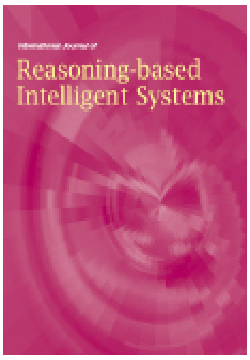


International Journal of Reasoning-based Intelligent Systems

ISSN online: 1755-0564 - ISSN print: 1755-0556
https://www.inderscience.com/ijris

Data mining techniques for intelligent educational management based on federated learning

Xiaoyin Liang

DOI: 10.1504/IJRIS.2025.10073544

Article History:

Received:	14 June 2025
Last revised:	10 August 2025
Accepted:	17 August 2025
Published online:	22 September 2025

Data mining techniques for intelligent educational management based on federated learning

Xiaoyin Liang

Academic Affairs Office,
Guangxi Vocational Normal University,
Nanning 530007, China
Email: LL271756476@126.com

Abstract: As intelligent education grows quickly, the amount of data about managing education is also expanding quickly. Finding useful information quickly while keeping data private is the key to making education management smarter. This work presents an intelligent education management data mining technique grounded in federated learning (FL) and formulates a multi-level system architecture. The system can safely share and intelligently analyse dispersed education data thanks to improvements to the FL algorithm and the addition of the DP protection mechanism. Experimental validation using the PISA dataset indicates that the suggested approach markedly enhances accuracy, F1 Score, and AUC index of the model, while safeguarding data privacy and security, hence exhibiting better performance and robust generalisation capability. The study's findings furnish a theoretical foundation and technological assistance for advancing the evolution of informatisation and intelligence within the realm of educational management.

Keywords: FL; educational administration; data mining; privacy protection; intelligent analysis.

Reference to this paper should be made as follows: Liang, X. (2025) 'Data mining techniques for intelligent educational management based on federated learning', *Int. J. Reasoning-based Intelligent Systems*, Vol. 17, No. 11, pp.12–25.

Biographical notes: Xiaoyin Liang received her Master's degree from Guilin University of Electronic Technology in 2016. She is currently an Associate Professor at Guangxi Vocational Normal University. Her research interests include vocational education technology and teaching management.

1 Introduction

1.1 Background and purpose of the study

As the digital transformation of global education speeds up, school administration is slowly evolving into a new phase of making decisions based on data and intelligence (Mohamed Hashim et al., 2022). Along with the building and upgrading of education informatisation systems, a lot of data about managing education has been systematically gathered, stored, and organised. This includes everything from how individual students learn to how teachers teach, as well as how to arrange teaching affairs and distribute educational resources. Not only is the amount of this data growing quickly, but its dimensions, structure, and content are also becoming more complicated.

But it isn't always easy to get useful information from schooling data. Data silos, privacy and security, and different platforms are becoming more and more common in real-world situations. First, education data is typically spread out and housed in many places, such schools, training centres, and education bureaux. This makes it hard to share and use data together, which makes it hard to get the full value out of it. Second, education data is incredibly private and sensitive, including things like student privacy, teacher assessment, management techniques, and other areas that need to be protected (Komljenovic, 2022). Without stringent protection measures, it's simple for data to leak, pose ethical problems, and even put people in danger of legal trouble. Also, different schools employ different management systems and data structures, which make it much harder to combine data and create a single model. In this situation, it's clear that the old way of doing data mining, which relied on centralised storage and processing, can't match the present need for efficient, safe, and collaborative data use in school administration.

FL, a novel framework for distributed machine learning, provides a forward-looking solution to the challenges mentioned above. The core concept of FL is that each participant, such as schools or educational bureaus, trains the model locally and only transmits the model parameters or gradients to a central server for aggregation. This approach enables multiple data sources to collaborate in developing a unified model. The original data remains on the user's device, thereby safeguarding their privacy by preventing storage in a centralised location. FL is particularly effective in the education sector, where there is an abundance of data, numerous stakeholders, and significant privacy concerns.

FL not only makes it easier to share data, but it also makes the system more stable and flexible. FL can lower the communication and computational load on the central node and speed up model training (Liu et al., 2022). Its modular design is also very flexible and can be easily added to the information systems that are already in place at different schools. FL is also in line with the current national policy guidance on data security, network security, and education informatisation because it respects data sovereignty in the system and guarantees data security in the technology. This means that FL has a lot of potential uses and strategic importance in the field of education.

FL has made a lot of progress in the last several years in data-sensitive fields like healthcare and finance, both in terms of practical results and new ideas. However, research in the field of education management is still very new. Education data has several common unstructured traits that set it apart from other fields. For example, learning logs, classroom interactions, assignment materials, and other types of data tend to have more text data, the quality of the data is not always the same, the frequency of data updates is not always the same, and it is harder to model. There are also a lot of different analysis tasks that may be done in school settings, like predicting how well a pupil will do, evaluating a teacher, and modelling academic warnings (Khan and Ghosh, 2021). There are big differences in how tasks need to be structured and trained. Consequently, a singular, generic FL solution is sometimes challenging to implement in intricate and evolving educational management contexts, necessitating immediate customised design and technical innovation tailored to educational data characteristics.

This study concentrates on FL-based intelligent education management data mining technology and constructs a distributed intelligent mining system architecture for multi-source education data, tailored to the specific requirements of the education management sector. The system design encompasses essential components, including data preparation technique, local model training process, parameter aggregation method, and privacy protection mechanism, with the objective of facilitating cross-institutional knowledge fusion while ensuring local data retention.

Compared with existing FL models in the education field, the system proposed in this paper makes incremental contributions in the following three aspects: First, in terms of aggregation strategy, it breaks through the traditional uniform weighting and dynamically adjusts weights based on local data volume and AUC performance, significantly alleviating the non-independent and identically distributed (non-IID) problem. Second, in terms of training mechanism, it is the first to jointly deploy adaptive regularisation and differential privacy (DP), which not only suppresses model drift but also ensures that the privacy budget is controllable. Third, in terms of application implementation, it establishes an end-to-end explainable process, incorporating SHAP analysis to identify key educational features, providing decision-makers with intuitive decision-making criteria, and addressing the current gap in educational FL research regarding explainability and risk assessment.

1.2 Status of the study

In recent years, the application of artificial intelligence in enterprise risk assessment has expanded rapidly, forming three common technical approaches: first, using a federated learning framework to solve data silo problems, and reducing model bias caused by non-independent identically distributed data through dynamic weighted aggregation strategies; second, introducing DP or homomorphic encryption to quantify operational risk while protecting sensitive information, achieving a balance between privacy and accuracy; third, using explainable models to extract key risk features, providing transparent decision-making basis for management. These approaches have been widely applied in scenarios such as financial credit, supply chain compliance, and cybersecurity, validating the effectiveness of dynamic aggregation, privacy computing, and explainability in risk assessment. This paper draws on these common technologies, adapting them for the first time to the educational management context to address the current gap in risk quantification research within the education sector.

