Federated Deep Learning for Privacy-Preserving Prostate Cancer Diagnosis from Multicenter Multiparametric MRI Data

Xuanxinyu Peng
Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA.
helloxuanxinyu@missouri.edu

Kegang Zhao
Department of Computer Science, University of Central Florida, Orlando, FL, USA.
kegangmail@ucf.edu

Abstract

Multiparametric magnetic resonance imaging has become a cornerstone in the diagnostic pathway for prostate cancer, yet the development of robust deep learning models for its interpretation is hampered by the fragmentation of sensitive imaging datasets across medical institutions. Federated learning offers a paradigm in which collaborative model training can occur without centralized data aggregation, preserving patient privacy while harnessing multicenter data diversity. This paper presents a comprehensive systems-level examination of federated deep learning architectures for privacy-preserving prostate cancer diagnosis using multiparametric MRI. We dissect the structural trade-offs inherent in designing federated topologies, aggregation strategies, and communication protocols that must contend with the high dimensionality and multimodal nature of prostate imaging. The analysis extends beyond algorithmic design to encompass privacy-enhancing technologies including differential privacy and secure aggregation, threat modeling against gradient leakage and inference attacks, and the intricate balance between utility and confidentiality. A substantial portion of the discussion is devoted to robustness and fairness across heterogeneous clinical sites, where variations in scanner vendors, imaging protocols, and patient demographics challenge the generalization guarantees of the global model. We further scrutinize the infrastructure demands for real-world multicenter deployment, addressing computational resource orchestration, longitudinal model maintenance, and integration with radiology workflows. Governance, regulatory compliance under GDPR and HIPAA, ethical accountability, and sustainable lifecycle management are framed as equally critical components of the federated ecosystem. By weaving together these technical and socio-technical dimensions, the paper articulates a systemic perspective that positions federated deep learning not merely as a distributed optimization technique but as a complex socio-technical infrastructure requiring careful architectural governance, transparent policy frameworks, and continuous cross-institutional stewardship to realize its potential for equitable and privacy-respecting prostate cancer diagnosis.

Keywords

federated learning, prostate cancer, multiparametric MRI, privacy preservation, deep learning, multicenter study, medical imaging.

1. Introduction

Prostate cancer remains one of the most prevalent malignancies among men worldwide, and its accurate and early diagnosis significantly influences treatment outcomes and patient survival. Multiparametric magnetic resonance imaging has emerged as a powerful noninvasive modality that combines anatomical T2-weighted sequences with functional diffusion-weighted imaging and dynamic contrast-enhanced sequences, thereby providing rich tissue characterization that aids in the detection and localization of clinically significant lesions. In recent years, deep learning methodologies have demonstrated substantial promise in automating and augmenting the interpretation of multiparametric MRI, often achieving diagnostic performance comparable to or surpassing that of experienced radiologists using standardized reporting systems. The success of these data-driven models, however, is predicated on access to large, diverse, and well-curated training datasets that encapsulate the full heterogeneity of disease manifestation, imaging protocols, and population variations. Such datasets are inherently distributed across numerous medical centers, each housing valuable patient data that cannot be easily pooled due to stringent privacy regulations, institutional policies, and the sheer logistical burden of transferring massive imaging volumes. This creates a fundamental tension between the clinical imperative to build generalizable diagnostic models and the legal and ethical obligation to safeguard patient confidentiality.

The advent of federated learning has opened a compelling avenue to reconcile the conflicting demands of collaborative model development and data sovereignty. In its canonical instantiation, federated learning enables multiple participating sites to jointly train a shared deep learning model by iteratively exchanging model parameters or gradients rather than raw data, with a central server orchestrating the aggregation. This paradigm has sparked considerable interest in medical imaging, where the sensitivity of radiological data has historically precluded the kind of large-scale multi-institutional studies that are routine in less regulated domains. Early proof-of-concept studies have illustrated that models trained in a federated fashion on brain tumor segmentation, mammography, and chest radiography can approach the performance of their centrally trained counterparts while offering strong privacy assurances. For prostate cancer specifically, pilot investigations have demonstrated the viability of federated learning in improving site-specific model robustness and enabling knowledge transfer across institutions without compromising patient identity. Nevertheless, translating these laboratory successes into a clinically deployable, multisite infrastructure for prostate MRI interpretation requires grappling with a host of systems-level challenges that extend well beyond the optimization of loss functions.

