Future Technology
Open Access Journal
ISSN 2832-0379

August 2026| Volume 05 | Issue 03 | Pages 263-275

Journal homepage: https://fupubco.com/futech

https://doi.org/10.55670/fpll.futech.5.3.23

Future Publishing LLC

Article

Predicting silver economy demand during population aging transition via federated learning-based multi-platform behavioral data collaboration

Qing He[1,2], Mustazar Mansur[1*], Hazrul Izuan Bin Shahiri[1]

[1]School of Economics and Management, The National University of Malaysia (UKM), Selangor 43600, Malaysia
[2]School of Economics, Sichuan University of Science & Engineering, Zigong 643000, China

ARTICLE INFO

Article history:
Received 15 December 2025
Received in revised form
10 May 2026
Accepted 16 June 2026

Keywords:
Federated learning, Silver economy, Demand prediction, Financial technology, Multi-platform behavioral data, Population aging

*Corresponding author
Email address:
mustazar@ukm.edu.my

DOI: 10.55670/fpll.futech.5.3.23

ABSTRACT

The accelerating global population aging has fueled a surge in financial demands in the silver economy, making it critical to forecast elderly financial demands accurately for product allocation and risk management in institutions. The study proposes the Federated Learning-based Silver Economy Prediction Framework (FL-SEPF), which is a collaborative framework for privacy-preserving prediction in the silver economy. FL-SEPF features a four-layer architecture with an adaptive weighted federated aggregation strategy that dynamically computes platform-specific weights based on data volume, data quality, and local validation loss to address Non-IID heterogeneity. Local models employ BiLSTM with attention mechanisms, and a cross-platform attention fusion module integrates multi-dimensional features for multi-task prediction covering demand type classification and intensity regression. Differential privacy and Top-K gradient sparsification ensure privacy protection and communication efficiency. Experiments on four-platform datasets, covering 185,000–203,000 elderly users, demonstrate that FL-SEPF achieves an F1-score of 0.8312±0.0047 and AUC-ROC of 0.8927±0.0038, outperforming FedAvg by 3.5% and XGBoost by 11.4% in F1-score, with only a 1.7% gap compared to centralized Transformer training. Ablation studies confirm that adaptive weighted aggregation contributes the largest performance gain (3.34% F1 drop upon removal). Under extreme Non-IID conditions, FL-SEPF shows only 5.5% F1 degradation versus 11.3% for FedAvg, and at a privacy budget $\varepsilon = 1.0$, performance loss is limited to approximately 1.0%. SHAP analysis reveals that financial behavior features, particularly portfolio diversity index and credit utilization rate, are the dominant predictors. This study provides a systematic federated learning solution for silver economy demand prediction under privacy-compliant conditions.

1. Introduction

The current global population is facing an unprecedented trend of aging. The United Nations Population Fund estimates that more than 2.1 billion people worldwide will be over 60 years old by 2050. The share of such people was reported at 13.5% of the overall population in 2020, and it is estimated to reach 21.3% in 2050. This structural shift carries far-reaching implications for pension wealth management, health insurance, consumer credit, and wealth management [1]. The silver economy has expanded beyond traditional retirement security to cover consumer finance, digital payments, robo-advisory services, and other fintech applications. Predicting the financial needs of the elderly population with reasonable accuracy has become a key basis for financial institutions seeking to optimize product allocation and for regulators assessing systemic risk [2]. Behavioral information about the elderly is distributed through various channels, including e-commerce, healthcare, interpersonal interactions, and financial services, thus creating demand information from different dimensions [3]. Privacy regulations such as the EU General Data Protection Regulation and China's Personal Information Protection Law explicitly restrict direct cross-platform data sharing and centralized processing, thereby preventing traditional centralized machine learning methods from effectively using multi-source, heterogeneous data [4]. Federated learning is a data privacy-preserving technology for distributed learning, allowing each learner to train the model locally and share only encrypted model parameters with other learners. This provides a technical solution to overcome the limitation of the data silo and achieve the collaboration of multi-platform behavioral data modeling [5]. In silver economy demand prediction, existing research has predominantly relied upon macro-level demographic factors and survey methods. These methods have limitations in making micro-level predictions and explaining heterogeneity in elderly demand patterns arising from differences in consumption behaviors and social networks [6]. Conventional approaches to financial demand forecasting using time series are constrained by assumptions of linearity and stationarity, which limit their ability to identify complex nonlinear behavioral patterns [7]. Deep learning models have shown higher predictive capability compared to conventional models in asset pricing as well as modeling consumer behavior [8]. While data-driven predictive models have been validated in financial fraud detection [9], they were developed for general consumers and not tailored to the specific behavioral patterns of the elderly population.

Federated learning has made great strides in financial technology for use cases such as credit assessment [10], fraud detection [11], and stock market prediction [12]. Improved learning mechanisms, such as FedProx, overcome convergence problems due to heterogeneity in the data set [13], and improved model architectures now include Transformers for analyzing financial sequences [14]. However, multi-platform federated learning still faces challenges including Non-IID-induced model drift, communication bottlenecks, and cross-platform heterogeneity, with existing solutions addressing individual challenges without systematic domain-specific optimization [15]. Based on the above analysis, three research gaps remain. First, federated learning has not been systematically applied to predicting demand in the silver economy. Second, existing privacy-preserving solutions lack aggregation strategies for heterogeneous multi-domain platform data [16]. Third, current federated recommendation systems do not account for the behavioral specificity of the elderly population [17]. Adaptive federated learning systems in supply chain scenarios have provided useful technical insights for dynamic information sharing, but their problem structure differs fundamentally from that of silver economy demand prediction [18]. To the best of our knowledge, no existing study has proposed a federated learning-based multi-platform collaborative framework specifically designed for predicting financial demand in the silver economy in the context of the demographic aging transition.

This study addresses these limitations using FL-SEPF. It also makes four major contributions. At the framework level, a four-layer federated collaborative prediction architecture that integrates data from e-commerce, healthcare, social, and financial platforms is proposed. At the algorithm level, an adaptive weighted aggregation strategy is designed for Non-IID elderly behavioral data. At the feature-engineering level, SHAP-based analysis reveals distinct contributions across four behavioral dimensions. At the experimental level, comprehensive validation of prediction accuracy, privacy protection, and communication efficiency is conducted. Several advanced federated aggregation strategies have been proposed to address Non-IID data heterogeneity beyond the standard FedAvg. The FedProx algorithm uses proximal regularization to constrain the local update process, ensuring no divergent updates away from the global parameter [13]. The algorithm FedNova scales the local updates by the number of local iterations performed, thereby resolving objective disagreement caused by heterogeneous computation [19]. The SCAFFOLD method applies control variates to compensate for client drift during local training [20]. Dynamic regularization is employed in FedDyn to ensure objective consistency [21]. In MOON, model-contrastive learning improves local training by contrasting representations between rounds [22]. These strategies address data or system heterogeneity primarily through regularization, normalization, or representation alignment. In contrast, the idea of adaptive weighted aggregation introduced in the FL-SEPF framework combines data-quality measurement with the local model's fitness to generate the required aggregation weights. Hence, a specific solution is provided for the silver economy problem, in which each platform can vary in its data distribution, feature space, completeness, and time period covered. Table 6 in Section 3.1 includes FedAvg and FedProx as baselines; the remaining methods are discussed comparatively in Section 3.6. The remainder of the study is arranged as follows: Section 2 describes the materials and methodology used; Section 3 discusses the findings and results; and Section 4 presents the conclusions and future directions for the study.

2. Materials and methods

2.1 Problem formulation

Silver economy demand forecasting is formulated as a multi-platform collaborative optimization problem. Given K platforms $P = \{P_1, P_2, \dots P_K\}$, each holding a local dataset $D_k\{(x_i, y_i)\}$ with feature vector x_i and demand label y_i (including categorical demand types and continuous intensity values), the federated learning objective is:

$$min_w F(w) = \sum_{k=1}^{K} \frac{n_k}{n} F_k(w) \tag{1}$$

Where $n = \sum_k n_k$ and $F_k(w) = \frac{1}{n_k} \sum_{i=1}^{n_k} l(f(x_i; w), y_i)$ is the local loss on platform P_k. Raw data D_k must remain local; only privacy-protected parameter updates Δw_k are transmitted. The task decomposes into multi-label demand-type classification and demand intensity regression, jointly optimized via multi-task learning.

2.2 System architecture overview

To solve the above optimization problem, this study proposes the FL-SEPF (Federated Learning-based Silver Economy Prediction Framework). Its overall architecture adopts a four-layer design, as shown in Figure 1. The data layer encompasses four heterogeneous platform data sources. The local computation layer handles model training, feature extraction, and differential privacy noise injection on each platform. The federated collaboration layer applies adaptive weighted aggregation to the encrypted parameters of local models to obtain the global model. The prediction output layer is responsible for classification, regression, and ranking activities. Encrypted channels connect all layers to ensure end-to-end privacy preservation.