As information technology and big data have advanced, data mining methods for managing education have gained significant popularity. These techniques can be utilised for various purposes, such as predicting student performance, optimising teaching resources, and providing early warnings about academic challenges. Many traditional data mining approaches rely on centralised learning, which requires data from different schools to be sent to a single server for uniform modelling. While this method can enhance data utilisation to some extent, it also presents several drawbacks, particularly concerning data privacy, complicating data sharing between institutions, and dealing with inconsistencies in the data. Educational data often contains sensitive information and storing it all in one location can easily result in privacy breaches. Additionally, discrepancies in data from various institutions hinder the effectiveness of unified modelling and complicate efforts to address the diverse needs of educational management across different contexts.

To tackle the issues, FL is introduced into the realm of educational management as an innovative distributed machine learning platform. The main idea behind it is that it doesn't need to retain raw data in one place (Ferrag et al., 2023). Instead, it uses local training and parameter aggregation to create cross-institutional joint models. This keeps data private and lowers the cost of sending data. In recent years, research on FL in educational management data mining has gradually developed, especially focused on usual problems such as student behaviour analysis, performance prediction and resource allocation. Initial research indicates that FL can maintain data security while considering model performance and generalisation capability, suggesting significant potential for widespread implementation.

Federated Averaging (FedAvg) is the most well-known and commonly used FL optimisation algorithm. It uses a local stochastic gradient descent (SGD) to train and periodically combines model parameters. This makes it the best choice for educational settings (Hu et al., 2022). Still, FedAvg has several real-world problems to deal with, such as non-IID data, system heterogeneity, and uneven training of participants. These problems make the model converge slowly and make it unstable. Researchers have developed a number of optimisation strategies to solve these problems:

The FedProx algorithm was proposed to address the issue of FedAvg's performance decline when dealing with non-IID data. This approach incorporates a regularisation term into the global model during each round of local training to prevent the local model from becoming overly biased. In educational contexts, the models and student demographics across different schools or districts demonstrate considerable variability, resulting in significant gradient discrepancies during model training. Researchers have investigated using FedProx in a system for analysing student behaviour across multiple schools (Palihawadana et al., 2022). This makes the model more stable and the aggregation better, and it gives a trustworthy base for high-quality individual analysis.

The FedNova algorithm, on the other hand, fixes the problem of clients not making steady progress in training. In the actual world, the fact that different schools have different levels of informatisation and processing capacity on their terminal devices can easily cause uneven local training rounds, which can change the aggregate impact. By normalising the local updates, FedNova lessens the aggregate bias that comes from variances in device performance. Researchers employed FedNova to promote the equity of contributions among schools and to improve the model's adaptability in contexts characterised by uneven resource distribution inside a regional instructional resource recommendation system founded on a federation structure.

FedOpt algorithms (like FedAdam and FedYogi) use first-order adaptive approaches in centralised optimisation to make the global model converge faster (Sun et al., 2023). The basic idea is to use the momentum mechanism or adaptive learning rate in the aggregation phase to improve the directionality and efficiency of the model update. FedOpt algorithm cleverly solves these problems by introducing the momentum mechanism or adaptive learning rate in the aggregation phase. The momentum mechanism can help the model better utilise the historical gradient information during the updating process, thus reducing the oscillation and enhancing the directionality of the update; the adaptive learning rate can dynamically adjust the learning rate according to the gradient change of each parameter, further improving the efficiency of the update. Some studies have tried to use FedAdam to model cross-region education assessment data. They found that it worked better in large-scale client environments and converged faster. This is especially useful for policy models that need to be quickly deployed and updated.

In the context of education management, the requirements of educational assignments vary significantly throughout schools, grades, and even among individual students. To address the issue of inadequate flexibility in universal global models, personalised federated learning (PFL) has been suggested. It has been confirmed in tasks such as learning route recommendation and teacher evaluation analysis that PFL considerably enhances local model performance while preserving global collaborative functionality, aligning with the educational principle of customised learning.

In recent years, researchers have also tried to bring the notions of multi-task learning and transfer learning into FL. This has led to new algorithms like FedMTL and FedTransfer that make the system more adaptable to different workloads and knowledge transfer situations. For instance, the migratory FL model is utilised to create an effective interface for knowledge migration of low-resource nodes in the task of interdisciplinary joint evaluation or cross-grade ability comparison.

On the other hand, because managing educational data needs a lot of security, technologies like DP, homomorphic encryption, and multi-party secure computing have been slowly added to the FL system architecture to make sure that the data transmission and model update process are more private. Some FL systems with better security have been used and tested in sensitive situations, like analysing data from university exams and keeping an eye on how students behave in online courses.

In general, the use of FL in educational management data mining is still in its early stages. Most research is focused on making FedAvg better and coming up with new algorithms. There are numerous ways that have tried to solve problems in different educational settings, but they still need to be looked at more closely in terms of how well they work with algorithms, how useful they are for the system, and how easy it is to understand their instructional worth.

Business operational risks are defined in this study as the possibility of educational decision-making biases or resource misallocations arising from data heterogeneity, privacy breaches, or model convergence anomalies during federated learning. To address the lack of risk assessment models, this study developed an 'educational operational risk assessment framework' tailored to educational management scenarios. This framework addresses three dimensions: data heterogeneity, model convergence stability, and the likelihood of privacy breaches. By combining the local data quality and model performance of participating institutions, it dynamically generates risk level labels (low, medium, high) to monitor potential decision-making risks in real time during federated training and trigger warning mechanisms when the system experiences abnormal fluctuations. This assessment mechanism has been embedded in the experimental process described in Section 5 to validate the system's robustness in real-world educational scenarios.

2 Federated learning

FL is a machine learning framework designed to train high-quality central models by leveraging training data distributed across a large number of clients. FL uses a distributed data storage approach, where the data is distributed across multiple local devices or data centres rather than being centralised in a single data centre or server (Bilal et al., 2018). These devices can be mobile devices, sensors, edge devices or data centres etc. The distributed data storage approach protects the data privacy of local devices. For model training, FL uses a form of local training followed by global aggregation. In local training, each device or data centre conducts model training locally and updates the model using its own data. These local model parameters are adjusted based on gradient descent or other optimisation algorithms applied to the local dataset. At the conclusion of each local training iteration, the device transmits the updated model parameters to a central server, which then aggregates or merges these parameters. This process allows the global model to incorporate updates from each device, resulting in a more accurate and comprehensive model. By combining distributed data storage with global aggregation of locally trained models, the benefits of data sharing are realised without actual data exchange, enabling each local device to share its trained model with others while preventing data leakage. FL supports incremental learning, which allows continuous improvement of the model without interrupting the model service (He et al., 2023). Each device can make incremental updates to the model locally without retraining the entire model.

Based on the difference in data distribution and the nature of learning tasks, FL can be classified into horizontal federation learning, vertical federation learning, and federated migration learning.

Horizontal federation learning is suitable for scenarios with less overlapping data samples and more overlapping features, such as banks in different regions (Wahab et al., 2021). In this model, each data owner aligns their respective data samples by user dimensions and trains the model together. In horizontal federation learning, the participants do not share the raw data directly, but exchange model parameters or gradients through a central server to achieve privacy protection.

Longitudinal federation learning is suitable for scenarios where data samples overlap more and features overlap less, such as banking and e-commerce in the same region. In longitudinal federation learning, the data owner aligns the data samples by feature dimensions and trains the model through encryption to ensure that data privacy is protected.