This paper undertakes a systematic and interdisciplinary analysis of federated deep learning for privacy-preserving prostate cancer diagnosis, deliberately shifting the discourse from pure algorithmic novelty to the structural, infrastructural, and governance dimensions that will determine real-world viability. We examine the architectural design space of federated systems adapted to the peculiarities of multiparametric MRI, which blends volumetric anatomical data with quantitative parametric maps and temporally resolved enhancement curves. The discussion navigates the trade-offs between centralized, decentralized, and hierarchical topologies, the impact of non-identically distributed data distributions across clinical sites, and the strategies for integrating differential privacy and secure aggregation into the training pipeline without catastrophically degrading diagnostic sensitivity. A sustained analytical thread is devoted to fairness and robustness, acknowledging that federated optimization processes can inadvertently perpetuate or even amplify demographic and equipment-based biases if not explicitly governed. Furthermore, we foreground the underappreciated dimensions of deployment infrastructure, computational sustainability, and

regulatory alignment, arguing that a successful federated learning ecosystem for prostate cancer diagnosis is as much a product of institutional cooperation and policy engineering as it is of machine learning research. By synthesizing these perspectives, we aim to provide a holistic reference framework that can guide the design, evaluation, and governance of forthcoming multicenter AI initiatives in prostate imaging.

2. Related Work and Foundational Concepts

The application of deep learning to prostate multiparametric MRI has evolved along multiple axes, with convolutional neural networks and vision transformers being harnessed for lesion detection, segmentation, and classification tasks. Landmark investigations have shown that deep networks trained on carefully annotated T2-weighted, diffusion-weighted, and apparent diffusion coefficient maps can differentiate clinically significant prostate cancer from indolent disease with high accuracy, often using architectures that fuse multimodal information at various levels of abstraction [1,2]. Simultaneously, other groups have advanced the state of the art by incorporating patient-level clinical variables and prior biopsy information into multimodal learning pipelines, demonstrating incremental improvements in specificity and reducing false-positive recommendations for biopsy [3]. These contributions have established a deep learning substrate that is ripe for multicenter extension, yet the vast majority of such studies have been constrained to single-institution datasets or small public repositories, limiting their external validity and translational impact.

Parallel to these domain-specific advances, the broader federated learning literature has matured rapidly, providing foundational algorithms and theoretical insights that underpin any medical imaging deployment. The federated averaging algorithm introduced by McMahan et al. remains a reference baseline, combining local stochastic gradient descent with periodic server-side weighted averaging of client model updates [6]. Recognizing the vulnerability of vanilla federated averaging to statistical heterogeneity across clients, subsequent proposals such as SCAFFOLD and FedProx introduced correction terms and proximal regularization to stabilize convergence under non-identical data distributions [13]. In the medical context, Sheller et al. provided one of the earliest demonstrations of brain tumor segmentation via federated learning across multiple hospitals, establishing the feasibility and potential performance parity with centralized data sharing [4]. This was followed by broader vision statements and multi-disease evaluations that underscored the privacy, regulatory, and practical advantages of the federated approach in digital health [5,17]. A comprehensive survey of open problems in federated learning has since catalogued issues ranging from communication efficiency and adversarial robustness to fairness and personalization, forming a research roadmap that remains highly relevant to prostate imaging [7].

Privacy-preserving mechanisms constitute an indispensable layer within any federated medical imaging system. Secure aggregation protocols ensure that a central server can compute the sum of client model updates without inspecting individual contributions, effectively protecting parameter-level information from honest-but-curious aggregators [8]. Meanwhile, differential privacy, originally formalized in the context of database querying, has been adapted to deep learning by clipping per-example gradients and injecting calibrated noise during training, furnishing rigorous information-theoretic privacy guarantees that are attractive for handling protected health information [9]. The prostate MRI domain introduces distinctive privacy threats because imaging data may contain facial profiles in the field of view, embedded metadata, and the potential for reconstruction of the original scan from model gradients through inversion attacks. The interplay between these privacy technologies

and the clinical acceptability of model performance forms a critical design axis, as aggressive privatization inevitably incurs a utility penalty that could manifest as missed cancers or unnecessary biopsies. Federated prostate cancer classification studies have begun to navigate this trade-off, illustrating that a carefully tuned combination of secure aggregation and lightweight differential privacy can yield models that satisfy both accuracy benchmarks and institutional data governance requirements [10].

