Federated Machine Learning for Monitoring Student Mental Health in Kazakhstan

Bakirova Gulnaz[1], Bektemyssova Gulnara[2], Nor'ashikin Binti Ali[3]
Department of Computer Engineering, International University of Information Technology, Kazakhstan [1, 2]
College of Computing and Informatics, Universiti Tenaga Nasional, Kajang, Selangor, Malaysia [3]

***Abstract*—Federated Learning (FL) offers a privacy-preserving and decentralized paradigm for machine learning, making it particularly suitable for analyzing sensitive psychological and physiological data. This study aims to develop and evaluate a federated learning framework for assessing the psycho-emotional well-being of students in Kazakhstani educational institutions, where data privacy and infrastructural constraints pose significant challenges. We benchmark three FL algorithms, such as FedAvg, FedOpt, and FedProx, on heterogeneous, institution-level datasets that combine sleep, dietary, activity, and self-reported emotional measures. Experiments simulate cross-device, non-IID deployments and evaluate convergence, accuracy, and stability across ten communication rounds. Results show that FedProx attains the best trade-off between accuracy and stability under non-IID conditions (peak accuracy is 99.9%), while FedOpt provides faster early convergence, and FedAvg performs well for more homogeneous partitions. The methodological contribution comprises optimized aggregation and adaptive client weighting to mitigate non-IID effects in resource-constrained educational settings. These findings validate FL as a scalable, privacy-preserving approach for mental health monitoring in education and support its use for early intervention and resilience tracking. The proposed framework contributes to data-driven mental health policy design in educational systems, addressing both ethical and infrastructural considerations. The study discusses limitations of the simulated setup and outlines directions for broader deployment and cross-silo validation.**

Keywords*—Federated Learning; data privacy; FedOpt; FedAvg; FedProx; mental health; non-IID data; educational data mining; psychological analytics

I. Introduction

Machine learning (ML), a subset of artificial intelligence (AI), has experienced rapid advancements, enabling powerful computational techniques to analyze and interpret vast amounts of data. Unlike static, rule-based systems, ML algorithms learn from data, identifying patterns and improving performance over time. This adaptability has facilitated breakthroughs in areas such as image recognition, natural language processing, recommendation systems, and predictive analytics [1]. Central to this progress is the availability of big data, characterized by its volume, velocity, and variety. These datasets, encompassing sources like IoT sensor readings and social media interactions, form the foundation for training complex models capable of delivering highly accurate insights [2].

Despite its potential, traditional ML methods often rely on centralized data collection, raising significant privacy and security concerns. Sensitive data aggregated on centralized servers is vulnerable to breaches and may conflict with regulations like the GDPR [3] and CCPA. These limitations have catalyzed the emergence of FL, a decentralized machine learning paradigm. FL enables collaborative model training across distributed devices such as smartphones, IoT nodes, and edge computing systems while keeping raw data localized. Instead of transferring data to a central server, each device trains a local model and transmits only model updates (e.g., gradients or weights) to a central server for aggregation. By preserving data privacy and minimizing communication overhead, FL has gained prominence in sensitive domains such as healthcare, education, and finance [4].

In this study, we focus on the analysis for monitoring student mental health in Kazakhstan. Applying the FL mental health [5] analysis presents unique challenges due to the country's diverse cultural, social, and educational contexts. Beyond behavioral and physiological differences, students' emotional well-being is influenced by various factors, such as academic workload, which includes lessons, homework, and the preparation for and completion of exams. These stressors contribute to emotional burnout and introduce additional variability into the data, complicating the analysis and modeling process [6].

Given these challenges, this study seeks to develop and evaluate a federated learning framework for analyzing the psycho-emotional well-being of students in Kazakhstan. By integrating privacy-preserving algorithms such as FedAvg, FedOpt, and FedProx, the proposed framework addresses critical challenges associated with data heterogeneity, infrastructural limitations, and confidentiality within educational environments.

This study is organized as follows: Section II provides a review of related studies on federated learning algorithms and their applications in privacy-preserving educational analytics. Section III outlines the dataset, data preprocessing methods, and experimental design, detailing the implementation of FedAvg, FedOpt, and FedProx. Section IV presents experimental results and offers a comparative evaluation of algorithm performance under heterogeneous data conditions. It also discusses the implications of the findings for enhancing psycho-emotional well-being monitoring in educational contexts. Section V concludes the study by summarizing key findings and outlining future research directions.

In summary, this study investigates how federated learning can enhance the monitoring of students' mental health in

Kazakhstan while preserving data privacy and addressing infrastructural limitations. This framework aims to demonstrate the feasibility of applying FL under heterogeneous educational conditions and to contribute to the development of ethical, data-driven mental health monitoring systems in the Kazakhstani context.

II. Related Work

The specific key challenges are as follows as data privacy and security. The analysis of psycho-emotional data involves sensitive personal information, including sleep patterns, dietary habits, physical activity levels, and emotional responses. Ensuring the privacy and security of this data is paramount to protect students' confidentiality. Data Heterogeneity Students' behavioral data is highly non-IID (non-independent and identically distributed) [7]. Emotional Burnout and Stress Academic-related stressors, such as homework, lesson loads, and exam preparation, exacerbate emotional burnout among students. These factors not only impact students' psycho-emotional states but also introduce temporal variability in the data. As a result, FL models must be adaptable to dynamic changes in students' emotional states over time. In our work, we have considered algorithms such as FedAvg, a better-known efficient algorithm for training machine learning models on multiple clients in a privacy-preserving manner. Traditionally, FedAvg uses a central server to distribute parameters to clients and aggregate updates, but it has difficulties with privacy and communication efficiency. FedOpt achieves faster convergence, lower communication overhead, and higher accuracy with fewer training epochs compared to benchmarks. Research objectives are: developing and evaluating privacy-preserving FL algorithms (e.g., FedAVG, FedOPT, FedProx) tailored to psycho-emotional state analysis. Address data heterogeneity through techniques such as regularization, clustering-based FL, or adaptive weighting of client updates. Optimize FL frameworks for resource-constrained devices by implementing efficient communication strategies (e.g., gradient sparsification, local updates). Validate the proposed framework using real-world datasets that include students' behavioral and psycho-emotional metrics.