Federated Migration Learning is suitable for scenarios where both data samples and feature overlap are small, such as banks and e-commerce in different countries. Federated Migration Learning enables knowledge sharing and model optimisation through migration learning strategies such as model migration, feature migration or instance migration, while protecting data privacy (Zhang et al., 2022).

FedAvg algorithm is the most used FL optimisation algorithm. In short, the main idea behind the FedAvg algorithm is to combine local model training with global model updating by having each client upload model parameters (like weight vectors) they got from training on local data to a central server. The server then does a weighted average of these parameters to make a global model, which is then sent back to each client so they can start the next round of training.

Same as the conventional optimisation algorithms, its essential idea is to perform stand-alone optimisation of the local SGD of the data holders and to perform clustering on a central server, and its training objective is to find out the optimal model parameter ω to minimise the total training loss $f(\omega)$. The objective function of FedAvg algorithm is defined as follows:

$$f(\omega)=\sum_{k=1}^{K}\frac{n_k}{n}F_k(\omega) \tag{1}$$

$$F_k(\omega)=\frac{1}{n_k}\sum_{i\in P_k}f_i(\omega) \tag{2}$$

$$f_i(\omega)=l\left(x_i,y_i,\omega\right) \tag{3}$$

In equation (1) and equation (2), K is the number of participants, P is the k^{th} participant local sample set, n_k is the number of k^{th} participant local samples, n is the sum of all participant sample numbers, and $F_k(\omega)$ is the k^{th} participant local objective function. In Equation (3), $f_i(\omega)$ is the loss function generated by the model parameter ω for the i^{th} sample (x_i, y_i) in P_k. FedAvg is a relatively basic federated optimisation algorithm that is relatively simple to deploy and has a wide range of applications.

It makes sense to break the process down into three steps: initialisation, local training, and global aggregation. In particular, the central server first sets up the global model parameters and sends them out to all the participants. Using local private data, each client does several rounds of local training on this initial model. Most of the time, they use small batch SGD or a similar technique (Jain et al., 2018). After the local training is done, each client gives the server the local model parameters (or changes to them). To update the new global model, the server takes a weighted average of the model parameters it gets from each client based on how much data each client sends. This goes on until the model converges or a certain number of iterations have been reached.

In fact, FedAvg has proved that it can scale well and communicate quickly. It works especially well in educational settings where resources are limited, or the network is unstable. The client only needs to upload parameters instead of raw data, which lowers the cost of communication and the risk to privacy. The local training, on the other hand, lets the system use all the terminal devices' computational power, which eases the load on the servers and works well with edge computing. These traits make FedAvg the simplest and most popular optimisation method in FL research and applications.

The FedAvg algorithm performs poorly when the data is non-IID and the system is heterogeneous. In education management, the data of different schools varies greatly, which can easily lead to local model conflicts and affect the global model convergence and performance. Meanwhile, inconsistent client resources can reduce training efficiency and even cause nodes to drop out, slowing down overall progress.

In short, FL is a new type of distributed machine learning that offers a good way to deal with data silos and privacy concerns in education management by making it possible for several parties to work together on training while yet protecting data privacy. At present, FedAvg and its many enhancement algorithms consistently enhance the flexibility of FL in heterogeneous data and non-IID contexts, hence facilitating the implementation of educational data mining tasks. However, for the varied requirements and intricate constraints of educational contexts, additional optimisation of algorithm performance and improvement of system stability and personalisation capabilities remain necessary. The following chapters will examine the design and implementation of FL-based intelligent data mining systems within the framework of intelligent education management applications.

3 Data mining techniques for education management

As information technology keeps becoming better, the field of education management has gathered a lot of different kinds of data, including students' learning behaviour, instructional activities, curricular resources, and test scores. Data mining technology for education management has arisen, seeking to accomplish intelligent monitoring, scientific assessment and precise decision-making in the education process through in-depth analysis of these data (Lutfiani and Meria, 2022). This technology includes important procedures like cleaning data, creating features, training models, and understanding results. It is commonly used to forecast student success, warn of academic risks, assess teaching quality, recommend courses, and analyse teacher performance. Most traditional data mining methods use centralised data aggregation and analysis. This makes the model work better to some extent, but when it comes to widely distributed and sensitive educational data, the centralised processing mode has problems like making it hard to protect privacy, limiting data sharing, and not being able to generalise the model enough.

In recent years, many machine learning and deep learning algorithms have been added to the field of school management data mining. Deep learning models often learn by using multi-layer nonlinear transformations to represent input characteristics (Chen et al., 2020). This can be written as:

$$h^{(l)} = \sigma\left(W^{(l)} h^{(l-1)} + b^{(l)}\right) \tag{4}$$

$$h^{(0)} = x \tag{5}$$

where $h^{(l)}$ is the hidden layer representation of layer l, $W^{(l)}$ is the weight matrix, $b^{(l)}$ is the bias, and $\sigma(\cdot)$ is the activation function. The model can automatically find important features in the data, learn complicated association rules, and improve prediction accuracy by transferring and learning from one layer to the next.

As education management systems become more sophisticated and varied, data heterogeneity, dynamism, and imbalance make modelling more difficult. Consequently, academics are concentrating on methodologies such as multi-source data fusion, multi-task learning, and transfer learning to improve the resilience and generality of models. For instance, multi-task learning makes models work better on distinct educational management subtasks by sharing the structure of the hidden layer and allowing numerous related tasks to be optimised at the same time (Dhaygude et al., 2024).

Still, educational management data mining has a lot of problems to deal with, like protecting data privacy, working together across institutions, responding in real time, and making systems that can grow. When numerous schools work together to make models, figuring out how to keep students' and instructors' personal information safe and follow all the rules and laws has become a major problem with technology. The variety of situations in which education management takes place also calls for a very flexible and tailored data mining system that can handle both macro-level policy decisions and micro-level individualised counselling. To address the issues, emerging technologies like distributed learning and FL have increasingly become significant tools for facilitating the transformation and enhancement of education management data mining.

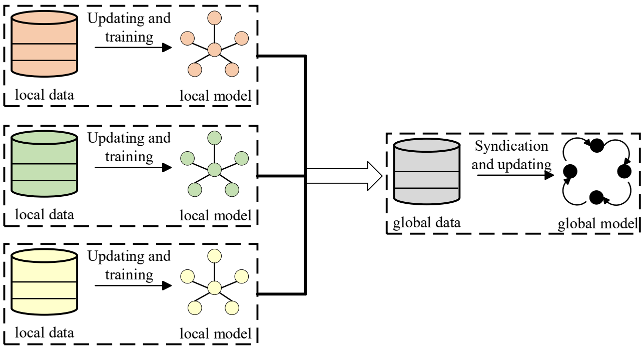
In short, school management data mining technology is moving away from the old, centralised model and towards a new one that values intelligence, collaboration, and privacy protection equally. In the future, integrating FL with other sophisticated distributed technologies to create a secure and efficient intelligent education management data mining system would furnish robust technical support for enhancing educational governance and advancing educational equity.

4 System architecture

4.1 *Intelligent educational data mining system architecture*

The goal of intelligent education data mining system architecture is to enable secure sharing, collaborative modelling, and intelligent analysis of education management data in a distributed setting. It also aims to maximise the privacy protection benefits of FL and fulfil the need for thorough mining of diverse education data from multiple institutions and types, as depicted in Figure 1.

