FedCMC: a federated learning model with contribution fairness based on multi-center core data extraction for assessing the myometrial invasion status of endometrial cancer

Yuping Li[1†], Bao Feng[2,3†], Yuan Chen[4†], Xiaohong Ruan[4,5],
Jiangfeng Shi[6], Ximiao Wang[3], Xianyan Wen[7], Peijun Li[7],
Junqi Sun[8], Changye Zheng[9], Yujian Zou[9], Mingwei Li[10],
Wansheng Long[3,7*], Yehang Chen[2*] and Dong Xie[11*]

[1]School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin,
Guangxi, China, [2]Laboratory of Intelligent Detection and Information Processing, Guilin University of
Aerospace Technology, Guilin, Guangxi, China, [3]Jiangmen Key Laboratory of Artificial Intelligence in
Medical Image Computation and Application, Jiangmen Central Hospital, Jiangmen,
Guangdong, China, [4]Department of Gynecology, Jiangmen Central Hospital, Jiangmen,
Guangdong, China, [5]Clinical Transformation and Application Key Lab for Obstetrics and Gynecology,
Pediatrics, and Reproductive Medicine of Jiangmen, Jiangmen Central Hospital, Jiangmen,
Guangdong, China, [6]School of Automation Science and Engineering, South China University of
Technology, Guangzhou, Guangdong, China, [7]Department of Radiology, Yuebei People's Hospital,
Shaoguan, Guangdong, China, [8]Department of Radiology, Jiangmen Central Hospital, Jiangmen,
Guangdong, China, [9]Department of Radiology, Affiliated Dongguan Hospital, Southern Medical
University, Dongguan, Guangdong, China, [10]Department of Gynecology, Kaiping Central Hospital,
Kaiping, Guangdong, China, [11]College of Science, Guilin University of Aerospace Technology, Guilin,
Guangxi, China

Background: Multi-center Federated Learning (FL) has played a significant role in disease prediction, offering a feasible solution to the challenges of cross-institutional collaboration. However, the fairness issues inherent in traditional FL frameworks have limited their further development in the medical field.

Methods: We propose a Contribution Fairness Federated Learning model based on Multi-center Core Data Extraction (FedCMC). This model accurately assesses the actual contributions of each center from both data and model perspectives using two fairness indicators: data information richness and model quality. In the data contribution assessment phase, we innovatively design a Multi-center Core Data Extraction Module (MCDEM). This module extracts representative core datasets from the original training pool, effectively filtering redundant information and enhancing the fairness of data contribution assessment and the model's generalization ability. Subsequently, weighted aggregation based on each center's contribution optimizes the benefits for high-contribution centers, incentivizing more users to participate in federated learning. Finally, a personalized federated learning strategy is adopted, enabling the model to fine-tune through each center's core dataset, thereby improving its prediction relevance and accuracy.

Results: We analyze data from 902 endometrial cancer (EC) patients across four independent medical institutions. In centers A, B, and C, the FedCMC model achieves areas under the ROC curve (AUC) of 0.8261, 0.8750, and 0.8964,

respectively. Comparative analysis with three traditional federated learning algorithms indicates that FedCMC offers significant advantages in both performance and fairness.

Conclusion: FedCMC effectively alleviates fairness issues in traditional FL frameworks and accurately predicts the myometrial invasion (MI) status of EC patients, supporting personalized treatment strategies.

1 Introduction

Endometrial cancer (EC) was the sixth most commonly diagnosed cancer among women worldwide in 2022, with its incidence and mortality rates continuing to rise. Reported deaths increased from 97,370 in 2020 to 97,704 in 2022, posing a significant threat to women's health (1, 2). EC is typically treated through hysterectomy (3). While this approach has improved patient survival rates, conservative methods are also safe and feasible for some patients with early-stage EC (4). Assessing the status of myometrial invasion (MI), which determines whether the tumor is confined to the endometrium or has invaded the myometrium, is crucial for risk stratification and helps develop personalized treatment plans (5). According to the 2021 ESGO guidelines, the choice of surgical staging in specific cases depends on MI status (6). Research further suggests that preoperative determination of MI status aids in selecting optimal treatment strategies, especially for younger EC patients wishing to preserve fertility (7). Thus, gynecologists urgently need to confirm MI status before treatment to adjust therapeutic approaches, improve patient outcomes, and optimize healthcare resources.

In recent years, deep learning technologies have been widely applied in the medical field, achieving notable results (8–10). As a data-driven technology, deep learning models trained on single-center medical data are often limited in scale, and their applicability is confined to specific scenarios, resulting in relatively poor generalization capabilities (11). To develop stable and effective general artificial intelligence models, multi-center research has become crucial. However, due to regulations on data privacy protection, data from different medical institutions in multi-center studies cannot be directly shared (12). Federated learning, as a distributed machine learning framework, enables the extraction of feature representations from different centers without sharing private data, offering a practical solution for cross-center collaboration in the medical field (13, 14).

Nevertheless, with the continuous growth in data scale, the accumulation of redundant information poses challenges to both model training efficiency and performance enhancement. The key issue now is how to extract a representative core dataset from the

massive amount of data to improve the training or fine-tuning of models. For example, in multi-center federated learning scenarios, how to rapidly extract a core dataset for training and personalized fine-tuning of models; or in complex application contexts, how to leverage a small core dataset to help large models quickly adapt to task requirements, are pressing issues that need to be addressed.

Ensuring sufficient client participation is key to guaranteeing the performance of federated learning models (15, 16). In a federated learning framework involving multi-party collaboration, the contributions of different participants to the learning process may vary significantly. These contributions are influenced by factors such as data scale and data quality. However, current federated learning systems often lack fair contribution evaluation. Some clients possess large-scale data and abundant local computational resources, resulting in high-quality uploaded models. Nevertheless, these contributions are not appropriately valued by the server. When aggregated with lower-quality local models, the overall model performance may degrade, and the distributed model may even perform worse than the local models (17). Moreover, some federated frameworks overly prioritize data scale while neglecting data quality, leading to clients with high-quality but small-scale data being overlooked (18). This neglect of contributors' efforts undermines their motivation to participate actively in federated learning. Therefore, to attract more participants to federated learning, it has become imperative to develop a framework that can fairly reflect the contributions of each participant.

To address the aforementioned challenges, this study proposes a Contribution-Fair Federated Learning Model Based on Multi-Center Core Data Extraction (FedCMC). First, we comprehensively evaluate the contribution of each center from both data and model perspectives. Specifically, we design an innovative Multi-Center Data Extraction Mechanism (MCDEM) to extract the core dataset, where data contribution is assessed based on the richness of local data information rather than simply relying on data volume. Second, based on the contribution evaluation results, we perform weighted aggregation to optimize the benefits of high-contribution centers, thereby incentivizing active participation in federated learning and enhancing collaborative motivation. Finally, to mitigate model

performance inconsistencies caused by data heterogeneity across centers, we adopt a personalized federated learning strategy (19). By fine-tuning the global model on the extracted core dataset, we further improve model performance across different centers and alleviate performance disparities. Through this approach, we achieve a fairer and more accurate prediction of myometrial invasion status in EC patients, providing technical support for clinical strategy optimization and facilitating more targeted and personalized treatment plans.

2 Materials and methods

2.1 Patients

This retrospective study was approved by the ethics committees of the four participating centers. Due to its retrospective nature, informed consent was waived. The study reviewed the medical records of 757 patients diagnosed with endometrial cancer (EC) via hysterectomy and confirmed by pathology at Center A between September 2010 and September 2022, along with their corresponding MRI imaging data. Additionally, Centers B, C, and D reviewed relevant data for 459, 374, and 40 patients, respectively, between December 2016 and February 2023. Given the small sample size at Center D, the data from Center D were merged into Center C for subsequent analyses.

The inclusion criteria were as follows: (1) uterine malignant epithelial tumors confirmed by total hysterectomy (including EEC, serous carcinoma, clear cell carcinoma, undifferentiated and dedifferentiated carcinoma, mixed carcinoma, and carcinosarcoma); (2) MRI examination containing sagittal T2-weighted imaging (T2WI) sequences; (3) pelvic MRI completed within 21 days before surgery; (4) complete clinical and pathological information. Exclusion criteria were: (1) poor image quality or significant artifacts affecting tumor assessment; (2) MRI performed more than 21 days before surgery; (3) insufficient pathological or clinical data; (4) patients who received neoadjuvant chemotherapy or radiotherapy prior to surgery; (5) inability to determine the location of the primary tumor. The patient inclusion and exclusion flow for the three centers is shown in Figure 1.

