Future Technology
Open Access Journal
ISSN 2832-0379

August 2026| Volume 05 | Issue 03 | Pages 164-174

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

https://doi.org/10.55670/fpll.futech.5.3.15

Future Publishing LLC

Article

A federated learning approach for predicting AI capability synergy effects in manufacturing mergers and acquisitions: engineering collaboration and economic value creation

Tengfei Fan, Minghao Huang*

Seoul Business School, aSSIST University, Seoul 03767, South Korea

ARTICLE INFO

Article history:
Received 15 January 2026
Received in revised form
20 April 2026
Accepted 29 May 2026

Keywords:
Federated learning, AI capability synergy effects, Manufacturing mergers and acquisitions, Engineering collaboration indicators, Privacy-preserving machine learning

*Corresponding author
Email address:
mhuang@assist.ac.kr

DOI: 10.55670/fpll.futech.5.3.15

ABSTRACT

Manufacturing mergers and acquisitions increasingly target AI capabilities, yet predicting synergy effects remains constrained by cross-enterprise data privacy barriers that render centralized approaches impractical. This study proposes a novel horizontal federated learning framework for AI capability synergy effect prediction in manufacturing M&A, integrating federated histogram-aggregated gradient boosting trees with FedAvg-optimized deep neural networks through a two-stage decoupled training strategy, alongside a systematically constructed engineering collaboration indicator system encompassing R&D compatibility, production system interoperability, and AI talent overlap. Empirical validation across 286 authentic manufacturing M&A cases from 23 enterprises demonstrates that the proposed framework achieves superior predictive performance to centralized machine learning and traditional econometric baselines while preserving complete data confidentiality, with engineering collaboration indicators contributing more substantially to synergy prediction than financial variables, and AI-intensive acquirers generating pronounced post-merger economic value premiums following a time-lagged pattern. These findings establish a methodological bridge between privacy-preserving machine learning and strategic management research, providing manufacturing executives with a comprehensive decision support toolkit for target screening, due diligence, and post-merger integration planning.

1. Introduction

As the global manufacturing industry accelerates toward Industry 4.0 and smart production, AI capabilities have been deeply integrated into the core competitive structure of manufacturing companies, thereby revolutionizing the strategic rationale for mergers and acquisitions in these organizations. In light of the steadily increasing number of AI-related M&A activities among manufacturing enterprises, the strategic rationale lies in the value-creation mechanism. Specifically, the synergies resulting from combining the technological assets, engineering architectures, and specialized personnel of the acquirer and target firms, which enable greater R&D innovation and value-added output than separate operations, are considered a vital component in achieving higher value creation through such deals. As a result, assessing and predicting the degree of AI capability synergy effects is the key to the entire process and decision-making chain associated with AI-enabled M&A in the manufacturing industry. Based on the principle proposed by the resource-based view theory that rare and inimitable technological capabilities form the basic source of competitive advantage in a business enterprise, the application of AI technologies has been proven to increase the flexibility and adaptability of the manufacturing industrial chain substantially [1], whereas empirical studies indicate that mergers and acquisitions based on digital transformation enable manufacturing companies to gain market competitiveness [2]. In light of these findings, this research aims to develop an AI capability synergy effect prediction model for manufacturing M&A within the strict framework of multi-enterprise data privacy. In the well-defined research line on predicting M&A synergy effects, the literature has seen a paradigm shift from conventional econometric methods to machine learning. Prior studies that introduced transfer learning and graph neural networks into a fuzzy comprehensive evaluation model of M&A synergy laid the methodological foundation for the use of deep learning in this area [3]. Subsequent research systematically demonstrated the effectiveness of ensemble-based machine learning in capturing complex nonlinear relationships among high-dimensional features in M&A transactions [4], and a comparative study showed that a hybrid machine learning approach produced significantly better results than traditional econometric methods for predicting the probability of successful M&As [5]. The use of deep learning methods that take into account the temporal dynamic structure of industry networks greatly improved the prediction accuracy of M&A behavior [6], and systematic machine learning modeling based on historical M&A activity datasets provided an important methodological reference [7]. However, the common assumption across the above research is that researchers have access to the key financial and business information of the enterprises involved in the transactions, which is clearly impossible in practice due to data sovereignty concerns and confidentiality requirements.

As a privacy-preserving distributed machine learning framework that enables cross-node cooperative learning without centrally aggregating raw data, federated learning has attracted substantial application-oriented research in manufacturing enterprises. Federated transfer learning schemes for cross-domain prediction applications proved the practicality of collaborative learning under the condition of data non-sharing among multi-factory scenarios [8], and federated predictive maintenance studies focusing on heterogeneous time-series data in manufacturing processes shed light on the fundamental principles behind how FL mitigates data distribution discrepancies in industrial scenarios [9]. Federated learning frameworks for manufacturing cooperation in complex part manufacturing and comprehensive studies on challenges in deploying FL in Industry 5.0 scenarios, together specified data heterogeneity and communication efficiency as the essential technical barriers in manufacturing federated learning [10, 11]. Large-scale federated learning implementations in additive manufacturing factories demonstrated the technical readiness of engineering prediction using high-quality data under strict privacy restrictions [12], and investigations into label synchronization schemes in federated learning systems provided significant engineering insights for implementing multi-node, heterogeneous federated learning systems [13]. The technical route to the simultaneous realization of privacy protection and computational efficiency in federated learning frameworks for smart manufacturing environments empowered by edge computing provides valuable engineering guidance for generalizing federated learning in cross-enterprise scenarios [14].

Regarding empirical studies on the relationship between digital transformation strategies and acquisition initiatives in AI-based M&A value creation, the interplay between digital transformation strategies and acquisition processes has become a major academic issue. On the one hand, the successful implementation of digital transformation strategies in mergers has been proven to play a key role in the value creation process for the merged entity [15]; on the other hand, it has been proven that the impact of digital M&A transactions on shareholder value creation strongly depends on the alignment of the acquiring and target firms' AI and digital capabilities [16]. Furthermore, in relation to the profound impact of digital acquisitions on the allocation of AI professionals as complementary assets, research revealed the importance of the structural compatibility of engineering talent for the process of merger integration [17]. Despite the significant contributions this research has made to the study of value-creation mechanisms in AI M&A, the analytical models involved lack sufficient systematic tools to quantify engineering-technical compatibility.

Combining the limitations of these three research directions, it becomes apparent that there is an important gap in establishing and testing a joint prediction model that accounts for the engineering-technical compatibility of AI capability synergy effects when predicting manufacturing mergers and acquisitions under data privacy concerns. To address this issue, this study proposes a novel horizontal federated learning approach to establish a joint prediction framework that combines federated histogram-aggregated gradient-boosted trees with deep neural networks optimized via the FedAvg algorithm through a decoupled training procedure. Meanwhile, an engineering collaboration indicator system encompassing R&D compatibility indices, production system interoperability scores, and AI talent overlap measures is established to incorporate engineering technical compatibility into the M&A synergy effect prediction framework. Using 286 real-world cases of manufacturing M&As involving 23 companies from 2015 to 2024, this study tests the dominant predictive role of engineering collaboration indicators and explores patterns of economic value premium after acquisitions by AI-based acquirer companies.