Figure 1 Intelligent education data mining system architecture (see online version for colours)



1 Data layer

The data layer is the most important part of the intelligent educational data mining system. It is responsible for gathering, storing, pre-processing, and managing the security of distributed educational data. As modern education becomes more digital, the amount of educational data is rising quickly. This includes logs of students' learning behaviours, records of classroom interactions, test scores, curricular resources, and teachers' teaching situations. There are many different sorts of data, such as structured, semi-structured, and unstructured data, as well as text, audio, video, and other multimedia content.

The data layer should first fix the problem of data heterogeneity because the data is spread out among multiple schools and management units. To achieve this, the system uses common data formats and standards, combines and changes data from diverse sources and formats, and makes sure the data is of high quality by cleaning it, handling missing values, finding anomalies, and other methods (Rao et al., 2019). The data layer also uses feature engineering technology to change the original data into feature representations that can be used as model input. This includes things like normalising numerical features and coding category features. This makes the data easier to understand and helps with model training.

The data layer also needs to be able to handle and schedule data well, process large amounts of data in real time or in batches, and make sure that the data is sent to the FL training module and intelligent analysis module on time and precisely. At the same time, it works with metadata management and data bloodline tracking technology to keep an eye on the entire process of changing and using the data source and make sure that the data can be traced and is reliable.

When it comes to education data, there is frequently a lot of personal and sensitive information, like students' names, grades, family history, and more. Because of this, the data layer must be built in a way that fully meets data security and privacy protection standards. To make sure that data is safe and follows the rules while it is being stored and used, specific steps include encrypting local data, controlling access rights, using multiple levels of authentication, and anonymising data. The data layer also allows for independent administration of local data, which keeps sensitive data from being stored in one place, lowers the danger of data leaks, and gives a strong security assurance for FL model training that comes after.

2 Federated learning co-training layer

The layer is a vital aspect of the intelligent educational data mining system. Its main job is to let various educational institutions share information and train models together while keeping each participant's data safe and private. This layer, through the FL framework, lets each node train models on its own using private data. This keeps sensitive data from being sent and stored in one place, which lowers the chance of data leaks.

Each participant node k in the system has its own local dataset D_k and model parameters w_k. This node changes the model parameters by making a local loss function $L_k(w_k)$ as little as possible:

$$\min_{w_k} L_k(w_k) = \frac{1}{|D_k|} \sum_{i \in D_k} l\left(f(x_i; w_k), y_i\right) \tag{6}$$

The loss function, like mean square error or cross entropy, is $l(\cdot)$, and $f(x_i; w_k)$ is the model's forecast for the input sample x_i. This keeps the local data's peculiarities and diversity intact and makes the model more flexible in diverse educational settings.

After the local training is done, each node sends the model parameters to the central server (AbdulRahman et al., 2020). The server then combines the model parameters based on the weights of each node's data volume to update the global model parameters w:

$$w \leftarrow \sum_{k=1}^{K} \frac{|D_k|}{\sum_{j=1}^{K} |D_j|} w_k \tag{7}$$

This weighted average technique makes sure that nodes with more data contribute more to the global model. This makes the model more accurate and better at generalising. The system gets closer to the best model with each round of local training and parameter aggregation. This fully uses the benefits of dispersed data resources.

The FL co-training layer can also work with multiple types of data settings, which means it can handle varied sizes and types of data from different schools and help schools use cross-institutional data more effectively. This layer, along with privacy protection technologies, keeps data safe and makes intelligent education management data mining far more useful and effective.

3 Intelligent analysis and mining layer

The layer is an important part of the education management data mining system. It is responsible for doing in-depth analysis and finding new information in the global model outputs from the FL layer and the education big data. This layer uses machine learning, data mining, and statistical analysis to find patterns and rules in the huge amounts of education data and give education management scientific help in making decisions.

It builds effective analysis models by using pre-processing techniques like feature selection and dimensionality reduction to model multi-dimensional information in educational data, such as student behaviour, use of teaching resources, and academic performance. Using the global model's prediction results, more critical indicators are taken out to measure the quality and efficacy of schooling (Azevedo et al., 2021). This layer's most typical analytical model can be shown as a mapping function:

$$y = g(x;\theta) \tag{8}$$

where x is the input feature vector, which includes data on students' learning behaviour, teachers' teaching logs, curriculum resources, and other sources of information. θ is the model parameter, and y is the prediction result, which could be the prediction of students' performance, the assessment of learning risk, or the suggestion of resource optimisation. The model $g(\cdot)$ can take on numerous shapes to fit different analysis demands.

Specifically, each educational institution uses a two-layer fully connected neural network for training locally: the input layer receives a 30-dimensional feature vector after standardisation, the hidden layer contains 128 ReLU activation units, and the output layer uses Softmax to complete the binary classification task. Local training uses cross-entropy as the loss function, with a batch size of 32, a fixed learning rate of 0.001, and a total of 5 local epochs. After completing local updates, the client uploads only the model parameters to the central server. The server first performs a weighted average based on the data volume of each client, then introduces a FedProx regularisation term to penalise local updates that deviate too far from the global model. The regularisation coefficient μ is determined by grid search to be 0.01, effectively mitigating the performance degradation caused by non-iIData.

The layer also supports rule-based knowledge discovery and visualisation techniques that turn complicated data into easy-to-understand reports and charts. This helps administrators understand how education is changing quickly and improve their strategies for managing teaching and learning. This layer's intelligent mining feature not only makes it easier to use data, but it also helps school management grow in a precise and customised way.

4 Application service layer

This layer, which is the top layer of the intelligent educational data mining system, is mainly in charge of turning the results of the underlying FL and intelligent analyses into a wide range of easy-to-use educational management services that directly meet the needs of teachers, educational administrators, and other decision makers. This layer makes intelligent programmes that use data to make school management more efficient and make decisions better.

It also enables flexible permission management and data access control. This makes sure that users in different roles may only access the data and services they are allowed to, which keeps the system safe and standardises it (Anciaux et al., 2019). The layer also uses multi-channel service delivery, such as WEB, mobile, and API interfaces, which makes it easy for school managers to make smart decisions at any time and from any place.

The intelligent education management system can use the application service layer to create a closed loop that goes from collecting data to training models to using the results. This will help education management become smarter and better, and in the end, it will help improve the quality of education.

The proposed intelligent educational data mining system introduces a novel ensemble mechanism within the federated learning framework, specifically tailored for heterogeneous educational data environments. Unlike conventional FL approaches that rely on simple parameter averaging, our system employs a dynamic weighted aggregation strategy that considers both the data distribution characteristics and the local model performance of each participating institution.

To further quantify potential risks, the system automatically records changes in model performance and parameter deviation for each client before each round of aggregation and generates a risk score using a weighted method: clients with greater performance degradation and more severe deviations contribute higher risk weights. The scoring results are fed back to the central server in real time in three tiers (low, medium, and high) to dynamically adjust privacy budgets and aggregation weights, thereby achieving risk visualisation and adaptive control without the need for additional modules.

Additionally, the system incorporates a local update regularisation mechanism during the training phase, inspired by FedProx and FedNova, but further optimised for educational data heterogeneity. This regularisation term is dynamically adjusted based on the degree of Non-IIDness in each client's dataset, as measured by the divergence of class distributions or feature distributions compared to the global average. These enhancements represent a significant technical innovation in FL-based educational data mining systems, enabling more robust and adaptive model training in real-world educational settings.