The clinical characteristics of each included patient were categorized into two groups. The first group comprised clinicopathological features, including age and post-hysterectomy pathological results such as histopathological type, grade, lymphovascular space invasion (LVSI) status, myometrial invasion (MI) status, and FIGO stage (2009) (20). The second group included subjective MRI assessments, such as the maximum tumor diameter (TMD) observed on MRI and the radiologists' evaluation of MI status. Ultimately, 812 EC patients were included in this multi-center study. Patients from Centers A, B, and C were 410, 294, and 108, respectively, and were randomly divided into training and test sets in a 6:4 ratio. Basic patient information is detailed in Table 1.
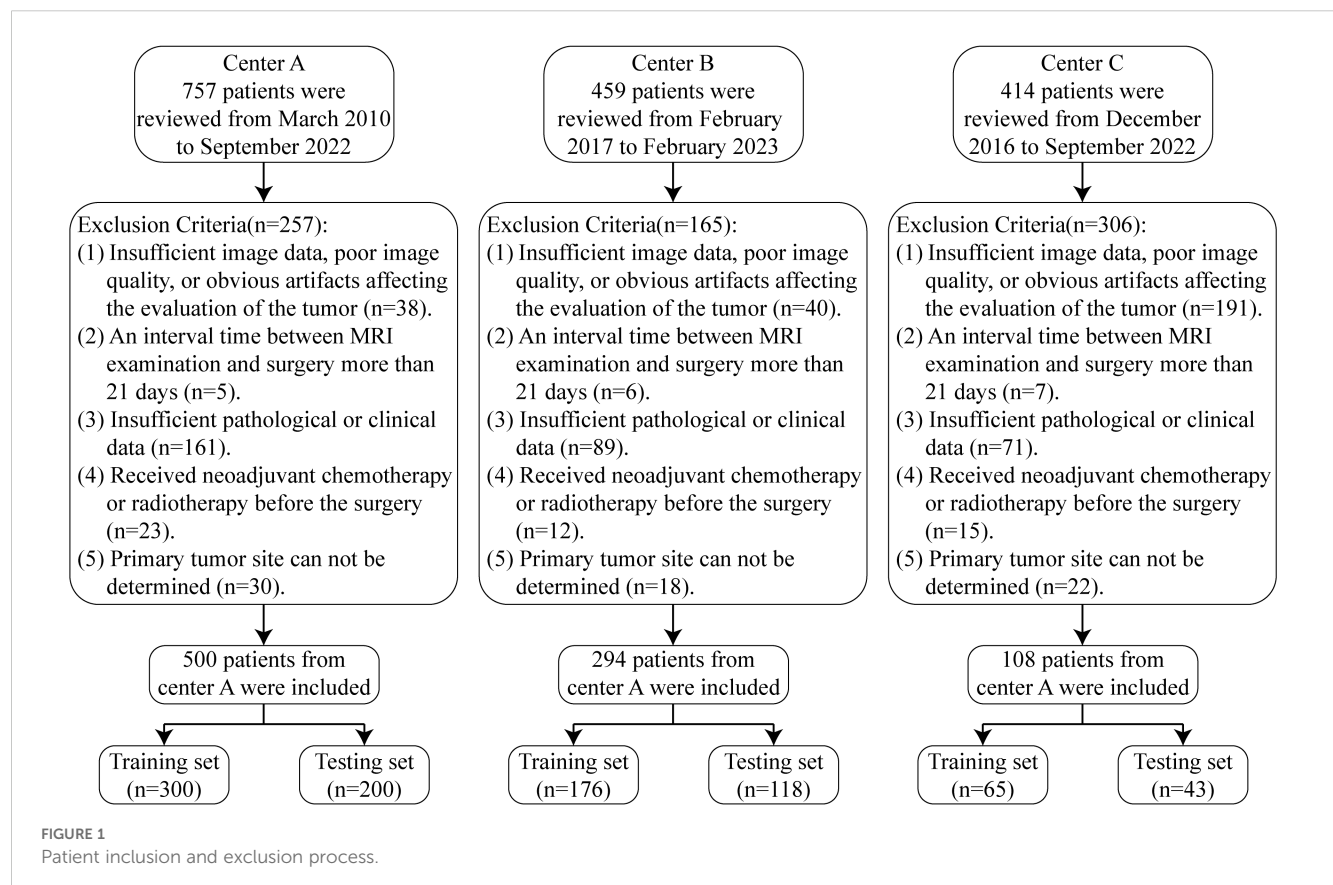

FIGURE 1
Patient inclusion and exclusion process.

TABLE 1 Basic patient information.

Center	Dataset	Status	Age (mean ± std)	TMD (mean ± std)	Histopathologic type		Histopathologic Grade			Read for MI status		FIGO stage			LVSI	
					EEC	Others	1	2	3	Absence	Presence	IA	IB	Others	Absence	Presence
Center A (500)	Train (300)	Non-MI(38)	47.76 ± 8.04	3.61 ± 1.89	35	3	25	9	4	9	29	36	0	2	37	1
		MI(262)	55.59 ± 8.06	4.33 ± 2.31	246	16	61	137	64	16	246	138	51	73	183	79
	Test (200)	Non-MI(25)	50.80 ± 8.63	2.76 ± 1.31	22	3	14	8	3	15	10	24	0	1	25	0
		MI(175)	55.13 ± 7.81	4.34 ± 2.17	153	22	33	92	50	15	160	87	37	51	113	62
Center B (294)	Train (176)	Non-MI(11)	51.27 ± 8.55	2.95 ± 2.59	11	0	11	0	0	3	8	11	0	0	11	0
		MI(165)	54.43 ± 8.07	4.05 ± 2.16	151	14	74	62	29	15	150	102	29	34	126	39
	Test (118)	Non-MI(8)	50.88 ± 10.19	3.70 ± 2.15	7	1	7	0	1	1	7	7	0	1	8	0
		MI(110)	54.84 ± 7.93	3.54 ± 1.59	98	12	49	38	23	13	97	70	17	23	93	17
Center C (108)	Train (65)	Non-MI(10)	46.20 ± 9.14	2.86 ± 1.60	10	0	10	0	0	2	8	10	0	0	10	0
		MI(55)	54.07 ± 9.04	4.27 ± 1.97	52	3	27	18	10	4	51	31	12	12	45	10
	Test (43)	Non-MI(6)	53.50 ± 9.85	4.05 ± 1.74	5	1	3	2	1	3	3	6	0	0	5	1
		MI(37)	60.11 ± 9.11	4.06 ± 1.97	35	2	8	23	6	9	28	20	5	12	36	1

Non-MI refers to endometrial cancer without myometrial invasion, MI refers to endometrial cancer with myometrial invasion; TMD, maximum diameter of the tumor on MRI images; mean indicates the average value, and std represents the standard deviation; Histopathologic type-EEC, endometrial endometrioid carcinoma; Histopathologic type-others, serous carcinoma, clear cell carcinoma, undifferentiated and dedifferentiated carcinoma, mixed carcinoma and carcinosarcoma; Reading for MI, radiologist's assessment of myometrial invasion; LVSI-Absence, negative lymphovascular space invasion; LVSI-Presence, positive lymphovascular space invasion.

2.2 MRI protocol and definition of MI

MRI examinations were performed using 1.5 or 3.0 T scanners, with patients positioned supine and breathing freely (see Supplementary Tables S1-S4 for MRI parameters). Two radiologists with specialized training in gynecologic imaging (with over 8 and 10 years of experience, respectively) independently reviewed the images, assessing the maximum tumor diameter and MI status on MRI. The radiologists were blinded to the patients' postoperative pathological results, and disagreements were resolved through discussion.