2. Methodology

2.1 Problem formulation and overall framework

As a distributed machine learning framework that enables multiple parties to collaboratively build a global prediction model without exchanging their raw data [18], federated learning offers a promising approach to addressing the inherent trade-off between protecting data privacy and enabling cross-party collaboration in the context of manufacturing M&A transactions. The formulation of the multi-party joint prediction problem is stated below. Let N manufacturing enterprises each maintain a local M&A sample set $\mathcal{D}_i = \{(x_j^{(i)}, y_j^{(i)})\}_{j=1}^{n_i}$, where the feature vector $x_j^{(i)} \in \mathbb{R}^d$ encompasses engineering collaboration indicators and financial operational variables, and $y_j^{(i)} \in \{0,1\}$ denotes the binary realization label of AI capability synergy effects. Under the constraint that data remains locally stored at each participating node, the participating nodes jointly optimize the following global objective function:

$$\min_w F(w) = \sum_{i=1}^{N} \frac{n_i}{n} F_i(w), \quad F_i(w) = \frac{1}{n_i} \sum_{j=1}^{n_i} \ell\left(f(\mathbf{x}_j^{(i)}; w), y_j^{(i)} \right) \tag{1}$$

where $n = \sum_{i=1}^{N} n_i$ denotes the total number of samples across all nodes, $\ell(\cdot)$ is the prediction loss function, and $F_i(w)$ measures the local empirical risk of node *i*. This formalization embeds the data privacy protection constraint within the structural design of the optimization objective itself, such that in each iteration, nodes upload only privacy-processed gradient information rather than raw samples to the central server, thereby establishing a mathematically rigorous information boundary for subsequent architectural design [19]. The three-tier architecture of the proposed framework includes the following: a data tier with locally stored M&A sample data from the manufacturing enterprises, a model tier incorporating federated gradient boosting trees and deep learning models using a two-step decoupled training approach, and a decision tier providing M&A suggestions for target selection, due diligence, and integration. Strict privacy constraints limit all inter-layer exchanges to gradient histogram statistics and model parameters rather than raw data, eliminating the risk of systemic data leakage at the architectural level. The architecture of the system is shown in Figure 1. The architecture of the horizontal federated learning is represented in Figure 1. The framework consists of three layers: a Data Layer that has eight enterprise nodes; a Model Layer that performs two-phase federated training via sequential GBDT histogram aggregation and DNN optimization facilitated by a central aggregation server; and a Decision Layer that provides synergy probability values for screening and due diligence.

2.2 Dataset construction and variable definition

The data used in this empirical research include 286 M&A cases involving 23 manufacturing enterprises from 2015 to 2024, drawn from acquisitions by companies engaged in automotive manufacturing, electronic equipment, and high-end equipment, and were obtained from the Wind database, Bloomberg terminal, and annual reports.

The M&A data for each enterprise is stored individually while sharing the same feature dimensions, thereby meeting the distributed data requirements of horizontal federated learning and laying the data foundation for the natural division of federated nodes. The sample screening includes the fulfillment of three conditions simultaneously: the merger was completed, there is enough information on AI-related assets, and there is adequate financial data for the three years after the merger. The dependent variable measures the binary synergy outcome from AI capabilities, operationalized as the joint increase in R&D output and Economic Value Added over the three years following the merger. Cases with performance values above the contemporaneous industry median are deemed synergies, while all others are considered non-synergies.

The independent variable set consists of an engineering cooperation group, comprising R&D compatibility, production system compatibility, and AI expertise overlap, along with financial operating controls, including deal value, acquisition premium ratio, target company Tobin's Q, and industry concentration. The data sample was split into training, validation, and test sets in a 7:1.5:1.5 ratio, using SMOTE and Focal Loss to balance the classes. All descriptive statistics are listed in Table 1. As shown in Table 1, two features of the data emerge that are essential to determining the methodology of this research. Firstly, the engineering collaboration metrics exhibit pronounced cross-node distributional differences, underscoring the need for horizontal federated learning rather than centralized learning. Secondly, the ratio of positive to negative samples is approximately 1:2.4, providing evidence for using SMOTE and Focal Loss together.

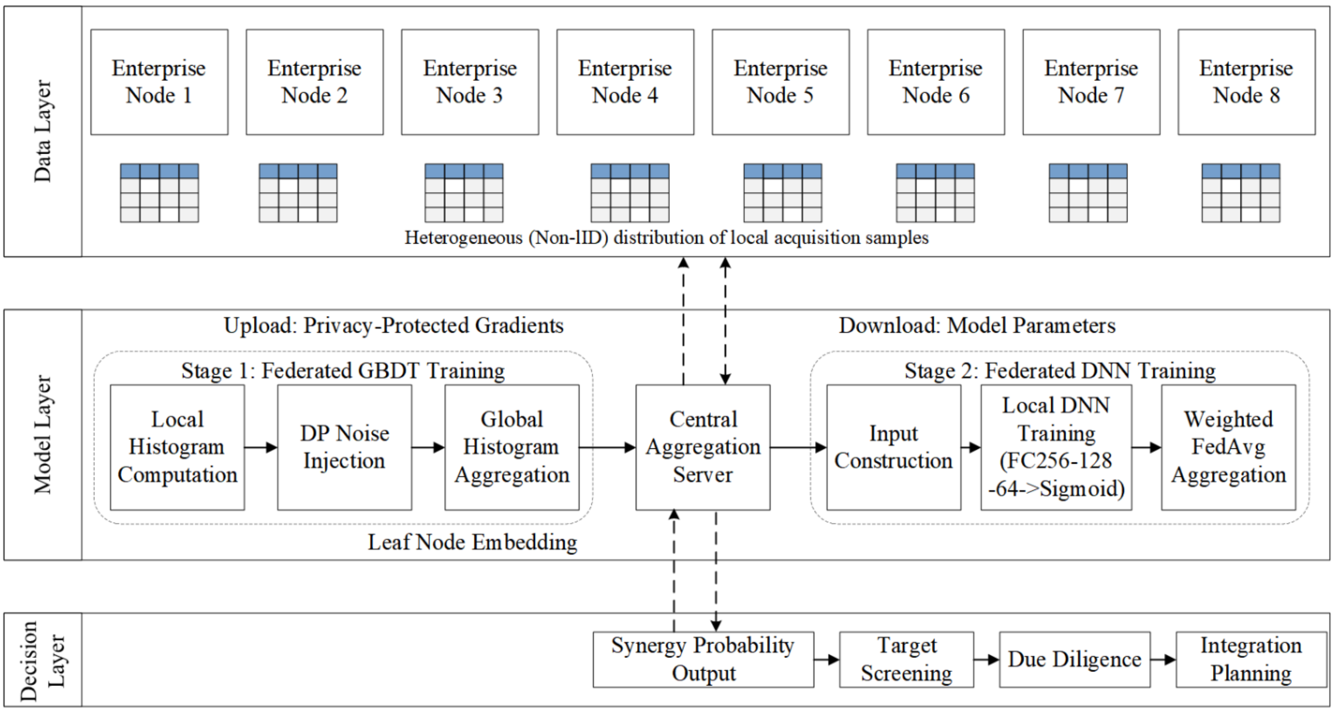

Figure 1. The proposed horizontal federated learning framework for AI capability synergy prediction in manufacturing M&A