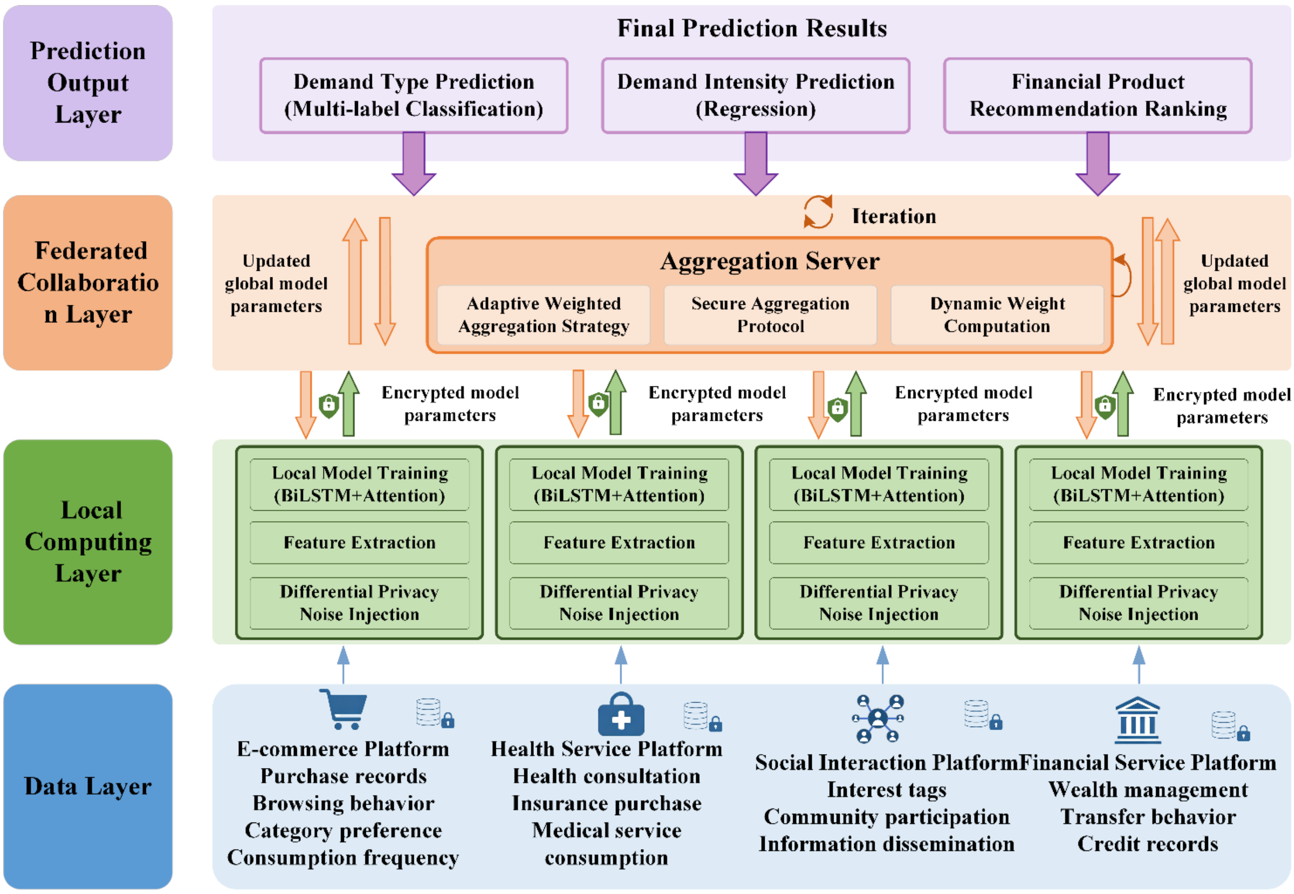

Figure 1. Overall architecture of the FL-SEPF framework

2.3 Data sources and feature engineering

The experimental dataset used in this study was provided by four partner institutions after anonymization and included elderly behavioral data across four domains: e-commerce, healthcare services, social interaction, and financial services. For the purpose of protecting trade secrets, the specific names of partner institutions are not disclosed. The four institutions include a large-scale e-commerce platform, a regional healthcare information service provider, a social media platform, and a commercial banking group, all operating within East China. The user base predominantly consists of urban elderly residents aged 60–85. All data were de-identified by the partner institutions prior to transfer, and written authorization for data use was obtained from each institution. As the study exclusively uses pre-existing anonymized commercial behavioral records with no direct interaction with human subjects, institutional ethics review board approval was not required. The data foundation of the FL-SEPF framework draws from four representative elderly behavioral data platforms, and the data volume, time span, and core fields of each platform are shown in Table 1.

The demand type classification task covers six categories: pension wealth management, health insurance, consumer credit, digital payment services, investment advisory, and comprehensive wealth management, defined based on existing silver economy product taxonomies and validated by domain experts from the partner institutions. Labels were assigned retrospectively based on each user's actual financial product purchase or inquiry records within the most recent six months of the observation period.

For the demand intensity regression task, intensity was quantified as a normalized composite score (0–1) combining inquiry frequency, normalized purchase amount, and service engagement duration, with component weights determined through expert consultation and validated against actual product adoption outcomes in a holdout time window. FL-SEPF operates under a horizontal federated learning paradigm without cross-platform user alignment. The four platforms manage separate user populations with distinct account systems, and no user linkage was performed. The cross-platform attention fusion module integrates platform-level feature representations rather than individual user-level records, avoiding the privacy risks associated with user matching in vertical federated learning. To measure heterogeneity across platforms, the pairwise Earth Mover's Distance (EMD) was calculated for the labeling distributions, as shown in Table 2. The distances range from 0.231 to 0.523, representing significant Non-IID heterogeneity. The e-commerce and financial services platforms show the smallest difference (EMD = 0.231), reflecting a higher level of business relevance to financial requirements compared to other platforms, whereas the social interactions and health services platforms show the largest difference (EMD = 0.523). The mean majority-to-minority class imbalance ratio ranges from 4.2:1 for financial services to 7.8:1 for social interaction platforms.

Table 1. Multi-platform data source description

Platform Type	Records	Time Span	Users	Feature Dimensions
E-commerce Consumption	2,850,000	2019.01–2024.06	185,000	126
Health Service	1,620,000	2019.01–2024.06	142,000	98
Social Interaction	3,210,000	2019.01–2024.06	203,000	87
Financial Service	1,980,000	2019.01–2024.06	168,000	115

Table 2. Pairwise EMD of label distributions across platforms

Platform Pair	EMD
E-commerce ↔ Health Service	0.347
E-commerce ↔ Social Interaction	0.478
E-commerce ↔ Financial Service	0.231
Health Service ↔ Social Interaction	0.523
Health Service ↔ Financial Service	0.394
Social Interaction ↔ Financial Service	0.461

Each channel generates unique behaviors. In particular, the e-commerce platform reveals spending power and purchasing behavior, while the healthcare platform tracks health management actions and insurance purchase decisions. The social interaction platform represents behavior related to the level of socializing and acquisition of information through the act of participating in communities and sharing information, and finally the finance platform directly offers information on financial behaviors such as risk preference, asset allocation, and credit status. Detailed data fields are listed in Table 1.

Data preprocessing includes iterative multiple imputation for handling missing data, outlier removal using the interquartile range (IQR), and Min-Max normalization. Min-Max normalization was opted for due to the previous elimination of outliers using the IQR method that significantly reduces the effect of outliers. Hence, using RobustScaler had negligible effects on performance (F1: 0.8307 vs. 0.8296). In the feature extraction phase, four feature sets are created based on the platform's four data types: consumption behavior, health demand, social activity, and financial behavior. Table 3 presents the dimensions and the representative features of each of the above categories. In the feature selection phase, the SHAP algorithm [23] is used to order the importance of all features. This method relies on the Shapley Value principle in cooperative game theory, which assesses the incremental gain each feature contributes to the outcome.

The top features per platform are retained by identifying the elbow point on the cumulative SHAP importance curve, defined as the point beyond which each additional feature contributes less than 1% to the cumulative importance. This yielded N = 42, 35, 28, and 38 for the e-commerce, health service, social interaction, and financial service platforms, respectively. SHAP-based selection was performed locally on each platform using its own training data to avoid cross-platform data leakage; only feature index lists (not feature values) were communicated to the server. After all preprocessing steps, 143 effective input features remain for the final model (Table 3). To address the class imbalance problem commonly found in multi-platform data—where the proportion of high-frequency financial service users among the elderly is significantly lower than that of low-frequency users—this study applies the Synthetic Minority Over-sampling Technique to augment minority-class samples at the local data level for each platform [24]. This mitigates the negative impact of class imbalance on model training while ensuring that the over-sampling process is completed entirely locally without violating privacy constraints. While local SMOTE may alter platform-specific label distributions, the primary source of Non-IID heterogeneity in this study is cross-platform differences in the feature space rather than label skew alone, limiting SMOTE's impact on inter-platform heterogeneity. Alternative imbalance-handling approaches were also evaluated: in preliminary experiments, SMOTE achieved an F1-score of 0.8305, compared with 0.8261 for focal loss and 0.8213 for class-weighted cross-entropy, supporting its selection for the final framework.