On T2WI images, EC was defined as a localized endometrial lesion with signal intensity (SI) lower than normal endometrium but higher than the myometrium [21]. MI status was categorized as no MI or MI, with MI further divided into superficial invasion (involving <50% of the myometrium) and deep invasion (involving ≥50% of the myometrium). The integrity of the junctional zone (JZ) was a critical criterion for MI assessment: disruption or interruption of the JZ indicated MI, whereas an intact JZ suggested no MI [22].

The surgical pathological diagnosis was performed by experienced clinical specialists in gynecological pathology. Pathological type, differentiation grade, MI status, and LVSI status were determined based on the 2020 WHO Classification of Tumors of Female Reproductive Organs. Staging was conducted according to the 2009 FIGO guidelines. Any diagnostic discrepancies were resolved through discussion.

2.3 ROI delineation

Regions of interest (ROI) were defined by a radiologist with >10 years of experience in gynecological imaging diagnosis. The delineation aimed to accurately contour the lesion. Subsequently, a rectangular bounding box covering the lesion's boundary was constructed, minimizing the impact of subjective ROI selection by clinicians [23, 24]. Deep learning methods were utilized to automatically extract ROI, eliminating the need for precise manual delineation. Details of the data preprocessing steps are provided in Supplementary Material S1.

2.4 Construction of the FedCMC framework

To address the fairness challenges faced by medical centers in federated learning, this study proposes a Contribution Fairness Federated Learning model based on Multi-center Core Data Extraction (FedCMC). First, we conduct a comprehensive evaluation of each center's contributions from both data and model perspectives. Given that varying degrees of information redundancy exist in local data across centers, directly using such data may lead to model bias and unfair data contribution assessment. To address this, we designed a Data Pruning Module to identify a core set of samples that represent the richness of each center's data information. These core samples are used to train local models and fairly evaluate data

contributions. Additionally, the model's accuracy, which intuitively reflects its quality, is incorporated to evaluate the overall contribution. Second, to address differences in contributions among centers, we introduce contribution-based fairness aggregation to optimize the actual benefits for high-contribution centers, thereby enhancing collaboration fairness across institutions. Finally, we adopt a personalized federated learning strategy to mitigate inconsistencies in global model performance across different centers. The algorithmic framework of FedCMC is illustrated in Figure 2.

2.4.1 Contribution evaluation based on model quality and core data

To accurately evaluate the contributions of each center, we propose two fairness metrics: data information richness and model quality.

Data information richness is measured by the size of the core sample set at each center. Since local data at different centers often contain varying degrees of redundancy, directly using such data can lead to model bias and unfair evaluation of data contributions. We designed a Data Pruning Module to eliminate samples rich in redundant information, retaining only core samples that represent the richness of each center's data. This approach shifts the data distribution toward underrepresented information, enhancing the model's adaptability to different categories of data and improving its generalization ability. Details of the Data Pruning Module can be found in Supplementary Material S2.

Model quality is primarily assessed based on the accuracy of each center's local model. Specifically, the accuracy of local models during each training round serves as the evaluation metric. This measure directly reflects how well a center's model adapts to and performs on its own data. By focusing on higher-quality local models during global aggregation, the global model's performance can be optimized while minimizing the adverse effects of low-quality models.

2.4.2 Contribution-based fair aggregation

We optimize the rewards for high-contribution centers through fair aggregation to motivate centers to actively participate in federated learning. Due to differences in data scale and quality among centers, their contributions to the global model vary. If high-contribution centers and low-contribution centers receive the same level of performance rewards, it would be unfair to the high-contribution centers, potentially discouraging their participation.

To address this issue, we propose an adaptive weighted aggregation method based on the contribution of each center. Contribution-weighted aggregation allows the global model to better fit the data distribution of high-contribution centers, improving its performance at these centers and increasing their rewards. At the same time, high contributions typically indicate higher-quality local models and richer data sources. Assigning greater weights to high-contribution centers not only enables the global model to learn more task-relevant features but also reduces the negative impact of low-performance centers, ultimately leading to a stronger global model. Details of the aggregation process are provided in Supplementary Material S3.
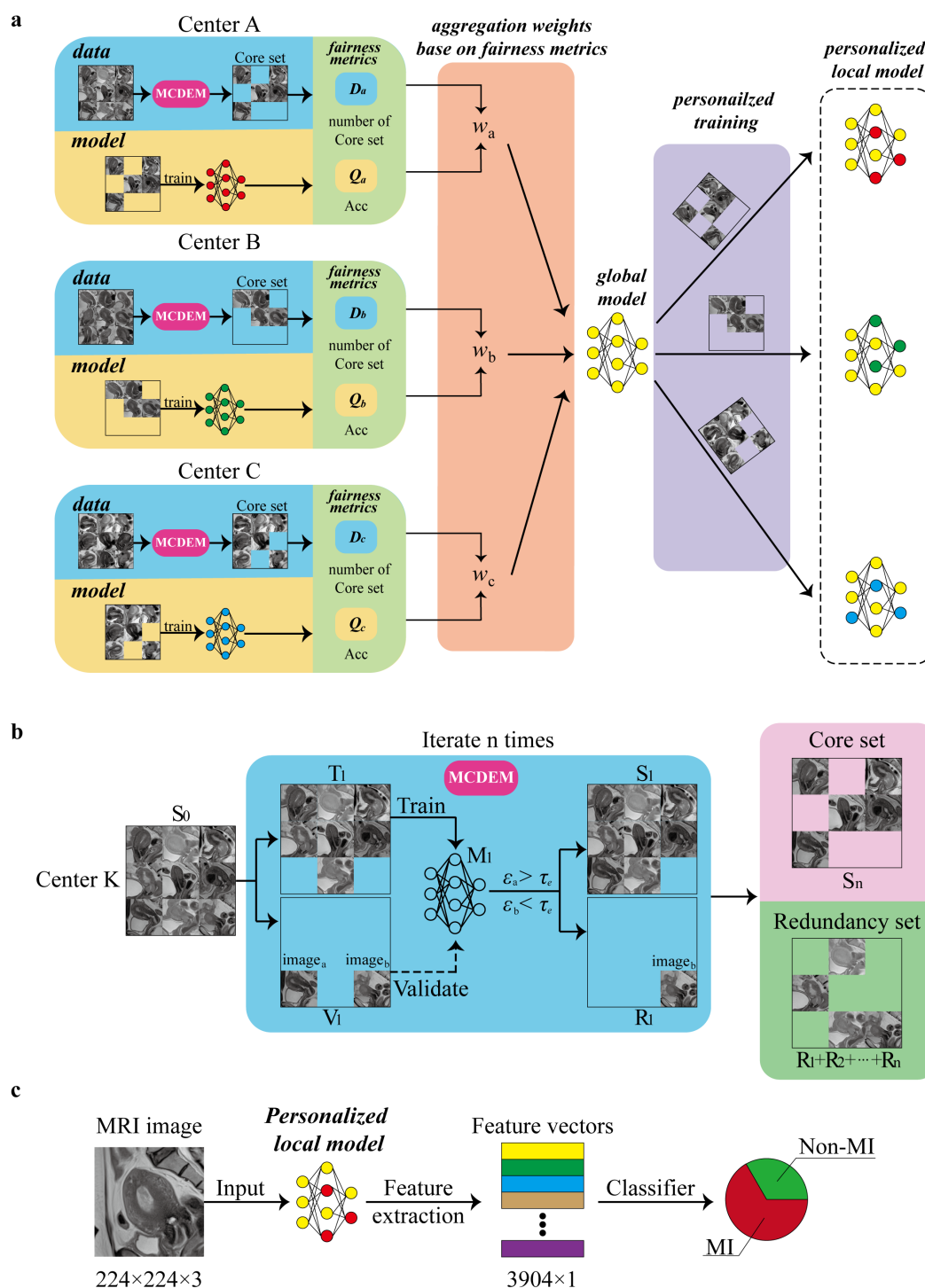
FIGURE 2

Framework of the FedCMC algorithm. **(a)** The process of constructing personalized local models in FedCMC. **(b)** Screening mechanism of the multi-center core data extraction module (MCDEM). **(c)** Feature extraction and classification in FedCMC. D represents the data richness evaluation metric; Q denotes the model quality evaluation metric; w is the contribution fairness aggregation weight; and S refers to the original training dataset; T and V represent the training and validation datasets obtained by randomly splitting S in an 8:2 ratio;is the prediction error of the samples in the validation set;is the error threshold used to identify redundant data (R); Non-MI endometrial cancer without myometrial invasion; MI endometrial cancer with myometrial invasion. A model is trained on the dataset T, and its prediction error is evaluated on the validation dataset V.