4.2 Model training and privacy preserving mechanisms

In smart educational data mining systems, model training and privacy protection are the most important parts that keep data safe and make models work better. Data on education management often contains a lot of private information, like students' names, grades, behaviour, and more. If this information gets out, it could be a big threat to privacy. So, the system needs to take strong steps to preserve privacy to keep data safe and follow the rules while also training models quickly.

This system uses the FL framework for model training. Each node trains the model on its own using local data and only sends the parameters or gradients it gets from training to the central server. The central server then combines these parameters to update the global model. This distributed training method efficiently prevents the centralised storage and transfer of sensitive data, thus lowering the chance of data leaks.

But releasing model parameters alone can still reveal some private information. Attackers might be able to figure out some things about a user's data by looking at the submitted model changes. This is why the system uses the DP mechanism, which adds random noise that has been carefully constructed to the uploaded model parameters. This makes it harder to see how individual samples affect model updates and protect individual privacy.

The mathematical definition of DP is that for any two datasets *D* and *D*' that are next to each other and differ by only one sample, the DP mechanism *M* holds.

$$\Pr\left[M(D)\in S\right]\leq e^{\epsilon}\cdot\Pr\left[M(D')\in S\right]+\delta \tag{9}$$

where ϵ is a privacy budget parameter that shows how strong privacy protection is; δ is the allowable chance of failure; and *S* represents any part of the mechanism's output. An attacker can't tell if a given piece of data is being used to train a model if there is enough noise, which protects the user's privacy.

The system uses secure multi-party computation (SMPC) technology in addition to DP to make sure that the process of aggregating model parameters is safe. SMPC lets each participant work together to finish the calculation of the global model without giving up their personal model parameters (Nigro et al., 2018). This makes sure that the server and other nodes can't see the training information of individual nodes, which makes the system even more secure.

The system also designs an adaptive privacy protection strategy to deal with the variety and changeability of educational data. This strategy changes the privacy budget and noise intensity based on the participants' computing power and privacy needs, so that privacy protection and model performance can be optimised at the same time. The system keeps making the global model better by using numerous rounds of iterative training and privacy protection together.

5 Experimental designs

5.1 Experimental data

The experimental dataset utilised in this study is sourced from the Programme for International Student Assessment (PISA) dataset, sponsored by the Organisation for Economic Co-operation and Development (OECD). It gives researchers a scientific basis for studying things like school management, education policy, and how to best use resources.

Table 1 shows the key parts of the PISA dataset.

Table 1 Information on the PISA dataset

Category	*Data fields/dimensions*	*Description*
Student characteristics	Age	Age of the student, typically 15 years old
	Gender	Gender of the student (male/female)
	Family background	Socioeconomic status, parents' education level, occupation, income, etc.
	Learning attitudes	Motivation, interest in learning, engagement with schoolwork
	Learning resources	Access to books, internet, study space, extracurricular tutoring, etc.
Academic performance	Mathematics score	Student's performance in the mathematics test
	Science score	Student's performance in the science test
	Reading score	Student's performance in the reading test
	Teacher quality	Teachers' qualifications, professional training, and teaching experience
Societal context	Regional disparities	Urban/rural or cross-national comparisons in educational performance
	Educational investment	Government or institutional spending on education per region or student

The PISA dataset was selected as the experimental dataset for this work primarily because of its comprehensive information on educational management and its capacity to robustly support FL-based educational data mining approaches. FL, as a new way to learn from many places at once, can let multiple schools share data and train models while keeping data safe and private. Using PISA information, we can model how different schools, regions, and countries might work together to learn and look for ways to make educational management more effective through FL.

In this study, the PISA dataset was used as the primary source for experimental validation. The dataset contains a wide range of features related to student characteristics, academic performance, and societal context. A detailed summary of the features is as follows:

- *Student characteristics:* This category includes features such as age (numerical, integer), gender (categorical, binary: male/female), family background (categorical and numerical, including parental education level, occupation, and income), learning attitudes (ordinal, derived from survey responses), and learning resources (binary and numerical, indicating access to books, internet, and study space).
- *Academic performance:* These are numerical features, including scores in mathematics, science, and reading. These scores are standardised and used as the primary target variables in classification and regression tasks.
- *Societal context:* Features such as regional disparities (categorical: urban/rural or cross-national) and educational investment (numerical, representing per-student government spending) were also included to reflect broader educational environments.

The dataset contains a mix of missing values, especially in the family background and learning resources categories. Missing values were handled using mean imputation for numerical features and mode imputation for categorical features. Numerical features such as mathematics score and educational investment were normalised using min-max scaling to ensure model stability and faster convergence. Categorical variables were one-hot encoded to allow compatibility with model training.

The dataset was partitioned across five simulated educational institutions (clients), each representing a different region or school type. Data distribution across clients was intentionally made non-IID to reflect real-world heterogeneity in educational environments. Specifically, each client contained a unique subset of students with different demographic and academic profiles.

For model training and evaluation, the data in each client was split into a 70% training set and a 30% test set, stratified by academic performance to maintain class balance. The global model was trained over 50 communication rounds, with all clients participating in each round. Model performance was evaluated using the aggregated global model on the test set of each client, and final results were averaged across clients to provide an overall performance estimate.

At the feature level, we extracted 30-dimensional inputs from PISA data: 23 numerical features (such as parents' years of education, household income, and number of learning resources) were standardised using z-scores; and 7 categorical features (gender, urban/rural area, school type) encoded using one-hot encoding. The label construction is based on students' mathematics scores, divided into 'high-scoring/low-scoring' categories according to the official threshold of 550 points. During the training phase, the local model uses logistic regression with L2 regularisation (λ = 1e-4); global aggregation employs FedAvg with 50 communication rounds, 5 local epochs per client, a batch size of 32, and a learning rate of 0.001. The entire process is implemented within the Flower framework and is fully reproducible.

5.2 Experimental environment and indicator design

This study utilises the simulation environment to develop a multi-node FL experiment framework that replicates the process of multi-agency collaborative modelling in actual educational management contexts. The experiment posits that each client symbolises a district-level or school-level educational unit possessing its own autonomous local data, which cannot be directly exchanged but may engage in global model training via the federation mechanism. The entire experimental platform is set up on a high-performance computing server with the Ubuntu 22.04 LTS operating system and hardware that includes an Intel Xeon Gold 6338 processor (32 cores), an NVIDIA A100 GPU (40GB graphics memory), and 256GB of RAM. This setup can handle high concurrency and low latency data processing and model computation needed for federation training. The experiment primarily utilises the NVIDIA A100 GPU (40GB graphics RAM) and 256GB of memory within the software environment.

Python 3.10 is the programming language used to make the software environment. PyTorch 2.0 is used to make the local training model, and the Flower framework is used to make the federated communication mechanism. Data processing and preprocessing mostly use popular tool libraries like NumPy, Pandas, and Scikit-learn. In the simulation deployment, there are 5 client nodes and 1 central server. The clients can only talk to the central server over the gRPC protocol; they can't talk to each other. Each round of federated training process includes: local model training → model upload → parameter aggregation → model dissemination, all nodes complete the training and communication in parallel, simulating the privacy protection and model collaboration scenarios of real distributed educational data. The federated training process consists of local model training, model upload, parameter aggregation, and model dissemination.