2.4 Federated learning framework design

The local model of the FL-SEPF framework adopts an architecture that combines Bidirectional Long Short-Term Memory (BiLSTM) networks with an attention mechanism to capture temporal dependencies in elderly users' behavioral sequences. Behavioral data were aggregated into monthly feature vectors over a sliding window of T = 12 months per user. Monthly granularity was chosen to balance temporal resolution with data sparsity for low-activity elderly users. BiLSTM was preferred over Transformer architectures for three reasons: (1) the short sequence length (T = 12) limits Transformer's advantage in long-range dependency modeling; (2) BiLSTM has fewer parameters, making it more communication-efficient in federated settings; and (3) in preliminary validation on our datasets, BiLSTM achieved comparable performance to a Transformer encoder (F1: 0.8305 vs. 0.8256), likely due to the moderate sequence length and structured temporal patterns in elderly behavioral data.

Table 3. Feature categories and dimensions

Feature Category	Source Platform	Number of Features	Representative Features
Consumption Behavior	E-commerce	42	Monthly spending amount, purchase frequency, category diversity index
Health Demand	Health Service	35	Consultation frequency, insurance policy count, medical expenditure ratio
Social Activity	Social Interaction	28	Community participation rate, content sharing frequency, interest tag entropy
Financial Behavior	Financial Service	38	Portfolio diversity index, credit utilization rate, risk preference score

BiLSTM captures bidirectional temporal dependencies through its gating mechanisms [25], and the concatenated forward and backward hidden states provide comprehensive contextual representations. An attention mechanism [26] is then applied to weight hidden states across time steps, enabling the model to focus on key behavioral events. The local model's loss function is a weighted composite of cross-entropy loss (for demand-type classification) and mean squared error loss (for demand-intensity regression).

$$L_{local} = \lambda . L_{CE} + (1 - \lambda) . L_{MSE} \tag{2}$$

where λ is a balancing coefficient that adjusts the relative influence of the classification and regression tasks during training. Its optimal value is determined through grid search on the validation set.

The conventional FedAvg algorithm weights platforms in proportion to their data volumes, which has theoretical guarantees under IID assumptions. However, under Non-IID conditions, this approach biases the global model toward large but poorly representative platforms [27]. To tackle this problem, this study proposes an adaptive weighted federated aggregation method in which each platform's weight for aggregation is determined in an adaptive way by considering three factors: the data volume factor $|D_k|$ represents the platform's sample contribution, the data quality score Q_k is defined in detail below, and the local validation loss L_k measures how well the current global model fits each platform's local validation set. Specifically, Q_k is computed as $Q_k = \frac{C_k + B_k + T_k}{3}$, where C_k is the feature completeness ratio (proportion of non-missing values across all samples and feature dimensions on platform P_k), $B_k = \frac{H(y_k)}{H_{max}}$ is the class balance score based on the ratio of the Shannon entropy of the local label distribution to the maximum entropy under uniform distribution, and T_k is the temporal continuity score measuring the proportion of users whose behavioral records cover at least 75% of the observation period. All three sub-indicators are normalized to [0, 1]. The adaptive weight is computed as:

$$\alpha_k = \frac{f(|D_k|, Q_k, L_k)}{\sum_{j=1}^{K} f(|D_j|, Q_j, L_j)} \tag{3}$$

where the fusion function is defined as:

$$f(|D_k|, Q_k, L_k) = (\frac{|D_k|}{n})^{\beta_1} . Q_k^{\beta_2} . \exp(-\beta_3 . L_k) \tag{4}$$

β_1, β_2, and β_3 are tunable hyperparameters that control the relative influence of the data volume, data quality, and model fitness factors, respectively. The power functions for data volume and quality capture diminishing returns and quality importance, respectively, while the negative exponential for validation loss down-weights platforms where the global model fits poorly. The global model is updated after each communication round according to the following rule:

$$w^{(t+1)} = \sum_{k=1}^{K} \alpha_k^{(t)} . w_k^{(t)} \tag{5}$$

The weights α_k are recalculated after each communication round based on the latest validation losses reported by each platform, allowing the aggregation strategy to adapt to changes in data distributions across platforms as training progresses. The global model is distributed by the server in each round of communication. Each platform executes E number of local training rounds, validates the loss and quality scores, adds differentially private noise, and uploads encrypted parameters. The server then computes adaptive weights via Equations (3)–(4) and aggregates the global model via Equation (5). This process repeats for T rounds. Regarding the privacy protection mechanism, the FL-SEPF framework introduces differential privacy noise injection before local model parameters are uploaded, achieving (ε, δ)-differential privacy guarantees by adding calibrated Gaussian noise to gradient vectors [28]. For any two neighboring datasets D and D' differing in a single record, the mechanism $\mathcal{M}$ satisfies:

$$Pr[\mathcal{M}(D) \in S] \le e^{\varepsilon} \cdot Pr[\mathcal{M}(D') \in S] + \delta \tag{6}$$

where the privacy budget ε controls the strength of privacy protection; a smaller ε provides stronger privacy guarantees but may have a greater impact on model performance [29]. The noise injection operation adds Gaussian noise before uploading local model parameters:

$$\tilde{w}_k = w_k + \mathrm{N}(0, \sigma^2 \cdot S^2 \cdot \mathbf{I}) \tag{7}$$

Where *S=1.0* is the gradient clipping norm applied to the L2 norm of the full gradient update vector before noise injection. The privacy parameter $\delta = \frac{1}{N_{total}} (N_{total} = 698{,}000\ total\ users)$, yielding $\delta \approx 1.43 \times 10^{-6}$. The cumulative privacy budget over *T=200* communication rounds with *E=5* local epochs per round was tracked using the Rényi Differential Privacy (RDP) accountant, yielding a final cumulative budget of $\varepsilon_{total} = 1.0$. To achieve a good balance between privacy protection and model utility, this study adopts an adaptive noise adjustment mechanism where the noise multiplier follows a cosine decay schedule:

$\sigma(t) = \sigma_{max} - (\sigma_{max} - \sigma_{min}) . \frac{1 - \cos(\frac{\pi t}{T})}{2}$, where $\sigma_{max} = 1.0$ and $\sigma_{min} = 0.4$. In the early training phase, the noise multiplier remains near σ_{max} for stronger privacy protection; as the model converges in later rounds, $\sigma(t)$ decreases toward σ_{min} to preserve prediction accuracy [30]. At the secure aggregation stage, only the central server receives encrypted updates from each platform and performs the aggregation, making it impossible to infer any local data from each platform based on the aggregated global parameters. In addition to encrypted model parameters, each platform uploads scalar values L_k and Q_k for adaptive weight computation. These represent dataset-level aggregate statistics rather than individual-level information, carrying substantially lower privacy risk than model gradients. To mitigate potential inference of coarse dataset properties through repeated observation across rounds, small-scale Laplacian noise (b=0.01) is added to both values before transmission. For communication optimization, Top-K gradient sparsification transmits only the K largest-magnitude gradient elements per round, reducing communication volume to K/d of the original size. The FL-SEPF framework supports a bounded-staleness asynchronous update mechanism, allowing each platform to upload parameters at its own pace. A staleness threshold of $\tau_{max} = 3$ rounds is enforced: updates based on a global model more than three rounds behind the current version are discarded, and the platform receives the latest global model for retraining. For updates within this bound, a decay factor $\gamma^{(t-t_k)}$ $(\gamma = 0.9)$ is applied to the platform's aggregation weight, where t is the current global round and t_k is the base round of the local update. In the current experiments, all four platforms operated synchronously for controlled comparison; however, the asynchronous mechanism is available for real-world deployment where platforms may operate at different speeds.

2.5 Silver economy demand prediction model

In the context of federated learning, this study proposes a multi-task demand forecasting approach for the silver economy domain, with its framework illustrated in Figure 2. In the proposed architecture (Figure 2), multi-task learning is employed, in which shared feature representations are learned via federated learning. The cross-platform attention fusion layer assigns adaptive weights to platform-specific features based on the tasks' objectives using decomposition-based approaches used in time series forecasting models [31]. The specific output layers are responsible for computing probabilities for classifying the demand type and regression values for demand intensity.

