Edge Computing Model based on Federated Learning for COVID-19 Clinical Outcome Prediction in the 5G Era

Ruochen Huang[1], Zhiyuan Wei[1], Wei Feng[1], Yong Li[3], Changwei Zhang[4], Chen Qiu[5] and Mingkai Chen[2*]

[1] School of Biomedical Engineering and Informatics, Nanjing Medical University
Nanjing, Jiangsu, 211166, China
[e-mail: huangrc@njmu.edu.cn, zhiyuanwei@njmu.edu.cn, fengwei@njmu.edu.cn]
[2] Key Laboratory of Broadband Wireless Communication and Sensor Network Technology, Nanjing University of Posts and Telecommunications, Nanjing, 210003, China
[e-mail: mkchen@njupt.edu.cn]
[3] The First Affiliated Hospital, Nanjing Medical University, Nanjing, Jiangsu, 210029, China
[e-mail: liyongmyzone@126.com]
[4] Pervasive Communication Center, Purple Mountain Laboratories, Nanjing, 211111, China
[e-mail: changwei.z@outlook.com]
[5] Peng Cheng Laboratory, Shenzhen, Guangdong, 518100, China
[e-mail: qiuch@pcl.ac.cn]
[*]Corresponding author: Mingkai Chen

Received December 6, 2023; revised Feburary 13, 2024; accepted March 27, 2024; published April 30, 2024

Abstract

As 5G and AI continue to develop, there has been a significant surge in the healthcare industry. The COVID-19 pandemic has posed immense challenges to the global health system. This study proposes an FL-supported edge computing model based on federated learning (FL) for predicting clinical outcomes of COVID-19 patients during hospitalization. The model aims to address the challenges posed by the pandemic, such as the need for sophisticated predictive models, privacy concerns, and the non-IID nature of COVID-19 data. The model utilizes the FATE framework, known for its privacy-preserving technologies, to enhance predictive precision while ensuring data privacy and effectively managing data heterogeneity. The model's ability to generalize across diverse datasets and its adaptability in real-world clinical settings are highlighted by the use of SHAP values, which streamline the training process by identifying influential features, thus reducing computational overhead without compromising predictive precision. The study demonstrates that the proposed model achieves comparable precision to specific machine learning models when dataset sizes are identical and surpasses traditional models when larger training data volumes are employed. The model's performance is further improved when trained on datasets from diverse nodes, leading to superior generalization and overall performance, especially in scenarios with insufficient node features. The integration of FL with edge computing contributes significantly to the reliable prediction

This research was funded by National Natural Science Foundation of China under Grant 62101293, China Postdoctoral Science Foundation under Grant 2021M701806, 2023M742672 and the Major Key Project of PCL

http://doi.org/10.3837/tiis.2024.04.001 **ISSN : 1976-7277**

of COVID-19 patient outcomes with greater privacy. The research contributes to healthcare technology by providing a practical solution for early intervention and personalized treatment plans, leading to improved patient outcomes and efficient resource allocation during public health crises.

Keywords: 5G; Edge Computing Model; COVID-19; Federated Learning; SHAP

1. Introduction

The COVID-19 pandemic has caused profound devastation across diverse regions globally. While the majority of patients have recovered, those with compromised constitutions or underlying conditions have tragically succumbed to the disease [1]. The novel coronavirus continues to undergo mutations. The situation calls for the urgent development of sophisticated predictive models to enhance outbreak tracking, early detection, and rapid deployment of medical resources, underscoring the critical need for advanced data analysis and machine learning in combating these challenges.

The pandemic's response has been hampered by several hurdles, notably the safeguarding of user privacy and the collection of comprehensive, real-time data sets for predictive modeling. Traditional machine learning approaches, while foundational for predicting clinical outcomes, grapple with significant issues [2]-[9]. Centralizing sensitive COVID-19 patient data raises serious privacy and security concerns, risking patient confidentiality and exposing data to potential breaches.

Additionally, the non-IID nature of COVID-19 data, shaped by diverse factors like regional healthcare practices and genetic variations, challenges traditional models that rely on IID assumptions. This can lead to biased or inaccurate predictions. Moreover, the sheer volume of data required for robust predictions may overwhelm these models, particularly in regions lacking the digital infrastructure to support such data analysis [10].

In this context, the study introduces a novel model supported by Federated Learning (FL) and edge computing, leveraging decentralized machine learning to address these challenges. This model, utilizing the FATE framework known for privacy-preserving technologies like homomorphic encryption and secure multi-party computation, is designed to enhance predictive precision while ensuring data privacy and managing data heterogeneity effectively.

This research, by integrating FL with edge computing, contributes significantly to predicting COVID-19 patient outcomes more reliably and with greater privacy. It presents a robust framework that not only facilitates broader participation in model training but also remains precise across varying data conditions.

The remainder of this paper is structured as follows: Section 2 provides an overview of the applications of machine learning and edge computing in the prevention, diagnosis, and treatment of COVID-19. Section 3 presents a comprehensive overview of the FL-supported edge computing model, including its structure and specific details. Section 4 presents the results obtained from applying the proposed model to a real-world dataset. Finally, Section 5 concludes the paper with a summary of the key findings and contributions.

2. Related Work

The advent of AI technology has had a profound impact on the healthcare industry, particularly in the areas of prevention, diagnosis, and treatment of COVID-19. Traditional machine learning models, such as Decision Trees (DT), Support Vector Machines (SVM), and Logistic Regression (LR), have been widely adopted due to their low algorithmic complexity and cost-effective hardware implementation [11]-[13]. These models have been crucial in preventing further infections by predicting the number of confirmed and death cases of COVID-19, predicting mortality and severity by combining cardiac markers with demographic and clinical features, and providing early risk assessment and accurate prediction of clinical endpoints [2][3][11].

On the other hand, deep learning models have shown exceptional capability in various medical scenarios, including the detection and diagnosis of COVID-19 through imaging [14]-[16]. The integration of these models with 5G and edge computing has led to the development of frameworks that leverage low latency, scalability, data protection, and reliable local edge servers for effective management of epidemics. However, it is crucial to acknowledge that deep learning models typically require massive amounts of data and high hardware requirements for training, which can be a limiting factor, especially when dealing with multimodal features in cloud-based scenarios.