A. Revised Research Contributions

1) Development of a localized Federated Learning (FL) framework specifically designed for psycho-emotional data analysis in educational settings. The proposed framework integrates cross-device horizontal FL with synchronous aggregation, optimized for limited network connectivity and heterogeneous client environments typical of Kazakhstani institutions.

2) Algorithmic enhancement of FedAvg, FedOpt, and FedProx through adaptive weighting and proximal regularization to mitigate non-IID effects. The study introduces modified update strategies that dynamically adjust client contributions based on data volume and distribution, improving global model convergence and stability compared to baseline implementations.

3) Integration of behavioral and emotional indicators into federated model evaluation, linking technical model performance with psycho-emotional metrics such as burnout and well-being. This enables data-driven validation of FL's applicability for educational mental health monitoring, and supports model interpretability through domain-relevant insights.

In FL, the global model is trained iteratively over several rounds of client-server interaction, without centralized data collection, where each interaction is defined as an FL round. This process continues until the model reaches the desired level of accuracy [8]. Fig. 1 shows the architecture of federated learning that was applied in our work. It takes several rounds to update the model to the global model [9]. FL approaches can be categorized into three types: synchronous, asynchronous, and semi-asynchronous. In our work, we propose synchronous FL, in which the parameter server must wait for clients, which leads to noticeable waiting time due to edge heterogeneity. It should be considered that the heterogeneity of edge nodes means that they have different computational and communication capabilities. Although this method introduces waiting time due to edge heterogeneity, it provides better model convergence and accuracy, making it suitable for sensitive applications like psycho-emotional state analysis in education.

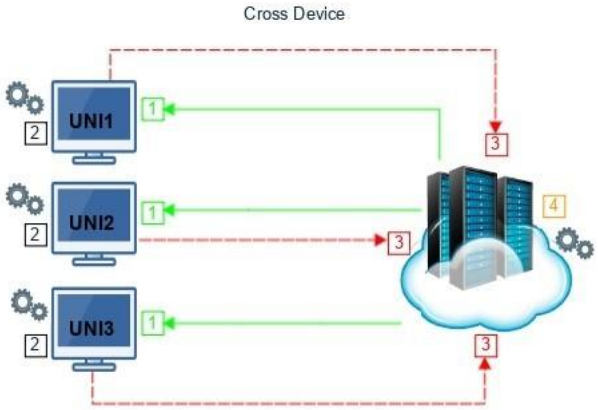


Fig. 1. The architecture of federated learning.

FL approaches can be categorized into three types by learning methods, architectures, and aggregation strategies, including synchronous FL, where the server waits for all clients to complete training before aggregating updates. While this method ensures consistency and high accuracy, it may lead to inefficiencies due to edge heterogeneity—variations in computational and communication capabilities among clients [10]. In asynchronous FL, clients communicate updates independently, reducing waiting times but potentially introducing inconsistencies in global model updates [11]. Semi-Asynchronous FL is a hybrid approach that balances the benefits of synchronous and asynchronous methods.

Experiments using synchronous method in FL [12], where clients must download global model updates at one point in time, and the server waits for clients to complete the training tasks. This synchronous optimization mechanism causes clients with limited network or insufficient hardware resources to lag; the server and other clients have to wait, and the learning efficiency decreases, i.e., there is an overload effect, but for our problem, we paid attention to the obvious advantages of this

synchronous update method in that the model converges easily and has high accuracy.

FL [13, 14] is a machine learning technique that focuses on mutual learning of a model by multiple clients, where the data remains decentralized. FL is a critical technology in the current era of big data and artificial intelligence, recognized for its critical role in protecting data privacy and eliminating the need to transfer and process huge amounts of data. FL allows the benefits of machines to remain intact. Unlike the traditional centralized learning approach, it facilitates collaborative training of statistical models sharing parameter updates instead of raw data [15, 16]. In our work, we have considered algorithms such as FedAvg, a better-known efficient algorithm for training machine learning models on multiple clients in a privacy-preserving manner. Traditionally, FedAvg uses a central server to distribute parameters to clients and aggregate updates, but has difficulties with privacy and communication efficiency [17]. FedOpt achieves faster convergence, lower communication overhead [18], and higher accuracy with fewer training epochs compared to benchmarks [19].

In [20], the authors present a comprehensive scheme for FL provisioning under time constraints in IoT environments using push-pull communication mode. By combining utility-based scheduling and efficient client selection, higher accuracy and lower latency are achieved compared to traditional methods. The approach is particularly effective in heterogeneous and resource-constrained networks, paving the way for practical implementation of FL in IoT and edge computing. In [21], ClipFL, a novel method for handling noisy labels in Federated Learning by identifying and excluding noisy clients, is discussed. The method demonstrates high accuracy, fast convergence, and reduced communication overhead, making it a practical solution for real FL systems with noisy data. The simplicity and scalability of the method make it suitable for large-scale deployment even in resource-constrained and heterogeneous environments. Traditional federated learning (FL) methods are enhanced, introducing a split learning approach where client-side and server-side models are trained together, which significantly improves the performance and efficiency of the interaction [22].

In [23], the importance of robustness in federated learning is emphasized, privacy and security concerns are addressed, and practical security mechanisms are described. By addressing these vulnerabilities, FL can be effectively deployed across industries, ensuring data integrity and model robustness in distributed environments. In the future, dynamic, scalable, and lightweight solutions should be prioritized to make FL both secure and efficient. One of the studies [24] shows that L2GDV (Loopless Local Gradient Descent with Varying step size) significantly improves the efficiency and performance of fuzzy learning by combining adaptive regularization step size methods. It provides robust convergence guarantees while reducing communication costs, which makes it suitable for real FL applications with heterogeneous data and limited resources. The study [25] emphasizes the vulnerability of FL to data poisoning attacks using unwanted samples, especially when using robust aggregation schemes such as Krum and Bulyan. While aggregation algorithms mitigate some attacks, they remain insufficient against adaptive adversary techniques.