3. Architectural Framework for Federated Prostate MRI Analysis

Designing a federated architecture for prostate MRI diagnosis demands a deliberate consideration of topological structure, coordination protocols, and multimodal fusion strategies that respect the physical distribution of data and compute resources. In the default star topology, a central parameter server orchestrates the training rounds by distributing the current global model to all participating centers, collecting locally updated weights after a few epochs of on-site training, and performing a weighted average to produce the next iteration of the global model. This topology simplifies synchronization and accountability but introduces a single point of trust and potential communication bottleneck, particularly when cohorts involve dozens of geographically dispersed institutions. A more resilient alternative is a hierarchical federated topology wherein regional aggregators first consolidate updates from clusters of nearby hospitals before transmitting intermediate summaries to a global coordinator, an arrangement that aligns naturally with existing health system structures such as regional university hospital networks. For prostate MRI, where imaging protocols often cluster by geographic or organizational practices, such intermediate aggregation may additionally serve as a domain normalization layer, absorbing local intensity and sequence variations before they propagate to the global model.

The choice of aggregation strategy directly impacts both convergence speed and the fairness properties of the resulting diagnostic model. While federated averaging weights each site proportionally to its dataset size, this scheme can inadvertently bias the global model toward institutions with larger volumes of scans, potentially masking rare cancer presentations that are concentrated in smaller, specialized clinics. Recent modifications to reweight updates according to clinically meaningful criteria, such as the prevalence of clinically significant cancer or the balance of biopsy-confirmed positive and negative samples, offer a pathway to incorporate domain knowledge into the optimization loop. Equally important is the handling of client heterogeneity that stems from the multimodal nature of multiparametric MRI, where not all sites acquire the same set of sequences. Some centers may perform biparametric protocols omitting dynamic contrast-enhanced imaging, while others collect high-resolution T2-weighted images with variations in field strength and coil configuration. Federated architectures must incorporate modality-agnostic backbones or partial model aggregation schemes that allow sites with missing sequences to contribute to the shared feature extractors without forcing the federation to reduce to the lowest common denominator. This structural flexibility can be realized through modular network designs in which a shared encoder is trained across all sites while sequence-specific branches are updated only by the sites that possess the corresponding data streams.

The communication fabric that underpins these federated operations strongly influences both the carbon footprint and the practical deployability of the system. Prostate MRI volumes are inherently large, and contemporary deep learning models for volumetric segmentation can easily contain tens of millions of parameters, leading to model update payloads on the order of hundreds of megabytes. In environments where hospitals operate behind firewalls with

limited outbound bandwidth, such transfers can be untenable if performed naively after every local epoch. A suite of communication-efficient techniques, including gradient compression, sparsification, and periodic rather than per-round synchronization, has been explored in general federated contexts [11]. For prostate diagnostics, one may further exploit the temporal stability of radiological practice: hospitals do not alter their imaging protocols on a weekly basis, so fine-tuning rounds can be scheduled with lower frequency once an initial convergence plateau is reached. Moreover, the separation of feature extractors from classifier heads allows lightweight transfer learning rounds where only the higher layers are exchanged, dramatically curtailing communication load while preserving diagnostic refinement. These considerations demand that architects of federated prostate AI platforms engage deeply with hospital IT departments to profile network characteristics and negotiate acceptable update windows that do not disrupt clinical picture archiving and communication system operations.

Beyond the core federated training loop, the architectural vision must accommodate continuous model updating and validation in a dynamic clinic environment. As new biopsy results accumulate and confirm or refute the radiological assessments, the ground truth labels associated with prior scans become available, enabling a form of delayed semi-supervised learning that can be integrated through asynchronous model updates. A well-designed system will implement a registry of model versions, each linked to the data distributions and labeling policies active at the time of training, so that audits can retroactively determine why a particular lesion was classified in a certain way. This versioning capability is crucial for regulatory compliance, as medical device software subject to FDA or CE marking requires rigorous software as a medical device lifecycle management. The federated architecture thereby transcends a simple distributed training protocol and evolves into a governed workflow that melds statistical efficiency with clinical accountability and device regulation.

4. Privacy-Enhancing Technologies and Threat Modeling

The privacy dimensions of federated prostate MRI learning extend far beyond the baseline promise of not transferring raw images, requiring a structured threat model that enumerates potential adversaries, their capabilities, and the specific data assets at risk. In a typical multicenter configuration, the central aggregator may be an honest-but-curious entity that faithfully executes the aggregation protocol yet attempts to infer properties of individual patient scans from the sequence of model updates. Malicious clients may attempt to reconstruct training images from the global model parameters distributed each round, or launch membership inference attacks to determine whether a particular patient’s scan was included in the training set of a given institution. Sophisticated adversarial scenarios further consider collusion between the server and a subset of clients, side-channel attacks that exploit timing and size metadata of communicated tensors, and the risk of model inversion at deployment time when the final trained model is disseminated back to all participants and possibly used outside the original federation. For prostate MRI, the sensitive nature of the anatomical region imaged amplifies the harm associated with any successful privacy breach, as scans can reveal incidental findings, surgical history, or body characteristics that are intensely personal and subject to both legal protection and social stigma.