The revolutionary integration of 5G technology into healthcare has marked a new epoch, particularly with the advent of edge computing. This technology stands out for its ability to provide low-latency and highly reliable data processing capabilities, which are imperative in critical healthcare applications [17]-[20]. One of the salient features of 5G edge computing in healthcare is its transformative impact on medical imaging. It has significantly expedited the process of image processing and data transmission, ensuring that high-quality medical images are readily available to healthcare professionals in real-time. This has not only augmented diagnostic precision but also dramatically increased the efficiency of the entire diagnostic process, resulting in faster and more reliable patient care [21].

Moreover, edge computing has paved the way for the implementation of innovative healthcare solutions such as remote patient monitoring and telesurgery [22][23]. These applications require Ultra-Reliable Low-Latency Communication (URLLC), which edge computing is uniquely positioned to provide. By processing data closer to the source, it ensures minimal delay, thereby facilitating real-time decision-making and interventions that could be life-saving.

However, it is imperative to acknowledge the challenges that come with these innovations. The consistency of network stability is crucial, especially when dealing with critical healthcare applications where any delay or disruption can have dire consequences. Additionally, as healthcare data is extremely sensitive, the decentralized nature of edge computing networks opens up potential security vulnerabilities. Protecting patient data from unauthorized access and ensuring the integrity of the data becomes paramount, and this is an area that requires continuous attention and improvement [24].

FL has emerged as a pivotal innovation within distributed machine learning, facilitating the cooperative training of models across an extensive network of devices while preserving the privacy of sensitive data. This revolutionary approach has found applications across a diverse range of sectors, from finance [25][26] to industrial Internet of Things (IoT) systems [27]-[29], demonstrating its versatility and effectiveness in harnessing decentralized data sources for machine learning. Among these applications, the healthcare sector stands out as one of the most critical areas where FL is making a significant impact.

In healthcare, where the protection of patient confidentiality is of utmost importance, FL offers a transformative solution. By enabling the application of FL to electronic health records, significant advancements have been made in predictive analytics, leading to the development of more personalized and precise treatment plans. This method not only enhances the accuracy and effectiveness of medical care but also ensures the protection of patient data, preventing any breaches of privacy. The widespread adoption of FL in various domains underscores its potential to revolutionize data analysis and machine learning, with the healthcare industry benefiting immensely from its ability to safeguard patient information while improving treatment outcomes [30][31].

In addition to preserving data confidentiality, FL also addresses the critical issue of data silos in healthcare. It enables the utilization of diverse datasets from various institutions for model training without needing to centralize the data, thereby enhancing the robustness and generalizability of the models. This is crucial in a field like healthcare where patient populations can vary significantly across different regions and institutions.

The unprecedented challenges posed by the COVID-19 pandemic necessitate innovative and technologically advanced solutions. The integration of FL with edge computing stands out as a strategic approach, synergistically harnessing the capabilities of both technologies to establish a system that is both responsive and efficient in crisis management.

By capitalizing on the low-latency and reliable data processing features of edge computing, this integration markedly enhances the performance of FL models, especially in scenarios that demand real-time responsiveness. Ensuring the promptness and precision of insights garnered from these models is imperative, as it underpins informed decision-making, a critical component in effectively combating the COVID-19 pandemic.

3. Materials and methods

3.1 Datasets

In this study, a real-world COVID-19 dataset was obtained from a hospital in Jiangsu province, China. Participants were diagnosed with COVID-19 using RT-PCR detection of SARS-CoV-2 nucleic acid or serum SARS-CoV-2-specific IgM and IgG antibodies. Severe cases were defined based on the criteria outlined in the "COVID-19 Diagnosis and Treatment Scheme (Trial Version 6)", which included the following: (1) blood oxygen saturation≤93%, (2) respiratory rate≥30 times / minute, and (3) arterial blood oxygen partial pressure (PaO2) / oxygen concentration (FiO2)≤300 mmHg.