Recent works [26–27] have explored centralized machine learning approaches for mental health prediction and early intervention, demonstrating promising results in healthcare and behavioral analytics. However, few studies have applied federated frameworks in educational contexts, particularly within developing regions, where data heterogeneity and privacy concerns remain major obstacles. However, these studies rely on centralized architectures, whereas our work introduces a federated paradigm suitable for privacy-constrained educational environments.

III. Proposed Methodology

FL utilizes synchronous, asynchronous, and semi-synchronous training methods, each suited to different network conditions. Synchronous learning ensures high accuracy by aggregating updates only after all clients submit them, but it requires stable connectivity and is sensitive to slow clients. Asynchronous learning allows clients to update independently, accelerating training in heterogeneous environments but risking model instability. Semi-synchronous learning balances these approaches by waiting for a subset of clients (e.g., 80%) before updating, reducing delays while maintaining stability. Given these factors, synchronous learning was chosen to ensure consistent model updates.

Although FL is classified into horizontal and vertical types, where horizontal FL involves similar features across different users, while vertical FL integrates different features for the same users. FL can also be Cross-Silo, applied in large organizations with stable connections, or Cross-Device, involving numerous heterogeneous personal devices with limited resources. Our study employs synchronous learning, horizontal FL, and Cross-Device FL for optimized distributed training. The structure of FL is illustrated in Fig. 2.

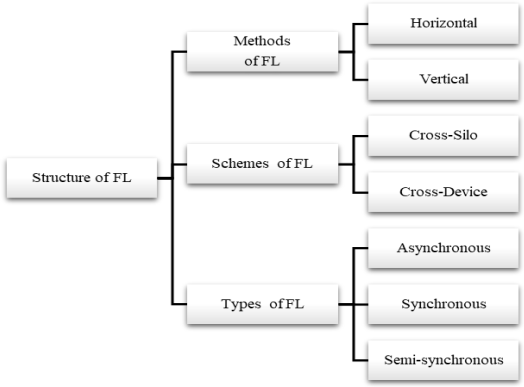


Fig. 2. Structure of FL.

A. Description of Method

These methods are used to predict students' psychological and emotional state based on nutrition, activity, and sleep data, such as interaction type. The distribution of the data presented below: Type interactions: When solving our task, we choose the Cross-Device type of interaction, since the task involves collecting data from devices where data is stored locally. The data are heterogeneous in terms of quality and distribution (different modes of sleep, nutrition, activity of students). Devices may be resource constrained (low processing power,

variable network). Cross-Device FL performs well in tasks where devices are limited in computational resources, and the data have a high level of heterogeneity. Selection: Horizontal. Distribution data: Each device has a similar data structure: sleep metrics, nutrition, activity, but the data belongs to different users. Horizontal training allows effective work with such scenarios, as it focuses on combining knowledge from similar data structures. Selection: synchronous. In our scenario, we will be able to organize synchronous student connection devices. Synchronous learning allows the server to produce an immediate update of the global model, which gives a good result for accuracy.

B. Selection of FL Algorithms

In addition, to effectively handle data heterogeneity in non-IID student datasets, we selected federated optimization (FO) algorithms that adapt to variability: FedAvg serves as a baseline approach, which can be enhanced by assigning different weights to clients with varying data distributions. FedProx incorporates regularization to mitigate the effects of dominant local updates, improving convergence in heterogeneous environments. FedOpt employs adaptive step sizes for both client and global models, facilitating faster convergence and enhanced scalability for complex, large-scale datasets.

FedAvg (Federated Averaging) [28] is a standard algorithm for federated learning that provides distributive learning on local devices, minimizing data exchange with the server. The algorithm works as follows: each client trains its model on local data and sends updates to the server. The server aggregates all local updates, computing an average for the model parameters. This method is repeated till convergence is achieved. The basic Formula (1):

$$\omega^{t+1} = \frac{1}{K}\sum_{k=1}^{K} \omega_k^{t+1} \quad (1)$$

where, ω_k^{t+1} updated model parameters from customer k, and K - number of customers.

FedOpt (Federated Optimization) [29] is an improvement to the FedAvg algorithm that incorporates the use of adaptive optimizers, such as Adam, for local model updates. Instead of using simple gradient descent as in FedAvg, FedOpt applies optimizers for each client, which helps to speed up convergence and reduce the impact of local errors. FedOpt can accommodate different learning rates for each client, making it more flexible and efficient. The basic Formula (2):

$$\omega^{t+1} = \omega_k^t - \mu_k \nabla L_k(\omega_k^t) \quad (2)$$

where, μ_k is the learning rate for client k, and ∇L_k is the gradient of the loss function for the client.

FedProx (Federated Proximal Optimization) [30] is an enhancement to the FedAvg algorithm designed to deal with heterogeneous data. It adds proximal regularization to minimize the impact of local differences in the data and improve learning stability. Instead of simply weighing local updates as in FedAvg, FedProx accounts for deviations of the local model from the global model using a proximal term that helps improve convergence and avoid overfitting. The basic Formula (3):

$$\omega_k^{t+1} = \omega_k^t - \mu_k \nabla L_k(\omega_k^t) - \mu(\omega_k^t - \omega) \quad (3)$$

where, μ is the regularization coefficient, and ω are the global parameters of the model.

The advantages of the algorithms are FedAVG - Easy implementation and low requirements for computing resources. FedOPT - use of adaptive optimizers accelerates convergence and helps avoid overfitting on unstable data. FedPROX is suitable for cases with heterogeneous data and high variability between clients. Proximal regularization helps to improve convergence.

The limitations of the algorithms are FedAVG - sensitivity to data heterogeneity, where data on different devices vary significantly, which can slow down the learning and degrade the quality of the model. FedOPT - Increased computational costs on the client, as more sophisticated optimization techniques must be used. FedPROX - requires tuning of the regularization parameter, which may add complexity to the optimization process.