2.6 Experimental setup

The hardware and software configurations of the experimental environment, along with the core hyperparameter settings, are shown in Table 4. The hyperparameters (β_1, β_2, β_3) = (0.3,0.4,0.5) were selected via grid search over {0.1, 0.3, 0.5, 0.7} for each parameter on the validation set. Table 5 reports the sensitivity analysis results, where each row varies one parameter while fixing the other two at their default values. F1-score varies within approximately 1.5% across the tested range. The model is most sensitive to β_2 (data quality weight), with a 1.4% F1 drop when β_2 is reduced to 0.1, while β_3 shows the least sensitivity—β_3 = 0.7 even slightly outperforms the default, though the difference is within the standard deviation. The final configuration (0.3, 0.4, 0.5) was selected as it yielded the best average performance across both classification and regression metrics on the validation set.

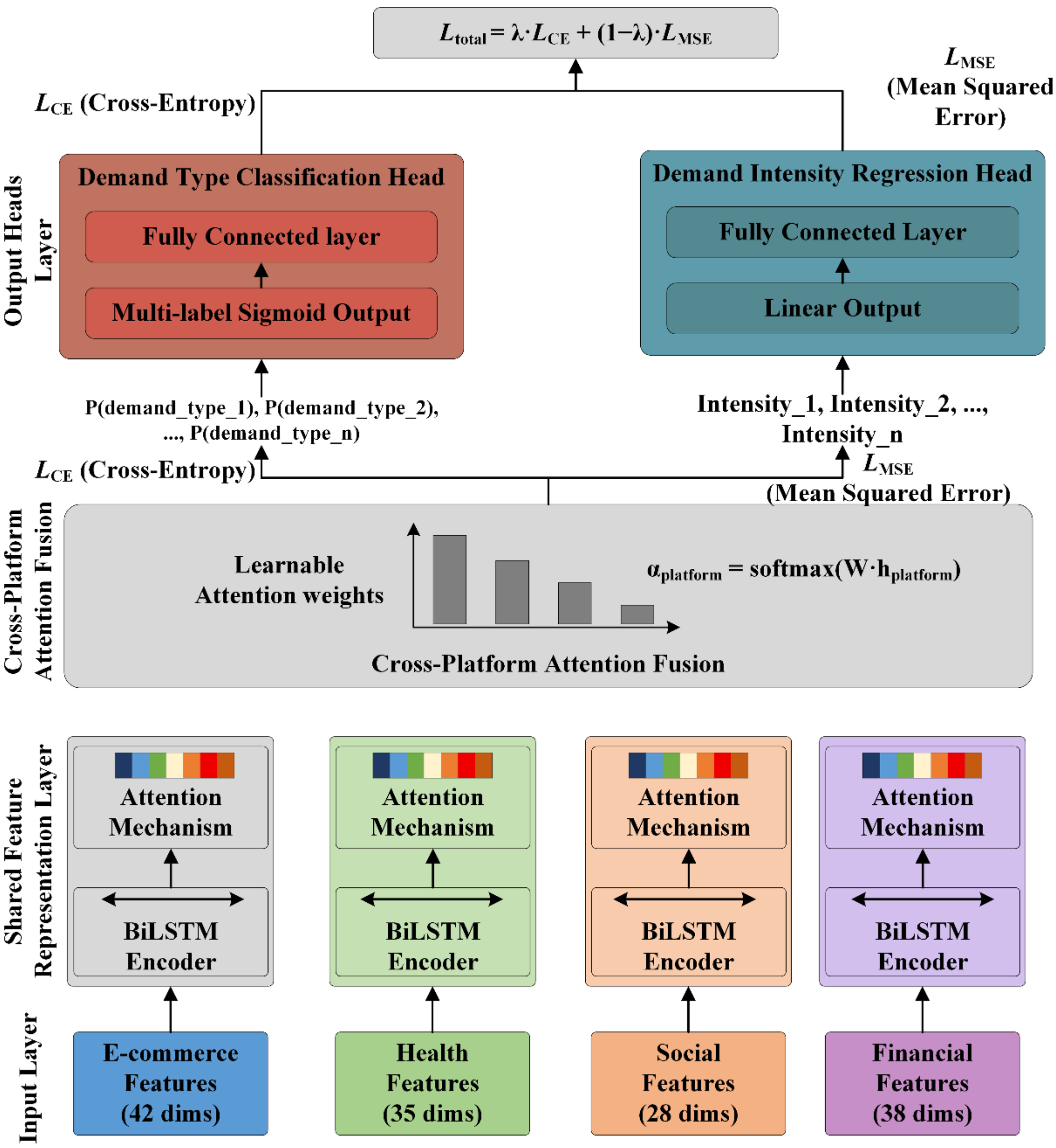

Figure 2. Architecture of the multi-task demand prediction model

All four platforms were simulated on a single physical server using separate data partitions and independent model instances to replicate the federated setting. Communication between platforms and the aggregation server was simulated via inter-process message passing. The implementation is based on PyTorch 2.1.0 with PySyft 0.8.0 for federated protocol simulation. The code and trained models are planned for public release upon acceptance to support reproducibility.

Table 4. Experimental configuration

Category	Specification
GPU	NVIDIA A100 80GB × 2
CPU	Intel Xeon Platinum 8358 (32 cores)
Memory	256 GB DDR4
Deep Learning Framework	PyTorch 2.1.0
Federated Learning Framework	PySyft 0.8.0 / FATE 1.11
Machine Learning Library	scikit-learn 1.3.2
Learning Rate	1e-3 (with cosine annealing scheduler)
Batch Size	256
Communication Rounds	200
Local Epochs	5
Privacy Budget ε	1.0 (default), tested range [0.1, 10.0]
Loss Balance Coefficient λ	0.6
Top-K Sparsification Ratio	10%
Aggregation Hyperparameters (β_1, β_2, β_3)	(0.3, 0.4, 0.5)

Table 5. Sensitivity Analysis of Aggregation Hyperparameters

Varied Parameter	Configuration (β_1, β_2, β_3)	F1-score	AUC-ROC
$\beta_1 = 0.1$	(0.1, 0.4, 0.5)	0.8227±0.0054	0.8849±0.0045
$\beta_1 = 0.3$ (default)	(0.3, 0.4, 0.5)	0.8305±0.0049	0.8921±0.0040
$\beta_1 = 0.5$	(0.5, 0.4, 0.5)	0.8283±0.0051	0.8904±0.0042
$\beta_1 = 0.7$	(0.7, 0.4, 0.5)	0.8191±0.0057	0.8838±0.0048
$\beta_2 = 0.1$	(0.3, 0.1, 0.5)	0.8189±0.0056	0.8835±0.0047
$\beta_2 = 0.4$ (default)	(0.3, 0.4, 0.5)	0.8305±0.0049	0.8921±0.0040
$\beta_2 = 0.7$	(0.3, 0.7, 0.5)	0.8271±0.0050	0.8893±0.0043
$\beta_3 = 0.1$	(0.3, 0.4, 0.1)	0.8238±0.0055	0.8861±0.0046
$\beta_3 = 0.5$ (default)	(0.3, 0.4, 0.5)	0.8305±0.0049	0.8921±0.0040
$\beta_3 = 0.7$	(0.3, 0.4, 0.7)	0.8318±0.0048	0.8913±0.0041

To comprehensively evaluate the performance of the FL-SEPF framework, this study sets up four groups of baseline methods. The traditional statistical and machine learning methods group includes the ARIMA time-series model, Logistic regression classification model, Random Forest ensemble model, and XGBoost gradient-boosting model [32]. The centralized deep learning models consist of a centralized LSTM model and a centralized Transformer model that are learned using data aggregated across all platforms. This is conducted to measure the performance degradation resulting from using federated learning in comparison to an ideal centralized setting. The centralized LSTM model uses the same BiLSTM+Attention architecture and hyperparameters as the federated local models to ensure a fair comparison of learning paradigms, while the centralized Transformer model uses a 4-layer Transformer encoder with the same input dimensions and training schedule to represent a potentially stronger architecture upper bound. The standard federated learning baselines include FedAvg and FedProx. Single-platform models train independently on each platform without collaboration. The evaluation metric system comprises three dimensions. Classification performance metrics are Accuracy, macro-averaged Precision, Recall, F1-score, and macro-averaged AUC-ROC (one-vs-rest). Macro-averaging is used to give equal weight to all demand types, thus avoiding the problem of majority classes dominating the metrics. Regression evaluation measures include Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination (R^2). The efficiency measures are denoted using the number of rounds required to achieve the desired level of performance. All comparative experiments use paired t-tests with Bonferroni correction for statistical significance. Furthermore, Cohen's d is used to assess the practical significance of the findings. The time-based method of splitting the data into sets is used, with the database split into training (70%), validation (15%), and testing (15%) sets. In this setup, the model undergoes five-fold cross-validation. Each fold was repeated with three different random seeds for model initialization and data shuffling, and the reported means and standard deviations are computed across all 15 runs (5 folds × 3 seeds). The experimental setup is designed to mimic a federated collaboration setting with $K = 4$.

