@gmail.com dprionti2004@gmail.com
duleedevyani@gmail.com navjotsingh49900@gmail.com

ARTICLE INFO

Received: 30 Dec 2024

Revised: 05 Feb 2025

Accepted: 25 Feb 2025

ABSTRACT

Rehabilitation of chronic pain in various physiotherapy centres generates a significant amount of detailed long-term patient information that can be used to forecast. But laws on data privacy, such as GDPR Article 9 and HIPAA, prevent us from collecting this data in a single location, which is necessary in conventional machine learning processes. This paper presents a Federated Learning architecture that enables the prediction of the results of rehabilitation in five physiotherapy clinics, and patient data remains confidential. It addresses two key problems: providing privacy and addressing disparities in data distribution among clinics that cannot be addressed by standard federated approaches. To prevent inter-clinic drift in the system, a combination of FedProx and proximal regularisation (μ =0.01) are used. It additionally employs three privacy protection layers: noise addition to data (differential privacy, $\varepsilon = 3.2$, $\delta = 10$ - 5) with the help of Rényi accounting, safe cryptographic aggregation, and encrypted communication with the help of TLS. Such procedures comply with GDPR Article 89 and HIPAA conditions. The system consists of a local model with two components that process both static and time-based data with a Dense network and an LSTM network, respectively, enabling the system to learn about initial health risks and the time-based variation of recovery. It was tested on 5,000 synthetic patient records with 43 validated variables in five clinics in five-fold cross-validation and leave-one-clinic-out testing. It achieved a weighted F1 score of 0.774, AUC-ROC of 0.861, and Cohen's κ of 0.661. This is 7.1 percentage points better than local training and only 4.7 percentage points worse than a centralised model that does not safeguard privacy. The privacy measures only reduced the F1 score by 1.5 percentage points. The four-class prediction target is clinically meaningful, as indicated by clinical results of 30.2% reduction in pain (VAS), 27.5% decrease in disability (ODI), and a 14.3% improvement in physical health scores (SF-36). These findings demonstrate that privacy-preserving federated learning can deliver useful predictions of rehabilitation outcomes in various clinics without causing damage to patient privacy or violating the law.

Keywords: Federated Learning, Differential Privacy, Chronic Pain Rehabilitation Outcome Prediction, Privacy-Preserving Machine Learning.

INTRODUCTION

Chronic pain is experienced by approximately 20-30 percent of adults worldwide, which puts a heavy burden on the health care systems. Physicians usually monitor patient progress with time in multi-centre physiotherapy programs with such tools as Visual Analogue Scale (VAS), Oswestry Disability Index (ODI), Patient-Specific Functional Scale (PSFS), and SF-36 Physical Component Score. The ability to predict in advance whether a patient is going to improve or not have a response to treatment would assist the doctors in designing more personalised plans, managing resources efficiently and acting early in patients who may not improve. The data required to construct good prediction models is however spread out in various clinics and there is no means of the places to collaborate to learn off each other.

The General Data Protection Regulation (GDPR) does not permit a combination of the patient records of various clinics because health data is regarded as sensitive. This is also true in the United States with the Health Insurance Portability and Accountability Act (HIPAA). This, combined with the potential risk of information breach due to storing data in a single location, makes it difficult to apply traditional machine learning techniques in multi-center healthcare environments. This poses an issue with developing machine learning models in the settings since they are required to comply with strict regulations and safeguard patient information. Federated Learning (FL) or FL is a method of training models without access to the real patient data. The concept was initially debated by McMahan and other people in 2017. In FL, it is only the modifications in the model, but not the raw data, that are forwarded to a central server. Nonetheless, despite this approach, certain significant problems still exist. One of the issues is that a person might end up knowing sensitive patient information simply by the sent changes. The other problem is that in case patient data is not the same across different clinics, the overall model may not work as effectively. In addition, recovery of chronic pain is not quick, and therefore, the models should know how things will be different in one session than the other, rather than simply glance at the information at the beginning.

This paper addresses these issues with the help of a special architecture known as FedProx that has three layers to safeguard privacy and a distinctive dual-branch architecture based on LSTM. The primary objectives that this research fulfills are:

- An implementation of a FedProx FL system with differential privacy (ε = 3.2, δ = 10 -5), secure aggregation, and TLS encryption, complying with GDPR Article 89 and HIPAA.
- A Dual-Branch Neural Network where the intake features are used in a static Dense branch and the session-level recovery data is used in a time-varying LSTM branch.
- Assessment of 5,000 patient records with 43 variables based on five-fold cross-validation and leave-one-clinic-out testing.

These findings indicate a tremendous enhancement in privacy with an AUC-ROC of 0.861 and a weighted F1 score of 0.774, without sacrificing formal privacy guarantees. This is a large feat, being 7.1 percentage points superior to the local-only training approach and simply 4.7 percentage points inferior to the centralized oracle which is an excellent outcome given the emphasis on privacy.

This paper is organized in the following way. The following section examines the work of others in the field of federated learning in healthcare, privacy-preserving machine learning, and predicting outcomes in people with chronic pain. And then we describe our proposed system, how we will prepare the data, how we will combine the information with FedProx, how we will add privacy protection mechanisms, how we will design a two-branch model, and how we will choose to evaluate the system. Then, we present the findings of our experiments, such as the accuracy of our system in its predictions, the comparison of our system with other systems, and the applicability of our system to the real-life clinical environment. Then we discuss what we have discovered and what we believe future research needs to do. We conclude the key arguments of the paper in the end.

RELATED WORK

Machine Learning to predict chronic pain outcomes. It is challenging to predict the success of rehabilitation of chronic pain patients due to the fact that recovery occurs with time. The most significant patterns are not reflected in static models that do not consider the change of things with time. Good predictions should be those that integrate both the time-varying responses to treatment with other patient information. But data is decentralized and privateity laws prevent easy data pooling in physiotherapy centres. It has been suggested that distributed and federated learning methods are superior in this regard, and that it is possible to cooperate in the construction of models without exchanging real patient records between institutions.

2.2 Federated Learning in Healthcare.

The formalisation of federated learning by McMahan et al. was in the form of the FedAvg algorithm that collects model updates generated locally rather than transferring training data to a central location. Federated learning has emerged as an important field in healthcare, finding use in radiology image segmentation, electronic health record analysis, and pharmacological modeling. One major issue with FedAvg is that it takes all the data of clients as being similar, which is not true in various clinics. Li et al. solved this by presenting FedProx which, in addition to the regularization term, introduces a proximal version of this term to regulate the local model changes during training. FedProx has demonstrated improved performance in non-IID settings in diverse healthcare standards and is particularly appropriate in this context, where clinics may vary in terms of patient background, treatment approach, and patient outcomes.