Table 1. Descriptive statistics of the manufacturing M&A dataset

Variable	Mean	SD	Min	Max
Dependent Variable				
AI Capability Synergy Effect (1 = realized, 0 = not realized)	—	—	0	1
Synergy Realized (*n* = 84) / Not Realized (*n* = 202)				
Engineering Collaboration Indicators				
R&D Compatibility Index	0.447	0.198	0.031	0.912
Production System Interoperability Score	0.513	0.221	0.044	0.956
AI Talent Overlap	0.382	0.176	0.018	0.871
Financial & Operational Control Variables				
Transaction Size (log, CNY million)	5.847	1.324	2.631	9.412
Acquisition Premium Rate (%)	28.63	14.27	2.10	81.40
Target Firm Tobin's Q	1.843	0.934	0.612	5.217
Industry Concentration (HHI)	1847.3	623.4	412.0	3986.0

Note: SD = standard deviation. All engineering collaboration indicators are Min-Max normalized to [0, 1]. The positive-to-negative sample ratio of approximately 1:2.4 motivates the joint adoption of SMOTE oversampling and Focal Loss. The dataset encompasses cases from 23 manufacturing enterprises spanning 2015–2024, with industry distribution comprising automotive manufacturing (38.5%, n = 110), electronic equipment (34.3%, n = 98), and high-end equipment (27.2%, n = 78). HHI = Herfindahl-Hirschman Index.

2.3 Horizontal federated learning architecture design

The feasibility of the horizontal federated learning framework is based on the data-structure property that the 23 corporations included in this study share consistent feature spaces but have independent M&A data sets. This condition meets all the basic requirements of horizontal data splitting. Considering some corporations do not have enough historical M&A data for efficient local model training, node merger is performed for corporations whose sample sizes are less than the threshold ($n_i < 5$). Consequently, 8 federated nodes were formed while maintaining the same characteristics as the original data set, so that the natural variation among samples from different nodes accurately reflects the non-uniform level of M&A activity among manufacturing firms. In the Non-IID data scenario, there are practical challenges to the convergence of federated optimization algorithms, and a comprehensive literature review on federated learning under communication-efficiency constraints and data heterogeneity is valuable for addressing this issue [20].

The federated training of the gradient boosting tree module is implemented through a histogram aggregation mechanism, where each node computes the first-order gradient $g_j = \partial\ell/\partial\hat{y}_j$ and second-order gradient $h_j = \partial^2\ell/\partial\hat{y}_j^2$ for local samples across each feature dimension k, and constructs gradient statistical histograms $\mathcal{H}_i^k = \{(\sum_{j\in\text{bin}_b} g_j, \sum_{j\in\text{bin}_b} h_j)\}_{b=1}^{B}$ over discrete binning intervals of continuous feature values. The central server performs bin-wise weighted aggregation of the histograms uploaded by each node:

$$\mathcal{H}^k = \sum_{i=1}^{N} \mathcal{H}_i^k = \{(\sum_{i=1}^{N} G_{i,b}^k, \sum_{i=1}^{N} H_{i,b}^k)_{b=1}^{B} \quad (1)$$

where $G_{i,b}^k$ and $H_{i,b}^k$ denote the first-order and second-order gradient statistics of node i within the b-th bin of feature k, respectively.

The server determines the optimal split points for each feature by maximizing the split gain function based on the global histogram $\mathcal{H}^k$ and distributes them to all nodes, which then synchronously construct the tree nodes locally. In essence, this approach benefits from the fact that, unlike raw data samples, the nodes only pass on gradient statistics in the form of histograms that cannot be used to restore the data distribution of any individual node, ensuring data integrity while still enabling global tree structure training [21]. The deep learning model is trained in a federated setting using weighted federated averaging, where the weighting factor accounts for the effective sample size of each participating node. This weighting scheme addresses the potential problem posed by unequal sample sizes across nodes by automatically compensating for any inherent bias introduced by data scaling. The algorithm for updating the global model parameters is:

$$w^{t+1} = \sum_{i=1}^{N} \frac{n_i}{n} w_i^{t+1} \quad (2)$$

where w_i^{t+1} denotes the model parameters uploaded by the node i upon completing local training in round t. The theoretical study on the progress and challenges in federated learning systematically demonstrates the efficiency of weighted aggregation techniques for dealing with the heterogeneous nature of data from different nodes [22]. Practical implementation using industrial federated learning systems in cross-industrial big data protection applications demonstrates the stable operation of the aggregation technique in real-world industrial environments [23]. At the privacy protection level, Gaussian noise satisfying (ε, δ)-differential privacy conditions is injected into the gradient histogram statistics and DNN parameter gradients uploaded by each node:

$$\tilde{g}_i = \frac{g_i}{max(1, ||g_i||_2/C)} + \mathcal{N}(0, \sigma^2 C^2 I) \quad (4)$$

where C is the gradient L2 norm clipping threshold and σ is the noise intensity coefficient. The gradient clipping operation first constrains each node's gradients within a sphere of radius C, then superimposes isotropic Gaussian noise with variance $\sigma^2 C^2$, a mechanism that mathematically guarantees that the raw data of any individual node cannot be probabilistically inferred from the aggregated results. From research conducted on the use of hybrid differential privacy and adaptive compression algorithms in industry-edge computing, it has been shown that the protection technique works well to prevent privacy leakage problems without compromising communication efficiency [24]. Further studies have been conducted on the security threat to gradient information in federated learning, and this supports the use of differential privacy noise injection along with secure aggregation techniques to provide adequate privacy in multi-party communication [25]. The Top-K sparsification algorithm is used for efficient communication by sending gradients containing only the top-K largest values, setting the rest to zero.