This study selects three representative assessment metrics from three dimensions: classification performance, training efficiency, and communication cost. These metrics are used to comprehensively evaluate the effectiveness and flexibility of FL-based data mining approaches in managing education.

First, accuracy is one of the most common methods used to quantify how well a model classifies targets in educational management problems, such as student performance and learning risk. The formula for calculating it is:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

where TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively. The model's high accuracy rate demonstrates its ability to make reliable predictions and provide educational managers with more dependable supplementary information to support their decision-making.

Second, this paper presents the F1 Score, the harmonic means of Precision and Recall, to address the issue of assessment bias caused by category imbalance in educational data (e.g., minority students, students facing academic challenges). The F1 Score more accurately reflects the model's balance between identification, accuracy and completeness:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \tag{11}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{12}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{13}$$

The model is more robust and reliable at managing uneven education data, the higher the F1 score.

Lastly, this paper talks about area under curve (AUC) as an important way to quantify how well the system classifies things. AUC is a general measure of how well the model can sort positive and negative data (Norton and Uryasev, 2019). The AUC value ranges from 0 to 1. A higher value indicates that the model is more effective at distinguishing between samples and performs better. The formula for its calculation is as follows:

$$\text{AUC} = \int_0^1 TPR(t)dFPR(t) \tag{14}$$

where $TPR(t)$ is the true positive rate, and $FPR(t)$ is the false positive rate at threshold t. The AUC is an effective metric for evaluating a model's performance across different thresholds. It can be used to assess how well a model performs in classification tasks within educational management data mining.

In conclusion, this paper thoroughly assesses the proposed method's practical application value in education management data mining by evaluating three key metrics: classification performance, model robustness, and system efficiency.

5.3 *Experimental results and analysis and discussion*

Experiment 1 seeks to assess the overall efficacy of the proposed intelligent FL-based data mining system for educational management within educational data contexts and to juxtapose it with other prevalent methodologies to ascertain if the system can maintain efficient and precise data mining operations while preserving data distribution independence.

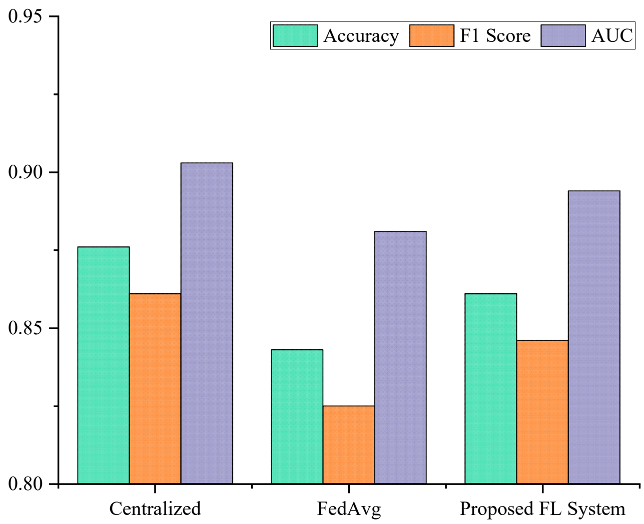
The experiments established three model training strategies for evaluation:

1. a centralised approach (centralised), where all data is trained in a single location
2. a traditional federated learning approach (FedAvg), which resembles a typical federated learning environment
3. the system architecture proposed in this paper (proposed FL system), which incorporates intelligent analysis modules and multi-layer co-training.

The same network structure and training parameters are used for each method, with cross-entropy as the loss function. All methods are trained until convergence.

Figure 2 shows the outcomes of the experiment.

Figure 2 Comparison of the performance of different methods in student performance (see online version for colours)



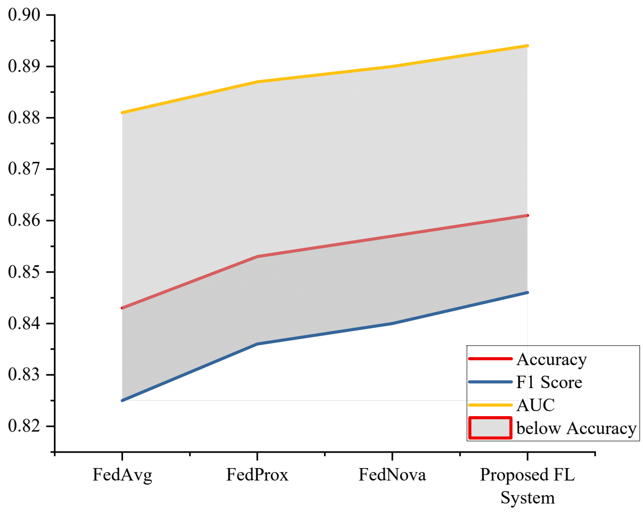
The experimental results indicate that the centralised technique attains the best accuracy and AUC values, demonstrating superior classification capabilities due to its comprehensive data availability. But the centralised model doesn't work well in real school management because data is spread out and privacy protection rules are quite rigors. The standard FL approach FedAvg, on the other hand, works well in distributed systems, but its model performance goes down when it has to deal with data heterogeneity.

The intelligent educational management data mining system described in this paper enhances the adaptability and resilience of FL by the implementation of a multi-layer collaborative training mechanism and an intelligent analysis module. Its accuracy and F1 score are better than the FedAvg approach and close to the centralised model. This shows that the system can still provide good mining results while keeping data private. It also has significant practical utility and might be promoted.

The goal of Experiment 2 is to see if the optimisation method employed in the FL co-training layer of the intelligent education management data mining system developed in this paper really does improve performance when it works with different types of education data. In particular, the traditional FedAvg algorithm is compared to two enhanced federated algorithms, FedProx and FedNova, as well as the enhanced algorithms incorporated in this paper's system, to assess the applicability and model performance of each algorithm in a Non-IID data environment.

The studies utilise the PISA dataset, focusing on the student academic performance classification task, while maintaining a consistent network structure and training hyperparameters. FedProx lessens the disparities across client models by adding regular terms to limit the local models. FedNova, on the other hand, speeds up convergence by changing the weights of the gradients of various clients. This paper's system builds on this by combining adaptive learning rate modification with a local update strategy to make the model even better at generalising. Figure 3 shows the outcomes of the experiment:

Figure 3 Performance comparison of federated algorithms (see online version for colours)



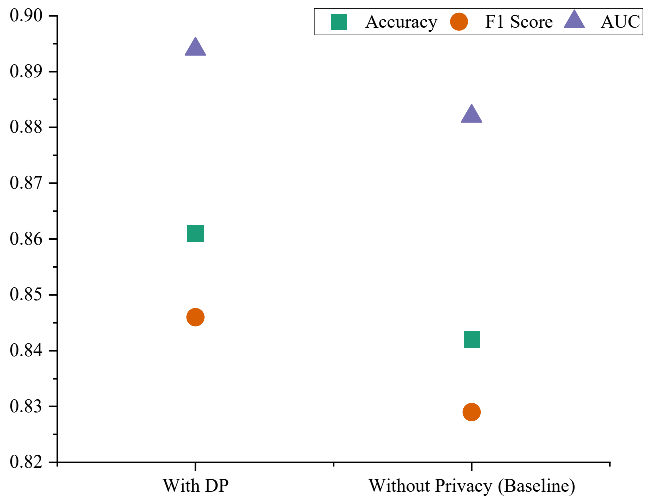
The experimental results demonstrate that the FedProx algorithm improves accuracy by one percentage point (from 0.843 to 0.853) and increases the F1 score and AUC by 1.1% and 0.6%, respectively, compared to the original FedAvg. This indicates that incorporating regularisation term constraints helps mitigate the adverse effects of non-IID data on model training. Consequently, the local model updates become more stable. Additionally, FedNova achieves further improvements, with accuracy reaching 0.857, an F1 score of 0.840, and an AUC of 0.890. This suggests that the weighted normalisation of the gradient enhances the balance of contributions from different clients, resulting in faster convergence and superior performance.