Secure multiparty computation and specifically secure aggregation have emerged as foundational countermeasures against gradient inspection by the central server. By employing cryptographic protocols based on Shamir secret sharing or threshold encryption, a group of clients can ensure that the server learns nothing beyond the aggregated sum of their model updates, so long as a minimum quorum of honest participants is maintained [8]. When applied

to deep models processing multiparametric MRI, computational overhead is a primary concern, because the dimensionality of gradients scales with the network size and the number of sequences modeled. However, modern system designs can mitigate this through a separation of operations: heavy-weight secure aggregation may be applied only to the final layers that carry the most sensitive information, while earlier feature extractors can be exchanged with lighter protection or through trusted execution environments. This tiered protection model acknowledges that the raw gradient of a pixel-level segmentation head carries a higher risk of image reconstruction than the gradient of a compressed feature embedding, allowing security engineers to allocate cryptographic resources where they offer the greatest privacy return per unit of computational cost.

Differential privacy introduces a mathematically rigorous alternative or complement, wherein a randomized mechanism guarantees that the presence or absence of any single patient's data does not significantly alter the probability distribution over the outputs observed by an adversary. In the federated setting, differential privacy can be enforced at the client level by clipping local gradient norms and adding Gaussian noise before transmission, thereby bounding the contribution of each site and individual. The privacy budget, parameterized by the scalar epsilon, becomes a key dial that system designers must tune in negotiation with clinical stakeholders [9]. Small epsilon values provide strong privacy guarantees but can erode the model's sensitivity to subtle imaging biomarkers, such as the faintly hypointense signal on apparent diffusion coefficient maps that characterizes early peripheral zone tumors. Empirical evidence from other medical domains suggests that moderate privacy budgets can preserve diagnostically acceptable performance, yet prostate cancer introduces unique challenges because the lesion conspicuity on MRI varies substantially with tumor grade and histological architecture. A privacy-utility analysis must therefore be stratified by cancer aggressiveness, verifying that the differential privacy noise does not disproportionately degrade the detection of clinically significant high-grade lesions, which would subvert the clinical purpose of the diagnostic system. Implementation-wise, the infrastructure must support cryptographic accumulation of privacy loss across multiple communication rounds, smoothly retiring clients whose epsilon budgets are exhausted, a mechanism requiring careful orchestration of client participation schedules and global hyperparameter management.

Consent management and data provenance form the procedural complements to these technical safeguards. A responsible federated prostate MRI network will institute a dynamic consent framework in which each participating institution defines a data usage policy specifying the permitted training purposes, the retention period of local model checkpoints, and the conditions under which influence from its patients' data may be traced via techniques such as Shapley values. These policies are coded into smart contracts or platform-level access control lists that are enforced before any model update is sent or aggregated. Audit logs that record the precise sequence of cryptographic operations and model exchanges across the federated nodes serve both as a deterrent against protocol violations and as evidence for regulatory inspections. In the European context, the General Data Protection Regulation recasts many technical decisions into legal obligations, specifying that data controllers must perform data protection impact assessments and that any international transfer of personal data, even in model form, requires adequate safeguards [15]. Similarly, HIPAA mandates administrative, physical, and technical safeguards for protected health information in the United States, with the hybrid entity status of many academic medical centers further complicating the designation of responsibilities [16]. The technical architecture of the federated system must therefore embed these legal categories within its control logic,

transforming regulatory compliance from an afterthought into a design requirement that shapes topology, encryption, and logging subsystems from their inception.

5. Robustness, Fairness, and Generalization across Heterogeneous Centers

The performance of a federated deep learning model for prostate cancer diagnosis is acutely sensitive to the distributional heterogeneity that characterizes real multicenter data. Unlike carefully controlled research cohorts, operational MRI scanners across different hospitals exhibit variations in magnetic field strength ranging from 1.5 to 3.0 Tesla, differences in coil configurations, acquisition parameter choices for echo time and repetition time, and inconsistent application of the Prostate Imaging Reporting and Data System guidelines for lesion annotation. The resulting feature distributions shift across sites, producing a non-identically distributed data landscape that violates the assumptions of standard stochastic optimization and leads to global models that may excel at high-resource centers while underperforming at sites with older equipment or different patient demographics. This phenomenon, often termed statistical heterogeneity, necessitates federated optimization algorithms that are explicitly robust to client drift. Methods that introduce controlled statefulness, such as SCAFFOLD's use of control variates to correct for local update biases, have been shown to improve convergence stability in strongly heterogeneous regimes [13]. In the prostate MRI context, an analogous correction could be informed by scanner-specific calibration phantoms or by incorporating radiomics features that serve as harmonization anchors across equipment generations.