Algorithms like FedSGD, FedAvg, FedProx, SCAFFOLD, and FedOpt were considered for this study, but FedAvg, FedProx, and FedOpt were selected for further study. It is shown in Table I.

TABLE I DETAILS OF ALGORITHMS

Method	Non-IID	Convergence rate	Flexibility to heterogeneity of clients	Use of server optimizers
FedAvg	Bad	Slowly	No	No
FedProx	Medium	Medium	Yes	No
FedOpt	Good	Fast	Yes	Yes

In summary, FedAvg is a basic algorithm for federated learning, effectively applicable if the data on the clients is similar. FedOpt improves FedAvg using adaptive optimizers, which accelerates convergence and improves model accuracy. FedProx focuses on heterogeneous data and incorporates proximal regularization, which improves training stability in the face of large differences between clients. Each of these methods has its own features and applications, and the choice of the appropriate algorithm depends on the data structure and computational resources.

In federated model training, FedAvg, FedOpt, and FedProx optimize learning on distributed clients while minimizing communication with the central server. FedAvg is effective when client data distributions are similar, offering a computationally efficient solution for homogeneous datasets, such as student learning or physical activity data. FedOpt adapts to heterogeneous data using adaptive optimizers (e.g., Adam), improving convergence speed and model accuracy, making it suitable for dynamic environments. FedProx addresses highly heterogeneous data by incorporating regularization, mitigating localized errors, and enhancing model stability and accuracy in diverse student datasets. In conclusion, FedAvg will be a simple and efficient solution when the data is homogeneous between clients, in case of heterogeneous data and different device types, we recommend using FedOpt or FedProx to improve model quality and convergence. Provided the problem requires regularization and consideration of local differences, FedProx will provide a stable solution. Thus, the choice of algorithm depends on the nature of the data and the goals: speed of

training and accuracy of the model or the ability to adapt to differences in customer data.

C. Experimental Settings

The tasks are focused on client-side data processing, where local models are applied directly on the client devices. Sending updates to the server. Server-side update processing, creating a global model. Sending global models to customers. Global model with client data is processed on the client side. These steps will ensure stability, privacy, and efficiency in implementing the study in real-world use cases. For our heterogeneous data problem, the choice of methods and approaches must consider the heterogeneous distribution of data across devices.

This study employs a client-side learning approach, ensuring that raw data remains on user devices, thus preserving privacy while still enabling effective global model updates. The following stages describe each step of this process in detail.

Stage 1: Data Preparation

At this stage, data is collected through a web platform where participants complete the Maslach Burnout Inventory (MBI) questionnaire. Once collected, the data is cleaned by removing duplicates, input errors, and outdated records, then normalized and standardized. Special attention is given to simulating real-world conditions with heterogeneous (non-IID) data to enhance model robustness and realism.

Stage 2: Local Training on Clients

Each client trains a random forest model locally on their own data without transmitting raw information to the server. The updated model parameters (e.g., tree weights and performance metrics) are sent to the central server to ensure data privacy. This approach allows the model to adapt to the unique characteristics of each client's dataset and improves prediction accuracy.

Stage 3: Model Aggregation on the Server

The server collects updates from all clients and aggregates them using algorithms such as FedAvg, FedProx, or FedOpt. Aggregation is weighted according to the volume of data on each client to ensure fair contribution. After aggregation, the server evaluates model convergence using metrics like MSE and R^2 and adjusts training parameters if needed.

Stage 4: Model Updating and Client Communication

The updated global model is sent back to clients via secure communication channels. Clients integrate the new model into their local environment and resume training on their data. This iterative process of model distribution and retraining ensures continuous improvement and adaptation to new data.

Stage 5: Model Evaluation

After several rounds of training, the model is evaluated using both classification (Accuracy, Precision, Recall, F1-score) and regression (MSE, R^2) metrics. The impact of data heterogeneity among clients is analyzed to assess its effect on model accuracy and convergence. This comprehensive evaluation helps identify weaknesses and optimize aggregation strategies and learning parameters. Fig. 3 illustrates these steps.

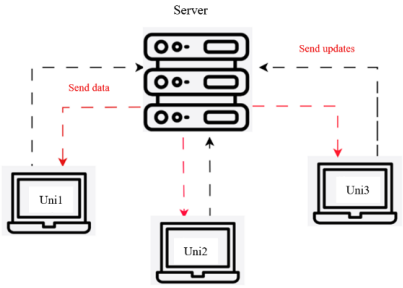
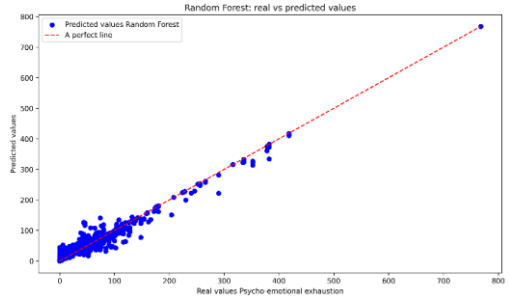


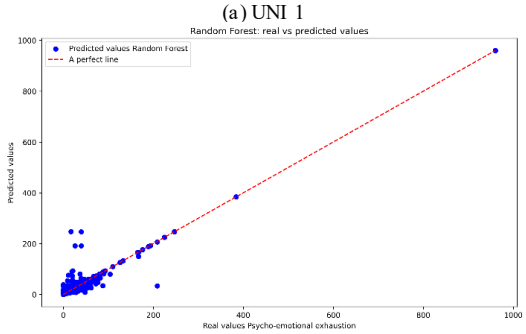
Fig. 3. Federated learning architecture: client-server interaction.

The Random Forest method was applied to the local model; this algorithm was determined by conducting experiments during which methods such as logical regression, linear regression, and time series were considered. The Random Forest method is an ensemble machine learning algorithm that uses multiple decision trees to solve classification and regression problems. The essence of the method is to train multiple trees on a Random Forest of training data and features, which improves the generalization ability of the model and reduces overfitting. Each tree is trained on a random sample of data with return (bootstrap method), and for classification, the final decision is made by a majority vote of the trees, and for regression, through averaging of predictions. Before sending them to the server after training on local devices, the following results were obtained, which are presented in Fig. 4.