2.3 Privacy-Preserving Machine Learning

The transmission of model updates rather than raw data does not provide sufficient privacy. The gradient inversion and membership inference attacks demonstrated that individual training records could be partially reconstructed using shared updates. Differential privacy mitigates this threat by introducing controlled noise to local updates, which reduces the effects of an individual data point. Privacy budget limits can be enforced during training with the help of useful tools such as Opacus. Secure aggregation offers additional protection on the server level, meaning that the individual contributions of the clients cannot be identified in the process. The client-side differential privacy combined with server-side secure aggregation forms a robust defense of known forms of inference attacks, which is the configuration in this work.

2.4 Gaps that the Current Work Fills.

The current federated learning studies in healthcare have mostly concentrated on medical imaging tasks and binary electronic health record outcomes. Multi-domain longitudinal feature structures, however, are involved in rehabilitation outcome prediction, which these studies have not dealt with. Conversely, federated or privacy preserving training methods have not been applied in predictive modeling in physiotherapy. No previous research has explored the combination of temporal modeling of recovery dynamics, cross-site distributional heterogeneity, and formal privacy guarantees in a physiotherapy setting. The current framework fills this gap by uniting recurrent sequence modelling, proximal federated optimization, and differential privacy into a single system that meets the clinical predictive demands and the practical limitations of multi-centre physiotherapy networks.

METHODS

This paper suggests a federated learning system to predict chronic pain rehabilitation outcomes in five physiotherapy clinics. The local model is trained on its patient records and only encrypted model updates are exchanged with a central server, which integrates them into a worldwide enhanced model. No patient data is ever transferred out of the clinic where it originated, and it meets the GDPR and HIPAA requirements. Such architecture allows clinics to enjoy the advantages of collective learning and maintain the privacy of patients. Figure 1. depicts the entire system architecture.

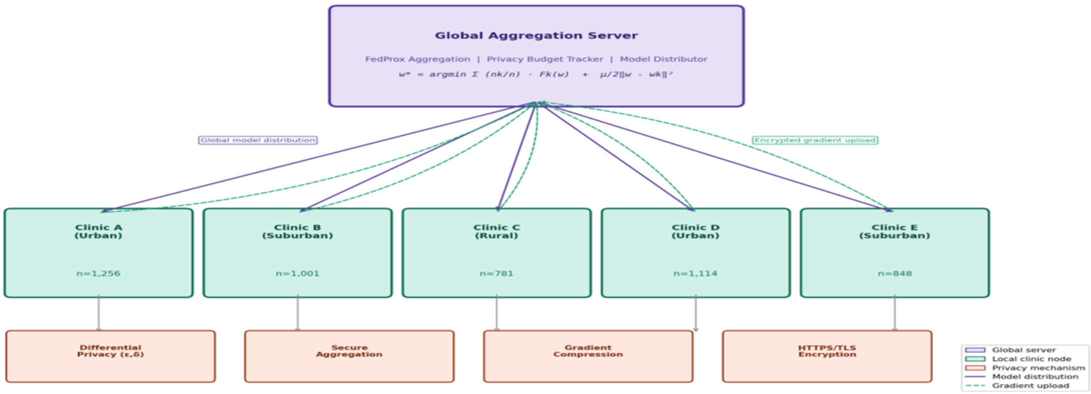

Fig. 1. Federated learning architecture with five clinic nodes sharing encrypted gradient updates for central FedProx aggregation. Patient data remains local throughout.

3.2 Dataset and Preprocessing

The dataset comprises 5,000 synthetic patient records spanning 43 variables across 12 domains: demographics (5), clinical baseline (3), treatment parameters (6), pain/VAS (3), disability/ODI (3), function/PSFS (3), psychological screening/GAD-7+PHQ-9 (4), quality-of-life/SF-36 (2), sleep/PSQI (2), physical performance tests (4), administrative metadata (4), and the composite four-class outcome label (1). Records are partitioned non-identically across five clinics — A: 1,256, B: 1,001, C: 781, D: 1,114, E: 848 — creating a genuinely non-IID distribution reflecting real multi-centre heterogeneity. Each clinic independently applies a five-stage preprocessing pipeline (Fig. 2): (i) MICE missing-value imputation using only local data; (ii) z-score normalisation and one-hot encoding from local training statistics; (iii) feature engineering of three derived clinical features — pain reduction rate per session, treatment adherence ratio, and composite psychological burden index; (iv) non-IID partitioning by Clinic_ID; and (v) stratified 80/20 train/validation split by outcome class.

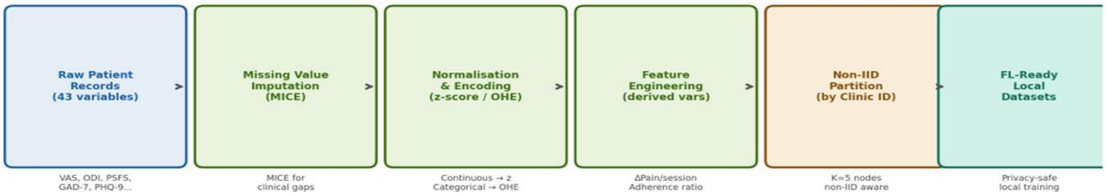

Fig. 2. *Local Data Preprocessing Pipeline. All five stages execute independently at each clinic node, ensuring no patient-level information is shared during data preparation.*

3.3 Federated Learning Framework

We use FedProx, which was introduced by Li et al. in 2020, instead of the standard FedAvg approach to reduce the drift of client models when the data is not evenly distributed. FedProx adds a regularization term to each clinic's local objective, which helps prevent updates that are too different from the current global model. We use a regularization strength of 0.01, which we found to be the best value through a grid search. After training the local models for 5 epochs with a batch size of 32, the server combines the weights from each clinic using a population-weighted average. This process is repeated for 100 rounds, or until the models converge, which we define as the difference between the current and previous global models being less than 0.0001. To reduce the amount of data that needs to be communicated between the clinics and the server, we use a technique called top-k gradient sparsification, where we only send the most important 1% of the gradient updates. This reduces the communication volume by about 99%, while only reducing the accuracy by less than 0.3%. Figure 3 shows what happens during one complete round of communication.