This paper proposes a system which achieves an accuracy of 0.861, which is 0.4 percentage points higher than FedNova. The F1 score and AUC are 0.846 and 0.894, respectively, which are also higher than those of all the other algorithms compared. It can better capture the different aspects of student performance, make the model more stable and efficient, and offer a more useful and valuable solution for data intelligence in education management.

Experiment 3 aims to evaluate the influence of the privacy protection mechanism implemented in the intelligent education management data mining system of this paper on model performance, and to confirm that the system can sustain high mining accuracy and stability while ensuring data security and privacy. The experiment employs the classification of students' academic performance utilising the PISA dataset and examines the balance between privacy protection and model performance by contrasting the efficacy of the baseline model, which lacks the privacy protection mechanism, with that of the model incorporating the DP technique.

The privacy-preserving module uses a noise injection method to add Laplace noise during the local model update phase to make sure DP guarantees are met. It also controls the privacy budget ε to make sure that the data isn't leaked. Figure 4 shows the outcomes of the experiment:

Figure 4 Impact of privacy protection mechanism on model performance (see online version for colours)



The experimental results demonstrate that the implementation of the DP mechanism enhances the model's accuracy by approximately 2.2 percentage points, while the F1 score and AUC improve by roughly 1.7% and 1.3%, respectively, indicating that privacy protection contributes to the model's accuracy performance to a certain degree. The protected model, on the other hand, performs worse than the unprotected baseline model. However, this is still good enough for real-world school management data mining.

The privacy protection method created by the system in this work strikes a good compromise between model performance and user data security and privacy. This makes intelligent education data mining reliable and useful. The method offers robust support for data security in distributed educational data environments and further advances the informatisation and intelligence initiatives in educational administration.

The intelligent FL-based educational management data mining system developed in this paper demonstrates considerable advantages across multiple dimensions through the rigors validation of three tests. Experiment 1 demonstrates that the system can attain high-precision classification results comparable to the centralised model while ensuring data privacy, hence validating the efficacy of the system's overall architecture. Experiment 2 further confirms the upgraded FL algorithm's ability to adapt and improve performance in the system for handling different types of educational material. This shows that the multi-layer collaborative training mechanism is better. Experiment 3 concentrates on the privacy protection method, demonstrating that the system sustains superior mining performance under DP guarantee, which addresses the requirements of data security and intelligent analysis. In short, the proposed system is very useful in real life and has a lot of potential for growth. It is also good for use in education management situations where privacy is important.

To enhance the interpretability of the proposed FL-based intelligent educational data mining system and provide deeper insights into the model's decision-making process, we have conducted a feature-important analysis using SHAP (SHapley Additive ExPlanations). This method is grounded in cooperative game theory and offers a unified measure of feature contributions to model predictions, making it particularly suitable for complex models trained on heterogeneous educational data.

In this analysis, we applied SHAP values to the final global model obtained from the proposed FL system. The results highlight the most influential features in predicting student academic performance. Specifically, features such as Family Background, Learning Attitudes, and Learning Resources were found to have the highest impact on model predictions. For instance, students with higher parental education levels and access to learning resources (e.g., books and internet) consistently showed better performance outcomes. Additionally, Teacher Quality and Educational Investment were also identified as significant contributors, indicating that institutional factors play a crucial role in shaping student achievement.

These findings not only validate the model's behaviour but also provide actionable insights for educational administrators. By identifying the most predictive variables, stakeholders can design more targeted interventions to support at-risk students and optimise resource allocation. This level of interpretability enhances both the scientific and practical value of the proposed system, supporting its application in real-world educational management scenarios.

6 Conclusions

6.1 Summary of the study

This project focuses on intelligent FL-based data mining techniques for educational management. The system creates and implements a series of architectures that integrate multi-level collaborative training with intelligent analysis. The suggested approach in this research greatly enhances the accuracy and reliability of educational management data mining, all while keeping data safe and private, as shown by experiments on the dataset. The experimental results demonstrate that the developed FL algorithm optimisation strategy and privacy protection mechanism can effectively address the challenges of data heterogeneity and uneven distribution, ensuring high efficiency and generalisation in model training, while facilitating the practical application of intelligent analysis of educational data.

This article also looks closely at how well different FL algorithms work in education management situations. It also incorporates DP technology to make things safer, which reduces the possibility of privacy leaks that come with standard centralised data processing approaches. The comprehensive study offers a novel data mining solution in the realm of education management, fosters equilibrium between data sharing and security, and establishes the theoretical framework and technological assistance for the development and implementation of forthcoming intelligent education systems.

Even if this study had good results, there are still some problems.

First, while the dataset utilised in this study is authoritative in international education assessment, it is relatively homogeneous, primarily concentrating on students' academic performance and certain background information. It does not encompass multidimensional data regarding other critical aspects of the education management system, such as teachers' behaviour, school management processes, and resource allocation.

Second, the DP technique used in this paper will always add noise while keeping data private and safe, which will make the model work less well. How to create a more sophisticated privacy protection plan that has the least influence on model performance while keeping private safe is a major problem that needs to be solved quickly in future research.

Lastly, FL itself needs a lot of computing power and network resources. In the education management system, where there are a lot of people involved and the network environment is complicated and changes often, frequent model updates and parameter exchanges can cause communication delays and put more strain on the system, which can lower the overall training efficiency.

6.2 Directions for subsequent research

To mitigate the constraints identified in this study, subsequent research may be conducted in the following areas. First, the system's adaptability to complex education scenarios and its level of generalisation should be improved by adding more dimensional and multimodal education management data sources and types. This could include data on teachers' teaching behaviours, how school resources are allocated and managed, and more. By adding rich data features, it can better show the real demands of education management and make data mining even more accurate and in-depth (Li et al., 2018).

Second, improving the privacy protection system is a key area for future research. In the future, it may be possible to build more efficient DP algorithms or combine them with FL, SMPC, homomorphic encryption, and other methods to improve privacy protection while having less of an effect on model performance. Also, optimising how privacy budgets and defence tactics are used will be an important aspect of making the system more secure and stable so that it can effectively withstand a variety of possible attacks in real-world situations.

One promising direction is the development of user-friendly interfaces and decision support modules that allow non-technical stakeholders to interact with the system and interpret its outputs. This includes visual dashboards, alert systems for at-risk students, and recommendation engines for resource allocation and intervention planning. Integrating such functionalities into daily educational practices will require close collaboration with domain experts, including educational psychologists, school administrators, and learning scientists.