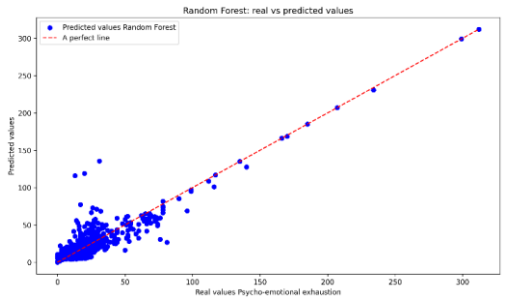
The graphs' description illustrates the relationship between actual and predicted values of Psycho-emotional exhaustion. The horizontal axis represents actual values, while the vertical axis shows predicted values. A red dashed line indicates a perfect match. For low and medium values, predictions closely align with the ideal line, demonstrating high accuracy. As values increase, a slight deviation appears, with greater scatter at higher levels (~1000). Despite this, predictions remain well-correlated with the actual data, maintaining stable performance across different ranges.



(a) UNI 1



(b) UNI 2



(c) UNI 3

Fig. 4. Comparison of the results of the random forest method in (a) UNI 1, (b) UNI 2, (c) UNI 3.

The pros of the method are high accuracy, especially for complex and high-dimensional data, robustness to overfitting due to the ensemble approach, handling of missing values, and estimation of feature importance, which aids in model interpretation. The cons are low interpretability compared to single trees, difficulties with prediction on very large data due to the large number of trees. Random forest is widely used in classification and regression tasks such as disease diagnosis, credit scoring, price prediction, and tasks requiring feature selection. In the federated learning task, random forest is an effective method for combining models from different clients, providing robustness to local data and improving overall performance. After numerous experiments on the server side, a local model was picked up and sent to the client side, where training began. The graphs compare the actual values and predicted values of the Random Forest model for the Psycho-emotional exhaustion index. For this purpose, data from different devices were used, which allows us to clearly demonstrate the effectiveness of the model in the conditions of heterogeneous data. The Random Forest model shows a good ability to predict Psycho-emotional exhaustion, especially in the small and medium ranges of values. Visualization confirms that the model successfully captures the data patterns, although some deviation from the ideal line is observed at higher values (see Table II).

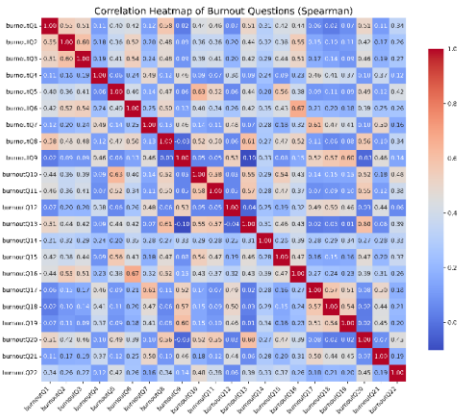
TABLE II ACCURACY BEFORE SENDING IT TO THE SERVER

	UNI 1	UNI 2	UNI 3
Accuracy	0.915676	0.554274	0.858384

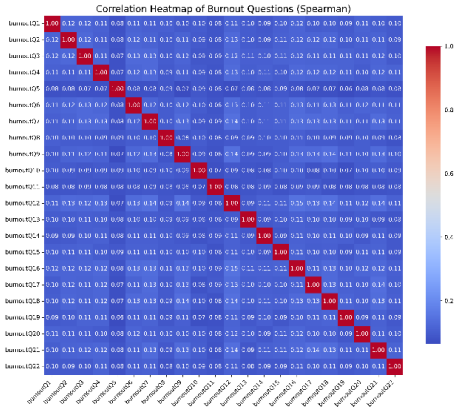
1) Description of the preparatory process: To solve the task, a website was created where the data collection of each individual student on their personal devices was organized. For example, it was decided to make a psycho-emotional and physical analysis of the state of students and schoolchildren based on the collected database. Currently, the FL has not yet started its march in Kazakhstan, so my work is of an introductory nature on the possibilities of using the FL. The purpose of the experiment is to make a practical demonstration of the whole process of organizing FL steps on a specific example described in the study.

Three educational institutions participated in the experiment; data were collected from the beginning of the learning process, from September 1, for 2 months before the first milestone control. A website was created for data collection, where all data were stored in a database based on the educational institution. The aim was to analyze the mental and emotional state of students from the moment of the beginning of the study and by the end of the first milestone control to determine how much their state had changed by the time of passing the first milestone control.

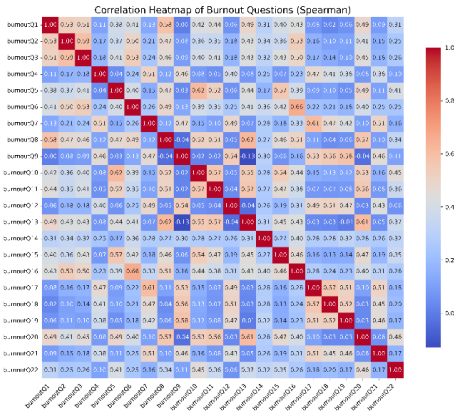
These heatmaps provide an overview of how different questions about burnout relate to each other, which can be useful for identifying patterns and determining which questions may share common underlying factors or experiences, as in Fig. 5.



(a) UNI 1



(b) UNI 2



(c) UNI 3

Fig. 5. Correlation heatmap of burnout questions in (a) UNI 1, (b) UNI 2, (c) UNI 3.

The heatmaps display Spearman correlation coefficients for burnout-related questions (burnoutQ1–burnoutQ22), with values ranging from -1 to 1. Red indicates strong positive correlations, blue represents weak or no correlations, and white signifies no correlation (0). Heatmap 1 highlights strong correlations between certain burnout questions. Heatmap 2 shows generally lower correlation values, with many closes to zero. Heatmap 3 reveals some strong correlations (e.g., burnoutQ1–Q3, burnoutQ6–Q9), but overall, the values remain weaker. These heat maps help identify patterns and relationships between burnout factors.