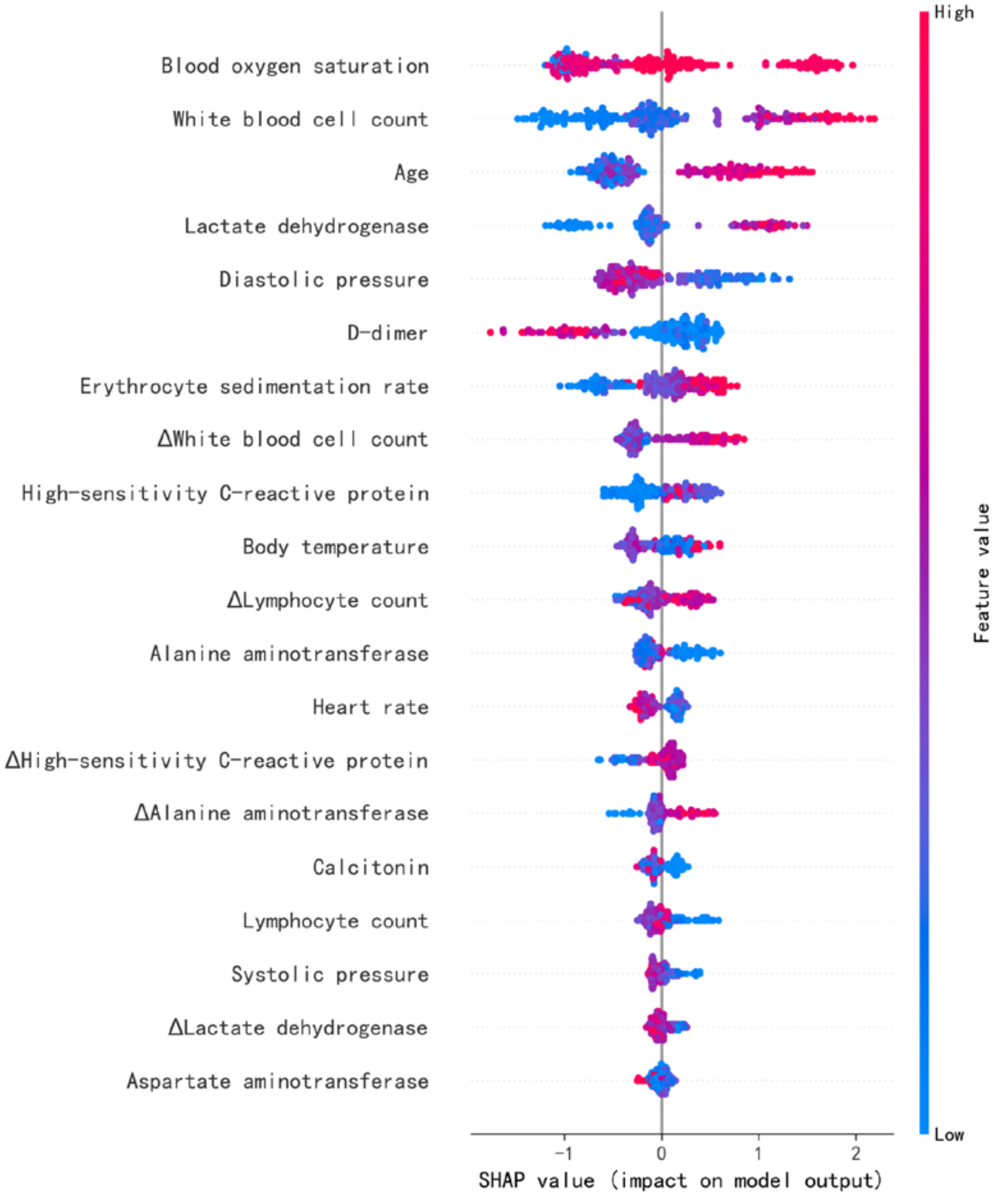


Fig. 1. SHAP values for the top 20 features.

The study conducted a thorough investigation, encompassing a diverse array of variables such as demographic details, symptoms, clinical test indicators (inclusive of CT indices), disease and treatment specifics, clinical attributes, and pathological factors. The utilization of SHapley Additive exPlanations (SHAP) values facilitated an evaluation of the individual contributions of features towards the model's efficacy, akin to assessing each player's impact on the final outcome in a game scenario. The SHAP values of the top 20 features, identified through the highest mean absolute SHAP values, are depicted as an extensive distribution in Fig. 1. In this representation, each dataset entry is denoted as a data point for its corresponding feature. The x-axis illustrates the feature's directional influence—positive or negative—on the model's predictive capability. The color scheme denotes feature values, scaled according to the range observed within the dataset.

3.2 FL-supported edge computing model

The preceding introduction underscores the significant potential of 5G in preventing and treating COVID-19 across various scenarios, with edge computing playing a pivotal role in realizing its unique capabilities. Cloud computing is distant from terminal devices, resulting in considerable bandwidth consumption and notable transmission delays. In addition to the considerable expenses associated with transmission and delays, edge devices must transmit sensitive data to cloud servers for processing, posing potential risks to data confidentiality. The concept of edge computing has emerged as a response to the growing need for enhanced privacy and cost-effective communication. Through the utilization of distributed computing, edge computing facilitates the transfer of computing power to the data collection point, leading to decreased data transmission delays, improved user experience, and heightened levels of security and privacy. In comparison to cloud computing, edge computing offers the following advantages: 1) It reduces bandwidth consumption during transmission by performing small transaction fragmentation data processing tasks at the forefront. 2) It ensures data security through local processing of privacy-sensitive data. 3) Edge computing enhances system robustness, enabling uninterrupted local data computing and transmission even in the presence of abnormal external network conditions.

The concept of FL has made a significant contribution to addressing the challenge of preserving user data privacy during machine learning training. By adopting a decentralized approach with a centralized server and distributed clients, FL enables collaborative model training through the transmission of model parameters. Each data island maintains its original data locally, ensuring data privacy throughout the training process, while the centralized server enhances model performance by aggregating and sharing the model parameters from the distributed clients. The emergence of FL has significantly lowered the barriers to data sharing, enabling a broader participation of data owners in the model training process. Additionally, it effectively mitigates the risks associated with data leakage and substantially reduces the costs associated with data concentration. The specific process of integrating FL and edge computing is outlined below.

In the first step, task initialization is performed. Initially, the cloud defines the training objectives and data format for the task, such as predicting the outcomes of hospitalized COVID-19 patients. Based on the characteristics of the edge nodes, the FL training process and model parameters, including the model parameter weights P^i, where $i = \{0, \dots, M\}$, are determined. Here, i denotes the current iteration round, and when $i = 0$, P^0 represents the initialized parameters. The number of iterations M, the number of all participating nodes N and local minibatch size B in the training process are also determined. Once these specifications are finalized, the cloud distributes the task requirements and model parameters to all edge nodes.

The second step involves edge node training and model updating. Let $n = \{1,2, \dots, N\}$ represent the edge nodes. The cloud randomly selects n' (where $n' \leq N$) edge nodes to distribute the model parameters. Upon receiving the global model parameters P^i, each edge node updates its local data to generate model parameters with local characteristics, denoted as P_n^i. The training objective is to minimize the loss function $L\left(P_n^i\right)$.

$$P_n^{i^*} = \arg\min_{P_n^i} L\left(P_n^i\right) \tag{1}$$

The local dataset D_n is initially partitioned into j subsets $B_j = \{1,2,\ldots,j\}$,based on local minibatch size B. For each subset B_j, iterate T rounds, updating the parameter P_n^i for each $b \in B_j$.

$$P_n^i \leftarrow P_n^i - \eta \nabla L(P;b) \tag{2}$$

Where η represents the learning rate, and ∇L represents the gradient. After completing the training, each edge node will upload the trained parameters P_n^i to the cloud.

In the third step, the cloud integrates and updates the parameters. The cloud consolidates the parameters uploaded by n' edge nodes and updates the global parameters. At the cloud, the optimization objective is to minimize the global loss function $L(P^i)$.

$$L(P^i) = \frac{1}{n'} \sum_{i=1}^{n'} L(P_n^i) \tag{3}$$

The cloud aggregates the parameters P_n^i from all edge nodes to compute the new global parameter P^{i+1}.

$$P^{i+1} = \frac{1}{\sum_{n \in N} D_n} \sum_{n=1}^{n'} D_n P_n^i \tag{4}$$

After multiple iterations of steps 2 and 3, the loss function converges or reaches the desired objective, signaling the completion of model training.