Fairness in federated prostate cancer diagnosis extends beyond the statistical properties of optimization to incorporate ethical and equity dimensions that are particularly acute in cancer care. If the federated model systematically achieves lower sensitivity for African American men, who bear a disproportionate burden of aggressive prostate cancer and are often underrepresented in academic cohorts, the very collaboration meant to democratize AI could inadvertently compound existing health disparities. Fairness-aware federated learning methods tackle this by reweighting the contributions of participating sites according to predefined equity metrics, such as minimizing the variance of per-site sensitivity or targeting a uniform lower bound on clinically significant cancer detection rate across demographic subgroups [12]. Implementing such fairness constraints in practice requires that each site diligently report disaggregated quality metrics stratified by self-reported race, age, and other protected attributes, a data governance step that some institutions may resist due to the same privacy concerns that motivated federation in the first place. Overcoming this tension calls for the integration of secure attribute-based reporting protocols that allow the global coordinator to compute fairness statistics without accessing individual-level demographic labels, perhaps through encrypted aggregation of confusion matrices stratified by binary protected attributes. Such technical accommodations, while complex, are essential to prevent a new form of digital redlining in which minority populations are systematically underserved by AI systems trained behind privacy firewalls.

Domain generalization is a closely related imperative that focuses on the model's ability to maintain diagnostic accuracy when deployed at a hospital that did not participate in any training round. The promise of federated learning is that exposure to ten heterogeneous sites during training will produce a model that adapts readily to an eleventh site's peculiarities, but this promise does not materialize automatically. Empirical studies in prostate MRI have demonstrated that models trained on data from high-volume academic centers can suffer marked performance degradation when applied to community hospitals owing to differences

in biopsy referral patterns, incidental findings, and reader expertise that shaped the labeling process [10]. Federated training reduces this fragility by exposing the model to a broader stylistic spectrum, yet deliberate interventions such as feature distribution matching, domain adversarial training within each client, or test-time adaptation layers that tune normalization statistics to the target scanner can further shore up generalization. These interventions must be coordinated across the federation without breaching privacy: each client can perform local adversarial domain alignment against a virtual reference distribution derived from the aggregated global feature statistics, achieving a form of collaborative domain invariance.

The robustness conversation must also encompass resilience to label noise and annotation inconsistency, which is rife in prostate MRI interpretation. Radiologists exhibit substantial inter-reader variability in assigning PI-RADS scores, delineating lesion boundaries, and determining the threshold for biopsy recommendation. A federated model trained on noisy labels from multiple sites can learn to either hedge its predictions or overconfidently latch onto spurious correlations between image texture and an arbitrary annotator's style. Research on federated label cleaning proposes that sites compute local confusion matrices against a small centrally curated validation set or that a student-teacher framework distills consensus knowledge across flips in annotation, but such schemes raise privacy flags if validation data must be shared. An architectural mitigation is to design the model to output not a single classification but a full predictive distribution that reflects the epistemic uncertainty stemming from annotator disagreement, enabling clinicians to gauge the trustworthiness of a given AI recommendation. Incorporating a Bayesian perspective, such as Monte Carlo dropout in the local models and aggregation of posterior parameter samples, can provide the necessary uncertainty quantification while remaining compatible with the federated averaging of probabilistic network weights.

Looking toward nascent computational paradigms, one may situate advanced model architectures within the federated prostate diagnostic ecosystem. Quantum machine learning has recently been explored for prostate cancer classification on multiparametric MRI, illustrating the potential of quantum-enhanced feature spaces to capture complex data relationships [14]. While such methods are in early stages and face practical barriers related to qubit coherence and embedding of high-dimensional imaging volumes, their trajectory highlights the need for federated frameworks that are architected for algorithmic extensibility. A mature infrastructure should accommodate heterogeneous model classes—classical convolutional networks, hybrid quantum-classical circuits, or future neuromorphic processors—within the same governance wrapper, provided that each module adheres to standardized interfaces for model update serialization, privacy budget accounting, and fairness metric reporting. The systemic viewpoint thus treats fairness and robustness not as static properties of a fixed model but as continuous design goals that must be re-evaluated with each evolution of the diagnostic pipeline.