2.4 Engineering collaboration indicator system quantification

The engineering technical compatibility factor under Resource-Based View Theory serves as the fundamental explanatory factor for AI capability synergy effect prediction, due to the creation of complementary technological knowledge, engineering architecture, and professional talent systems through the spill-over and recombination of capabilities, leading to synergy effects much larger than those from financial resources alone. This study develops an engineering collaboration index system comprising three factors, where each factor is quantified in a manner that can be measured using public information. The R&D compatibility index measures the structural compatibility of technological knowledge bases using a linear combination of the overlap between technological categories and the R&D investment structures of both parties involved. Let P_A and P_T denote the valid patent sets of the acquiring and target firms, respectively, with patent technology overlap measured using the Jaccard similarity coefficient:

$$\text{PatOvlp} = \frac{|\mathcal{P}_A \cap \mathcal{P}_T|}{|\mathcal{P}_A \cup \mathcal{P}_T|} \tag{3}$$

R&D investment structure similarity is measured by the cosine similarity between the R&D sub-category investment proportion vectors $\mathbf{r}_A, \mathbf{r}_T \in \mathbb{R}^m$ of both parties, where m denotes the number of R&D sub-categories:

$$\text{RDSim} = \frac{\mathbf{r}_A \cdot \mathbf{r}_T}{||\mathbf{r}_A||_2 \cdot ||\mathbf{r}_T||_2} \tag{6}$$

A composite indicator for R&D compatibility can be computed via the convex combination of the two constituent sub-indicators using weight coefficient $\alpha \in [0,1]$ chosen from the hyperparameters by maximizing the area under the ROC curve (AUC). This approach guarantees that the weight allocation is data-driven and objective:

$$\text{RDC} = \alpha \cdot \text{PatOvlp} + (1-\alpha) \cdot \text{RDSim} \tag{4}$$

Production system interoperability is a comprehensive assessment of engineering compatibility between the production systems of both M&A firms, covering three sub-dimensions: communication protocol compatibility of manufacturing equipment (measured by the adoption ratio of OPC-UA and MQTT standards); factory digitalization maturity (scaled continuously from 0 to 1 based on the five-level industrial digitalization maturity model); and ERP/MES interface compatibility (based on expert scoring). The weight vector$\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)^\top$for the three sub-dimensions is determined through the Analytic Hierarchy Process (AHP) by a panel of manufacturing M&A domain experts, with internal logical consistency of the weight assignment verified through Consistency Ratio (CR < 0.1) testing, and the composite score computed as the inner product of the weight vector and the normalized sub-dimension score vector $s \in [0,1]^3$:

$$\text{PSI} = \omega^\top s \tag{8}$$

The adoption of the AHP weighting mechanism can more accurately reflect the varying strategic significance of each engineering compatibility component during the integration of AI capabilities, compared with the equal-weight summation method. The AI talent overlap measure quantifies the compatibility of the AI professional talent structures of both M&A parties, using the cosine similarity between the skill map vectors for AI-related positions at both firms as the metric. The skill map vectors $s_A, s_T \in \mathbb{R}^q$, where q denotes the total number of skill dimensions, which are constructed from job requirement data on public recruitment platforms including LinkedIn, Zhaopin, and BOSS Zhipin. They encompass demand frequencies for algorithm engineering, data engineering, and AI application positions across skill dimensions such as deep learning, computer vision, and natural language processing. Similarity is computed after TF-IDF weighted normalization:

$$\text{AITalent} = \frac{\mathbf{s}_A \cdot \mathbf{s}_T}{||\mathbf{s}_A||_2 \cdot ||\mathbf{s}_T||_2} \tag{9}$$

After Min-Max normalization to the [0,1] interval, the three engineering collaboration indicators (RDC,PSI,AITalent) are concatenated with the financial operational control variables to form the complete model input feature vector $x = [x_{\text{eng}}^\top, x_{\text{fin}}^\top]^\top \in \mathbb{R}^d$, where $x_{\text{eng}} \in \mathbb{R}^3$ *is the engineering collaboration feature sub-vector and* $x_{\text{fin}} \in \mathbb{R}^{d-3}$ is the financial operational feature sub-vector, with the joint input design directly embodying the core methodological position of the engineering-financial dual-dimensional collaborative prediction framework proposed in this study.

2.5 Federated gradient boosting and deep neural network fusion model

GBDTs have been extensively used for high-dimensional prediction thanks to their excellent nonlinear modeling ability with respect to structured tabular data [26]. However, when GBDTs and DNNs are utilized in federated distributed learning, the conflict between the non-differentiability of GBDT tree structure parameters and the continuously optimized gradients of DNN parameters makes federated joint optimization infeasible. To address this problem, the proposed method adopts a two-stage decoupling federated training strategy, which involves performing federated GBDT training in the first stage via cross-node gradient histogram aggregation and delivering leaf node embeddings as core inputs for the federated DNN optimization stage, with complete parameter space decoupling for both stages to eliminate gradient conflicts and enable independent parameter tuning. Cross-node parameter-space decoupling training in federated heterogeneous neural networks has been implemented in practice [27], and federated reinforcement learning optimization in the context of industry IoT digital twins offers a critical methodological reference for the two-stage design [28]. Following convergence of federated GBDT training, each node extracts leaf node embedding vectors $e_j \in \{0,1\}^T$ for local samples, where T is the total number of GBDT trees, with each tree corresponding to a one-hot encoded leaf node index, and these vectors encode high-order nonlinear partition representations of samples in GBDT decision space in sparse binary form. The DNN input vector is constructed by concatenating the leaf node embeddings with the original engineering collaboration indicator vectors:

$$z_j = \left[e_j^{\mathrm{T}} || x_{\mathrm{eng},j}^{\mathrm{T}}\right] \in \mathbb{R}^{T+3} \quad (5)$$

where || represents the vector concatenation operation. The rationale for including the engineering collaboration indicator x_{eng} in the DNN input layer rather than relying solely on leaf node embeddings alone is that leaf node embeddings are a discrete compression representation of original data, and thus the semantic precision of engineering collaboration indicators may be lost during this stage. Direct concatenation ensures that the entire continuous-representation capability of this critical feature is preserved during further deep nonlinear transformations. This DNN model uses a fully connected architecture with hidden layer sizes of 256, 128, and 64 neurons, respectively, where each hidden layer undergoes batch normalization before being fed into another layer, and a dropout rate of 0.3 is applied to reduce federated training overfitting risks. The output layer employs sigmoid activation for mapping outputs into probabilities of realizing AI capability synergy effects through the following formula:

$$\hat{y} = \sigma(W_L h_{L-1} + b_L), \quad \sigma(z) = \frac{1}{1+e^{-z}} \quad (6)$$

where h_{L-1} denotes the hidden layer output of the $(L-1)$-th layer, W_L and b_L are the weight matrix and bias vector of the output layer, respectively, and binary classification prediction is performed using 0.5 as the decision threshold.

To intuitively illustrate the iterative algorithm of the proposed two-stage federated learning approach, Figure 2 presents the procedure from federated GBDT histogram training to federated DNN gradient optimization in a horizontally divided, two-stage manner, along with the separation interface between stages and the types of data exchanged between nodes and the central server.