Application of FL was aimed at predicting further psycho-emotional and physical state of students for the next boundary control and readiness to pass the winter session. The heterogeneity of the data can be determined based on the following characteristics shown in Table III.

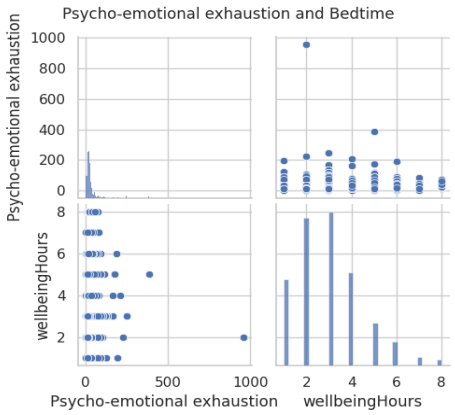
TABLE III CHARACTERISTICS OF THE DATA

Data type	Heterogeneous	Homogeneous
Nutrition	'Breakfast', 'Dinner', 'Lunch', 'total_meals'.	
Physical activity	'intensity', 'activityType', 'duration'	
Sleep and rest	'wellbeingHours0', 'wellbeingHours0', 'wellbeingHours1'.	
Emotional state	'Psycho-emotional exhaustion', 'burnoutQ1'- 'Q22'.	
Identifiers and meta-data		'student_id', 'entry_date'

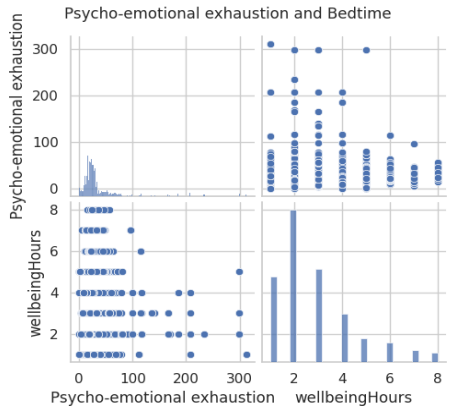
2) Heterogeneity analysis: Students' data exhibits heterogeneity due to variations in behavioral patterns, including nutrition, physical activity, sleep, and emotional state. For instance, dietary habits differ among students, with some skipping meals while others follow a structured diet. Similarly, physical activity levels vary, ranging from active sports participation to a sedentary lifestyle. Emotional burnout scores also fluctuate based on individual stress levels and coping mechanisms. Conversely, homogeneous attributes include 'entry date' (consistent format across devices) and 'student_id' (a unique identifier that does not impact target metrics). These attributes remain stable regardless of individual behaviors in Fig. 6. To address data heterogeneity, preprocessing is required to normalize activity and nutrition data, ensuring consistency across samples. Additionally, leveraging homogeneous attributes (e.g., timestamps, unique identifiers) aids in standardization and enhances model reliability.

The pair plots illustrate a weak negative correlation between Psycho-emotional exhaustion and wellbeingHours. Most students show low sleep duration regardless of exhaustion level, indicating that increased rest does not consistently reduce emotional fatigue. The presence of outliers reflects the heterogeneity of behavioral data among students.

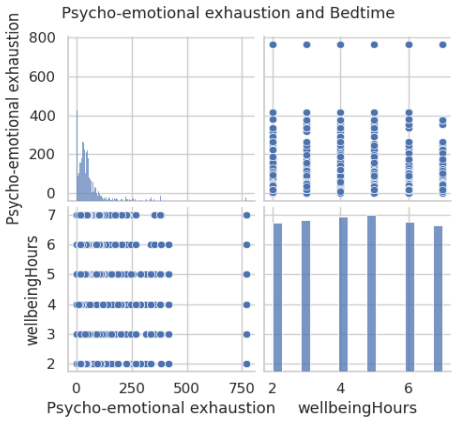
The following results highlight the comparative performance and convergence behavior of the selected FL algorithms under non-IID educational data conditions.



(a) UNI 1



(b) UNI 2



(c) UNI 3

Fig. 6. Fluctuations in psycho-emotional exhaustion over one month in (a) UNI 1, (b) UNI 2, (c) UNI 3.

IV. RESULTS AND DISCUSSION

In this study, the performance of three federated learning algorithms, namely FedAVG, FedOPT, and FedProx, was benchmarked on a classification task aimed at predicting students' psycho-emotional states. The primary evaluation metric was Accuracy, measured on test data after each of ten training rounds. Table IV and Fig. 7 illustrate the dynamics of model accuracy across all rounds, providing insight into the learning stability and convergence behavior of each algorithm.

TABLE IV ACCURACY DYNAMICS FOR ALGORITHMS

Rounds	Accuracy		
	FedAVG	FedOpt	FedProx
1	0.793071234	0.8383071568	0.9682882882
2	0.850712345	0.8707123457	0.9793333333
3	0.9366666666	0.942457545	0.989333333
4	0.95956756756	0.9895675675	0.9866633333
5	0.97666666666	0.926666666	0.999936
6	0.9956756756	0.9695675675	0.99657878888
7	0.97666666666	0.9876666666	0.988888999
8	0.96666666666	0.9866666666	0.9968288288
9	0.995675675	0.995675675	0.9933333333
10	0.999675699	0.999675699	0.998888999

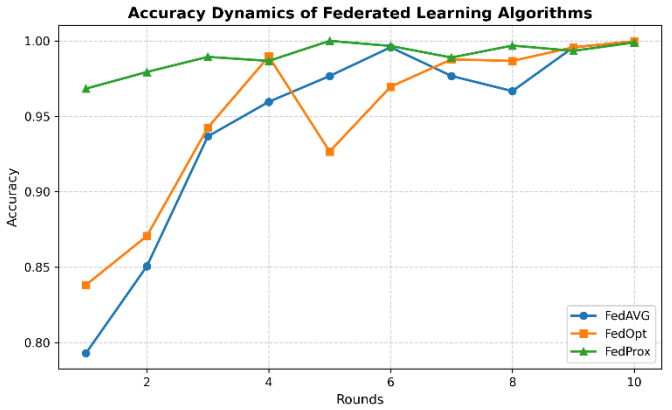


Fig. 7. Accuracy dynamics for algorithms.