6. Infrastructure, Deployment, and Sustainability

The transition from an experimental federated learning script running on simulated partitions to a live, cross-continental production system for prostate cancer diagnosis exposes a constellation of infrastructural demands that are often underestimated in the academic literature. Each participating hospital must provision dedicated computational resources, typically GPU-accelerated servers, that sit inside the institutional firewall and have controlled connectivity to the central aggregation service. These on-premise nodes must be compatible with the clinical picture archiving and communication system to fetch anonymized

multiparametric MRI studies and the corresponding structured reports, a process that requires careful engineering of DICOM routers, FHIR API endpoints, and database triggers that strip protected health information before the images reach the training pipeline. The federated platform software, whether built on frameworks such as NVIDIA FLARE, OpenFL, or custom orchestration layers, must be containerized and packaged with minimal dependency footprints so that it can be deployed across sites with heterogeneous IT policies. Site onboarding thus involves a rigorous qualification protocol encompassing security penetration testing, network latency benchmarking, and verification that the local compute node meets the memory and throughput requirements for processing volumetric MRI batches without affecting clinical radiology operations.

Longitudinal sustainability introduces a temporal dimension that challenges the project-based funding models typical of academic research. A prostate cancer diagnostic model degrades over time if not regularly updated, as imaging protocols evolve, new scanner models are installed, and the patient population shifts due to changing screening guidelines. The federated infrastructure must therefore support continuous learning cycles that can run for years, sustained by institutional commitments rather than grant cycles. This necessitates automated monitoring dashboards that track model performance metrics such as area under the receiver operating characteristic curve and recall for clinically significant cancer across each site, flagging distributional drift that may indicate a need for retraining. To minimize the burden on hospital staff, the system should support push-button participation modes where local updates are triggered by predefined accumulation thresholds—for example, whenever fifty new biopsy-confirmed cases are available—with the aggregation server dynamically scheduling rounds when a quorum of clients signals readiness. The operational cost, including electricity for GPU clusters and network bandwidth fees, must be transparently accounted for and shared among consortium members through fair-share models, possibly weighted by the relative benefit each site derives from the global model. Without such economic sustainability planning, many promising federated networks risk abandonment once initial enthusiasm wanes.

The carbon footprint of federated deep learning for prostate MRI merits explicit attention within the sustainability calculus. Training large 3D convolutional models across dozens of sites for hundreds of rounds can consume energy on the order of thousands of kilowatt-hours, rivaling the footprint of centralized data center training when the sum of distributed computations is tallied. However, federated architectures offer opportunities for carbon-aware scheduling that centralized paradigms lack. Aggregation rounds can be timed to coincide with periods of high renewable energy availability on regional grids, and sites with carbon-intensive energy mixes can be assigned smaller local training budgets or participate less frequently, with their contributions compensated by more frequent updates from low-carbon sites. The communication subsidies inherent in federated training—the transfer of model updates rather than raw images—already confer a significant reduction in data movement energy compared to the hypothetical alternative of shipping terabytes of DICOM data to a central repository. This comparative advantage should be explicitly quantified in environmental impact assessments of AI projects, positioning federated learning as a green alternative when the full lifecycle from data acquisition to inference is evaluated.

The integration of the federated model into clinical workflows introduces human-centered infrastructural requirements that extend beyond the technical stack. For radiologists to trust and adopt AI-assisted prostate MRI interpretation, the model’s outputs must be presented

within the existing reporting environment as overlays on multiparametric sequences, accompanied by interpretability heatmaps that highlight the image regions driving the classification decision. Techniques such as Grad-CAM have proven valuable for generating such visual explanations in medical imaging contexts and can be computed locally at each site without transmitting interpretability masks to the central server, thereby preserving privacy while enhancing user confidence [19]. The infrastructure must also support a feedback loop wherein radiologists can correct model predictions and those corrections are captured as high-quality training labels for future federated rounds, closing the gap between training-time supervision and deployment-time refinement. This human-in-the-loop paradigm transforms the federated system into a learning organization that continuously aligns its diagnostic behavior with expert consensus, provided that the governance structures incentivize clinicians to invest time in the corrective interactions.