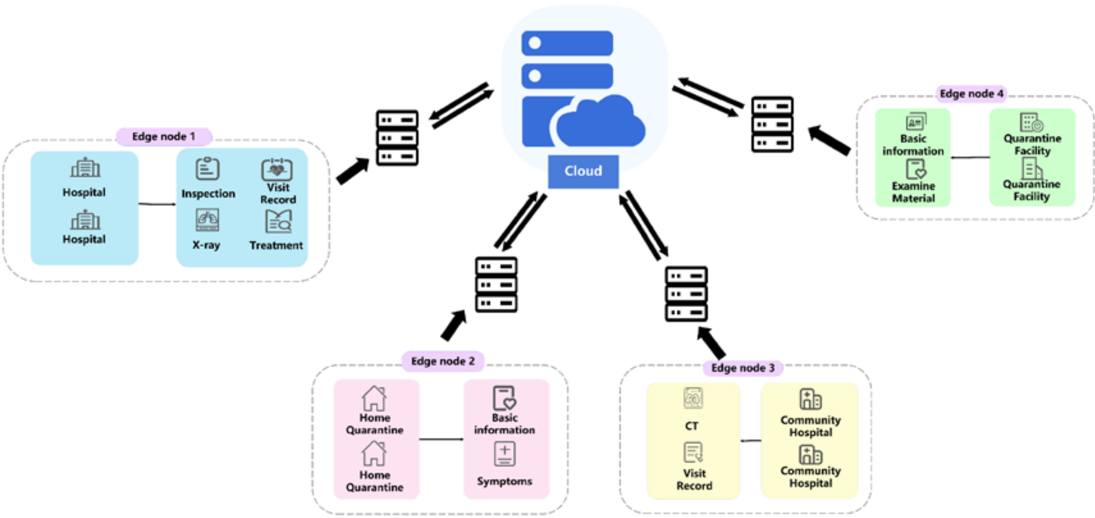


Fig. 2. Framework of FL-supported edge computing model for COVID-19.

As illustrated in **Fig. 2**, the FL-supported edge computing model distinguishes itself from traditional machine learning models by effectively leveraging the advantages of both edge computing and FL. By utilizing the computational capabilities of edge nodes for local data model training, the system optimizes resource utilization due to each node's limited data volume. Additionally, FL's parameter synchronization aligns the computational capabilities of local and cloud models, enhancing the delivery of AI services directly at the edge. This approach not only significantly reduces system latency and bandwidth consumption but also ensures rapid and efficient services for COVID-19 patients. Importantly, with patient data remaining local during model training, privacy concerns are mitigated, encouraging greater participation from medical practitioners and thereby enriching the dataset for developing more robust models.

Building upon this foundation, the study meticulously recreated the technological environment of various departments within a hospital in Jiangsu Province, China. The deployment of the FATE framework on cloud-based virtual machines within a designated sandbox environment enabled the analysis of partial COVID-19 case data. This setup demonstrated the FL-supported edge computing model's capability for sophisticated data analysis and machine learning model training within a secure, decentralized healthcare setting, highlighting its potential to revolutionize data-driven decision-making in critical sectors.

The integration of FATE, a leading collaborative learning platform designed for industrial-scale applications, further empowers organizations and institutions to develop machine learning models while prioritizing data privacy. Developed by Webank's AI department and made open-source in 2019, FATE now contributes to the Linux Foundation under the Apache 2.0 license. Its key features include a distributed architecture that supports various FL configurations, advanced security protocols for user privacy protection, and performance optimization tools for efficient large-scale model training. These attributes make FATE an ideal framework for implementing our FL-supported edge computing model, reinforcing the model's capabilities in managing complex, privacy-sensitive datasets across healthcare and other sectors.

4. Experiment results

In this study, the proposed model was compared with seven independent machine learning models: SVM, LR, DT, Catboost, Random Forest, Extra Trees, and Ada Boost. This rigorous approach was underpinned by clinical data meticulously gathered from a specified hospital in Jiangsu Province, China, ensuring that the process adhered to the highest clinical standards. The data collection process received rigorous evaluation and approval by the hospital's ethics committee, under the ethical approval designation number 2020-SR-106, with all pertinent documents related to this clearance submitted as part of the supplementary materials.

The dataset, comprising data from 303 patients confirmed to have COVID-19 through diagnostic methods such as the RT-PCR test and the assay for serum SARS-CoV-2-specific IgM and IgG antibodies, was categorized into two groups based on their admission detection indicators and initial severity classification: severe outcome (n=132) and mild outcome (n=171). These diagnostic approaches ensure a comprehensive identification of infected individuals, encompassing both active infections and previous exposure to the virus. Furthermore, the classification of severe COVID-19 cases adheres rigorously to the criteria outlined in the "COVID-19 Diagnosis and Treatment Scheme (Trial Version 6)", which include detailed clinical indicators such as respiratory distress, oxygen saturation levels, and other critical parameters. This methodology allows for a precise segmentation of the study population based on the severity of their condition, facilitating a nuanced analysis of the disease's impact and the effectiveness of various treatment approaches.