2.4.3 Personalization strategy

To address the issue of inconsistent global model performance caused by data heterogeneity among centers, we adopt a personalized federated learning strategy. Specifically, the global model is fine-tuned on each center's local data, resulting in a personalized model that better aligns with the local data distribution. This strategy effectively enhances the performance of models at all centers, particularly for low-contribution centers and those with highly heterogeneous data.

To maximize feature utilization, each center used its local model's convolutional kernels as feature extractors, extracting multiple feature maps from MRI images for each patient. The mean of these feature maps was calculated to create unified deep learning features (Supplementary Figure S3). A total of 3904 deep learning features were extracted using the 3904 convolutional kernels, which were then used to construct a classifier (Supplementary Figure S4). SBELM (25) integrates an L1-norm into the optimization of the extreme learning machine to automatically select the most informative features, yielding a sparse solution. In high-dimensional, small-sample medical settings, it demonstrates superior generalization performance compared with other classifiers; therefore, SBELM was chosen as the classifier for this study (classifier comparison in Supplementary Table S5). To address the class imbalance in this binary classification problem, focal loss (26) was used for all loss functions to alleviate potential bias caused by the imbalance between positive and negative samples.

2.5 Model evaluation and comparison

To comprehensively evaluate the performance of FedCMC in a multi-center setting, this study conducted comparative analyses with three classic federated learning models: FedAvg (18), FedProx (27), and Moon (28). To systematically assess the performance of each algorithm, quantitative metrics such as the AUC, specificity, sensitivity, accuracy (ACC), area under the Precision-Recall curve (PR-AUC), and F1-score were employed to validate the predictive outcomes of the models. The AUC metric was calculated with a 95% confidence interval (CI) to more accurately measure the robustness of the models across different datasets. These evaluation metrics provide a comprehensive perspective on the predictive capabilities of the algorithms and offer an objective basis for statistical comparisons between models. Additionally, decision curve analysis (DCA) was used to assess the clinical utility of the models in predicting MI status in endometrial cancer.

2.6 Statistical analysis

Statistical analyses were conducted using two-tailed tests, with a p-value < 0.05 considered statistically significant. All analyses were performed using R software (version 4.2.2) and IBM SPSS Statistics (version 26.0).

3 Result

3.1 FedCMC prediction of myometrial invasion in endometrial cancer patients

The experimental results demonstrate that the proposed FedCMC algorithm achieved high diagnostic performance across centers A, B, and C (Table 2). The threshold corresponding to the maximum Youden index was chosen as the optimal diagnostic threshold for the model of each center. Subsequently, we used these

TABLE 2 Diagnostic performance of FedCMC in multiple centers.

Center	DataSet	Radio	AUC (95% CI)	Accuracy	Sensitivity	Specificity	PR-AUC	F1
A	Training Cohort (n=164)	57.4% (2179/3795)	0.8429 (0.7674 - 0.9185)	0.8171 (134/164)	0.8254 (104/126)	0.7895 (30/38)	0.9312	0.8739
	test Cohort (n=200)	–	0.8261 (0.7343 - 0.9178)	0.7800 (156/200)	0.8000 (140/175)	0.6400 (16/25)	0.9583	0.8642
B	Training Cohort (n=74)	43.5% (843/1940)	0.9062 (0.8138 - 0.9986)	0.8919 (66/74)	0.9048 (57/63)	0.8182 (9/11)	0.9655	0.9344
	test Cohort (n=118)	–	0.8750 (0.7576 - 0.9924)	0.8729 (103/118)	0.8818 (97/110)	0.7500 (6/8)	0.9802	0.9282
C	Training Cohort (n=58)	88.7% (613/691)	0.9271 (0.8457 - 1.0000)	0.9138 (53/58)	0.9375 (45/48)	0.8000 (8/10)	0.9626	0.9474
	test Cohort (n=43)	–	0.8964 (0.7110 - 1.0000)	0.9535 (41/43)	0.9730 (36/37)	0.8333 (5/6)	0.9499	0.9730

Radio refers to the proportion of core dataset relative to the original training pool. For instance, in Center A, the pruned dataset contains 2,179 images compared to 3,795 images in the original dataset, accounting for 57.4%.

thresholds to calculate other performance metrics for each center. The sizes of the pruned training datasets in centers A, B, and C were 57.4%, 43.5%, and 88.7% of their original sizes, respectively. Using only the pruned training data, the models achieved AUC values of 0.8238, 0.8830, and 0.9130 on the test sets for centers A, B, and C, respectively.

3.2 Comparison of FedCMC with other algorithms

To further evaluate the performance of FedCMC, it was compared against three federated learning algorithms (Fedavg, Fedprox, and Moon). Figures 3, 4 illustrate the ROC curves (results of DeLong test, NRI, and IDI are provided in Supplementary Table S6) and DCA curves for the four algorithms across the three centers. FedCMC consistently achieved the highest AUC results in all three centers (Table 3). The results indicate that FedCMC outperformed the other three algorithms while using only

56.6% (3635/6426) of the training data, demonstrating the effectiveness of the proposed approach.

3.3 Ablation study of FedCMC

To validate the effectiveness of different components within FedCMC, ablation studies were conducted on the MCDEM and the fair aggregation mechanism. The specific AUC results are shown in Table 4.

In the ablation experiments, the comparison between Group 1 and Group 2 demonstrates that the data pruning module effectively improves the model's diagnostic performance by removing redundant data. The comparison between Group 2 and Group 3 indicates that the fairness aggregation mechanism further optimizes the global weight distribution of the model, leading to overall performance improvement across the four centers. Notably, the performance enhancement is more pronounced in high-contribution centers, aligning with the fairness principle that higher-contribution
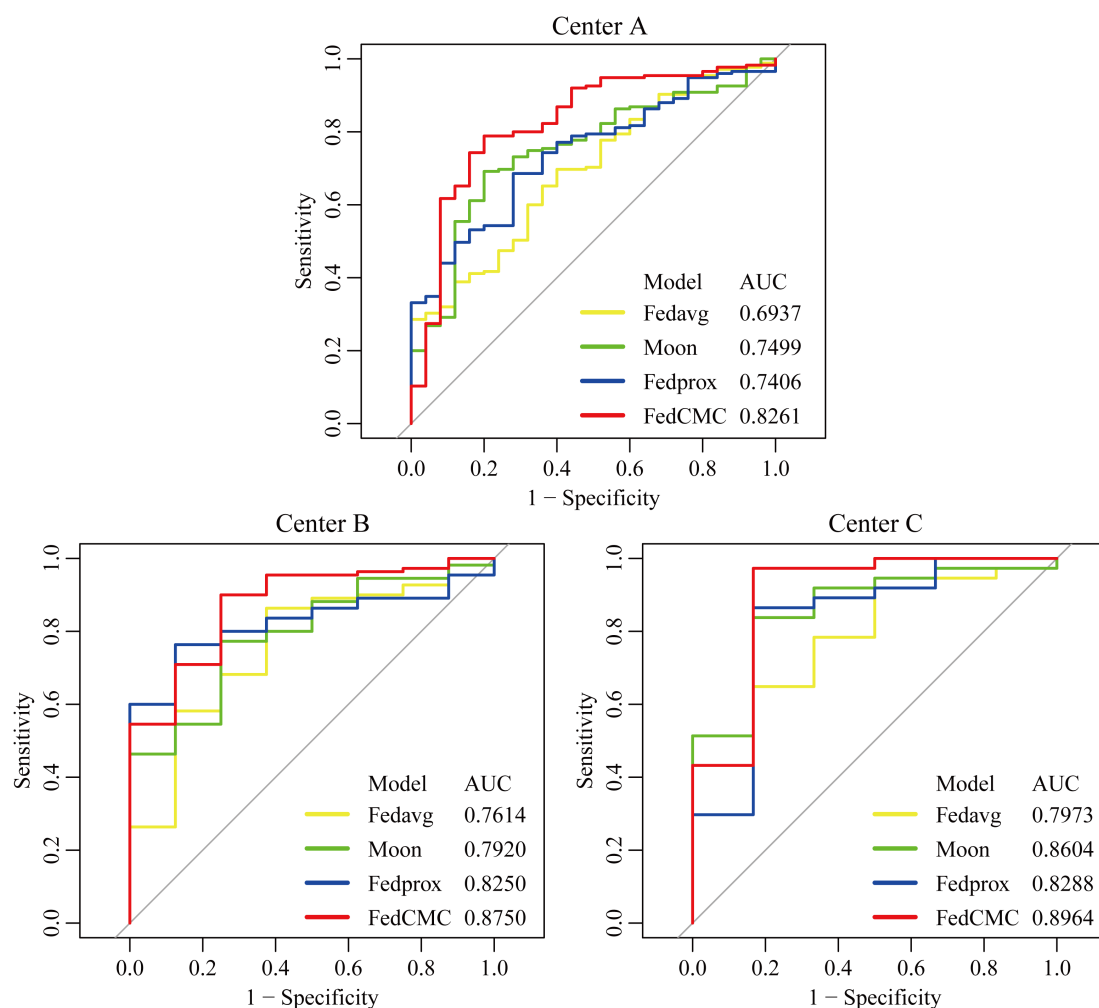