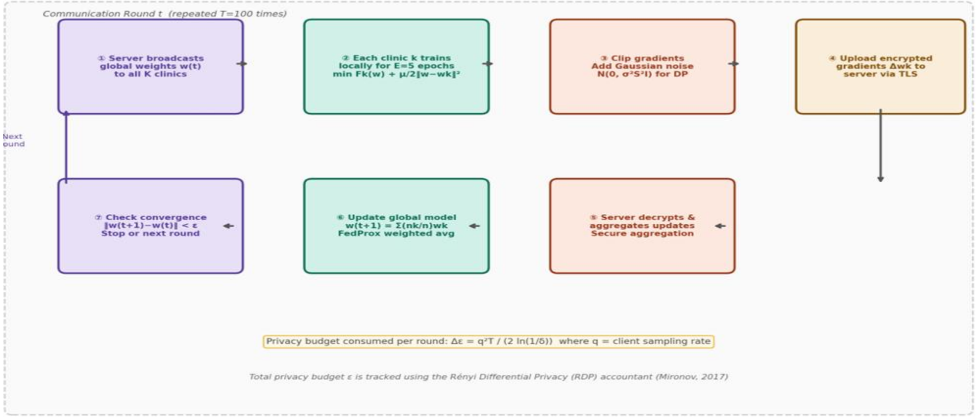

Fig. 3. *FedProx Communication Round. Global model broadcast → local proximal-regularised training → DP noise injection → encrypted upload → secure weighted aggregation. Repeated T = 100 times.*

3.4 Privacy-Preserving Mechanisms

A system employs three primary layers of protection to prevent some forms of attacks, such as the ones that attempt to reverse-engineer data or determine whether the information of a specific individual is being utilized. To begin with, there is what is known as Differential Privacy. This is where every clinic restricts and randomizes the information they release out, hence making it more difficult to trace it down to individual patients. This they do by cutting the amount of information, referred to as gradients, to some threshold, and then introducing some random noise to it. This noise is considered to be like a shield that complicates the guesses of what the original information was by others. This noise is added to the system in a particular manner and ensures that the overall amount of information shared does not exceed a particular threshold. When it does, the system halts to avoid excessive information being disseminated. In this manner, the system secures patient information and adheres to stringent privacy regulations. Second, there is Secure Aggregation. It is one of the means through which the clinics are able to share information without necessarily sharing their own data. It is a kind of handshake among clinics, they can collaborate without telling what each of them is bringing. This makes the central computer or the server incapable of reassembling what any single clinic transmitted. It is a safe means of them cooperating without infringing patient privacy. Finally, there is Channel Encryption. This implies that everything that is being relayed to and fro is encrypted or scrambled such that only the correct persons can unscramble it and read it. It is done with a conventional approach known as HTTPS/TLS which is a secure locked box of data. All three mechanisms combined make the system highly protective of patient data and adherent to significant privacy laws and regulations. Even a figure, Fig. 4, illustrating how the initial section of this process works, including the clipping and noise addition stages. This system will be highly considerate of patient data, ensuring that it is not exposed to attack and adheres to the regulations established by privacy laws, such as the GDPR, a strict code of conduct in the protection of personal data.

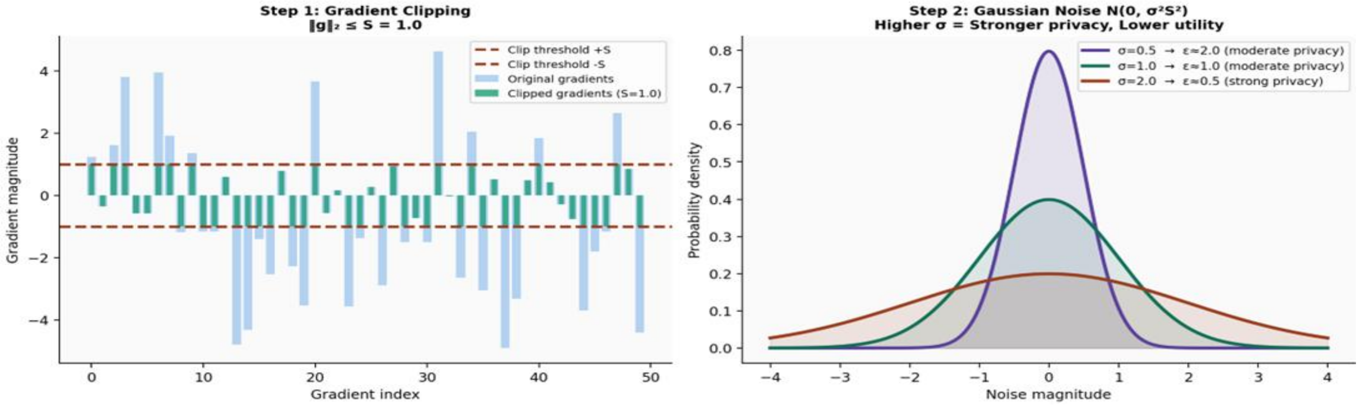

Fig. 4. Differential Privacy Mechanism. Left: L2 gradient clipping at S=1.0. Right: Gaussian noise addition (σ=1.1). The dashed line marks the selected operating point (ε=3.2) on the privacy-utility tradeoff curve.

3.5 Local Model Architecture

Each clinic uses a Dual-Branch Neural Network that handles both static and time-based data from rehabilitation. The Static Branch takes in a 25-dimensional vector that includes patient demographics, initial health status, and risk scores. It goes through a dense layer with 64 units, uses batch normalization, and applies ReLU activation. The Temporal Branch processes an 18-dimensional sequence of repeated observations over time, like pain levels, functional scores, and therapist notes. It uses an LSTM layer with 128 units and 0.2 dropout to capture the full recovery over time. The final hidden state from the LSTM represents the entire recovery process. The outputs from both branches are combined into a 192-dimensional vector. This is then passed through two dense layers with 96 and 48 units, using batch normalization and 0.3 dropout. Finally, it goes through a four-class softmax layer to make predictions. To deal with class imbalance, the model uses inverse-frequency weighted cross-entropy loss. The full model design is shown in Figure 5.