Moreover, future research should assess the system's reliability and generalisability across diverse educational and cultural contexts. This includes evaluating its performance in low-resource settings, where data quality and infrastructure may be limited. Addressing potential data biases – such as those arising from socioeconomic or demographic disparities – will also be crucial to ensure equitable outcomes and avoid reinforcing existing inequalities in education.

Finally, to deal with FL's communication overhead and computational efficiency, we should work on making model training algorithms that are lighter and more efficient, improving the system's communication protocols and resource management strategies, lowering bandwidth needs and latency, and making the system more scalable and useful. It can also be used with edge computing and other technologies to process data in a distributed way. This will make training more real-time and responsive overall and set the stage for the widespread use of large-scale intelligent education management systems.

Acknowledgements

This work is supported by the Guangxi Higher Education Undergraduate Teaching Reform Project (No. 2024JGB507), the Guangxi Higher Education Undergraduate Teaching Reform Project (No. 2024JGB509), and the Guangxi Vocational Education Teaching Reform Research Project (No. GXGZJG2024B190).

Declarations

All authors declare that they have no conflicts of interest.

References

AbdulRahman, S., Tout, H., Ould-Slimane, H., Mourad, A., Talhi, C. and Guizani, M. (2020) 'A survey on federated learning: The journey from centralized to distributed on-site learning and beyond', *IEEE Internet of Things Journal*, Vol. 8, No. 7, pp.5476–5497.

Anciaux, N., Bonnet, P., Bouganim, L., Nguyen, B., Pucheral, P., Popa, I.S. and Scerri, G. (2019) 'Personal data management systems: the security and functionality standpoint', *Information Systems*, Vol. 80, pp.13–35.

Azevedo, J.P., Hasan, A., Goldemberg, D., Geven, K. and Iqbal, S.A. (2021) 'Simulating the potential impacts of COVID-19 school closures on schooling and learning outcomes: A set of global estimates', *The World Bank Research Observer*, Vol. 36, No. 1, pp.1–40.

Bilal, K., Khalid, O., Erbad, A. and Khan, S.U. (2018) 'Potentials, trends, and prospects in edge technologies: fog, cloudlet, mobile edge, and micro data centers', *Computer Networks*, Vol. 130, pp.94–120.

Chen, Q., Zhang, W. and Lou, Y. (2020) 'Forecasting stock prices using a hybrid deep learning model integrating attention mechanism, multi-layer perceptron, and bidirectional long-short term memory neural network', *IEEE Access*, Vol. 8, pp.117365–117376.

Dhaygude, A.D., Ameta, G.K., Khan, I.R., Singh, P.P., Maaliw III, R.R., Lakshmaiya, N., Shabaz, M., Khan, M.A., Hussein, H.S. and Alshazly, H. (2024) 'Knowledge-based deep learning system for classifying Alzheimer's disease for multi-task learning', *CAAI Transactions on Intelligence Technology*, Vol. 9, No. 4, pp.805–820.

Ferrag, M.A., Friha, O., Kantarci, B., Tihanyi, N., Cordeiro, L., Debbah, M., Hamouda, D., Al-Hawawreh, M. and Choo, K-K.R. (2023) 'Edge learning for 6G-enabled internet of things: a comprehensive survey of vulnerabilities, datasets, and defenses', *IEEE Communications Surveys & Tutorials*, Vol. 25, No. 4, pp.2654–2713.

He, X., Chen, Q., Tang, L., Wang, W., Liu, T., Li, L., Liu, Q. and Luo, J. (2023) 'Federated continuous learning based on stacked broad learning system assisted by digital twin networks: an incremental learning approach for intrusion detection in UAV networks', *IEEE Internet of Things Journal*, Vol. 10, No. 22, pp.19825–19838.

Hu, Z., Shaloudegi, K., Zhang, G. and Yu, Y. (2022) 'Federated learning meets multi-objective optimization', *IEEE Transactions on Network Science and Engineering*, Vol. 9, No. 4, pp.2039–2051.

Jain, P., Kakade, S.M., Kidambi, R., Netrapalli, P. and Sidford, A. (2018) 'Parallelizing stochastic gradient descent for least squares regression: mini-batching, averaging, and model misspecification', *Journal of Machine Learning Research*, Vol. 18, No. 223, pp.1–42.

Khan, A. and Ghosh, S.K. (2021) 'Student performance analysis and prediction in classroom learning: A review of educational data mining studies', *Education and Information Technologies*, Vol. 26, No. 1, pp.205–240.

Komljenovic, J. (2022) 'The future of value in digitalised higher education: why data privacy should not be our biggest concern', *Higher Education*, Vol. 83, No. 1, pp.119–135.

Li, L., Chi, T., Hao, T. and Yu, T. (2018) 'Customer demand analysis of the electronic commerce supply chain using Big Data', *Annals of Operations Research*, Vol. 268, pp.113–128.

Liu, W., Chen, L. and Zhang, W. (2022) 'Decentralized federated learning: Balancing communication and computing costs', *IEEE Transactions on Signal and Information Processing over Networks*, Vol. 8, pp.131–143.

Lutfiani, N. and Meria, L. (2022) 'Utilization of big data in educational technology research', *International Transactions on Education Technology*, Vol. 1, No. 1, pp.73–83.

Mohamed Hashim, M.A., Tlemsani, I. and Matthews, R. (2022) 'Higher education strategy in digital transformation', *Education and Information Technologies*, Vol. 27, No. 3, pp.3171–3195.

Nigro, C., Nigro, L. and Sciammarella, P.F. (2018) 'Modelling and analysis of multi-agent systems using UPPAAL SMC', *International Journal of Simulation and Process Modelling*, Vol. 13, No. 1, pp.73–87.

Norton, M. and Uryasev, S. (2019) 'Maximization of AUC and buffered AUC in binary classification', *Mathematical Programming*, Vol. 174, pp.575–612.

Palihawadana, C., Wiratunga, N., Wijekoon, A. and Kalutarage, H. (2022) 'FedSim: similarity guided model aggregation for federated learning', *Neurocomputing*, Vol. 483, pp.432–445.

Rao, T.R., Mitra, P., Bhatt, R. and Goswami, A. (2019) 'The big data system, components, tools, and technologies: a survey', *Knowledge and Information Systems*, Vol. 60, pp.1165–1245.

Sun, Y., Shen, L., Sun, H., Ding, L. and Tao, D. (2023) 'Efficient federated learning via local adaptive amended optimizer with linear speedup', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 45, No. 12, pp.14453–14464.

Wahab, O.A., Mourad, A., Otrok, H. and Taleb, T. (2021) 'Federated machine learning: survey, multi-level classification, desirable criteria and future directions in communication and networking systems', *IEEE Communications Surveys & Tutorials*, Vol. 23, No. 2, pp.1342–1397.

Zhang, C., Xu, Y., Elahi, H., Zhang, D., Tan, Y., Chen, J. and Zhang, Y. (2022) 'A blockchain-based model migration approach for secure and sustainable federated learning in IoT systems', *IEEE Internet of Things Journal*, Vol. 10, No. 8, pp.6574–6585.