3. Results and discussion

3.1 Overall performance comparison

Table 6 presents the performance comparison between the FL-SEPF framework and all baseline methods on the silver economy demand prediction task, covering both classification and regression evaluation metrics. To ensure fair comparison, all federated learning methods (FedAvg, FedProx, FL-SEPF) are evaluated under the same privacy budget . The experimental data in Table 6 show that the FL-SEPF framework outperforms standard federated learning baselines and traditional methods across all evaluation metrics, while maintaining a small performance gap relative to centralized deep learning models. Compared with traditional statistical and machine learning methods, FL-SEPF achieves 11.4%, 9.9%, and 23.1% improvements in F1-score, AUC-ROC, and RMSE respectively over XGBoost. This performance gap confirms the structural advantage of deep learning models in capturing complex nonlinear behavioral patterns of the elderly population, and shows that collaborative integration of multi-platform behavioral data can supply complementary information that a single data source cannot cover. Compared with centralized models, FL-SEPF's F1-score is only 1.7% lower than the Centralized Transformer (0.8312 vs. 0.8456), demonstrating that federated learning maintains near-centralized performance under strict privacy constraints.

Among the federated learning methods, FL-SEPF achieves a 3.5% improvement in F1-score over FedAvg (from 0.8034 to 0.8312) and a 2.3% improvement over FedProx (from 0.8126 to 0.8312), with R^2 increasing from 0.6834/0.6942 to 0.7198. Paired t-test results show that the differences noted above are statistically significant at $p < 0.01$ (Cohen's $d = 0.82$ and 0.64, respectively), confirming the systematic advantage of the adaptive weighted aggregation strategy over fixed-weight schemes when handling Non-IID elderly behavioral data. The 95% confidence intervals for the F1-score differences are [0.0218, 0.0338] for FL-SEPF vs. FedAvg and [0.0127, 0.0245] for FL-SEPF vs. FedProx, confirming that the improvements are not only statistically significant but also practically meaningful.

Table 6. Overall performance comparison across all methods

Method	Accuracy	F1-score	AUC-ROC	RMSE	MAE	R^2
ARIMA	0.6238±0.0118	0.5847±0.0124	0.6512±0.0109	0.4821±0.0095	0.3956±0.0087	0.4127±0.0102
Logistic Regression	0.6845±0.0098	0.6531±0.0105	0.7189±0.0091	0.4463±0.0082	0.3612±0.0076	0.4685±0.0094
Random Forest	0.7412±0.0079	0.7185±0.0083	0.7836±0.0072	0.3847±0.0071	0.3089±0.0065	0.5463±0.0078
XGBoost	0.7689±0.0071	0.7462±0.0076	0.8124±0.0064	0.3615±0.0068	0.2876±0.0059	0.5812±0.0073
Centralized LSTM	0.8534±0.0048	0.8347±0.0052	0.8951±0.0041	0.2734±0.0054	0.2145±0.0046	0.7236±0.0051
Centralized Transformer	0.8621±0.0043	0.8456±0.0046	0.9038±0.0037	0.2651±0.0049	0.2067±0.0042	0.7385±0.0047
Single-Platform (Best)	0.7856±0.0085	0.7623±0.0089	0.8247±0.0078	0.3428±0.0074	0.2743±0.0067	0.6124±0.0082
FedAvg	0.8215±0.0061	0.8034±0.0065	0.8672±0.0054	0.2986±0.0062	0.2387±0.0055	0.6834±0.0063
FedProx	0.8298±0.0057	0.8126±0.0059	0.8753±0.0049	0.2914±0.0058	0.2315±0.0052	0.6942±0.0059
FL-SEPF (Ours)	0.8489±0.0052	0.8312±0.0047	0.8927±0.0038	0.2781±0.0064	0.2178±0.0051	0.7198±0.0058

The Single-Platform Best model achieves an F1-score of only 0.7623, confirming that single-source data is insufficient for characterizing elderly financial demand. The performance comparison visualization in Figure 3 shows the distribution of differences among methods across various metric dimensions.

3.2 Ablation study

To understand how each module in the FL-SEPF framework functions independently, a series of ablation studies is conducted to measure the performance effects of eliminating or replacing specific modules. The outcomes of these studies are detailed in Table 7. The ablation results reveal varying contributions of each FL-SEPF component to prediction performance. Removing adaptive weighting causes the largest single-module decline (−3.34% F1), confirming its critical role in addressing Non-IID distributional drift. Removing multi-task learning reduces F1 by 2.68%, confirming the complementarity of knowledge between classification and regression tasks. Replacing attention fusion with concatenation reduces F1 by 1.99%, indicating that adaptive cross-platform feature weighting is beneficial.

Removing differential privacy improves F1 by 1.01%, consistent with the inherent accuracy-privacy trade-off. This modest cost confirms that the adaptive noise mechanism effectively limits the performance impact of privacy protection. The results of the platform reduction experiment also validate the marginal returns of multi-platform collaboration, where reducing the number of platforms from 4 to 3 results in a drop in F1-score by only 0.94%, whereas the drop in F1-score increases to 2.84% with only 2 platforms, and to 8.29% with only 1 platform. To systematically show the effect of platform reduction, Table 8 presents the marginal effect of including platforms, starting with the most efficient one (Financial Service), and then adding further platforms according to their contribution. The nonlinear trend indicates that the benefits of multi-platform collaboration stem primarily from the complementarity of behavioral data across dimensions rather than from volume alone.

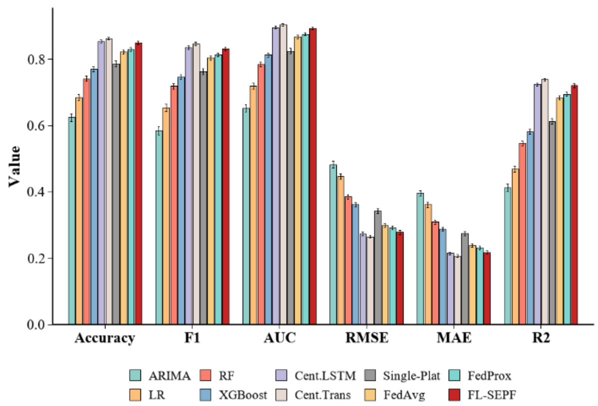

Figure 3. Performance comparison across all methods and evaluation metrics

The results show diminishing marginal returns: the addition of E-commerce data yields the largest single-platform gain (+4.53% F1), while Social Interaction contributes the smallest increment (+0.78%), consistent with the SHAP analysis in Section 3.5, which shows that social features rank lowest in importance.

3.3 Privacy-utility trade-off and non-IID robustness

The privacy-utility trade-off is critical for the practical deployment of federated learning. Figure 4 is a diagram that shows, using a two-panel format, the performance of FL-SEPF under various privacy budget values ε and its robustness under various non-IID data distributions. In terms of the privacy-utility trade-off, as ε decreases from 10.0 to 0.1, F1-score drops from 0.8378 to 0.7684, and AUC-ROC decreases from 0.8973 to 0.8342 (Figure 4A). At the default $\varepsilon = 1.0$, performance loss is only 1.0% compared to no privacy protection. Degradation remains below 3% for $\varepsilon \in [0.5, 2.0]$, but accelerates to 8.3% at $\varepsilon = 0.1$ due to excessive noise affecting convergence. The Non-IID robustness experiments use the Dirichlet distribution parameter α to control heterogeneity (smaller α = stronger Non-IID), differing from the real-data setting in Table 6.

Table 7. Ablation study results

Model Variant	F1-score	AUC-ROC	RMSE	R^2	ΔF1 (%)
FL-SEPF (Full Model)	0.8312±0.0047	0.8927±0.0038	0.2781±0.0064	0.7198±0.0058	—
w/o Adaptive Weighting → FedAvg	0.8034±0.0065	0.8672±0.0054	0.2986±0.0062	0.6834±0.0063	-3.34
w/o Differential Privacy	0.8396±0.0044	0.8985±0.0035	0.2723±0.0059	0.7264±0.0054	+1.01
w/o Attention Fusion → Concatenation	0.8147±0.0056	0.8789±0.0047	0.2876±0.0061	0.7023±0.0060	-1.99
w/o Multi-task Learning → Single-task	0.8089±0.0062	0.8714±0.0051	0.2953±0.0066	0.6912±0.0064	-2.68
4 Platforms → 3 Platforms	0.8234±0.0051	0.8867±0.0042	0.2823±0.0063	0.7108±0.0057	-0.94
4 Platforms → 2 Platforms	0.8076±0.0068	0.8698±0.0056	0.2947±0.0069	0.6878±0.0066	-2.84
4 Platforms → 1 Platform	0.7623±0.0089	0.8247±0.0078	0.3428±0.0074	0.6124±0.0082	-8.29