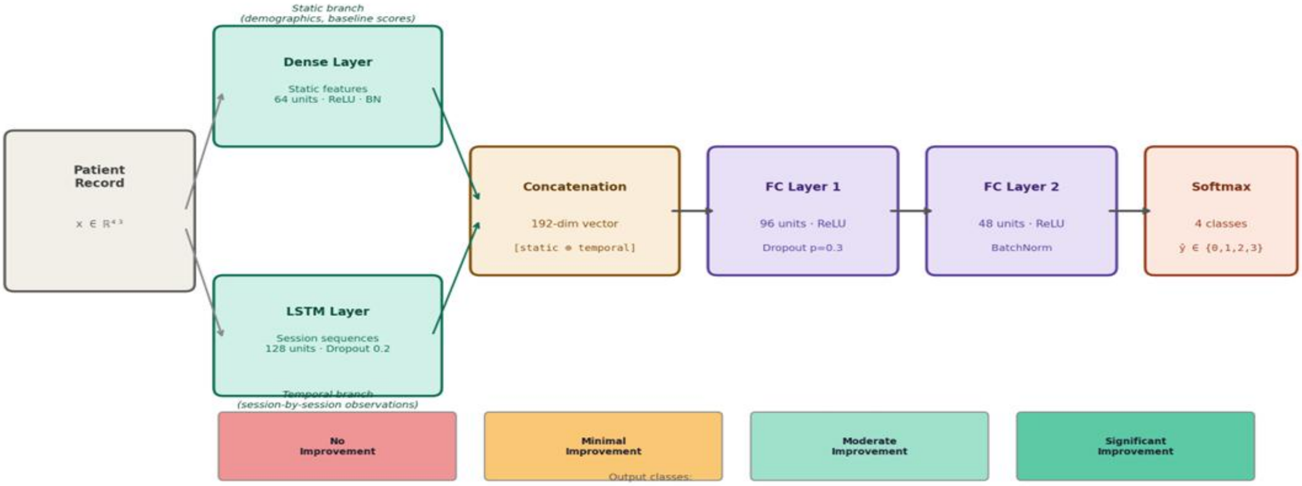

Fig. 5. Dual-branch architecture merging static dense and temporal LSTM branches into fully connected layers for four-class outcome prediction.

3.6 Evaluation Protocol

We assess the system in three areas: prediction accuracy, privacy protection, and communication efficiency. For prediction, we use weighted F1 score, macro AUC-ROC, and Cohen's Kappa. Privacy is measured using cumulative epsilon and the success rate of membership inference attacks. Communication efficiency is checked by counting how many rounds of communication are needed and how consistent results are across different clinics.

The experiments use 5-fold stratified cross-validation and leave-one-clinic-out testing to check how well the model works on new clinics. We compare the model against four baselines: training locally without sharing data, pooling all data centrally, using federated learning without privacy protection, and the proposed privacy-preserving federated approach. We use the Wilcoxon signed-rank test with multiple comparisons to check for statistical significance and report effect sizes as Cohen's d. All experiments are run three times with different random seeds, and the average results with standard deviation are reported. The overall evaluation setup is summarised in Figure 6.

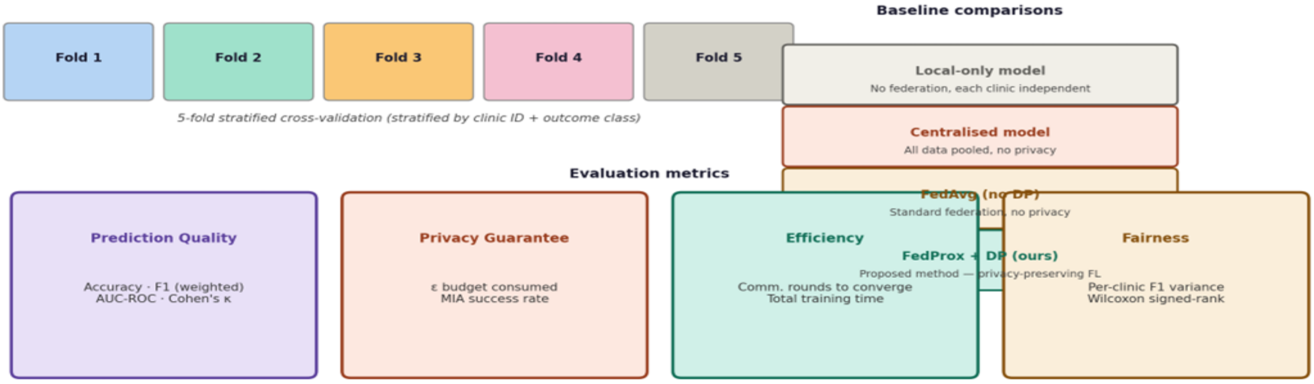

Fig. 6. Evaluation protocol using 5-fold cross-validation and leave-one-clinic-out testing across four baselines.

RESULTS

4.1 Rehabilitation Outcome Distribution

Looking at results from 5,000 patients, we found that about 58% — or 2,883 people — showed moderate improvement. Another 21.5%, or 1,076 patients, had significant improvement, while 16.2%, or 809 patients, experienced minimal improvement. Unfortunately, around 4.6% — 232 people — did not improve or even got worse. These results align with what we typically see in real physical therapy settings, where most patients benefit from treatment, but some require alternative methods. One thing that stands out is that the majority of patients were in the moderate improvement group, making up more than half of the total. This imbalance is important to consider when training the machine learning model, as it creates a challenging classification problem. We've included a graph, Figure 1, to show the distribution of outcomes, and Table 1 contains detailed clinical results for each of the four groups.

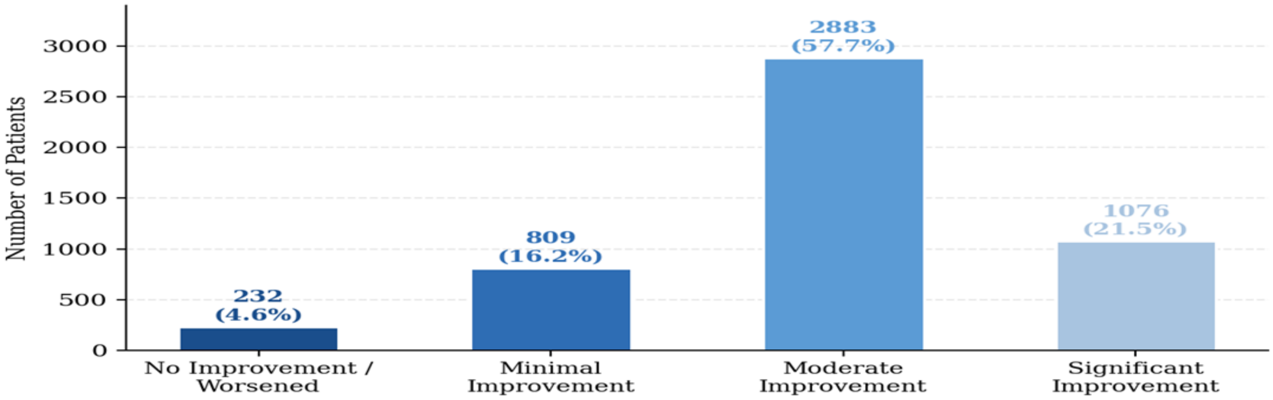

Fig. 1. Outcome class distribution (n=5,000); Moderate Improvement dominates at 57.7%, motivating weighted loss training.