Figure 2 depicts the fundamental engineering principle behind the proposed two-stage decoupling framework. In the first stage, nodes calculate local gradient histograms, add differential privacy noise, and submit protected gradient information to the central server to perform global histogram merging operations, where the server allocates optimal split points, enabling nodes to locally construct trees and form leaf node representations, based on theoretical analysis of federated tree model security and gradient-free boosting convergence [29,30]. In the second stage, leaf node representations and engineering collaboration vectors are concatenated and passed to a three-layer feed-forward neural network, where Top-K sparse gradient representations are transmitted for weight-based FedAvg aggregation. The training process stops if the AUC change on the validation set remains below the tolerance limit for 10 iterations; otherwise, it loops back to the first stage for another round. The global training loss function adopts Focal Loss to address class imbalance:

$$\mathcal{L}_{\mathrm{focal}} = -\frac{1}{n}\sum_{j=1}^{n}\left[\alpha_t(1-p_t)^{\gamma}\, log(p_t)\right] \quad (7)$$

where $p_t = \hat{y}$ when $y = 1$ and $p_t = 1 - \hat{y}$ when $y = 0$.
The focusing parameter $\gamma = 2$ substantially suppresses the gradient weights of high-confidence easily classified samples through modulation of the $(1-p_t)^{\gamma}$ factor, shifting the model optimization focus toward difficult-to-classify cases near the synergy effect discrimination boundary, while the class weight coefficient α_t is adaptively set according to the inverse of the positive-to-negative sample ratio in the training set and operates in conjunction with SMOTE oversampling.

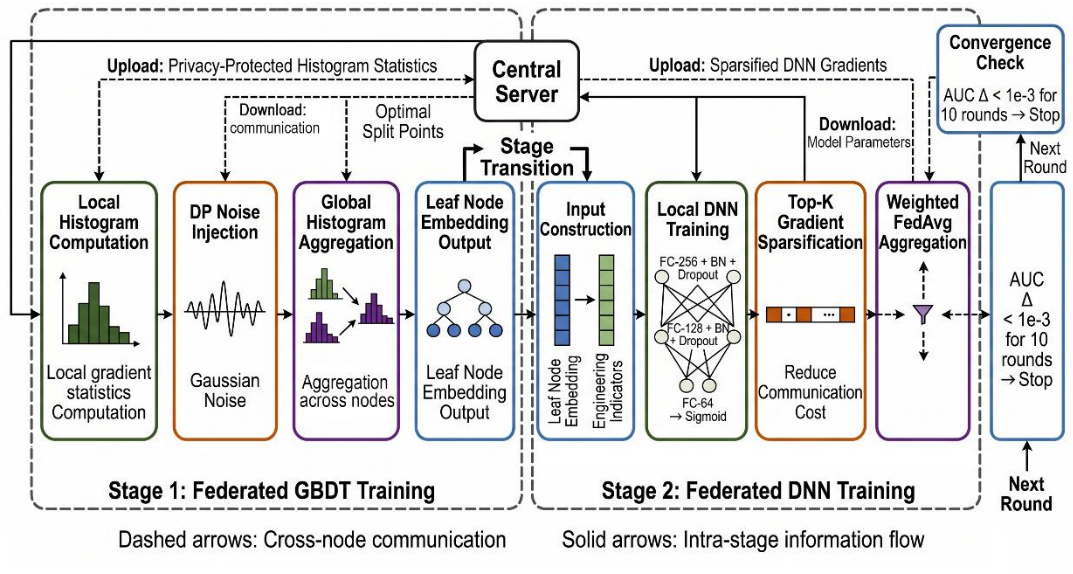

Figure 2. Two-stage federated training procedure integrating gradient boosting trees with deep neural networks

The choice of an appropriate loss function to ensure stable federated model convergence under heterogeneous data distribution, informed by optimization techniques for designing adaptive vertical federated learning models with unbalanced features [31], directly supports the rationale for using Focal Loss over cross-entropy in the proposed federated framework. Convergence is defined as a change of less than 10^{-3} in global validation set AUC within 10 rounds, and the overall process of federated training is outlined in Algorithm 1.

Algorithm 1. Two-stage federated training for AI capability synergy prediction

Input: Local datasets $\mathcal{D}_i$ for N nodes; tree number T; privacy parameters C, σ; threshold τ
Output: Global model w^*; synergy probability $\hat{y}$ for each M&A case
Initialize global model w^0; apply SMOTE on each local $\mathcal{D}_i$
// Stage 1: Federated GBDT Training
for each tree $t = 1$ to T do
Each node computes local gradient histograms $\mathcal{H}_i^k$ from local M&A samples
Inject Gaussian noise satisfying (ε, δ)-DP; upload $\tilde{\mathcal{H}}_i^k$ to central server
Server aggregates global histogram $\mathcal{H}^k$; determines optimal split points
Broadcast split points to all nodes; each node constructs local tree nodes
Each node generates leaf node embedding $e_j \in \{0,1\}^T$
end for
// Stage 2: Federated DNN Training
Repeat
Each node concatenates e_j *with* x_{eng} to form input z_j
Forward pass: FC-256 + BN + Dropout → FC-128 + BN + Dropout → FC-64 → Sigmoid
Compute Focal Loss L_{focal} with $\gamma = 2$; apply Top-K gradient sparsification
Upload sparsified gradients; server executes weighted FedAvg: $w^{t+1} = \sum_{i=1}^{N} \frac{n_i}{n} w_i^{t+1}$
Broadcast w^{t+1} to all nodes; evaluate global validation AUC
until AUC change $< \tau$ for 10 consecutive rounds
return w^*; output $\hat{y} = \sigma(W_L h_{L-1} + b_L)$ for M&A synergy prediction

This is illustrated in Algorithm 1, which outlines the entire process of two-stage federated learning in pseudocode, including computation of the local gradient histogram under differential privacy guarantees, global histogram aggregation at the server, distribution of the split points, creation of leaf embeddings, formation of the inputs to the DNN model through the combination of embeddings and engineering variables, optimization via Focal Loss using weighted FedAvg, and early stopping based on the AUC criterion.

3. Results

3.1 Experimental setup

Experiments are conducted using a dataset comprising 286 real manufacturing M&A cases drawn from 23 companies, segmented into 8 real federated nodes depending on the nature of data ownership of each company, with sample sizes per node varying from 20 to 52, genuinely reflecting the Non-IID heterogeneity due to varying levels of M&A activities in manufacturing firms. Baseline models are configured based on the following assumptions. A centralized XGBoost model is built on the entire dataset without any limitations due to privacy protection requirements. It serves as a benchmark for gauging the algorithm's maximum achievable performance without data restrictions. Logistic regression is a well-established econometric baseline that measures improvements in the federated learning method compared to established approaches to analyzing M&A activities. A single-node local DNN is trained only on the data from that node, without federated training across other nodes. This will help in determining the marginal added value of the federated training methodology. Evaluation metrics encompass five categories: prediction accuracy, AUC-ROC, F1-Score, Precision, and Recall. Regarding hyperparameter configuration, the number of federated communication rounds is set to 100, local training epochs per node to 5, learning rate to 0.01, GBDT tree depth to 6, and DNN hidden layer dimensions sequentially to 256, 128, and 64, with the differential privacy noise coefficient initialized at $\sigma = 0.1$. All configurations are determined through grid search on the validation set to ensure experimental reproducibility. The sample size of 286 cases is determined by the availability of complete three-year post-merger financial records and sufficient AI-related asset disclosure across the 23 participating enterprises, representing the full population of qualifying manufacturing M&A transactions accessible through the designated data sources within the study period rather than a randomly drawn subset.