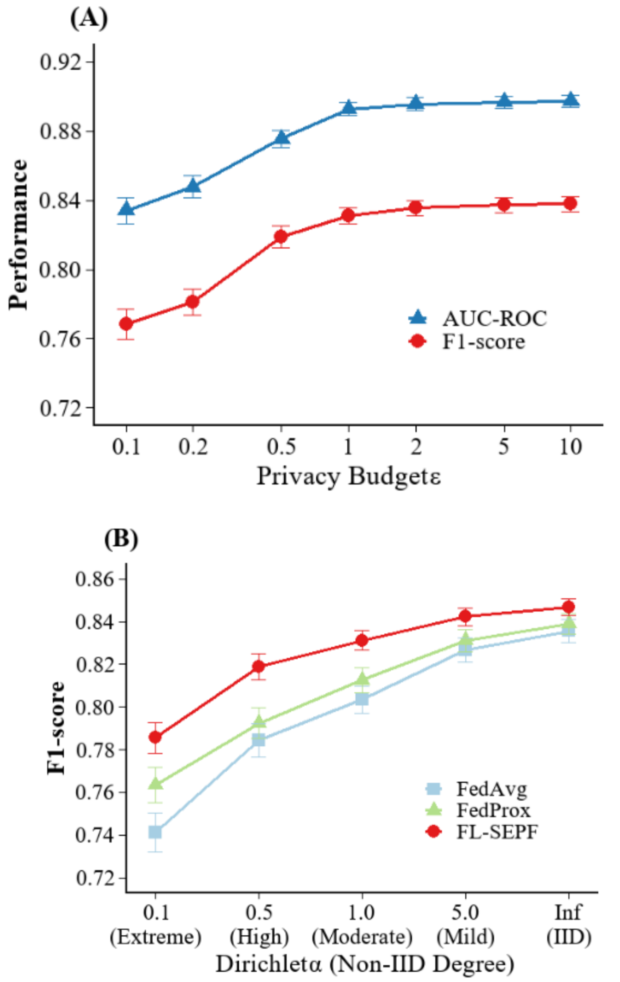

Figure 4. Privacy-utility trade-off and Non-IID robustness analysis. (A) F1-score and AUC-ROC under varying privacy budget ε . (B) F1-score under varying Non-IID degrees (Dirichlet α)

Under extreme Non-IID conditions ($\alpha = 0.1$), FedAvg's F1-score drops by 11.3%, FedProx by 8.7%, while FL-SEPF drops by only 5.5%. Under moderate conditions ($\alpha = 0.5$), FL-SEPF maintains a stable advantage (F1: 0.8189 vs. 0.7845 vs. 0.7923), confirming the adaptive strategy's consistent effectiveness across varying heterogeneity levels.

3.4 Convergence and communication efficiency

Communication efficiency and convergence speed are critical for system deployability. As shown in Figure 5, the comparison of the convergence of the FL-SEPF and the conventional federated learning baselines, as well as the reduction in communication overhead using the gradient compression strategy, is represented in a two-panel format.

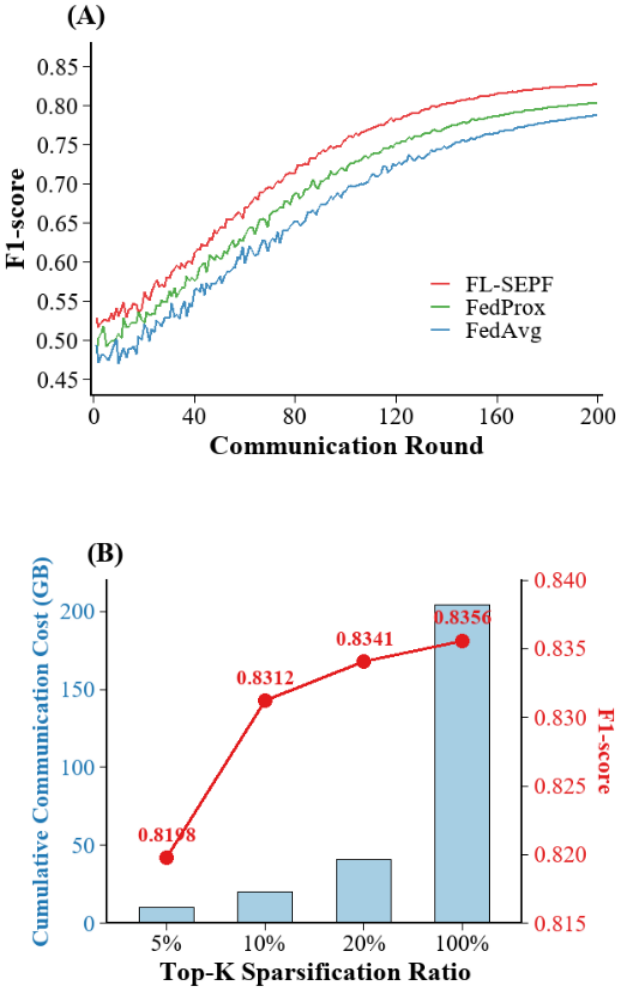

Figure 5. Convergence and communication efficiency analysis. (A) F1-score convergence over 200 communication rounds. (B) Communication cost versus F1-score under different sparsification ratios

Table 8. Incremental platform addition results

Platform Configuration	F1-score	AUC-ROC	ΔF1 vs. Previous
Financial Service only	0.7623±0.0089	0.8247±0.0078	—
+ E-commerce	0.8076±0.0068	0.8698±0.0056	+4.53%
+ Health Service	0.8234±0.0051	0.8867±0.0042	+1.58%
+ Social Interaction (Full)	0.8312±0.0047	0.8927±0.0038	+0.78%

3.5 Feature importance analysis

To reveal the distinct contributions of behavioral characteristics across dimensions in predicting the financial demands of the silver economy, this study uses the SHAP method to conduct a feature-importance interpretability analysis of the FL-SEPF global model. As shown in Figure 6, the SHAP beeswarm plot of the top 20 features is presented. Table 9 groups and ranks feature importance by platform source, presenting the core features contributed by each platform along with their mean SHAP values.

Table 9. Top 15 feature importance rankings by platform source

Rank	Feature Name	Source Platform	Mean SHAP Value	Direction
1	Portfolio diversity index	Financial Service	0.1847	Positive
2	Monthly spending amount	E-commerce	0.1623	Positive
3	Credit utilization rate	Financial Service	0.1489	Negative
4	Insurance policy count	Health Service	0.1356	Positive
5	Risk preference score	Financial Service	0.1278	Positive
6	Purchase frequency	E-commerce	0.1145	Positive
7	Medical expenditure ratio	Health Service	0.1089	Positive
8	Average holding period	Financial Service	0.0967	Negative
9	Category diversity index	E-commerce	0.0923	Positive
10	Community participation rate	Social Interaction	0.0856	Positive
11	Health concern intensity	Health Service	0.0812	Positive
12	Transaction frequency	Financial Service	0.0789	Positive
13	Price sensitivity coefficient	E-commerce	0.0734	Negative
14	Content sharing frequency	Social Interaction	0.0698	Positive
15	Chronic disease management engagement	Health Service	0.0654	Positive

The rankings of feature importance in Table 9 provide some interesting and practically significant results. The financial service platform contributes three of the top five features, confirming that existing financial behavior is the most direct signal for future demand. The top-ranked portfolio diversity index indicates that elderly users with diversified holdings are more likely to adopt new financial services. The negative SHAP direction of credit utilization rate suggests that conservative financial behavior is associated with stronger demand for protective instruments such as pension management and health insurance. Monthly e-commerce spending ranks second overall (SHAP = 0.1623), indicating that consumption capacity is an effective proxy for financial service demand. Healthcare features (insurance policy count and medical expenditure ratio, ranked 4th and 7th) confirm the importance of health-related factors in driving demand for insurance and pension products. The features of the social interaction platform are ranked lower (from the tenth position), but the community participation rate, as the most important feature of the social interaction platform, has a significant SHAP value (0.0856), indicating that elderly people with higher social activity levels have stronger acceptance of digital financial services. Social interaction data has value as a supplement to profiling the financial demands of the elderly. It should be noted that the SHAP analysis was performed on the global model after federated training, yielding global-level feature importance rankings. Local SHAP values computed on individual platform data may differ from the global rankings due to platform-specific data distributions. Investigating the alignment between local and global SHAP explanations in federated settings is a topic for future research.