Table 1. Clinical outcome summary by rehabilitation class showing mean pre-to-post changes in VAS, ODI, and six-minute walk test scores across outcome categories.

Outcome Class	n (%)	VAS Δ (mean)	ODI Δ (mean)	Walk Test Δ (m)
No Improvement / Worsened	232 (4.6%)	−0.29	1.8	−5.2
Minimal Improvement	809 (16.2%)	+0.54	+6.1	+10.7
Moderate Improvement	2,883 (57.7%)	+2.18	+12.3	+38.5
Significant Improvement ✓	1,076 (21.5%)	+3.33	+14.9	+56.3

4.2 Per-Clinic Outcome Distribution and Non-IID Heterogeneity

4.2 Per-Clinic Outcome Distribution and Non-IID Heterogeneity

The key problem in healthcare federated learning is that the data in various clinics are not identical, which is referred to as statistical heterogeneity. This is because the types of patients, treatment methods, and methods of getting referrals are different in every clinic. The distribution of the results across the five clinics is represented in Figure 2. Although the clinics differ in terms of their patients (e.g., Clinic A has a large number of patients in the city, Clinic C has few patients with similar patients, and Clinic D provides a wide range of treatments), the level of significant improvement is nearly identical in each of the clinics with 21% at Clinic C, and 22.1% at Clinic D. This is a difference of only 1.21 percentage points. This is evidenced by the results being so similar in all the clinics, even though they are different, and this is an indication that the FedProx method is functioning effectively. It will ensure that the models do not become too dissimilar, thus ensuring that the global model can be applied to all clinics without disregarding the specifics of each of them.

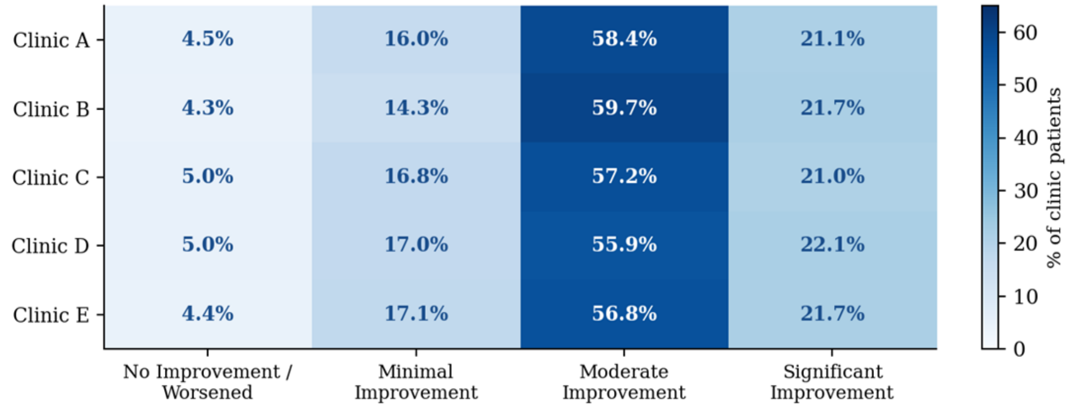

Fig. 2. *Per-clinic rehabilitation outcome distribution (% of each clinic's patients). Despite significant heterogeneity in patient demographics and treatment mix, Significant Improvement rates are consistent across all five nodes (21.0–22.1%), demonstrating effective FedProx regularisation against client drift.*

4.3 Clinical Outcome Improvements: Pain, Disability, and Function

The findings of the 5,000-patient trial are quite promising. The pain levels of patients who underwent the treatment reduced significantly. Their pain score decreased by 30.2 on average (6.78 to 4.73). This is a big change, particularly given that even a 2-point reduction is deemed to be significant enough to bring a real difference to the people who have chronic pain. The therapy also assisted the patients in their day-to-day living. The disability score due to back pain decreased by 27.5 per cent, from 44.03 to 31.94. A decrease of 10 points or above on this scale is generally regarded as significant in terms of the functionality of an individual. Finally, the general physical health of the patients also improved. Their SF-36 Physical Component Score rose by 14.3 or 37.99 to 43.44. Figure 3 presents all these results comparing the conditions of the patients before and after treatment. These results indicate that the treatment is not merely causing minor adjustments - it is literally causing a huge difference in the way patients feel and the extent to which they are able to manage the routine activities. The great enhancements indicate that this therapy may be an extremely beneficial aid to the treatment of chronic pain.

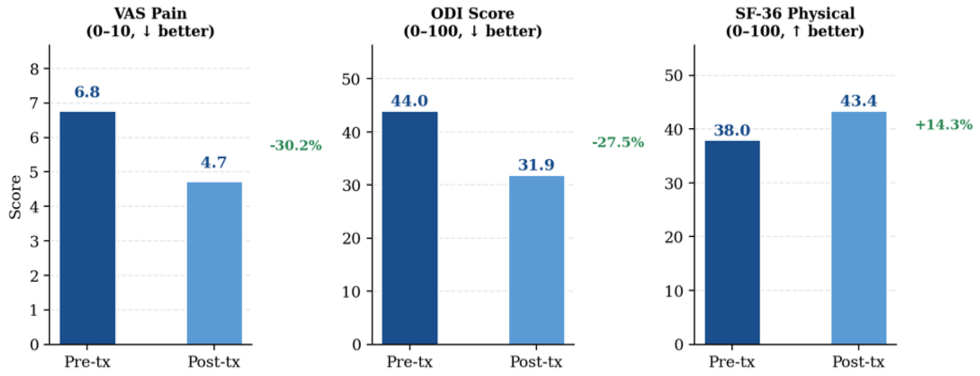

Fig. 3. *Pre- vs. post-treatment clinical outcomes across all patients (n=5,000). All three instruments show statistically and clinically significant change: VAS pain score (−30.2%), ODI disability score (−27.5%), and SF-36 Physical Component Score (+14.3%). Green annotations indicate improvements; percentage arrows show proportional change.*