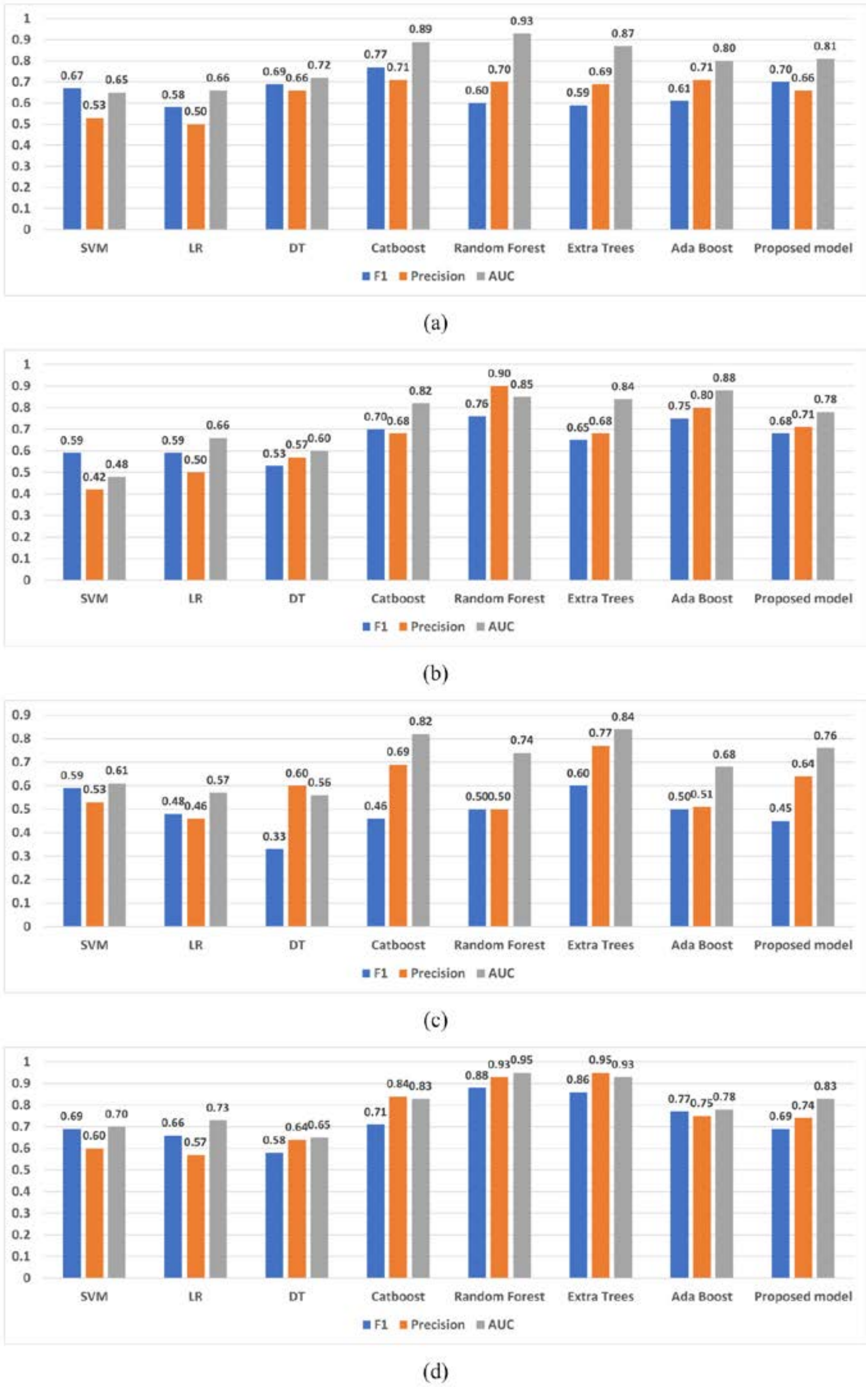


Fig. 3. F1 score, precision, and AUC comparison of DT, SVM, LR, Catboost, Random Forest, Extra Trees, Ada Boost and Proposed model. (a) shows the validation results after training all models on edge node 1. (b) shows the validation results after training all models on edge node 2. (c) shows the validation results after training all models on edge node 3. (d) shows the validation results after training all models on edge node 4.

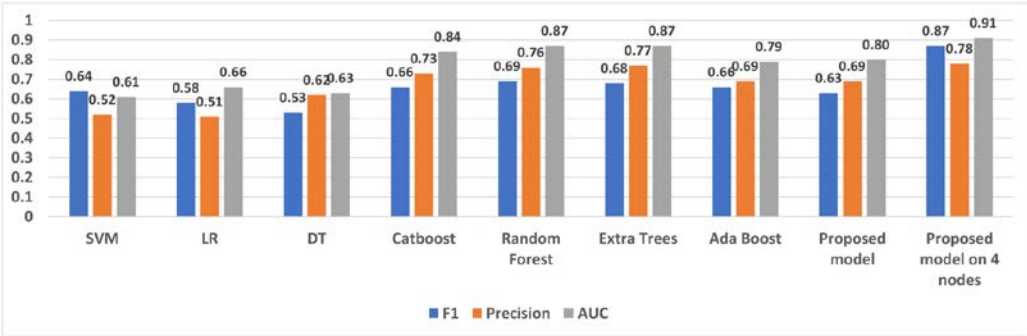


Fig. 4. Average validation results of all models trained on edge nodes 1 to 4 and the proposed model's validation on aggregated data from all four nodes.

Fig. 3(a) to **(d)** comparatively assessed the predictive precision of the machine learning models in edge computing environments, including a detailed simulation that evaluated the efficacy of the FL-supported model. Additionally, datasets from all nodes were aggregated to enhance the proposed model, and its performance was compared against the averaged results of the seven models, as depicted in **Fig. 4**. The assessment used performance metrics such as the F1 score, Area Under Curve (AUC), and precision, employing a 5-fold cross-validation method to ensure robust evaluation.

The proposed model exhibited remarkable performance, notably outperforming SVM, LR, and DT in precision metrics. Its performance was competitive against sophisticated ensemble methods such as Catboost. This demonstrated its efficacy in balancing sensitivity and specificity, particularly in the nuanced context of COVID-19 severity classification.

A critical observation was the proposed model's exceptional efficacy when leveraging the amalgamated dataset from all nodes. It achieved an F1 score of 0.87, precision of 0.78, and an AUC of 0.91, highlighting its scalability and adaptability, as well as its enhanced predictive performance over both traditional and ensemble classifiers. Further analysis revealed the varied performance of models like SVM, LR, and DT across different nodes could be attributed to their distinct reactions to the feature space and distribution. For instance, SVM and LR may struggle with non-linear separations in high-dimensional spaces, while DT is prone to overfitting amidst complex structures. Conversely, ensemble methods, by aggregating decisions from multiple weak learners, exhibit superior performance, indicating robust handling of data's complexity and diversity. The standout performance of the proposed model in an aggregated setting is likely attributed to the FL framework's capacity to harness diverse datasets without compromising privacy. Its design, tailored for the decentralized nature of edge computing, facilitates more generalized representations.

Model efficacy variations across nodes (E1 to E4) reflect the heterogeneity in data structure and distribution, impacting model suitability. The complexity of data structures, including feature count, correlation, and non-linear relationships, significantly influences model performance. Ensemble models like Catboost and Random Forest excel in managing complex, high-dimensional data structures through their aggregate learning approach.

The primary advantage of the proposed model lies in its ability to enhance data utilization efficiency while safeguarding privacy. By processing data locally and sharing only model updates, this approach reduces privacy concerns, promoting data sharing among individuals and medical entities. This is particularly crucial in addressing global health emergencies like COVID-19, where diverse data is essential in developing effective strategies.