7. Governance, Ethical Considerations, and Policy Implications

Governance of a federated prostate cancer diagnostic network operates at the intersection of law, institutional policy, and professional ethics, demanding structures that can equitably reconcile the interests of diverse stakeholders. A foundational governance decision is the selection of a cooperation model, which may range from a purely academic consortium governed by a memorandum of understanding to a legal entity such as a cooperative or trust that holds the global model and licenses it back to members under defined terms. The choice implicates liability allocation when the model contributes to a missed cancer diagnosis or an unnecessary biopsy: is the responsibility borne by the hospital that applied the model, by the consortium that trained it, or by the software developer who implemented the aggregation protocol? The emerging regulatory framework for artificial intelligence in medical devices, including the proposal for an EU Artificial Intelligence Act and the FDA's evolving stance on software as a medical device, categorizes continuously learning systems as presenting unique challenges that demand predetermined update validation plans and real-world performance monitoring [20]. A federated governance board comprising representatives from each member institution, as well as independent ethicists and patient advocates, is indispensable for deliberating these risks and setting the conditions under which a new model version is released into clinical practice.

Data sovereignty and community consent constitute ethically charged dimensions of cross-border federated learning for prostate MRI. Even though raw images do not leave a hospital, certain jurisdictions interpret the transfer of model parameters derived from patient data as a form of data export that triggers cross-border transfer restrictions. Indigenous and rural communities may demand collective consent processes that go beyond individual opt-outs, asserting communal rights over the use of health data generated within their facilities. The governance regime must accommodate such requirements through a tiered participation model that permits communities to contribute to a local or regional sub-federation without mandating global sharing, while still benefiting from the broader model through knowledge distillation techniques that transfer soft labels rather than parameters. This graduated sovereignty framework respects cultural and legal heterogeneity while maintaining scientific collaboration. It also requires meticulous technical implementation of geofencing logic that restricts the flow of updates based on the declared jurisdictional status of each node, an unusual requirement for machine learning platforms that underscores the deeply socio-technical nature of federated health AI.

Accountability mechanisms should be embedded within the federated protocol to enable retrospective audits and causal attribution. When a model error pattern is identified, investigators need to determine whether the error originated from a specific site's noisy labels, an aggregation round where a malicious update was injected, or a systematic underrepresentation of a particular prostate zone pathology in the entire coalition. Computational forensic techniques, such as secure logging of hashed model updates and tamper-evident aggregation records stored on a distributed ledger, can provide an immutable trace of the training history without exposing individual patient data. Coupled with statistical methods for influence function estimation, these tools allow federations to identify and potentially exclude problematic clients in subsequent rounds, though exclusion criteria must be specified transparently in governance documents to avoid arbitrary disenfranchisement. The ethical principle of procedural justice demands that any site facing expulsion has a right to understand the evidence and appeal the decision before a neutral arbitration panel, again elevating the governance apparatus to a level of sophistication commensurate with the life-impacting nature of prostate cancer care.

Policy implications stretch into the realm of reimbursement, certification, and global health equity. Health technology assessment bodies increasingly require evidence of real-world effectiveness across diverse settings, which aligns naturally with the distributed evidence generation capability of federated networks. Consortia can design prospective observational trials embedded within the federated infrastructure, comparing AI-assisted prostate MRI readings against standard-of-care interpretation across multiple centers and producing the kind of multisite evidence that regulators and payers demand. Certification of a federated model as a medical device may necessitate a novel regulatory pathway in which the continuing safety and effectiveness is monitored through the same federated channels used for training, collapsing the distinction between development and post-market surveillance. From a global health perspective, the strong bias of current prostate MRI AI toward data from high-income countries threatens to create a technological dependency that overlooks the different epidemiology and resource constraints of low- and middle-income settings. Policy interventions, such as weighted participation incentives, subsidized compute grants for under-resourced hospitals, and conditional licensing that mandates validation on geographically diverse datasets, are essential to steer the federated ecosystem toward inclusive clinical value. The ultimate metric of success is not the elegance of the aggregation algorithm but the equitable improvement in prostate cancer outcomes across all communities that the system purports to serve.

8. Conclusion

Federated deep learning stands at a pivotal moment in its evolution from a compelling theoretical construct to a deployable infrastructure for privacy-preserving prostate cancer diagnosis with multiparametric MRI. This paper has argued that realizing its clinical potential requires a deliberate expansion of the analytical frame from the narrow optimization of model weights to a comprehensive systems perspective that encompasses architectural topology, privacy engineering, robustness under heterogeneity, equity of diagnostic performance, infrastructural sustainability, and rigorous governance. We have articulated how the structural choices in aggregation strategy, modality handling, and communication scheduling directly shape both the accuracy and the fairness of the federated model, and how these technical decisions are inseparable from the legal and ethical fabric within which hospitals operate. The interplay between differential privacy budgets and the detection of high-grade lesions serves

as a microcosm of the larger systemic tension between data protection and clinical benefit, one that must be navigated through transparent tuning and multidisciplinary consensus rather than unilateral engineering decree. Robustness to scanner-specific domain shifts and annotation noise is not a mere statistical nuisance but a core requirement for safe deployment in community settings, demanding federated optimization algorithms and uncertainty quantification methods that are co-designed with clinical workflow constraints.