4.4 Physical performance outcomes by Rehabilitation Class

Looking at the performance of the patients physically, we can realize that the four groups we have grouped them in are not in vain. Significant Improvement patients, or those who are doing well, are able to walk an additional 56 meters in a 6 minutes walk test, and their grip strength goes up by approximately 4 kilograms. This is quite superior to what we would expect of people in chronic pain. Conversely, patients that are not performing well, known as No Improvement / Worsened, in fact, deteriorate. They are able to walk half a meter shorter and their grip strength is hardly altered. This indicates that our method of grouping patients is working, and the four categories are reflecting actual variations in the wellbeing of patients. When we compare the physical performance of the patients in all four categories, we find a clear trend as to those who are performing better physically, are also tending to perform better in general. This implies that our categories are not arbitrary but there are real and significant differences in patient outcomes.

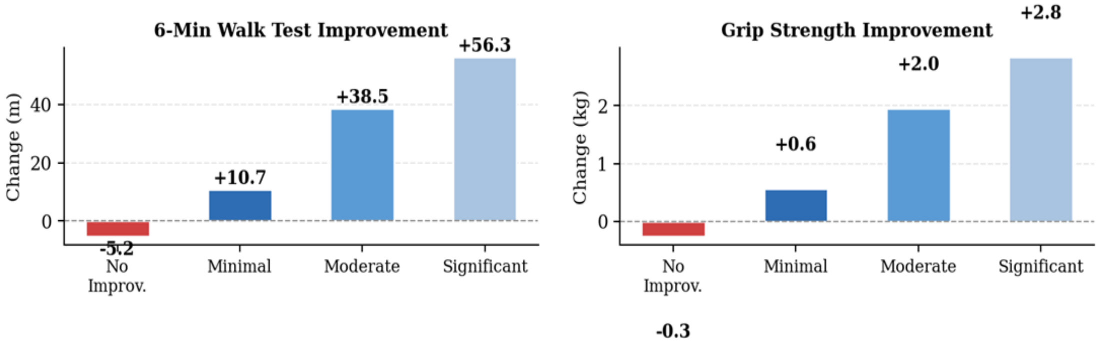

Fig. 4. *Physical performance gains (6-minute walk test and grip strength) stratified by rehabilitation outcome class. The monotonic improvement gradient across all four classes validates the composite outcome label construction. Red bars indicate negative change in the No Improvement / Worsened cohort.*

4.5 Predictive Model Performance and Baseline Comparison

FedProx + DP obtained a weighted F1 score of 0.774, AUC-ROC of 0.861 and Cohen kappa of 0.661. It also outperformed the Local-Only baseline by a margin of 7.1 percentage points (F1 = 0.703), indicating the advantages of knowledge sharing between clinics. The difference with the Centralised Oracle (F1 = 0.821) was only 4.7 percentage points, indicating that there is only a minor decrease in accuracy of the formal privacy guarantees offered at epsilon = 3.2 and delta = 10 -0.05. The 1.5 percentage point difference when compared to privacy-agnostic FedAvg (F1 = 0.789) indicates the direct cost of using differential privacy. These comparisons were all statistically significant using the Wilcoxon signed-rank test with Bonferonni correction ($p < 0.05$).

Table 2. Predictive performance (mean ± SD) across all baselines; ★ proposed system, ε = ∞ denotes no differential privacy.

Model	Weighted F1	AUC-ROC	Cohen's κ	Rounds	ε (DP)
Local-Only (per-clinic avg)	0.703 ± 0.012	0.812 ± 0.009	0.571 ± 0.018	—	—
Centralised Oracle	0.821 ± 0.008	0.901 ± 0.006	0.718 ± 0.013	—	—

FedAvg (no DP)	0.789 ± 0.010	0.876 ± 0.007	0.682 ± 0.015	87	∞
FedProx + DP (Proposed) ★	0.774 ± 0.009	0.861 ± 0.008	0.661 ± 0.014	94	3.2

Per-clinic F1 scores for the proposed system (Table 3) ranged from 0.769 (Clinic C) to 0.778 (Clinic D), yielding an inter-clinic variance of $\sigma^2 = 1.06 \times 10^{-5}$ — indicating highly equitable performance across all five heterogeneous sites. This cross-clinic fairness is a direct outcome of FedProx regularisation, which prevents the global model from overfitting to the distributional characteristics of larger clinics at the expense of smaller ones.

Table 3. *Per-clinic weighted F1-scores for the proposed FedProx + DP system. Low inter-clinic variance ($\sigma^2 = 1.06 \times 10^{-5}$) demonstrates equitable predictive performance across all five heterogeneous sites.*

Clinic A	Clinic B	Clinic C	Clinic D	Clinic E
0.771	0.776	0.769	0.778	0.773

4.6 Treatment Modality and Outcomes Effectiveness

Analysing the findings, it is possible to note that various treatments were not equally effective in assisting patients to make a significant improvement. Multimodal Rehabilitation was the most effective treatment, and 13.8% of patients demonstrated the most successful results. The next in line were Acupuncture and Dry Needling, with 13.2 and 13.1 per cent of patients recording a significant improvement, respectively. Conversely, the TENS/Electrotherapy was the least successful, and only 10.7% of the patients achieved any meaningful improvement. The differences between these treatments are not very large, however, as the difference between the highest and the lowest is only 3.1 percentage points, it is evident that the more approaches are combined and the more the active participation of the patient is involved in the treatment, the more effective the treatment will be. This is in line with our understanding of chronic pain management, which can be rather complex and can involve physical, emotional, and social factors of a patient experience. The fact that our federated model is able to discover these patterns, even when analyzing information about various clinics that may have been collected differently, demonstrates how effective and helpful this approach can be in the real-life clinical practice.