3.2 Ablation study

To systematically verify the independent role of each important module within the proposed framework, four model variants are created using the same partitioned dataset and performance assessment settings, with each variant tested by deleting or replacing a single major module. In particular, the four model variants are composed of deleting the DNN fusion module to reduce the framework back to the federated GBDT structure, deleting the engineering collaboration indicators to keep only the financial variable indicators, replacing the federated learning scheme to conduct centralized training using the same amount of data to exclude the effects of privacy-preserving limitations, and reducing the framework to individual local machine learning to exclude the benefits brought by data complementation between different companies. This allows for systematic dissection of the predictive power of the engineering collaboration indicators, the modeling improvements in the DNN fusion framework, and the information-exchange advantages of the federated learning scheme compared with local learning. The full test results for all model variants across five evaluation criteria are shown in Table 2. As illustrated in Table 2, omitting engineering collaboration indicators results in a maximum reduction of 6.7 percentage points, confirming the irreplaceable predictive ability of these indicators, whereas reducing the model to a single node yields the minimum accuracy of 79.2 percent.

In relation to the quantitative analysis of privacy protection effectiveness costs, the differential privacy noise parameter σ is an essential engineering parameter when deploying federated learning platforms. To illustrate the quantitative trade-off between privacy protection effectiveness and prediction accuracy, which provides valuable guidance for parameters in practical applications, the experiment varied σ from 0.05 to 0.5 while keeping other hyperparameters fixed, and recorded the validation accuracy at each value of σ. The trade-off between the effectiveness of privacy protection and prediction accuracy is illustrated in Figure 3.

Table 2. Ablation study results: performance comparison across model variants with individual modules removed

Model Variant	Accuracy (%)	AUC-ROC	F1-Score	Precision	Recall
Complete FL Framework (Proposed)	87.3	0.921	0.869	0.883	0.856
w/o DNN Fusion (FL-GBDT Only)	83.1	0.879	0.824	0.836	0.812
Centralized Training (No FL)	82.1	0.874	0.813	0.831	0.796
w/o Engineering Indicators	80.6	0.853	0.791	0.808	0.774
Single-Node Local Training	79.2	0.834	0.768	0.776	0.761

Table 3. Predictive performance comparison between the proposed FL framework and baseline models

Model	Accuracy (%)	AUC-ROC	F1-Score	Precision	Recall
Proposed FL Framework	87.3	0.921	0.869	0.883	0.856
Centralized XGBoost	82.1	0.874	0.813	0.831	0.796
Single-Node Local DNN	76.8	0.812	0.754	0.763	0.746
Logistic Regression	74.5	0.791	0.738	0.744	0.731

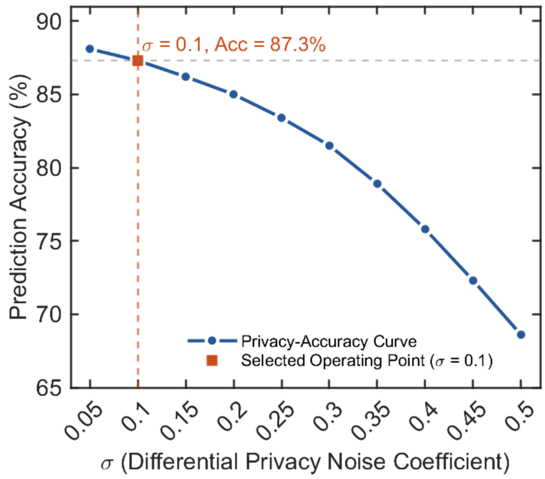

Figure 3. Privacy-accuracy trade-off curve under varying differential privacy noise coefficients σ

Figure 3 shows the privacy-utility curve for different values of the differential privacy noise parameter. The curve is monotonic: accuracy degrades slightly at low noise values and more rapidly at high noise values. The chosen operating point at $\sigma = 0.1$ corresponds to an accuracy of 87.3%. This indicates that the developed framework offers comparable accuracy while maintaining differential privacy properties.

3.3 Predictive performance comparison

Building on the analysis of individual modules, the entire federated learning model is thoroughly evaluated through comparative analyses across three models and five parameters. This evaluation assesses the predictive capabilities of the federated learning model under stringent data privacy constraints, quantifying performance differences among federated collaborative learning, centralized machine learning, and conventional econometrics techniques.

Experimental conditions strictly ensure consistency in terms of the data split and parameter measurement. The results from the full comparison of the performance metrics are shown in Table 3.

Table 3 shows that the accuracy of the proposed federated learning model is 87.3%, which is 5.2 % higher than that of the centralized XGBoost model trained without privacy constraints, demonstrating the value of cross-node collaboration in providing information complementarity beyond data sufficiency and model sophistication.

3.4 Feature importance analysis

The substantial improvement in prediction accuracy indicates the framework's validity. Furthermore, a deep exploration of how each aspect of engineering collaboration contributes to synergy effect prediction provides the essential empirical basis for testing the primary methodological assumption of this study: the dominant role of engineering compatibility in predicting AI capability synergy effects. For this experiment, the federation strategy for calculating SHAP values uses an approach in which each node computes local SHAP values and transmits the average and variance of each SHAP feature to the server. The server performs weighted averaging, using each node's sample size as its weight, to compute the global feature importance ranking. This strategy ensures privacy-preserving SHAP analysis in a federated environment without leaking any original sample data throughout the procedure. The global feature importance ranking based on SHAP values is shown in Figure 4. As shown in Figure 4, the group importance of engineering collaboration measures is 0.42, which is significantly higher than that of the financial and operational measures, with a group importance of 0.30. The highest-ranked individual attribute is the R&D Compatibility Index.

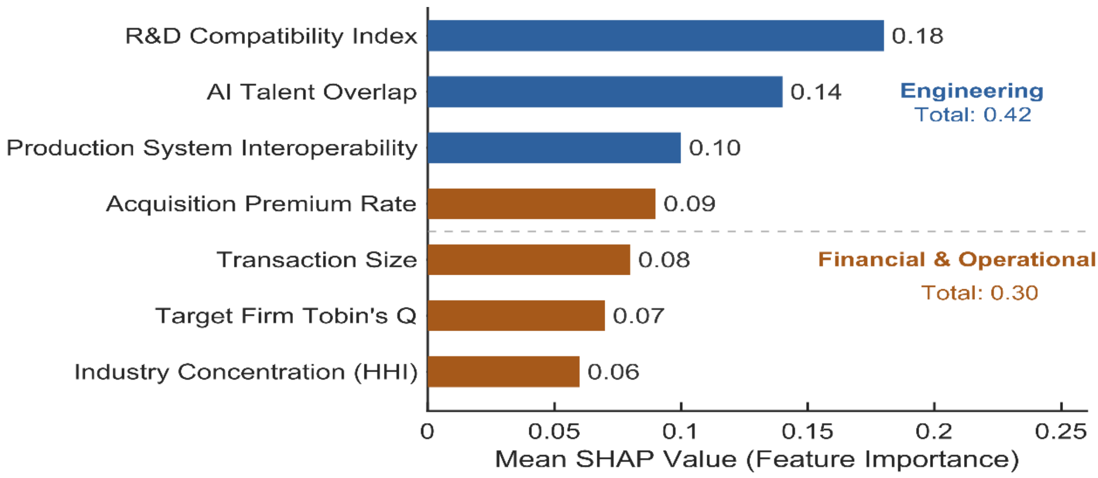

Figure 4. SHAP-based feature importance ranking of engineering collaboration indicators and financial variables