As shown in Fig. 7, all three algorithms exhibit an upward trend in accuracy as training progresses, demonstrating the effectiveness of the federated learning approach in distributed educational data environments. Among them, FedProx consistently achieved the highest accuracy, particularly in later rounds where it reached 99.99%, indicating its robustness in managing heterogeneous, non-IID data. This superior performance is attributed to its proximal regularization, which stabilizes updates and reduces divergence caused by local data variability across clients:

a) FedOpt, which incorporates adaptive optimizers (e.g., Adam), also demonstrated strong convergence characteristics, achieving accuracy levels close to FedProx while maintaining lower communication overhead. However, its slightly lower accuracy in the final rounds suggests that additional regularization could further enhance its performance under highly heterogeneous conditions.

b) FedAVG, while computationally efficient, displayed slower convergence and lower final accuracy compared to the other two methods. Its performance plateaued in the later rounds, emphasizing its sensitivity to client data heterogeneity. Nonetheless, FedAVG remains a practical solution for homogeneous datasets or resource-constrained environments, where computational simplicity is prioritized over precision.

c) Overall, the results confirm that FedProx is the most effective algorithm for addressing data heterogeneity and achieving high model accuracy in psycho-emotional state prediction. FedOpt offers a balanced trade-off between speed and accuracy, whereas FedAVG is better suited for uniform data distributions.

These findings support the study's objectives by demonstrating that federated learning algorithms can be effectively applied to privacy-preserving educational analytics. The results offer actionable insights for educational institutions, researchers, and policymakers, emphasizing the potential of federated learning to enhance the monitoring of students' psycho-emotional well-being while ensuring data confidentiality and accommodating diverse learning environments across regions.

Moreover, the results directly align with the study's objective of developing a federated learning framework for monitoring mental health in Kazakhstan, confirming that such systems can be implemented effectively under heterogeneous educational conditions.

V. CONCLUSION AND FUTURE SCOPE

This study aimed to develop and evaluate a privacy-preserving federated learning framework for analyzing the psycho-emotional well-being of students in Kazakhstan, particularly under conditions of data heterogeneity and non-IID distributions. To achieve this, three federated learning algorithms, FedAvg, FedOpt, and FedProx, were benchmarked across ten communication rounds in a classification task. The results revealed distinct performance dynamics, convergence rates, and stability patterns under federated conditions.

FedAvg showed steady improvement in accuracy from 0.793 in round 1 to 0.9997 in round 10, indicating reliable convergence despite minor fluctuations in intermediate stages. FedOpt achieved slightly higher accuracy in the early rounds (0.9425 and 0.9895 in rounds 3 and 4) and demonstrated faster convergence than FedAvg. FedProx, starting with the highest initial accuracy (0.9683), maintained consistent progress and reached 0.9989 by the final round, confirming its robustness and effectiveness in handling heterogeneous and non-IID data distributions.

These findings validate the study's objective by demonstrating that federated optimization algorithms can effectively manage data heterogeneity and ensure secure, distributed model training for sensitive educational analytics. However, the experiments were conducted in a simulated federated environment involving a limited number of institutions, which may not fully represent large-scale or real-world deployments.

Future work will address these limitations by incorporating a broader range of educational institutions, exploring asynchronous and cross-silo FL configurations, and integrating additional behavioral and contextual variables to enhance predictive accuracy, adaptability, and robustness. Previous studies have explored centralized machine learning for mental health analysis and early intervention [31]; however, these relied on centralized data aggregation, leading to privacy and scalability challenges.

This research advances the field by introducing a federated learning-based approach that ensures data confidentiality while

maintaining model accuracy and adaptability across heterogeneous educational environments. Overall, it establishes a foundation for ethical, data-driven, and privacy-aware educational analytics in Kazakhstan and beyond, supporting early intervention strategies and strengthening national efforts to promote student mental health and academic resilience.

REFERENCES

[1] G. P. Reddy and Y. V. Pavan Kumar, "A beginner's guide to federated learning," in 2023 Intelligent Methods, Systems, and Applications (IMSA), 2023, pp. 557–562. doi:10.1109/IMSA58542.2023.10217383.

[2] J. -W. Lin, T. Kim, H. -J. Tu and P. -H. Kuo, "Adaptive Model Transfers and Aggregations for Efficient Federated Learning in IoT Edge Systems with Non-IID Data," 2023 6th International Conference on Data Science and Information Technology (DSIT), Shanghai, China, 2023, pp. 56-60, doi: 10.1109/DSIT60026.2023.00017.

[3] S. M. Jalal, M. R. Hasan, M. A. Haque and M. G. R. Alam, "A Horizontal Federated Random Forest for Heart Disease Detection from Decentralized Local Data," 2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC), Hyderabad, India, 2022, pp. 191-196, doi: 10.1109/R10-HTC54060.2022.9929490.

[4] J. Mills, J. Hu and G. Min, "Multi-Task Federated Learning for Personalised Deep Neural Networks in Edge Computing," in IEEE Transactions on Parallel and Distributed Systems, vol. 33, no. 3, pp. 630-641, 1 March 2022, doi: 10.1109/TPDS.2021.3098467.

[5] Pranto, M.A., & Al Asad, N. (2021). A Comprehensive Model to Monitor Mental Health based on Federated Learning and Deep Learning. 2021 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON), 18-21.

[6] J. -W. Lin, T. Kim, H. -J. Tu and P. -H. Kuo, "Adaptive Model Transfers and Aggregations for Efficient Federated Learning in IoT Edge Systems with Non-IID Data," 2023 6th International Conference on Data Science and Information Technology (DSIT), Shanghai, China, 2023, pp. 56-60, doi: 10.1109/DSIT60026.2023.00017.