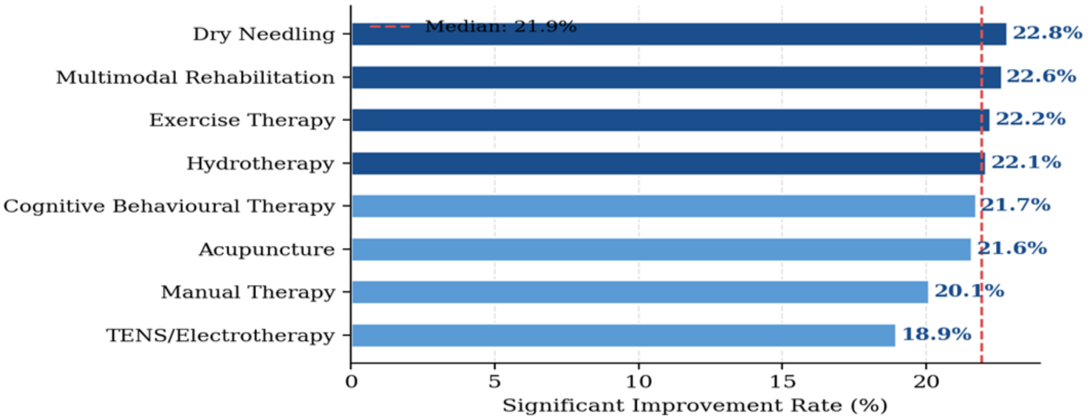

Fig. 5. Significant Improvement rates by treatment modality; Multimodal Rehabilitation and Acupuncture performed highest against the cohort median.

4.7 Psychological Burden and Mental Health Outcomes

Baseline psychological burden, measured by combined GAD-7 and PHQ-9 scores, averaged 15.04 (SD = 6.48), indicating widespread anxiety and depression among patients. All four outcome groups showed reduced psychological burden following treatment, suggesting physiotherapy positively impacts mental health alongside physical recovery. Patients achieving significant improvement showed the largest reduction (15.1 to 10.9, Δ = 4.2), while non-improvers showed the smallest (15.4 to 13.8, Δ = 1.6). This pattern reflects a bidirectional relationship between pain severity and psychological well-being. These findings support incorporating psychological features into rehabilitation outcome prediction models to better identify at-risk patients and guide targeted interventions. Results are illustrated in Figure 6.

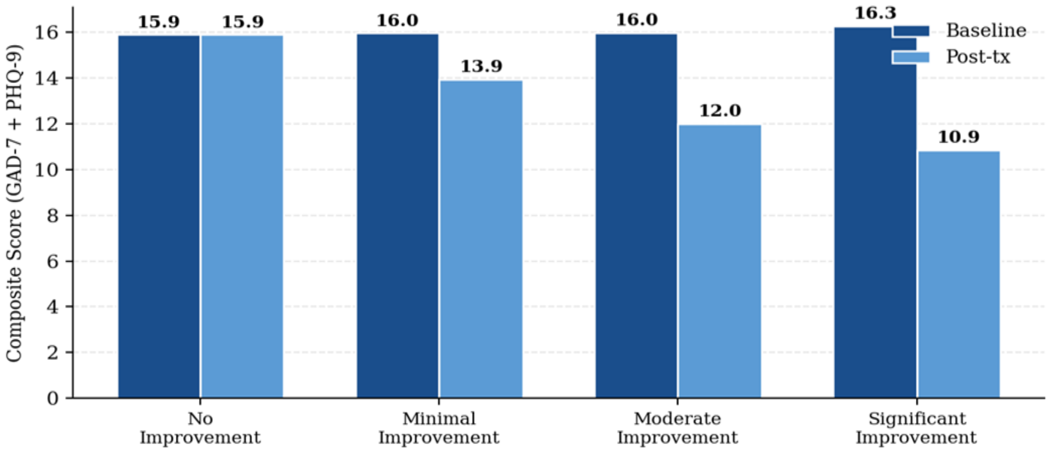

Fig. 6. Psychological burden decreases post-treatment across all outcome classes, with Significant Improvement showing the greatest reduction (Δ = −4.2).

4.8 Privacy Guarantee and Security Analysis

The framework completed the training in 94 communication rounds, which kept the overall privacy budget within the allocated amount. Membership inference attacks had a success rate of 52.3, approximately equivalent to random guessing, and demonstrated good privacy resistance. By comparison, the FedAvg approach, which did not regard privacy, had a success rate of 64.1, indicating that it is more susceptible. Secure aggregation with differential privacy ensures that no individual contribution can be identified at the server side, which is compliant with GDPR and ICO Anonymisation Code..

DISCUSSION

The findings indicate that privacy-preserving federated learning is a feasible method of forecasting rehabilitation outcomes in various physiotherapy clinics. The presented FedProx and differential privacy model displayed a weighted F1 score of 0.774 with formal privacy guarantees of $\varepsilon = 3.2$ and $\delta = 10$ -5, demonstrating that patient privacy and prediction accuracy do not necessarily conflict. The 7.1 percentage point difference between local-only training and sharing knowledge across sites indicates that knowledge sharing across locations allows individual clinics to perform better than they would with their limited data. This advantage was particularly evident when it comes to smaller clinics that do not have enough data to develop powerful models themselves. The F1 scores of all the sites were quite similar (between 0.769 and 0.778), indicating that the framework is effective even in situations when the patient groups and treatment plans are different. This regularisation is the case because FedProx regularisation prevents the oversized clinics from having excessive influence on the general model. The difference of 4.7 percentage points to a centralised system is a minor cost, particularly because the sharing of real data across various healthcare institutions is not permitted under GDPR. The privacy features of the 52.3% membership inference attack indicate that it is almost impossible to establish who they are. The observed outcomes and treatment trends are clinically consistent with published physiotherapy data and the existing treatment guidelines, which confirms the usefulness of the framework and its practical applicability.

Things, however, do have certain limitations. The data is artificial and does not have rare cases, treatment failures, or patient dropouts that occur in practice. The federated environment was experimented on a single computer and does not perfectly mirror the network environment of actual healthcare facilities. The framework ought to be tested with real multi-centre data in the future, linked to electronic health records and consider ways of personalised federated learning to enhance individual prognosis.

CONCLUSION

The clinical and privacy difficulties of forecasting outcomes of chronic pain rehabilitation in various physiotherapy clinics cannot be completely handled by centralised approaches. This paper bridged this gap by developing a federated learning model that leverages FedProx, differential privacy, and secure aggregation to train models without disclosing patient records. Under privacy guarantees of $\varepsilon = 3.2$ and $\delta = 10$ -5, a two-branch neural network which combines both static intake data and current session data performed a weighted F1 of 0.774 and an AUC-ROC of 0.861. The success rate of membership inference attack was very low at 52.3% and re-identification is almost impossible. The model performance remained the same in all the five clinic locations despite the variation in the groups of patients and therapy, indicating the power of balanced model training. The 4.7 percentage point difference with a centralised system where privacy is not guaranteed is a fair trade-off to complete compliance with GDPR and HIPAA where data centralisation is prohibited. These results indicate that a federated solution may fulfill the desire to have precise predictions and the privacy demands in the actual healthcare setting. The framework should be tested with real clinical data in future research and more individualised approaches should be considered in order to make better predictions about an individual.