3.5 Post-merger economic value effects

Feature importance analysis confirms the predominant role of engineering collaboration in the predictive mechanism. Furthermore, an investigation of post-merger economic value creation trends among AI-intensive acquirers provides the final empirical verification of this study's main argument through a value realization perspective. In this experiment, the data are partitioned into AI-intensive and non-AI-intensive acquirers based on the sample median of AI-related R&D costs to total R&D costs. Differences in post-merger economic value increments between the two groups are statistically confirmed via two-sided independent-samples t-tests. To assess the evolution of economic value increments following mergers, this study analyzes value premium dynamics over the three-year post-merger period (Years 1, 2, and 3) and applies the same analysis to the high engineering collaboration group, defined as the top 30% of engineering collaboration scores. Economic value increment dynamics over three-year post-merger periods for both acquirer groups are depicted in Figure 5.

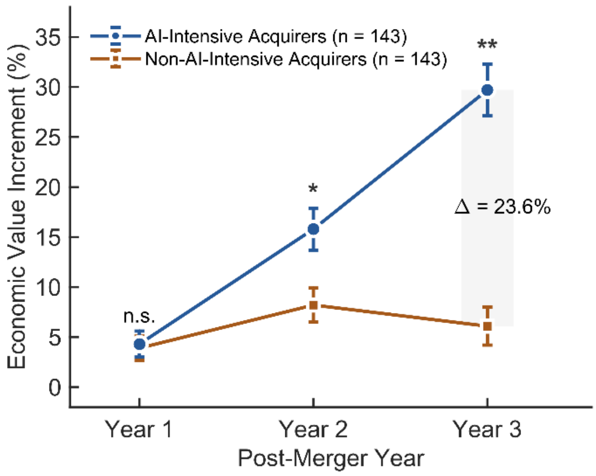

Figure 5. Post-merger economic value premium trajectories of AI-intensive versus non-AI-intensive acquirers across three years
Note: Error bars represent ±1 standard error. Significance markers indicate between-group differences at each time point: * $p < 0.05$, ** $p < 0.01$, n.s. not significant.

Post-acquisition economic value increments of AI-intensive and non-AI-intensive acquiring firms over three years are shown in Figure 5. AI-intensive firms exhibit a consistently increasing trend, while non-AI-intensive firms maintain a stable trend before slightly declining. There is no significant difference between the two groups in Year 1, followed by statistical significance in Year 2 and the greatest deviation in Year 3 ($\Delta = 23.6\%$).

4. Discussion

The total feature importance for engineering collaboration indicators is 0.42, notably higher than the feature importance of financial indicators at 0.30, with the R&D compatibility index ranking first at 0.18. Such results are confirmed by research showing that innovations, product improvements, and the market growth they facilitate through AI investments cannot be achieved through financial resource integration alone; rather, they require deeper integration of technological capabilities [32]. The positive impact of M&As on innovation also depends on the successful recombination of R&D assets and the accumulation of digital knowledge [33]. Moreover, the positive influence of R&D collaboration associated with AI applications on firm performance further supports the validity of the predictive information conveyed by engineering compatibility [34]. The value premium of AI-intensive acquirers is further enhanced in the high-engineering-collaboration sample, providing strong evidence of a time-lagged value-creation phenomenon that aligns with theoretical assumptions about the gradual nature of technical adaptation in AI capability integration [35]. Digital M&A benefits for firm performance, however, depend on the fulfillment of capability-matching prerequisites [36]. This study provides quantitative evidence for the boundary condition of capability matching, which applies specifically to the engineering compatibility dimension. The conditional effects of digital M&A on driving innovation upstream provide the basis for a complementary effect [37]. The proposed approach improves upon the centralized XGBoost baseline by 5.2% at $\sigma = 0.1$, while the drop in precision remains within an acceptable range. Furthermore, the success of high-quality predictive collaboration among enterprises under privacy-preserving conditions provides a meaningful reference for the technical feasibility of multi-enterprise collaborative modeling in manufacturing M&A contexts [38]. The primary limitation of this study is that the dataset is predominantly drawn from Chinese and East Asian manufacturing markets, and the validity of the results obtained in other cultural contexts, namely European and North American institutions, remains unknown. In addition, because the metrics of engineering collaboration are calculated using publicly available data, missing information from some enterprises may lead to measurement errors. Further research could include applying vertical federated learning to address heterogeneity in feature spaces across enterprises.

5. Conclusion

In this study, a horizontal federated learning approach for predicting the synergistic effects of AI capabilities on manufacturing M&A is developed using a two-stage, decoupled training scheme. The approach employs a two-stage, decoupled training strategy that combines a federated histogram-aggregated gradient-boosted tree with a FedAvg-based deep neural network, enabling cross-enterprise collaborative modeling without any raw data leakage. Additionally, an engineering collaboration index system is established, comprising an R&D compatibility index, a production system interoperability index, and an AI personnel complementarity index. Empirical analysis of 286 real M&A events across 23 manufacturing enterprises shows that the horizontal federated learning model achieves 87.3% prediction accuracy, outperforming the centralized XGBoost model (82.1%) and the logistic regression model (74.5%). Furthermore, the overall importance of engineering collaboration indicators is 0.42, substantially higher than that of financial indicators (0.30). Moreover, the value premium of AI-focused acquiring firms exhibits a time-lagged increasing trend, peaking at $\Delta = 23.6\%$ in the third post-M&A year. These conclusions provide a methodological link between privacy-preserving machine learning and strategic management research on manufacturing M&A, illustrating the critical importance of engineering compatibility for forecasting AI-based M&A synergy effects. The suggested decision support tool can be directly employed at the three major decision-making phases of target selection, due diligence, and integration planning, offering a comprehensive chain decision support solution that considers both data compliance and the precision of AI-based M&A decisions in the age of intelligent manufacturing.

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] He, Qiujie, et al. "Research on the impact of artificial intelligence on the manufacturing industry chain resilience." Scientific Reports 15.1 (2025): 33432.

[2] H. D. Tang, S. H. Fang, and D. C. Jiang, "Market performance in digital transformation: Can digital M&As enhance manufacturing firm's market power?," Journal of Quantitative and Technological Economics, vol. 39, no. 12, pp. 90-110, 2022.

[3] M. Zhu and Z. Meng, "Fuzzy comprehensive evaluation model of M&A synergy based on transfer learning graph neural network," Computational Intelligence and Neuroscience, vol. 2021, no. 1, p. 6516722, 2021.

[4] K. Routhu, V. Velaga, C. S. Moore, S. B. Boppana, P. C. R. Chinta, and K. M. Jha, "Leveraging machine learning techniques for predictive analysis in merger and acquisition (M&A)," Journal of Artificial Intelligence and Big Data, vol. 3, no. 1, pp. 56-71, 2023.