Building on the extensive performance evaluation of the proposed model alongside seven independent machine learning models, the next phase of the analysis delves into the computational complexity. The comparison, based on training times, is summarized in the **Table 1**:

Table 1. Comparative analysis of training times across models

Model	Training Time (s)
SVM	0.55
LR	2.5
DT	0.8
Catboost	54.1
Random Forest	38.2
Extra Trees	34
Ada Boost	20
Proposed model	1862

The results indicate that the proposed model, which incorporates privacy-preserving technologies such as homomorphic encryption and secure multi-party computation within the FATE framework, incurs a significant computational overhead in terms of encrypted data communication and computation during the training process. However, it is important to note that the overhead associated with these data processing techniques does not scale linearly. The distributed nature of the FATE platform can effectively mitigate this overhead, especially when handling large datasets. This mitigation strategy significantly narrows the gap in training times between the proposed model and other models, underscoring its efficiency and feasibility in processing large-scale datasets.

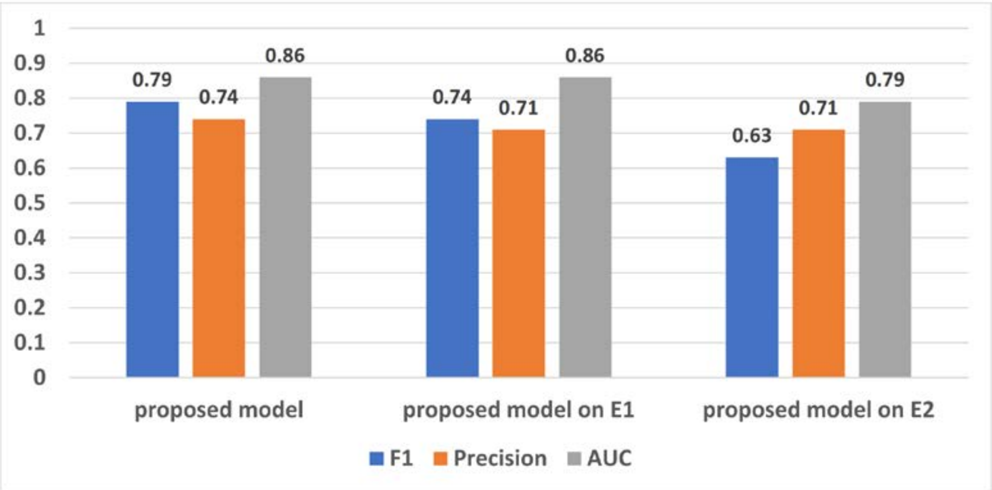


Fig. 5. F1 score, AUC, and precision comparison of proposed model on E1 and E2

Table 2. Feature in E1 and E2

Feature name	E1	E2
Age	Yes	Yes
Sex	Yes	Yes
Diseases of the endocrine system	Yes	No
Cardiovascular disease	Yes	No
Respiratory diseases	Yes	No
Diseases of the immune system	Yes	No
Tumor	Yes	No
Fever	Yes	No
Cough	Yes	No
White blood cell count	No	Yes
Systolic pressure	No	Yes
Diastolic pressure	No	Yes
Temperature	No	Yes
Lactate dehydrogenase	No	Yes
Blood oxygen saturation	No	Yes
High-sensitivity C-reactive proteins	No	Yes
Blood sedimentation rate	No	Yes
D-dimers	No	Yes
Alanine aminotransferases	No	Yes
Aspartate aminotransferases	No	Yes
ESR	No	Yes
ΔChanges in lymphocyte count	No	Yes
ΔChanges in lactate dehydrogenase	No	Yes
ΔHigh-sensitivity C-reactive proteins	No	Yes
ΔAlanine aminotransferases	No	Yes
ΔAspartate aminotransferases	No	Yes

Based on an analysis from some of the authors' experiences treating COVID-19 patients on the front lines, it was noted that during the initial outbreak, inconsistencies in medical conditions led to variations in the feature data collected by different hospitals (edge nodes). For instance, community hospitals, due to limited resources, struggled to provide CT scans for patients. Consequently, building predictive models with limited feature data posed a significant challenge. This study attempted to integrate edge computing with FL, taking into account the use of different features by edge nodes for training and prediction. The dataset was divided into two edge nodes, E1 and E2, with only age and gender as overlapping features, as shown in **Table 2**. Such a division not only caters to edge nodes with varying medical capabilities but is also applicable to model training scenarios across different countries. Given that different countries often adopt varied medical standards, it was challenging to rapidly standardize all features during the early stages of the outbreak, resulting in significant feature disparities. After dividing into E1 and E2 nodes, 80% of the data was used for training the models, while the remaining 20% was allocated for testing.

Fig. 5 demonstrates the model's performance on the test set, where the "proposed model" represents the performance of the model after federated training on the global dataset, while "proposed model on E1" indicates the performance after training and prediction on the E1 dataset alone. Similarly, "proposed model on E2" signifies the model's performance after training and prediction exclusively on the E2 dataset. **Fig. 5** illustrates that, despite the presence of diverse features, the proposed model maintains performance, resulting in a substantial enhancement in the performance of the cloud-based model. Compared to training the E1 and E2 nodes separately, the F1 scores for the E1 and E2 nodes increased by 0.05 and 0.16, respectively, with a corresponding 0.03 improvement in precision. The performance of the proposed model tends to improve with the incorporation of more training data, ensuring privacy preservation. This contrasts with traditional machine learning models, which often suffer from poor generalization due to insufficient training data. In such scenarios, the performance of the proposed model surpasses that of traditional machine learning models. However, the proposed model's drawback is the bandwidth cost incurred from transmitting model parameters.

In the proposed method, the FL framework utilizes XGBoost as the local algorithm, which constructs a model comprising 80 trees, with each tree having a depth and leaf node count of 3. The resultant single model approximates a size of 300KB. Considering the architecture comprises four sub-nodes, the communication cost for a complete round of parameter updates per node is calculated as 300KB * 4 * 2, equating to 2.4MB. Factoring in the 78 iterations required for model training, the cumulative communication cost is 187.2MB. When evaluated in the context of 5G technology, which is characterized by high data transfer rates and minimal latency, this incurred cost is relatively negligible, affirming the transmission's feasibility and efficiency.

5. Conclusion

In conclusion, the study has successfully demonstrated the potential of integrating 5G technology with FL to create a robust predictive model for COVID-19 clinical outcomes. This integration not only enhances the precision of patient prognosis during hospitalization but also addresses critical privacy concerns by ensuring data remains localized, a feature that is particularly valuable in the sensitive healthcare domain.

The model's ability to generalize across diverse datasets, even in the presence of varying node features, underscores its adaptability and resilience in real-world clinical settings. This adaptability is further amplified by the use of SHAP values, which streamline the training process by identifying influential features, thus reducing computational overhead without compromising predictive precision.