FIGURE 3
ROC curves of four algorithms across three centers. ROC receiver operating characteristic curve.
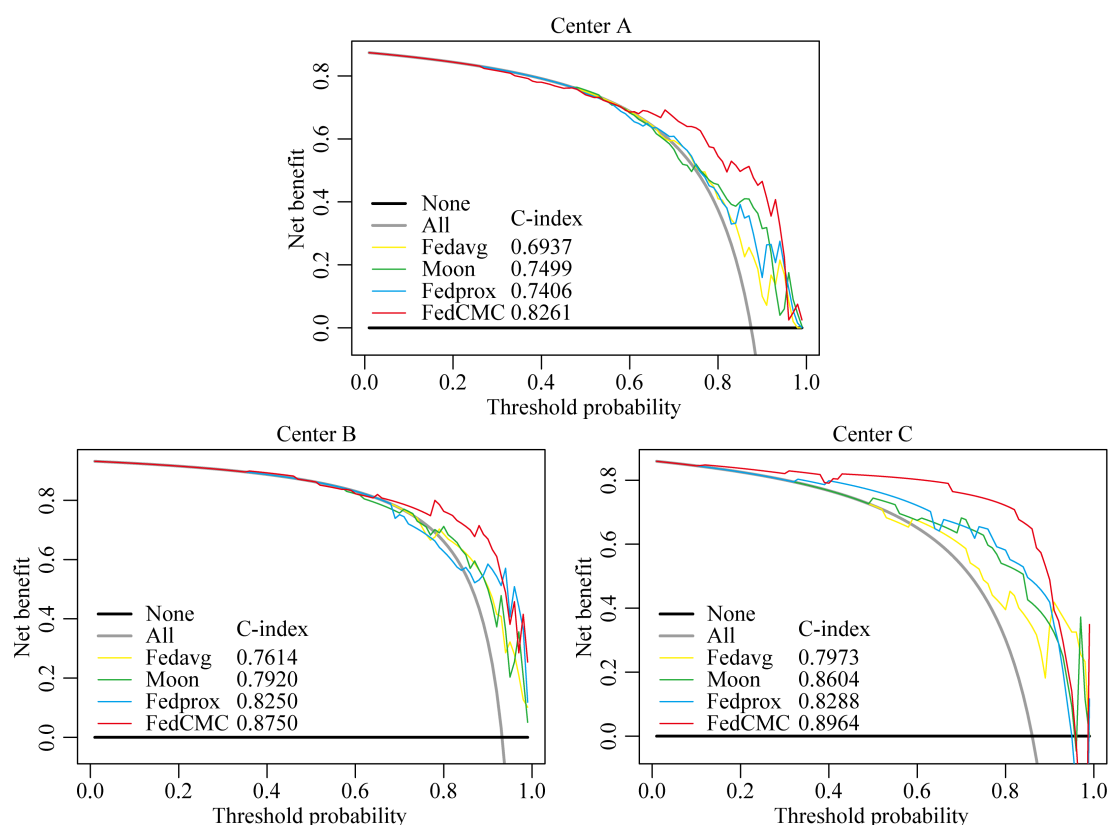
FIGURE 4

DCA curves of three models using data from three centers. The gray solid line represents the assumption that all patients belong to the MI group of endometrial cancer, while the black line assumes no patients belong to the MI group. Threshold probability represents the point where the expected benefits of treatment and avoidance of treatment are equal. FedCMC, ours algorithm; Fedavg, Moon, and Fedprox, comparison algorithms.

centers receive greater rewards. Meanwhile, the performance of low-contribution centers is not compromised. In the ablation experiments, when the fairness aggregation mechanism is not employed, the aggregation strategy defaults to average aggregation.

3.4 Analysis of FedCMC algorithm results

To visually observe the distribution of redundant samples in local data, we used Principal Component Analysis (PCA) to map the sample distributions into a two-dimensional space (29) (Figure 5). The overlapping distribution of redundant and original data indicates high similarity between them. This suggests that redundant samples may not provide additional information to the model, and their removal can mitigate model bias by reducing over-representation of repetitive information without compromising representational capacity. This visual comparison validates the effectiveness of our MCDEM, which preserves the core characteristics of the dataset while reducing redundancy, revealing the redundancy in local data at each center. Figure 6 illustrates the prediction scores of FedCMC for two categories, with p < 0.05 indicating significant differences in the model's predictions between the two labels. This demonstrates the model's effectiveness in distinguishing between MI and Non-MI statuses of EC patients.

4 Discussion

In recent studies on endometrial cancer prediction, Coada et al. extracted 107 radiomics features from contrast-enhanced CT scans of 81 patients in a single-center cohort and employed LASSO-Cox, CoxBoost, and random forest survival models to stratify postoperative recurrence risk, achieving an AUC of 0.86 – 0.90 on the test set (30). Li et al. utilized a multi-center cohort of 415 patients, integrating T2-weighted MRI radiomics features with clinical information to construct a multi-classification model for preoperative prediction of deep myometrial invasion, high-risk classification, histological subtype, and lymphovascular space invasion (LVSI), with test-set AUCs ranging from 0.79 to 0.91 (31). However, single-center studies often suffer from limited sample sizes, making it challenging to train models that are both robust and generalizable. Moreover, centralized training is difficult to implement in practice due to medical data privacy concerns, highlighting the importance of conducting multi-center collaborative studies based on federated learning. Traditional federated learning typically performs weighted aggregation of the global model based on the performance of each center's model (32). This approach may lead to the global model overfitting the data distribution of certain centers with easily distinguishable data categories, while neglecting the more complex and diverse data from other centers, thus limiting the global model's

TABLE 3 Performance comparison of four algorithms on test sets of three centers.