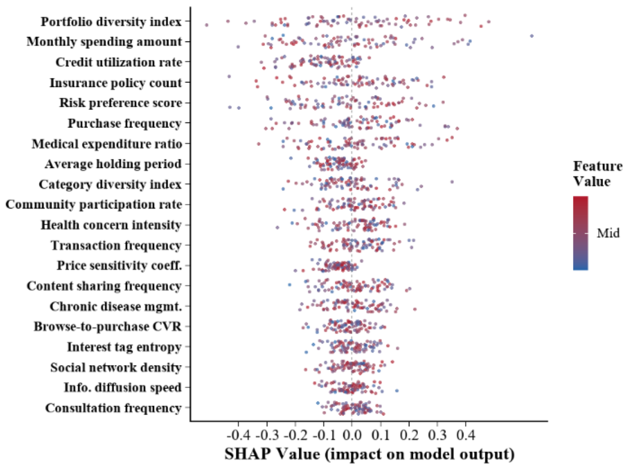

Figure 6. SHAP summary plot of top-20 feature importance rankings

3.6 Discussion

The effectiveness of the adaptive weighted aggregation strategy lies in its ability to adapt to Non-IID data by incorporating data quality, model fitness, and volume, thereby reducing the impact of poorly representative platforms. This mechanism aligns with recent trends in adaptive federated learning for heterogeneous environments [33], and FL-SEPF's robustness under extreme Non-IID conditions echoes findings on enhancing Non-IID data handling [34]. It is worth noting that the FL-SEPF aggregation strategy differs in design philosophy from other advanced federated methods. FedNova [19] and SCAFFOLD [20] focus on correcting objective inconsistency or client drift through normalized updates or control variates, while FedDyn [21] aligns local and global optima via dynamic regularization. MOON [22] operates at the representation level through contrastive learning. These methods are largely architecture-agnostic and address heterogeneity at the optimization level. FL-SEPF's adaptive weighting, by contrast, is application-driven: it explicitly incorporates domain-specific quality signals (data completeness, class balance, temporal continuity) into the aggregation process, which is particularly well-suited to the silver economy scenario, where platforms differ fundamentally in data modalities and reliability. The limitations of the current comparative experiment are the exclusion of FedNova, SCAFFOLD, FedDyn, and MOON due to implementation constraints with the multi-task BiLSTM+Attention architecture. Future studies should integrate these approaches into a single baseline for a proper benchmarking study.

One limitation of the present assessment is the absence of subgroup assessments by age cohort, income level, or digital literacy. Such an approach would enable one to determine whether the FL-SEPF model is operating fairly across various subgroups in the population of older individuals, particularly given the problem of discrimination against those who engage in fewer online activities and therefore leave behind fewer behavioral traces. Regarding the privacy-utility trade-off, FL-SEPF loses only about 1.0% F1-score at $\varepsilon = 1.0$, attributable to the adaptive noise mechanism [35]. The centralized Transformer slightly outperforms FL-SEPF (F1 gap of 1.7%), but requires centralized data storage infeasible under current privacy regulations. The integration of SHAP interpretability with federated learning extends prior work on explainable AI in financial fraud detection [36]. The BiLSTM+Attention architecture builds on temporal modeling strategies validated in supply chain scenarios [37] and on probabilistic modeling approaches for portfolio risk assessment [38], which offer relevant references for incorporating uncertainty into future demand prediction. The multi-platform collaboration shares methodological commonalities with privacy-preserving demand prediction in transportation [39]. Unlike these works, FL-SEPF addresses greater data-source heterogeneity across fundamentally different platform domains. In the threat model studied in this work, an honest but curious server is assumed to follow the protocol while attempting to derive information locally by analyzing the provided updates. Under this assumption, differential privacy ($\varepsilon = 1.0$) and Top-K gradient sparsification (10%) provide complementary defenses. The former masks the possibility of sample differentiation across gradients, whereas the latter method removes 90 percent of gradient components, thereby significantly increasing the difficulty of reversing gradients. Testing the methods formally on specific attack vectors, including membership inference and model reconstruction, would be an important research direction for the future.

From a practical perspective, the SHAP-based rankings have useful implications: banks should focus on their internal financial data when conducting demand analysis, use e-commerce purchase power as a proxy for actual demand, and apply the relationship between health and finance to recommend appropriate insurance products. These findings extend data-driven financial decision support paradigms [40] to the silver economy domain, complementing research on AI-enhanced financial platform optimization [41] and privacy-preserving credit assessment [42]. To assess robustness under realistic deployment conditions, supplementary robustness tests incorporated configurable network latency (uniformly sampled from 50–500 ms per round), a 10% per-round platform dropout probability, and heterogeneous computation time ratios (1.0:1.3:0.8:1.5 across the four platforms). Under these conditions, FL-SEPF maintained an F1-score of 0.8278±0.0054, a decrease of 0.41% relative to the ideal setting, confirming its robustness to moderate disturbances. Nevertheless, full deployment validation on real distributed infrastructure remains necessary and is planned as future work.

4. Conclusion

This study proposes FL-SEPF, a federated learning-based framework that aggregates elderly behavioral data from e-commerce, healthcare, social interaction, and financial service platforms for privacy-preserving demand prediction. FL-SEPF achieves an F1-score of 0.8312 and an AUC-ROC of 0.8927, outperforming FedAvg by 3.5% and XGBoost by 11.4%, while maintaining a gap of less than 1.7% from centralized Transformer training. The contributions of this study span the framework, algorithm, and application levels, as detailed in Section 1. While acknowledging the above contributions, the present study also has certain limitations to be taken into account. FL-SEPF was validated in a simulated environment; real-world deployment validation remains necessary to address gaps in network latency variability, dynamic platform participation, and device heterogeneity. Considering geographic applicability, this dataset identifies behavioral patterns and financial preferences of the elderly in a given region. Elderly people in different parts of the world may show significant variance in terms of their digital adoption, consumption behavior, risk-taking, and financial product utilization. Such variation may also stem from cultural, economic, and policy differences that affect the relative importance of different behavioral feature categories across regions. Moreover, profiling elderly financial behavior raises concerns about potential exploitation of vulnerable populations through over-targeting, and the digital divide among elderly users may lead to unequal prediction accuracy for digitally less active subgroups. These considerations should be addressed through deployment safeguards in practical applications. Therefore, the framework's ability to generalize over regions must be tested by conducting further experiments on a multi-regional dataset. Similarly, scaling FL-SEPF to $K \geq 10$ platforms would require empirical validation of aggregation efficiency and weight calibration under higher-dimensional platform heterogeneity. With respect to the model's interpretability, SHAP analysis offers global interpretability at the feature level; however, the internal decision mechanisms for adaptive weight allocation and cross-platform feature fusion in a federated learning paradigm are insufficiently transparent for financial regulatory auditing. Several future research directions merit exploration. Additional data modalities such as location trajectories, wearable health data, and voice interactions could enrich demand profiling. Graph neural networks could be integrated with federated learning to capture demand propagation through social and financial product networks. Cross-institutional federated learning pilots among financial, e-commerce, and healthcare institutions are essential for real-world validation. Extending FL-SEPF from horizontal to vertical federated learning would enable modeling cross-platform user journeys at the cost of increased privacy engineering complexity. Integrating large language models could extend the framework beyond structured behavioral patterns to capture semantic demand signals. Finally, incorporating concept-drift detection and incremental-update mechanisms would enhance robustness during long-term deployment, thereby supporting lifecycle-oriented financial product management.

Ethical issue
The authors are aware of and comply with best practices in publication ethics, specifically regarding authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with research ethics policies. The authors adhere to publication requirements that the submitted work is original and has not been published elsewhere.

Data availability statement
The manuscript contains all the data. However, additional data will be provided by the corresponding author upon reasonable request.

Conflict of interest
The authors declare no potential conflict of interest.

References

[1] X. Yang and M. Qi, "Inhibiting or promoting: Population aging and economic development in China," PLOS ONE, vol. 19, no. 5, p. e0303197, 2024, doi: 10.1371/journal.pone.0303197.

[2] T. Egeland, K. Heinonen, and I. G. Stensaker, "The Silver Economy: A Business Perspective on the Aging Society," The Journal of Applied Behavioral Science, vol. 0, no. 0, p. 00218863251397563, 2025, doi: 10.1177/00218863251397563.

[3] H. Berg and K. T. Liljedal, "Elderly consumers in marketing research: A systematic literature review and directions for future research," International Journal of Consumer Studies, vol. 46, no. 5, pp. 1640-1664, 2022, doi: 10.1111/ijcs.12830.

[4] P. Kairouz and H. B. McMahan, "Advances and Open Problems in Federated Learning," Foundations and Trends in Machine Learning, vol. 14, no. 1-2, pp. 1-210, 2021, doi: 10.1561/2200000083.

[5] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated Machine Learning: Concept and Applications," ACM Trans. Intell. Syst. Technol., vol. 10, no. 2, p. Article 12, 2019, doi: 10.1145/3298981.

[6] X. Yin, Y. Li, R. Gao, J. Li, and H. Wang, "Understanding the Purchase Decisions of Silver Consumers in Short-Form Video Platforms from the Perspective of Existence, Relatedness, and Growth Needs," Behavioral Sciences, vol. 13, no. 12, p. 1011, 2023, doi: 10.3390/bs13121011.