The infrastructural and governance dimensions highlighted here call for a reimagining of federated projects as long-lived socio-technical institutions, not ephemeral research experiments. The design of onboarding protocols, the sharing of computational burdens, the orchestration of carbon-aware aggregation schedules, and the creation of liability and arbitration frameworks are equally constitutive of system success as are innovations in neural network architecture. As the field continues to incorporate emerging computational paradigms—including the early exploration of quantum machine learning for MRI classification—the federated architecture must remain flexible enough to absorb such advances without retracting its privacy and fairness guarantees [14]. Looking forward, the convergence of federated learning with synthetic data generation, federated reinforcement learning for adaptive screening policies, and edge computing on MRI scanner consoles promises to further transform the diagnostic landscape. Progress will depend on the willingness of computer scientists, radiologists, ethicists, hospital administrators, and patient advocates to engage in sustained interdisciplinary dialogue that elevates governance and policy to the same level of attention as algorithm development. It is through this holistic integration that federated deep learning can fulfill its promise as a trustworthy, equitable, and effective partner in the global effort to reduce the burden of prostate cancer.

References

1. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.

2. Schelb, P., Kohl, S., Radtke, J. P., Wiesenfarth, M., Kickingereder, P., Bickelhaupt, S., Kuder, T. A., Stenzinger, A., Schlemmer, H. P., Maier-Hein, K. H., & Bonekamp, D. (2019). Classification of cancer at prostate MRI: Deep learning versus clinical PI-RADS assessment. Radiology, 293(3), 607–617.

3. Arif, M., Schoots, I. G., Castillo Tovar, J., Bangma, C. H., Krestin, G. P., Roobol, M. J., Niessen, W. J., & Veenland, J. F. (2020). Clinically significant prostate cancer detection and segmentation in low-risk patients using a convolutional neural network on multi-parametric MRI. European Radiology, 30, 6582–6592.

4. Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., Kotrotsou, A., Milchenko, M., Xu, W., Marcus, D., Colen, R. R., & Bakas, S. (2020). Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data. Scientific Reports, 10(1), 12598.

5. Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maaß-Hein, K., Ourselin, S., Sheller, M., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. NPJ Digital Medicine, 3(1), 119.

6. McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS) (pp. 1273–1282). PMLR.

7. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., El Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., … Zhao, S. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210.

8. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS) (pp. 1175–1191).

9. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS) (pp. 308–318).

10. Sarma, K. V., Harmon, S., Sanford, T., Roth, H. R., Xu, Z., Tetreault, J., Xu, S., Flores, M. G., Raman, A. G., Kulkarni, K., Wood, B. J., Choyke, P. L., Priester, A. M., Marks, L. S., Raman, S. S., Enzmann, D., Turkbey, B., Speier, W., & Arnold, C. W. (2021). Federated learning improves site performance in prostate MRI classification. Radiology: Artificial Intelligence, 3(4), e200193.

11. Konecny, J., McMahan, H. B., Yu, F. X., Richtarik, P., Suresh, A. T., & Bacon, D. (2016). Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492.

12. Li, T., Sanjabi, M., Beirami, A., & Smith, V. (2020). Fair resource allocation in federated learning. In International Conference on Learning Representations (ICLR).

13. Karimireddy, S. P., Kale, S., Mohri, M., Reddi, S. J., Stich, S. U., & Suresh, A. T. (2020). SCAFFOLD: Stochastic controlled averaging for federated learning. In International Conference on Machine Learning (ICML) (pp. 5132–5143). PMLR.

14. Chen, P., Safari, M., Barker-Clarke, R., Yang, X., & Scott, J. G. (2026, April). Prostate cancer classification using quantum machine-learning on multiparametric MRI. In Medical Imaging 2026: Imaging Informatics (Vol. 13930, pp. 320-327). SPIE.

15. European Commission. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union, L119, 1–88.

16. U.S. Department of Health and Human Services. (1996). Health Insurance Portability and Accountability Act of 1996. Public Law 104-191.

17. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology, 10(2), 1–19.

18. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.

19. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 618–626).

20. Cutillo, C. M., Sharma, K. R., Foschini, L., Ghosh, S., Mandl, K. D., & The Digital Health & AI Ethics & Regulatory Working Group. (2020). Machine intelligence in healthcare—Perspectives on trustworthiness, explainability, usability, and transparency. NPJ Digital Medicine, 3(1), 47.