Center	Algorithm	AUC (95% CI)	Accuracy	Sensitivity	Specificity	PR-AUC	F1
A	Fedavg	0.6937 (0.5900 - 0.7974)	0.7000 (140/200)	0.7314 (128/175)	0.4800 (12/25)	0.9372	0.8101
	Moon	0.7499 (0.6553 - 0.8446)	0.7150 (143/200)	0.7143 (125/175)	**0.7200 (18/25)**	0.9477	0.8143
	Fedprox	0.7406 (0.6512 - 0.8300)	0.7500 (150/200)	0.7714 (135/175)	0.6000 (15/25)	0.9492	0.8438
	FedCMC	**0.8261 (0.7343 - 0.9178)**	**0.7800 (156/200)**	**0.8000 (140/175)**	0.6400 (16/25)	**0.9583**	**0.8642**
B	Fedavg	0.7614 (0.5871 - 0.9356)	0.6525 (77/118)	0.6455 (71/110)	**0.7500 (6/8)**	0.9671	0.7760
	Moon	0.7920 (0.6498 - 0.9343)	0.7091 (78/110)	0.7091 (78/110)	**0.7500 (6/8)**	0.9725	0.8211
	Fedprox	0.8250 (0.7310 - 0.9190)	0.7881 (93/118)	0.7909 (87/110)	**0.7500 (6/8)**	0.9770	0.8744
	FedCMC	**0.8750 (0.7576 - 0.9924)**	**0.8729 (103/118)**	**0.8818 (97/110)**	**0.7500 (6/8)**	**0.9802**	**0.9282**
C	Fedavg	0.7973 (0.6263 - 0.9683)	0.7209 (31/43)	0.7297 (27/37)	0.6667 (4/6)	0.9357	0.8182
	Moon	0.8604 (0.7028 - 1.0000)	0.8372 (36/43)	0.8649 (32/37)	0.6667 (4/6)	0.9468	0.9014
	Fedprox	0.8288 (0.6078 - 1.0000)	0.7907 (34/43)	0.7838 (29/37)	**0.8333 (5/6)**	0.9336	0.8657
	FedCMC	**0.8964 (0.7110 - 1.0000)**	**0.9535 (41/43)**	**0.9730 (36/37)**	**0.8333 (5/6)**	**0.9499**	**0.9730**

Bold font indicates the best metric in each column for each center.

performance. Another common method is to allocate weights based on the data volume of participating centers (18). Although this method emphasizes the importance of data scale, it overlooks differences in data quality. Redundant information at the local level can negatively impact the global model and make data contribution evaluation less accurate.

Research on fair federated learning models has made some progress in recent years but still has many shortcomings. Hosseini et al. (33) proposed Prop-FFL, which incorporates fairness constraints into the optimization objective to reduce performance gaps between different participants. Although this method partially alleviates fairness issues caused by non-independent and identically distributed (non-IID) data, its optimization objective overly emphasizes balance, potentially sacrificing the global model's

performance. Jiang et al. (34) introduced the FedCE algorithm, which estimates client contributions in both gradient and data spaces, using these estimates to assign aggregation weights while considering both model performance fairness and collaboration fairness. However, in its evaluation of contributions in the data space, this method still fails to account for redundant information that offers low or even negative contributions to the global model. Relying solely on data volume as a metric makes it difficult to evaluate each client's data contribution accurately and fairly, while redundant information can limit the overall model performance.

To address the limitations of traditional federated learning and existing fair federated frameworks, this study proposes a Contribution-Fair Federated Learning model based on multi-

TABLE 4 AUC results of the FedCMC ablation study.

MCDEM	Fair Aggregation	Data Set	Center A	Center B	Center C
		Train	0.7831	0.8562	0.9145
		Test	0.7355	0.8148	0.8649
✓		Train	0.8076	0.8470	0.9229
✓		Test	0.7721	0.8409	0.8874
✓	✓	Train	0.8429	0.9062	0.9271
✓	✓	Test	0.8261	0.8750	0.8964

MCDEM, multi-center core data extraction module; Fair Aggregation, fairness-based weighted aggregation mechanism.
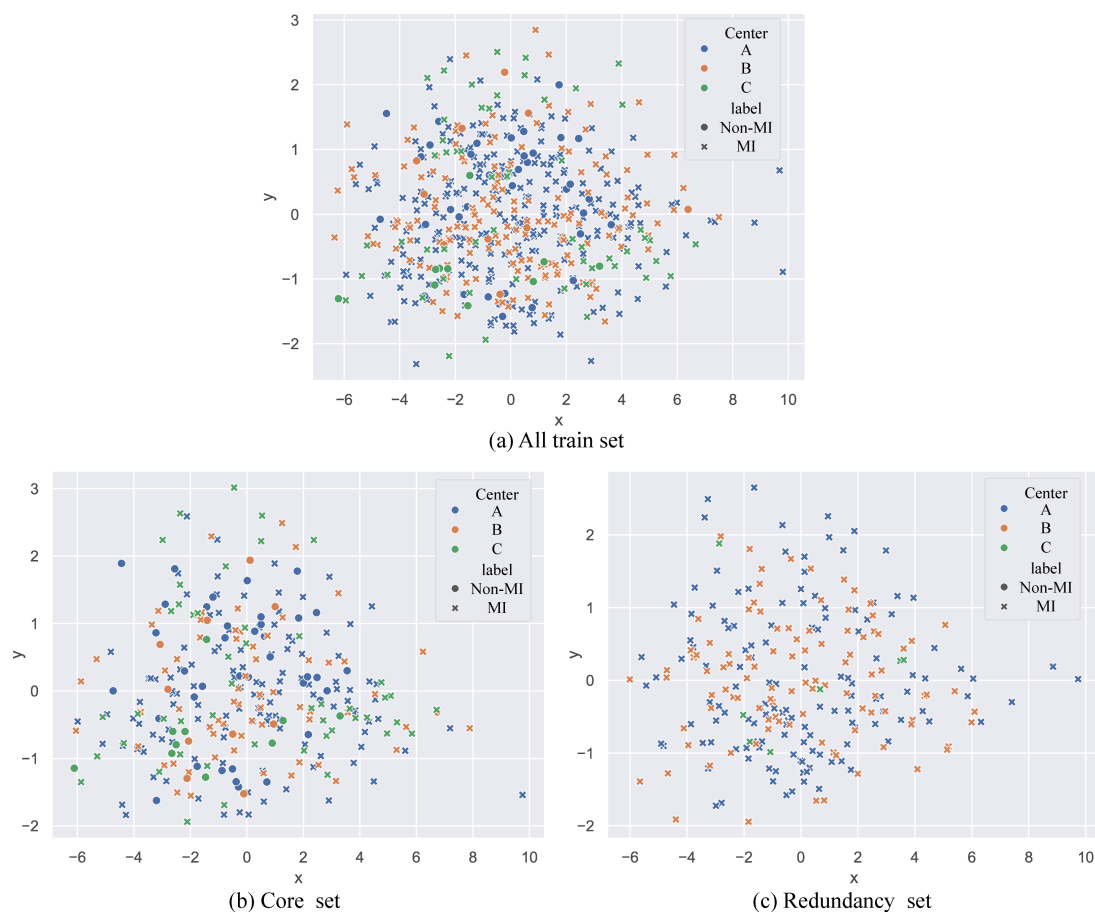
FIGURE 5
Scatter plots of training samples from three centers after PCA dimensionality reduction. **(a)** Spatial distribution of training samples in the original randomly split training sets of three centers. **(b)** The sample space distribution of the core dataset, representing the sample diversity of each center after pruning by the MCDEM. **(c)** The sample space distribution of the redundant dataset for each center after pruning by the MCDEM. Non-MI refers to endometrial cancer without myometrial invasion, and MI refers to endometrial cancer with myometrial invasion.

center core data extraction (FedCMC). The test set AUCs for Centers A, B, and C reached 0.8261, 0.8750, and 0.8964, respectively, demonstrating good generalization performance across centers. Compared with traditional algorithms such as FedAvg, Moon, and FedProx, FedCMC achieved the highest AUC performance at all three centers. On average, the performance and model fairness improved by 10.61% and 31.24%, respectively, compared to traditional federated learning algorithms (details in Supplementary Table S7). These findings indicate that FedCMC not only prioritizes fairness in multi-center medical settings, but also significantly enhances the diagnostic performance for predicting myometrial invasion status in endometrial cancer patients.