[7] Q. Ma et al., "FedSA: A Semi-Asynchronous Federated Learning Mechanism in Heterogeneous Edge Computing," in IEEE Journal on Selected Areas in Communications, vol. 39, no. 12, pp. 3654-3672, Dec. 2021, doi: 10.1109/JSAC.2021.3118435.

[8] Z. Jiang et al., "Computation and Communication Efficient Federated Learning With Adaptive Model Pruning," in IEEE Transactions on Mobile Computing, vol. 23, no. 3, pp. 2003-2021, March 2024, doi: 10.1109/TMC.2023.3247798.

[9] T. Vu et al., "User selection approaches to mitigate the straggler effect for federated learning on cell-free massive MIMO networks," arXiv preprint arXiv:2009.02031, 2020.

[10] J. Le et al., "Federated Continuous Learning With Broad Network Architecture," in IEEE Transactions on Cybernetics, vol. 51, no. 8, pp. 3874-3888, Aug. 2021, doi: 10.1109/TCYB.2021.3090260.

[11] J. P. Albrecht, "How the GDPR will change the world," Eur. Data Prot. Law Rev., vol. 2, pp. 287–289, 2016. doi: 10.1093/IDPL/IPX023.

[12] X. Zhang, Q. Wang, Z. Ye, H. Ying and D. Yu, "Federated Representation Learning With Data Heterogeneity for Human Mobility Prediction," in IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 6, pp. 6111-6122, June 2023, doi: 10.1109/TITS.2023.3252029.

[13] K. Bonawitz et al., "Practical secure aggregation for privacy-preserving machine learning," in Proc. 2017 ACM SIGSAC Conf. Comput. Commun. Secur., 2017 doi:10.1145/3133956.3133982.

[14] J. Konecný et al., "Federated learning: Strategies for improving communication efficiency," arXiv preprint arXiv:1610.05492, 2016.

[15] O. R. A. Almanifi et al., "Communication and computation efficiency in federated learning: A survey," Internet Things, vol. 22, p. 100742, 2023. doi: 10.1016/j.iot.2023.100742.

[16] H. Zhang, L. Liu, and L. Lu, "Federated scientific machine learning for approximating functions and solving differential equations with data heterogeneity," arXiv preprint arXiv:2410.13141, 2024. doi: 10.48550/arXiv.2410.13141.

[17] T. Sun, D. Li, and B. Wang, "Decentralized federated averaging," IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, pp. 4289–4301, 2021. doi: 10.1109/TPAMI.2022.3196503.

[18] S. J. Reddi et al., "Adaptive federated optimization," arXiv preprint arXiv:2003.00295, 2020.

[19] M. A. Hidayat, Y. Nakamura and Y. Arakawa, "Privacy-Preserving Federated Learning With Resource-Adaptive Compression for Edge Devices," in IEEE Internet of Things Journal, vol. 11, no. 8, pp. 13180-13198, 15 April15, 2024, doi: 10.1109/JIOT.2023.3347552.

[20] H. Ko et al., "Joint Client Selection and Bandwidth Allocation Algorithm for Federated Learning," in IEEE Transactions on Mobile Computing, vol. 22, no. 6, pp. 3380-3390, 1 June 2023, doi: 10.1109/TMC.2021.3136611.

[21] C. Wu, Z. Li, F. Wang and C. Wu, "Learning Cautiously in Federated Learning with Noisy and Heterogeneous Clients," 2023 IEEE International Conference on Multimedia and Expo (ICME), Brisbane, Australia, 2023, pp. 660-665, doi: 10.1109/ICME55011.2023.00119.

[22] M. Favero, T. Marchioro and C. Schiavo, "Federated Forests With Differential Privacy for Distributed Wearable Sensors," 2024 IEEE International Conference on Aerospace Electronics and Remote Sensing Technology (ICARES), Yogyakarta, Indonesia, 2024, pp. 1-6, doi: 10.1109/ICARES64249.2024.10767941.

[23] X. Feng, Q. Shen, C. Li, Y. Fang and Z. Wu, "Privacy Preserving Federated Learning from Multi-Input Functional Proxy Re-Encryption," ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Korea, Republic of, 2024, pp. 6955-6959, doi: 10.1109/ICASSP48485.2024.10446283.

[24] L. Liu and D. Zhou, "Analysis of regularized federated learning," Neurocomputing, vol. 611, p. 128579, 2024. doi: 10.1016/j.neucom.2024.128579.

[25] Koh, Y., Lee, C., Ku, Y., & Lee, U. (2023). Data Visualization for Mental Health Monitoring in Smart Home Environment: A Case Study. 2023 Workshop on Visual Analytics in Healthcare (VAHC), 53-55.

[26] Wang, P. (2024). Analysis of Adolescent Mental Health Based on Machine Learning Algorithm. 2024 3rd International Conference for Advancement in Technology (ICONAT), 1-5.

[27] L. Shi et al., "Data poisoning attacks on federated learning by using adversarial samples," 2022 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI), Shijiazhuang, China, 2022, pp. 158-162, doi: 10.1109/ICCEAI55464.2022.00041.

[28] W. Qiu, W. Ai, H. Chen, Q. Feng and G. Tang, "Decentralized Federated Learning for Industrial IoT With Deep Echo State Networks," in IEEE Transactions on Industrial Informatics, vol. 19, no. 4, pp. 5849-5857, April 2023, doi: 10.1109/TII.2022.3194627.

[29] Y. Deng and X. Yan, "Federated Learning on Heterogeneous Opportunistic Networks," 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), Nanjing, China, 2024, pp. 447-451, doi: 10.1109/AINIT61980.2024.10581819.

[30] Tang and T. -H. Chang, "FedLion: Faster Adaptive Federated Optimization with Fewer Communication," ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Korea, Republic of, 2024, pp. 13316-13320, doi: 10.1109/ICASSP48485.2024.10447045.

[31] Kataru, S., King, K., & Fernando, L. (2024). Machine Learning-Based Early Detection and Intervention for Mental Health Issues in Children. 2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC), 2001-2007.