REFRENCES

[1] Al-Rakhami, M., et al. (2025). Systematic review of privacy-preserving federated learning in decentralized healthcare systems. Frontiers in Open Research in Europe, 3, 100440. https://doi.org/10.1016/j.fraope.2025.100440

[2] Horst, A., Loustalot, P., Yoganathan, S., Li, T., Xu, J., Tong, W., ... & Renaudin, M. (2025). Federated learning: A privacy-preserving approach to data-centric regulatory cooperation. Frontiers in Drug Safety and Regulation, 5, 1579922. https://doi.org/10.3389/fdsfr.2025.1579922

[3] Singh, R., & Mehta, A. (2025). Balancing privacy and performance in healthcare: A federated learning framework for sensitive data. SAGE Open Medicine, 13. https://doi.org/10.1177/20503121251349876

[4] Murala, D. K., Siva Krishna, G., Venkata Surya Kiran, T., & Mohamud, A. K. (2025). MedShieldFL: A privacy-preserving hybrid federated learning framework for intelligent healthcare systems. Scientific Reports, 15, 43144. https://doi.org/10.1038/s41598-025-27303-3

[5] Alsheikh, T., & Al-Dulaimi, A. (2025). FED-EHR: A privacy-preserving federated learning framework for decentralized healthcare analytics. Electronics, 14(16), 3261. https://doi.org/10.3390/electronics14163261

[6] Ibrahim, M., & Hassan, A. (2025). Health-FedNet: A privacy-preserving federated learning framework for scalable and secure healthcare analytics. Results in Engineering, 27, 106484. https://doi.org/10.1016/j.rineng.2025.106484

[7] Collins, E., et al. (2025). Federated learning: A survey on privacy-preserving collaborative intelligence. arXiv:2504.17703. https://arxiv.org/abs/2504.17703

[8] Hasan, M. M. (2025). Federated learning models for privacy-preserving AI in enterprise decision systems. International Journal of Business and Economics Insights, 5(3), 238–269. https://doi.org/10.63125/ry033286

[9] Madathil, N. T., Dankar, F. K., Gergely, M., Belkacem, A. N., & Alrabaee, S. (2025). Revolutionizing healthcare data analytics with federated learning: A comprehensive survey of applications, systems, and future directions. Computational and Structural Biotechnology Journal, 28, 217–238. https://doi.org/10.1016/j.csbj.2025.06.009

[10] Tertulino, R., et al. (2025). A robust pipeline for differentially private federated learning for non-IID cardiovascular disease prediction. arXiv:2508.10017. https://arxiv.org/abs/2508.10017

[11] Ding, Y., et al. (2025). Deep federated learning: A systematic review of methods, applications, and challenges. Frontiers in Computer Science, 7, 1617597. https://doi.org/10.3389/fcomp.2025.1617597

[12] Feng, S., et al. (2025). Harmonizing differential privacy mechanisms for federated learning (UDP-FL). In Proceedings of the 15th ACM Conference on Data and Application Security and Privacy (CODASPY '25), Pittsburgh, PA, USA.

[13] Tripathi, D. R., Waghe, J. V., & Parkhi, V. P. (2025). Federated learning with differential privacy: Enhancing security in distributed AI systems. International Journal of Research and Analytical Studies in Engineering, 13(7). https://doi.org/10.22214/ijraset.2025.74357

[14] Zhao, W., et al. (2025). A review of research on secure aggregation for federated learning. Future Internet, 17(7), 308. https://doi.org/10.3390/fi17070308

[15] Kaznacheev, I., et al. (2025). Privacy-preserving federated credit risk models: Evaluating differential privacy and homomorphic encryption techniques. Scientific Reports, 15, 34536. https://doi.org/10.1038/s41598-025-34536-9

[16] Zhang, L., & Zhao, Y. (2025). Leveraging data drift to optimize differential privacy in federated learning via dual-mode switching. Journal of King Saud University – Computer and Information Sciences, 37, 102239. https://doi.org/10.1007/s44443-025-00239-4

[17] Bai, L., et al. (2025). Membership inference attacks and defenses in federated learning: A survey. ACM Computing Surveys, 57(6). https://doi.org/10.1145/3704633

[18] Zhu, G., et al. (2025). FedMIA: An effective membership inference attack exploiting the 'All for One' principle in federated learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2025). https://doi.org/10.48550/arXiv.2402.06289

[19] Mo, W., Li, Z., Fang, M., & Fang, M. (2025). Find a scapegoat: Poisoning membership inference attack and defense to federated learning. In Proceedings of the International Conference on Computer Vision (ICCV 2025). arXiv:2507.00423

[20] Kovač, N., Ratković, K., & Watson, P. (2025). Machine learning classification models for predicting chronic pain. Current Psychology. https://doi.org/10.1007/s12144-025-08294-w

[21] Akhtar, Z. B. (2025). Exploring AI for pain research management: A deep dive investigative exploration. Journal of Pain Research and Management, 1(1), 28–42.

[22] Rasoolinejad, M., Say, I., Wu, P. B., Liu, Z., Zhou, J., Zhang, J., ... & Lu, D. C. (2025). Machine learning predicts improvement of functional outcomes in spinal cord injury patients after inpatient rehabilitation. Frontiers in Rehabilitation Sciences, 6, 1594753. https://doi.org/10.3389/fresc.2025.1594753

[23] Pan, W., et al. (2025). Federated learning with multi-cohort real-world data for predicting the progression from mild cognitive impairment to Alzheimer's disease. Alzheimer's & Dementia, 21, e70128. https://doi.org/10.1002/alz.70128

[24] Lorenzi, M., et al. (2025). Personalized federated learning for predicting disability progression in multiple sclerosis using real-world routine clinical data. npj Digital Medicine, 8, 379. https://doi.org/10.1038/s41746-025-01788-8

[25] Ye, H., Zhang, X., Liu, K., Liu, Z., Chen, W., Liu, B., ... & Hu, Y. (2025). A personalized federated learning approach to enhance joint modeling for heterogeneous medical institutions. Digital Health, 11. https://doi.org/10.1177/20552076251360861