Compared with other federated algorithms, FedCMC has the following advantages: 1) More accurate and fair contribution evaluation. This study comprehensively evaluates each center's contribution based on data information richness and model quality. At the data level, instead of using data size to evaluate contributions, an innovative MCDEM was designed. The MCDEM eliminates redundant data, selecting core datasets for each center. By reducing the over-representation of redundant information, the model focuses on under-represented data, improving generalization

while enabling fairer contribution evaluations. After processing with the MCDEM, the core datasets of Centers A, B, and C accounted for 57.4%, 43.5%, and 88.7% of the original data, respectively, removing 43.4% of redundant data overall. The first and second groups in the ablation experiment in Table 4 show that using only the core datasets for model training improved the AUCs of the three centers by 3.66%, 2.61%, and 2.25%, respectively, fully validating the effectiveness of MCDEM. At the model level, the local model accuracy was used to evaluate its contribution, emphasizing high-quality models and reducing the adverse impact of low-quality models on the global model. 2) Fair and reasonable reward distribution through weighted aggregation based on two fairness metrics, D and Q. High-contribution centers often possess richer data information and better model quality. Increasing their weight in the global model aggregation improves overall performance and aligns the global model more closely with the data distribution of high-contribution centers. This ensures that high-contribution centers gain more significant performance benefits, establishing a fair and reasonable federated incentive mechanism to address collaboration fairness issues. Analysis of the second and third groups in the ablation experiment in Table 4 shows that weighted
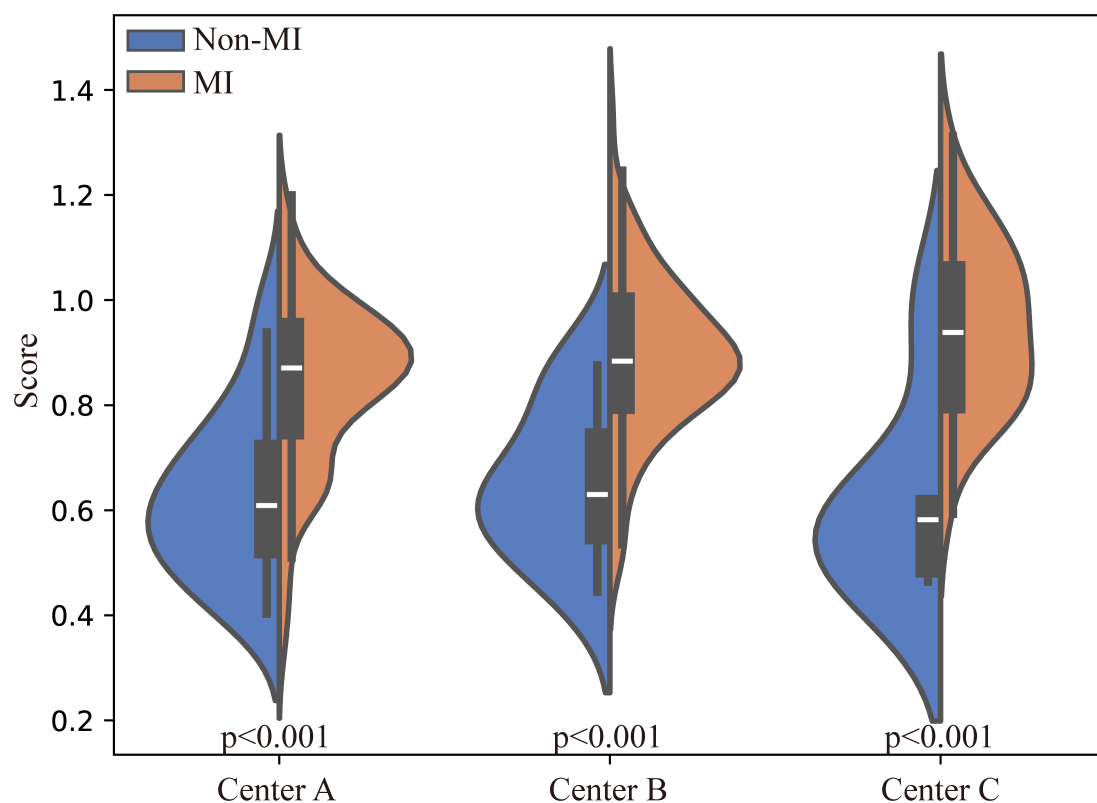
FIGURE 6
Violin plots. Illustrating the distribution of positive and negative samples in three data centers evaluated by FedCMC. Statistical test: independent t-test (two-tailed).Non-MI, no myometrial invasion; MI, myometrial invasion; p, significance value.

aggregation based on the two fairness metrics significantly improved the performance of high-contribution centers, with Center A showing the most pronounced improvement—a 5.40% increase in AUC. 3) A simple and effective personalized federated learning strategy. Each local endpoint fine-tunes the model after acquiring prior knowledge from the global model. This improves the model's adaptability to local data, mitigating inconsistencies in model performance under highly heterogeneous data scenarios.

Existing reviews categorize fairness challenges in federated learning into collaboration fairness and model fairness (35, 36). From the perspective of collaboration fairness, FedCMC improves the benefits for high-contribution centers through accurate contribution evaluation at the data and model levels and fair aggregation based on contributions, enhancing multi-center collaboration incentives. From the perspective of model fairness, FedCMC improves performance consistency compared to three traditional federated learning algorithms through a simple personalized federated learning strategy. Additionally, we believe that collaboration fairness and model fairness are not entirely independent but mutually reinforcing under a fair federated framework. Addressing collaboration fairness promotes consistency in model performance, while tackling model fairness in turn incentivizes more institutions to participate in federated learning. This mutual reinforcement is crucial for building more

robust federated learning models, thereby offering stronger support for clinical preoperative prediction and personalized treatment.

Despite the achievements of this study, there are still limitations. The current research mainly focuses on client-level fairness and has not deeply explored attribute-level fairness at the sample level (35), nor has it investigated the issue of communication overhead in federated learning. Additionally, the dataset used in this study is relatively limited in scale. Future research could explore attribute-level fairness based on larger and more diverse datasets to improve the existing fairness framework and design more robust and fair federated learning models.

5 Conclusion

To address preoperative myometrial invasion prediction in endometrial cancer and fairness concerns in medical federated learning, we propose FedCMC, a contribution-fair federated learning model based on multi-center core data extraction. By more accurately and fairly quantifying each center's contribution at both the data and model levels, and employing a contribution-based adaptive aggregation strategy, FedCMC places greater emphasis on high-contributing centers to enhance fairness and incentivize broader participation. Experiments on preoperative

myometrial invasion prediction demonstrate that FedCMC yields more pronounced performance gains for high-contributing centers. Compared with three traditional federated learning algorithms, FedCMC not only alleviates fairness issues in federated learning but also enhances classification performance in preoperative myometrial invasion prediction, offering potential technical support for personalized treatment of EC patients.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors (jmlws2@163.com).

Ethics statement

This study was approved by the Medical Ethics Committee of Jiangmen Central Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

YL: Methodology, Writing – review & editing, Investigation, Writing – original draft, Software. BF: Methodology, Writing – review & editing, Funding acquisition, Resources. YC: Data curation, Resources, Writing – review & editing. XR: Data curation, Writing – review & editing, Resources. JFS: Writing – review & editing, Software. XMW: Data curation, Writing – review & editing. XYW: Data curation, Writing – review & editing. PL: Writing – review & editing, Data curation. JQS: Writing – review & editing, Data curation. CZ: Data curation, Writing – review & editing. YZ: Writing – review & editing, Data curation. ML: Writing – review & editing, Data curation. WL: Resources, Data curation, Writing – review & editing. YHC: Writing – review & editing, Methodology. DX: Funding acquisition, Writing – review & editing, Supervision.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2025.1648502/full#supplementary-material

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2024) 74:229–63. doi: 10.3322/caac.21834

2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2021) 71:209–49. doi: 10.3322/caac.21660

3. Lu KH, Broaddus RR. Endometrial cancer. *N Engl J Med*. (2020) 383:2053–64. doi: 10.1056/NEJMra1514010

4. Leone Roberti Maggiore U, Khamisy-Farah R, Bragazzi NL, Bogani G, Martinelli F, Lopez S, et al. Fertility-sparing treatment of patients with endometrial cancer: A review of the literature. *J Clin Med*. (2021) 10:4784. doi: 10.3390/jcm10204784

5. Jónsdóttir B, Marcickiewicz J, Borgfeldt C, Bjurberg M, Dahm-Kähler P, Flöter-Rådestad A, et al. Preoperative and intraoperative assessment of myometrial invasion in endometrial cancer—A Swedish Gynecologic Cancer Group (SweGCG) study. *Acta Obstet Gynecol Scand*. (2021) 100:1526–33. doi: 10.1111/aogs.14146

6. Concin N, Matias-Guiu X, Vergote I, Cibula D, Mirza MR, Marnitz S, et al. ESGO/ESTRO/ESP guidelines for the management of patients with endometrial carcinoma.