[5] H. Ramezani, S. Ghasemi, and M. Gonzalez, "Predicting mergers and acquisitions success with machine learning," Accounting, Finance and Computational Intelligence, vol. 1, no. 2, pp. 66-77, 2024.

[6] Yang, Dayu. "A Deep Learning Method for Predicting Mergers and Acquisitions: Temporal Dynamic Industry Networks." arXiv preprint arXiv:2404.07298 (2024).

[7] Q. Liu, "Predicting mergers and acquisitions: A machine learning-based approach," International Review of Financial Analysis, vol. 99, 2025.

[8] I. Kevin, K. Wang, X. Zhou, W. Liang, Z. Yan, and J. She, "Federated transfer learning based cross-domain prediction for smart manufacturing," IEEE Transactions on Industrial Informatics, vol. 18, no. 6, pp. 4088-4096, 2022.

[9] J. Ahn, Y. Lee, N. Kim, C. Park, and J. Jeong, "Federated learning for predictive maintenance and anomaly detection using time series data distribution shifts in manufacturing processes," Sensors, vol. 23, p. 7331, 2023.

[10] T. Deng, Y. Li, X. Liu, and L. Wang, "Federated learning-based collaborative manufacturing for complex parts," Journal of Intelligent Manufacturing, vol. 34, no. 7, pp. 3025-3038, 2023.

[11] Islam, Farzana, Ahmed Shoyeb Raihan, and Imtiaz Ahmed. "Applications of federated learning in manufacturing: Identifying the challenges and exploring the future directions with industry 4.0 and 5.0 visions." arXiv preprint arXiv:2302.13514 (2023).

[12] M. Mehta, M. V. Bimrose, D. J. McGregor, W. P. King, and C. Shao, "Federated learning enables privacy-preserving and data-efficient dimension prediction and part qualification across additive manufacturing factories," Journal of Manufacturing Systems, vol. 74, pp. 752-761, 2024.

[13] Llasag Rosero, Raúl, et al. "Label synchronization for Hybrid Federated Learning in manufacturing and predictive maintenance." Journal of Intelligent Manufacturing 35.8 (2024): 4015-4034.

[14] H. Liu, S. Li, W. Li, and W. Sun, "Efficient decentralized optimization for edge-enabled smart manufacturing: A federated learning-based framework," Future Generation Computer Systems, vol. 157, pp. 422-435, 2024.

[15] Yang, Nan, et al. "The role of digital transformation in mergers and acquisitions." The North American Journal of Economics and Finance 75 (2025): 102306.

[16] G. Huang, F. Ye, and L. Chen, "When does digital merger and acquisition create shareholder value? An empirical investigation in the Chinese context," Heliyon, vol. 10, no. 14, p. e34237, 2024.

[17] Liang, Peng, Lin Liang, and Xinhui Tang. "The impact of digital-oriented mergers and acquisitions on enterprise labor demand." International Review of Financial Analysis 96 (2024): 103778.

[18] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," ACM Transactions on Intelligent Systems and Technology, vol. 10, no. 2, pp. 1-19, 2019.

[19] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A survey on federated learning," Knowledge-Based Systems, vol. 216, p. 106775, 2021.

[20] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," IEEE Signal Processing Magazine, vol. 37, no. 3, pp. 50-60, 2020.

[21] Li, Qinbin, et al. "Fedtree: A federated learning system for trees." Proceedings of Machine Learning and Systems 5 (2023): 89-103.

[22] P. Kairouz et al., "Advances and open problems in federated learning," Foundations and Trends in Machine Learning, vol. 14, pp. 1-210, 2021.

[23] Y. Liu, T. Fan, T. Chen, Q. Xu, and Q. Yang, "FATE: An industrial grade platform for collaborative learning with data protection," Journal of Machine Learning Research, vol. 22, no. 226, pp. 1-6, 2021.

[24] B. Jiang, J. Li, H. Wang, and H. Song, "Privacy-preserving federated learning for industrial edge computing via hybrid differential privacy and adaptive compression," IEEE Transactions on Industrial Informatics, vol. 19, no. 2, pp. 1136-1144, 2023.

[25] J. Zhang, H. Zhu, F. Wang, J. Zhao, Q. Xu, and H. Li, "Security and privacy threats to federated learning: Issues, methods, and challenges," Security and Communication Networks, vol. 2022, p. 2886795, 2022.

[26] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785-794.

[27] H. Sun, Y. Zhang, M. Li, and Z. Xu, "FLFHNN: An efficient and flexible vertical federated learning framework for heterogeneous neural networks," in Wireless Algorithms, Systems, and Applications (WASA 2022), LNCS 13471, 2022: Springer.

[28] W. Yang, W. Xiang, Y. Yang, and P. Cheng, "Optimizing federated learning with deep reinforcement learning for digital twin empowered industrial IoT," IEEE Transactions on Industrial Informatics, vol. 19, no. 2, pp. 1884-1893, 2022.

[29] B. Chen, R. Zhang, and X. Hou, "Secure and efficient federated gradient boosting decision trees," Applied Sciences, vol. 13, no. 7, p. 4283, 2023.

[30] X. Qiu, H. Pan, W. Zhao, C. Ma, P. P. B. de Gusmao, and N. D. Lane, "Gradient-less federated gradient boosting trees with learnable learning rates," in arXiv preprint arXiv:2304.07537, 2023.

[31] J. Zhang, S. Guo, and Z. Qu, "Adaptive vertical federated learning on unbalanced features," IEEE Transactions on Parallel and Distributed Systems, vol. 34, pp. 2074-2085, 2023.

[32] T. Babina, A. Fedyk, A. X. He, and J. Hodson, "Artificial intelligence, firm growth, and product innovation," Journal of Financial Economics, vol. 151, p. 103745, 2024.

[33] Y. Bai and H. Zhang, "Research on the impact of enterprise mergers and acquisitions on technological innovation: An empirical analysis based on listed Chinese enterprises," PLoS One, vol. 19, no. 11, p. e0309569, 2024.

[34] Ming, Xin, Qiang Wang, and Yan Liu. "The performance implications of R&D collaborations on artificial intelligence." Technovation 145 (2025): 103271.

[35] Liu, Dewen, et al. "Artificial intelligence as a merger and acquisition catalyst: A resource-based view." International Review of Financial Analysis (2026): 105163.

[36] A. Hanelt, S. Firk, B. Hildebrandt, and L. M. Kolbe, "Digital M&A, digital innovation, and firm performance: An empirical investigation," European Journal of Information Systems, vol. 30, no. 1, pp. 3-26, 2021.

[37] Chen, Wenting, and Wenxin Liu. "Digital M&A triggers innovation: When the going gets tough, the tough get going." Research in International Business and Finance 78 (2025): 102957.

[38] Loftus, Tyler J., et al. "Federated learning for preserving data privacy in collaborative healthcare research." Digital Health 8 (2022): 20552076221134455.