[7] D. Pattnaik, S. Ray, and R. Raman, "Applications of artificial intelligence and machine learning in the financial services industry: A bibliometric review," Heliyon, vol. 10, no. 1, 2024, doi: 10.1016/j.heliyon.2023.e23492.

[8] S. Gu, B. Kelly, and D. Xiu, "Empirical Asset Pricing via Machine Learning," The Review of Financial Studies, vol. 33, no. 5, pp. 2223-2273, 2020, doi: 10.1093/rfs/hhaa009.

[9] W. Hilal, S. A. Gadsden, and J. Yawney, "Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances," Expert Systems with Applications, vol. 193, p. 116429, 2022/05/01/ 2022, doi: 10.1016/j.eswa.2021.116429.

[10] Z. Wang, J. Xiao, L. Wang, and J. Yao, "A novel federated learning approach with knowledge transfer for credit scoring," Decision Support Systems, vol. 177, p. 114084, 2024/02/01/ 2024, doi: 10.1016/j.dss.2023.114084.

[11] M. Abdul Salam, K. M. Fouad, D. L. Elbably, and S. M. Elsayed, "Federated learning model for credit card fraud detection with data balancing techniques," Neural Computing and Applications, vol. 36, no. 11, pp. 6231-6256, 2024/04/01 2024, doi: 10.1007/s00521-023-09410-2.

[12] S. Pourroostaei Ardakani, N. Du, C. Lin, J.-C. Yang, Z. Bi, and L. Chen, "A federated learning-enabled predictive analysis to forecast stock market trends," Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 4, pp. 4529-4535, 2023/04/01 2023, doi: 10.1007/s12652-023-04570-4.

[13] B. Liu, N. Lv, Y. Guo, and Y. Li, "Recent advances on federated learning: A systematic survey," Neurocomputing, vol. 597, p. 128019, 2024/09/07/ 2024, doi: 10.1016/j.neucom.2024.128019.

[14] C. Zhang, N. N. A. Sjarif, and R. Ibrahim, "Deep learning models for price forecasting of financial time series: A review of recent advancements: 2020–2022," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 14, no. 1, p. e1519, 2024, doi: 10.1002/widm.1519.

[15] X. Ma, J. Zhu, Z. Lin, S. Chen, and Y. Qin, "A state-of-the-art survey on solving non-IID data in Federated Learning," Future Generation Computer Systems, vol. 135, pp. 244-258, 2022/10/01/ 2022, doi: 10.1016/j.future.2022.05.003.

[16] R. Gosselin, L. Vieu, F. Loukil, and A. Benoit, "Privacy and Security in Federated Learning: A Survey," Applied Sciences, vol. 12, no. 19, p. 9901, 2022, doi: 10.3390/app12199901.

[17] P. Chatterjee, D. Das, and D. B. Rawat, "Federated Learning Empowered Recommendation Model for Financial Consumer Services," IEEE Transactions on Consumer Electronics, vol. 70, no. 1, pp. 2508-2516, 2024, doi: 10.1109/TCE.2023.3339702.

[18] G. Zheng, D. Ivanov, and A. Brintrup, "An adaptive federated learning system for information sharing in supply chains," International Journal of Production Research, vol. 63, no. 11, pp. 3938-3960, 2025/06/03 2025, doi: 10.1080/00207543.2024.2432469.

[19] J. Wang, Q. Liu, H. Liang, G. Joshi, and H. V. Poor, "Tackling the Objective Inconsistency Problem in Heterogeneous Federated Optimization," arXiv e-prints, p. arXiv: 2007.07481, 2020, doi: 10.48550/arXiv.2007.07481

[20] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "SCAFFOLD: Stochastic Controlled Averaging for Federated Learning," presented at the Proceedings of the 37th International Conference on Machine Learning, Proceedings of Machine Learning Research, 2020. [Online]. Available: https://proceedings.mlr.press/v119/karimireddy20a.html.

[21] D. A. E. Acar, Y. Zhao, R. M. Navarro, M. Mattina, P. N. Whatmough, and V. Saligrama, "Federated learning based on dynamic regularization," arXiv preprint arXiv:2111.04263, 2021, doi: 10.48550/arXiv.2111.04263.
[22] Q. Li, B. He, and D. Song, "Model-Contrastive Federated Learning," arXiv preprint arXiv:2103.16257, 2021, doi: 10.48550/arXiv.2103.16257.
[23] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," Advances in neural information processing systems, vol. 30, 2017, doi: 10.48550/arXiv.1705.07874
[24] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," Journal of artificial intelligence research, vol. 16, pp. 321-357, 2002, doi: 10.1613/jair.953.
[25] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
[26] A. S. Vaswani, Noam ; Parmar, Niki ; Uszkoreit, Jakob ; Jones, Llion ; Gomez, Aidan N. ; Kaiser, Lukasz ; Polosukhin, Illia, "Attention is all you need," Advances in neural information processing systems, vol. 30, 2017, doi: 10.48550/arXiv.1706.03762
[27] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," Proceedings of Machine learning and systems, vol. 2, pp. 429-450, 2020, doi: 10.48550/arXiv.1812.06127
[28] K. Wei et al., "Federated Learning With Differential Privacy: Algorithms and Performance Analysis," IEEE Transactions on Information Forensics and Security, vol. 15, pp. 3454-3469, 2020, doi: 10.1109/TIFS.2020.2988575.
[29] C. Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy," Foundations and Trends in Theoretical Computer Science, vol. 9, no. 3-4, pp. 211-487, 2014, doi: 10.1561/0400000042.
[30] R. Xue et al., "Differentially Private Federated Learning With an Adaptive Noise Mechanism," IEEE Transactions on Information Forensics and Security, vol. 19, pp. 74-87, 2024, doi: 10.1109/TIFS.2023.3318944.
[31] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," Advances in neural information processing systems, vol. 34, pp. 22419-22430, 2021, doi: 10.48550/arXiv.2106.13008.
[32] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, 2016. [Online]. Available: https://doi.org/10.1145/2939672.2939785.
[33] S. Wang et al., "Adaptive Federated Learning in Resource Constrained Edge Computing Systems," IEEE Journal on Selected Areas in Communications, vol. 37, no. 6, pp. 1205-1221, 2019, doi: 10.1109/JSAC.2019.2904348.
[34] V. S. Mai, R. J. La, and T. Zhang, "A Study of Enhancing Federated Learning on Non-IID Data With Server Learning," IEEE Transactions on Artificial Intelligence, vol. 5, no. 11, pp. 5589-5604, 2024, doi: 10.1109/TAI.2024.3430250.
[35] M. Talaei and I. Izadi, "Enhancing federated learning with adaptive differential privacy and priority-based aggregation," arXiv preprint arXiv:2406.18491, 2024, doi: 10.48550/arXiv.2406.18491.
[36] T. Awosika, R. M. Shukla, and B. Pranggono, "Transparency and Privacy: The Role of Explainable AI and Federated Learning in Financial Fraud Detection," IEEE Access, vol. 12, pp. 64551-64560, 2024, doi: 10.1109/ACCESS.2024.3394528.
[37] H. D. Nguyen, K. P. Tran, S. Thomassey, and M. Hamad, "Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management," International Journal of Information Management, vol. 57, p. 102282, 2021/04/01/ 2021, doi: 10.1016/j.ijinfomgt.2020.102282.
[38] G. Fatouros, G. Makridis, D. Kotios, J. Soldatos, M. Filippakis, and D. Kyriazis, "DeepVaR: a framework for portfolio risk assessment leveraging probabilistic deep neural networks," Digital Finance, vol. 5, no. 1, pp. 29-56, 2023/03/01 2023, doi: 10.1007/s42521-022-00050-0.
[39] A. Munawar and M. Piantanakulchai, "A collaborative privacy-preserving approach for passenger demand forecasting of autonomous taxis empowered by federated learning in smart cities," Scientific Reports, vol. 14, no. 1, p. 2046, 2024/01/24 2024, doi: 10.1038/s41598-024-52181-6.
[40] C. Dai, "Deep Learning-based anomaly detection in stock markets and business decision support," Future Technology, vol. 4, no. 4, pp. 296-310, 09/15 2025. [Online]. Available: https://fupubco.com/futech/article/view/410
[41] T. Wang and G. R. Tobias, "Research on intelligent optimization mechanisms of financial process modules through Machine Learning-enhanced collaborative systems in digital finance platforms," Future Technology, vol. 4, no. 4, pp. 240-254, 09/02 2025. [Online]. Available: https://fupubco.com/futech/article/view/503.
[42] F. Wu, N. Meekaewkunchom, C. Muangmee, and T. Sattabut, "Research on the impact mechanism of AI-driven supply chain creditworthiness assessment on commercial banks' credit policies for SMEs," Future Technology, vol. 4, no. 3, pp. 45-53, 05/30 2025. [Online]. Available: https://fupubco.com/futech/article/view/363.