The research contributes to the field of healthcare technology by providing a practical solution that can aid healthcare professionals in making informed decisions for early intervention and personalized treatment plans. This, in turn, can lead to improved patient outcomes and more efficient resource allocation during public health crises.

Furthermore, the study offers policymakers a powerful tool for developing targeted strategies to combat COVID-19 and future pandemics, by providing insights into the effectiveness of various treatment approaches based on predictive analytics.

However, the study's limitations, particularly the challenges posed by data disparities among nodes, highlight the importance of future research in data balance and training weight optimization. Looking forward, the work paves the way for more sophisticated, efficient, and patient-focused healthcare solutions, marking a significant step in the integration of 5G and FL in healthcare.

Acknowledgement

Ruochen Huang contributed to conception and design of the study and wrote the first draft of the manuscript. Zhiyuan Wei and Wei Feng made significant contributions to the experimental section and partial formatting editing. Changwei Zhang and Chen Qiu performed the statistical analysis. Yong Li conducted the process of data collection and organized the database. Mingkai Chen supervised the research activity planning and execution. All authors contributed to manuscript revision, read, and approved the submitted version.

The studies involving human participants were reviewed and approved by the Ethics Committee of the First Affiliated Hospital, Nanjing Medical University (2020-SR-106). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

References

[1] M. Ciotti, M. Ciccozzi, A. Terrinoni, W.-C. Jiang, C.-B. Wang, and S. Bernardini, "The COVID-19 pandemic," *Crit. Rev. Clin. Lab. Sci.*, vol. 57, no. 6, pp. 365–388, Aug. 2020. Article (CrossRef Link).

[2] J. S. Kim, "COVID-19 Prediction and Detection Using Machine Learning Algorithms: Catboost and Linear Regression," *Am. J. Theor. Appl. Stat.*, vol. 10, no. 5, pp. 208-215, Oct. 2021. Article (CrossRef Link).

[3] E. A. Toraih, R. M. Elshazli, M. H. Hussein, A. Elgaml, M. Amin, M. El-Mowafy, M. El-Mesery, A. Ellythy, J. Duchesne, M. T. Killackey, K. C. Ferdinand, E. Kandil, and M. S. Fawzy, "Association of cardiac biomarkers and comorbidities with increased mortality, severity, and cardiac injury in COVID-19 patients: A meta-regression and decision tree analysis," *J. Med. Virol.*, vol. 92, no. 11, pp. 2473–2488, 2020. Article (CrossRef Link).

[4] Dairi, A., Harrou, F., Zeroual, A., Hittawe, M. M., and Sun, Y., "Comparative study of machine learning methods for COVID-19 transmission forecasting," *J. Biomed. Inf.*, vol. 118, p. 103791, Jun. 2021. Article (CrossRef Link).

[5] Sujath, R., Chatterjee, J. M., and Hassanien, A. E., "A machine learning forecasting model for COVID-19 pandemic in India," *Stoch. Environ. Res. Risk A.*, vol. 34, no. 7, pp. 959-972, 2020. Article (CrossRef Link).

[6] Zoabi, Y., Deri-Rozov, S., and Shomron, N., "Machine learning-based prediction of COVID-19 diagnosis based on symptoms," *NPJ Digit. Med.*, vol. 4, no. 1, pp. 1-5, 2021. Article (CrossRef Link).

[7] Huyut, M. T., "Automatic Detection of Severely and Mildly Infected COVID-19 Patients with Supervised Machine Learning Models," *IRBM*, vol. 44 no.1, p. 100725, 2023. Article (CrossRef Link).

[8] A. H. Sodhro, A. I. Awad, J. van de Beek, and G. Nikolakopoulos, "Intelligent authentication of 5G healthcare devices: A survey," *Internet of Things*, vol. 20, p. 100610, Nov. 2022. Article (CrossRef Link).

[9] P. P. Ray, D. Dash, and D. De, "Edge computing for Internet of Things: A survey, e-healthcare case study and future direction," *J. Netw. Comput. Appl.*, vol. 140, pp. 1-22, Aug. 2019. Article (CrossRef Link).

[10] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated Learning: Challenges, Methods, and Future Directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50-60, May 2020. Article (CrossRef Link).

[11] F. S. Heldt, M. P. Vizcaychipi, S. Peacock, M. Cinelli, L. McLachlan, F. Andreotti, S. Jovanović, R. Dürichen, N. Lipunova, R. A. Fletcher, A. Hancock, A. McCarthy, R. A. Pointon, A. Brown, J. Eaton, R. Liddi, L. Mackillop, L. Tarassenko, and R. T. Khan, "Early risk assessment for COVID-19 patients from emergency department data using machine learning," *Sci. Rep.*, vol. 11, no. 1, p. 4200, Feb. 2021. Article (CrossRef Link).

[12] V. Singh, R. C. Poonia, S. Kumar, P. Dass, P. Agarwal, V. Bhatnagar, and L. Raja, "Prediction of COVID-19 corona virus pandemic based on time series data using support vector machine," *J. Discrete Math. Sci. Cryptogr.*, vol. 23, no. 8, pp. 1583-1597, Nov. 2020. Article (CrossRef Link).

[13] W. C. Roda, M. B. Varughese, D. Han, and M. Y. Li, "Why is it difficult to accurately predict the COVID-19 epidemic?," *Infect. Dis. Model.*, vol. 5, pp. 271-281, 2020. Article (CrossRef Link).

[14] M. S. Hossain, G. Muhammad, and N. Guizani, "Explainable AI and mass surveillance system-based healthcare framework to combat COVID-I9 like pandemics," *IEEE Netw.*, vol. 34, no. 4, pp. 126-132, Jul. 2020. Article (CrossRef Link).

[15] M. R. Islam and M. Nahiduzzaman, "Complex features extraction with deep learning model for the detection of COVID19 from CT scan images using ensemble based machine learning approach," *Expert Syst. Appl.*, vol. 195, p. 116554, Jun. 2022. Article (CrossRef Link).

[16] S. Aslani and J. Jacob, "Utilisation of deep learning for COVID-19 diagnosis," *Clin. Radiol.*, vol. 78, no. 2, pp. 150-157, Feb. 2023. Article (CrossRef Link).