Int J gynecological cancer: Off J Int Gynecological Cancer Soc. (2021) 31:12–39. doi: 10.1136/ijgc-2020-002230

7. Contreras NA, Sabadell J, Verdaguer P, Julià C, Fernández-Montolí ME. Fertility-sparing approaches in atypical endometrial hyperplasia and endometrial cancer patients: current evidence and future directions. *Int J Mol Sci*. (2022) 23:2531. doi: 10.3390/ijms23052531

8. Sharma AK, Tiwari S, Aggarwal G, Goenka N, Kumar A, Chakrabarti P, et al. Dermatologist-level classification of skin cancer using cascaded ensembling of convolutional neural network and handcrafted features based deep neural network. *IEEE Access*. (2022) 10:17920–32. doi: 10.1109/ACCESS.2022.3149824

9. Li B, Chen H, Zhang B, Yuan M, Jin X, Lei B, et al. Development and evaluation of a deep learning model for the detection of multiple fundus diseases based on colour fundus photography. *Br J Ophthalmol*. (2022) 106:1079–86. doi: 10.1136/bjophthalmol-2020-316290

10. Li F, Wang Y, Xu T, Dong L, Yan L, Jiang M, et al. Deep learning-based automated detection for diabetic retinopathy and diabetic macular oedema in retinal fundus photographs. *Eye*. (2022) 36:1433–41. doi: 10.1038/s41433-021-01552-8

11. Bleker J, Yakar D, van Noort B, Rouw D, de Jong IJ, Dierckx RAJO, et al. Single-center versus multi-center biparametric MRI radiomics approach for clinically significant peripheral zone prostate cancer. *Insights into Imaging*. (2021) 12:1–11. doi: 10.1186/s13244-021-01099-y

12. Voigt P, Von dem Bussche A. *The eu general data protection regulation (gdpr). A Practical Guide. 1st Ed* Vol. 10. . Cham: Springer International Publishing (2017) p. 10–5555. doi: 10.1007/978-3-319-57959-7

13. Pacheco SAB. (2024). A comprehensive survey on federated learning and its applications in health care, in: *2024 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, Kota Kinabalu, Malaysia: Institute of Electrical and Electronics Engineers (IEEE). pp. 407–12. doi: 10.1109/IICAIET62352.2024.10730687

14. Upreti D, Yang E, Kim H, Seo C. A comprehensive survey on federated learning in the healthcare area: Concept and applications. *CMES–Comput Modeling Eng &Sciences*. (2024) 140(3):2239–74. doi: 10.32604/cmes.2024.048932

15. Li T, Sahu AK, Talwalkar A, Smith V. Federated learning: Challenges, methods, and future directions. *IEEE Signal Process magazine*. (2020) 37:50–60. doi: 10.1109/MSP.2020.2975749

16. Wang S, Ji M. A unified analysis of federated learning with arbitrary client participation. *Adv Neural Inf Process Syst*. (2022) 35:19124–37. doi: 10.48550/arXiv.2205.13648

17. Deng Y, Lyu F, Ren J, Chen YC, Yang P, Zhou Y, et al. Improving federated learning with quality-aware user incentive and auto-weighted model aggregation. *IEEE Trans Parallel Distributed Syst*. (2022) 33:4515–29. doi: 10.1109/TPDS.2022.3195207

18. McMahan B, Moore E, Ramage D, Hampson S, Arcas BAY. (2017). Communication-efficient learning of deep networks from decentralized data, in: *Artificial intelligence and statistics*, . pp. 1273–82.

19. Tan AZ, Yu H, Cui L, Yang Q. Towards personalized federated learning. *IEEE Trans Neural Networks Learn Syst*. (2022) 34:9587–9603. doi: 10.1109/TNNLS.2022.3160699

20. Pecorelli S. Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. *Int J Gynaecology Obstetrics*. (2009) 105:103–104. doi: 10.1016/j.ijgo.2009.02.012

21. Nougaret S, Horta M, Sala E, Lakhman Y, Thomassin-Naggara I, Kido A, et al. Endometrial cancer MRI staging: updated guidelines of the European Society of Urogenital Radiology. *Eur Radiol*. (2019) 29:792–805. doi: 10.1007/s00330-018-5515-y

22. Nakao Y, Yokoyama M, Hara K, Koyamatsu Y, Yasunaga M, Araki Y, et al. MR imaging in endometrial carcinoma as a diagnostic tool for the absence of myometrial invasion. *Gynecologic Oncol*. (2006) 102:343–347. doi: 10.1016/j.ygyno.2005.12.028

23. Prakash NB, Murugappan M, Hemalakshmi GR, Jayalakshmi M, Mahmud M. Deep transfer learning for COVID - 19 detection and infection localization with superpixel based segmentation. *Sustain cities Soc*. (2021) 75:103252. doi: 10.1016/j.scs.2021.103252

24. Dong S, Yang Q, Fu Y, Tian M, Zhuo C. RCoNet: Deformable mutual information maximization and high-order uncertainty-aware learning for robust COVID - 19 detection. *IEEE Trans Neural Networks Learn Syst*. (2021) 32:3401–3411. doi: 10.1109/TNNLS.2021.3086570

25. Luo J, Vong CM, Wong PK. Sparse Bayesian extreme learning machine for multi-classification. *IEEE Trans Neural Networks Learn Syst*. (2013) 25:836–843. doi: 10.1109/TNNLS.2013.2281839

26. Ross TY, Dollár GKHP. (2017). Focal loss for dense object detection, in: *proceedings of the IEEE conference on computer vision and pattern recognition*. Institute of Electrical and Electronics Engineers (IEEE). pp. 2980–8. doi: 10.48550/arXiv.1708.02002

27. Li T, Sahu AK, Zaheer M, Sanjabi M, Talwalkar A, Smith V. Federated optimization in heterogeneous networks. *Proc Mach Learn Syst*. (2020) 2:429–450. doi: 10.48550/arXiv.1812.06127

28. Li Q, He B, Song D. (2021). Model-contrastive federated learning, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Institute of Electrical and Electronics Engineers (IEEE). pp. 10713–22. doi: 10.48550/arXiv.2103.16257

29. Abdi H, Williams LJ. Principal component analysis. *Wiley Interdiscip reviews: Comput Stat*. (2010) 2:433–459. doi: 10.1002/wics.101

30. Coada CA, Santoro M, Zybin V, Di Stanislao M, Paolani G, Modolon C, et al. A radiomic-based machine learning model predicts endometrial cancer recurrence using preoperative CT radiomic features: A pilot study. *Cancers*. (2023) 15:4534. doi: 10.3390/cancers15184534

31. Li X, Dessi M, Marcus D, Russell J, Aboagye EO, Ellis LB, et al. Prediction of deep myometrial infiltration, clinical risk category, histological type, and lymphovascular space invasion in women with endometrial cancer based on clinical and T2–weighted MRI radiomic features. *Cancers*. (2023) 15:2209. doi: 10.3390/cancers15082209

32. Huang C, Chen W, Chen Y, Yang S, Zhang Y. DearFSAC: an approach to optimizing unreliable federated learning via deep reinforcement learning. *arXiv preprint arXiv:2201.12701*. (2022) 5279–84. doi: 10.48550/arXiv.2201.12701

33. Hosseini SM, Sikaroudi M, Babaie M, Tizhoosh HR. Proportionally fair hospital collaborations in federated learning of histopathology images. *IEEE Trans Med Imaging*. (2023) 42:1982–1995. doi: 10.1109/TMI.2023.3234450

34. Jiang M, Roth HR, Li W, Yang D, Zhao C, Nath V, et al. (2023). Fair federated medical image segmentation via client contribution estimation, in: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Institute of Electrical and Electronics Engineers (IEEE). pp. 16302–11. doi: 10.1109/CVPR52729.2023.01564

35. Zhu Z, Si S, Wang J, Cheng N, Kong L, Huang Z, et al. A survey on the fairness of federated learning. *Big Data Res*. (2024) 10:62–85. doi: 10.11959/j.issn.2096-0271.2022088

36. Huang W, Ye M, Shi Z, Wan G, Li H, Du B, et al. Federated learning for generalization, robustness, fairness: A survey and benchmark. *IEEE Trans Pattern Anal Mach Intell*. (2024) 46:9387–406. doi: 10.1109/TPAMI.2024.3418862