[17] M. Chen, Y. Hao, L. Hu, M. S. Hossain, and A. Ghoneim, "Edge-CoCaCo: Toward joint optimization of computation, caching, and communication on edge cloud," *IEEE Wirel. Commun.*, vol. 25, no. 3, pp. 21-27, Jun. 2018. Article (CrossRef Link).

[18] Alshammari, N., Sarker, M. N. I., Kamruzzaman, M. M., Alruwaili, M., Alanazi, S. A., Raihan, M. L., and AlQahtani, S. A., "Technology-driven 5G enabled e-healthcare system during COVID-19 pandemic," *IET Commun.*, vol. 16, no. 5, pp. 449-463, 2022. Article (CrossRef Link).

[19] Lin, X., Wu, J., Bashir, A. K., Yang, W., Singh, A., and AlZubi, A. A., "FairHealth: Long-Term Proportional Fairness-Driven 5G Edge Healthcare in Internet of Medical Things," *IEEE Trans. Ind. Inform.*, vol. 18, no. 12, pp. 8905-8915, 2022. Article (CrossRef Link).

[20] Qureshi, H. N., Manalastas, M., Ijaz, A., Imran, A., Liu, Y., and Al Kalaa, M. O, "Communication Requirements in 5G-Enabled Healthcare Applications: Review and Considerations," *Healthcare*, vol. 10, no. 2, p. 293, 2022. Article (CrossRef Link).

[21] A. Ahad, M. Tahir, M. A. Sheikh, K. I. Ahmed, A. Mughees, and A. Numani, "Technologies trend towards 5G network for smart health-care using IoT: A review," *Sensors*, vol. 20, no. 14, p. 4047, Jan. 2020. Article (CrossRef Link).

[22] Pham, Q.-V., Fang, F., Ha, V. N., Piran, M. J., Le, M., Le, L. B., Hwang, W.-J., and Ding, Z, "A Survey of Multi-Access Edge Computing in 5G and Beyond: Fundamentals, Technology Integration, and State-of-the-Art," *IEEE Access*, vol. 8, pp. 116974-117017, 2020. Article (CrossRef Link).

[23] McEnroe, P., Wang, S., and Liyanage, M., "A Survey on the Convergence of Edge Computing and AI for UAVs: Opportunities and Challenges," *IEEE Internet of Things J.*, vol. 9, no. 17, pp. 15435-15459, 2022. Article (CrossRef Link).

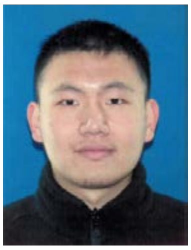
[24] A. M. Peralta-Ochoa, P. A. Chaca-Asmal, L. F. Guerrero-Vásquez, J. O. Ordoñez-Ordoñez, and E. J. Coronel-González, "Smart healthcare applications over 5G networks: A systematic review," *Appl. Sci.*, vol. 13, no. 3, p. 1469, Jan. 2023. Article (CrossRef Link).

[25] Zhan, Y., Li, P., Qu, Z., Zeng, D., and Guo, S, "A Learning-Based Incentive Mechanism for Federated Learning," *IEEE Internet of Things J.*, vol. 7, no. 7, pp. 6360-6368, 2022. Article (CrossRef Link).

[26] Wang, Z., Xiao, J., Wang, L., and Yao, J, "A novel federated learning approach with knowledge transfer for credit scoring," *Decis. Support Syst.*, vol. 177, p. 114084, 2024. Article (CrossRef Link).

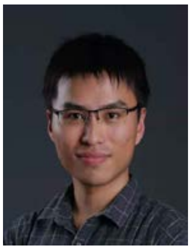
[27] Zeng, P., Liu, A., Xiong, N. N., Zhang, S., and Dong, M., "TD-MDB: A Truth-Discovery-Based Multidimensional Bidding Strategy for Federated Learning in Industrial IoT Systems," *IEEE Internet of Things J.*, vol. 11, no. 3, pp. 4274-4288, 2024. Article (CrossRef Link).

[28] Guo, J., Wu, J., Liu, A., and Xiong, N. N., "LightFed: An Efficient and Secure Federated Edge Learning System on Model Splitting," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 11, pp. 2701-2713, 2022. Article (CrossRef Link).
[29] Savazzi, S., Nicoli, M., and Rampa, V., "Federated Learning With Cooperating Devices: A Consensus Approach for Massive IoT Networks," *IEEE Internet of Things J.*, vol. 7, no. 5, pp. 4641-4654, 2020. Article (CrossRef Link).
[30] T. S. Brisimi, R. Chen, T. Mela, A. Olshevsky, I. Ch. Paschalidis, and W. Shi, "Federated learning of predictive models from federated electronic health records," *Int. J. Med. Inform.*, vol. 112, pp. 59-67, Apr. 2018. Article (CrossRef Link).
[31] Zhu, J., Cao, J., Saxena, D., Jiang, S., and Ferradi, H., "Blockchain-empowered Federated Learning: Challenges, Solutions, and Future Directions," *ACM Comput. Surv.*, vol. 55, no. 11, pp. 1-31, 2023. Article (CrossRef Link).



Ruochen Huang received his B.E. degree in communication engineering and Ph.D degree in information and communication systems from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2012 and 2019. Now, he is lecturer at the department of medical informatics, school of biomedical engineering and informatics, Nanjing Medical University, Nanjing, China. His current research interests include large language model, machine learning and natural language processing.

Zhiyuan Wei received the B.S. degree in Computer Science from Nanjing University of Posts and Telecommunications, Nanjing City, China, in 2015. From 2016 to 2019, he served as a Data Structure Engineer at Headquartered of Suning Holdings Group in Nanjing, Jiangsu, China. Since 2020, he been pursuing his Ph.D. degree in biomedical engineering at Nanjing Medical University, Nanjing, Jiangsu, China. His research focus on clinical big data analysis and comparative effectiveness research.



Wei Feng received the B.S. degree in information management from China Pharmaceutical University, Nanjing, China, in 2011. He is currently pursuing the M.S. degree in biomedical engineering at Nanjing Medical University, Nanjing, Jiangsu, China. From 2016 to 2019, he was a Software Engineer for Oracle China in Suzhou, Jiangsu, China. Since 2019, he has been a postgraduates at Nanjing Medical University.

Yong Li received his Ph.D. degree in the field of internal medicine from Nanjing Medical University between 2015 and 2018. Currently, he serves as the Deputy Director, Associate Professor, and Postdoctoral Researcher at the Department of Cardiovascular Medicine, Chongqing Hospital of Jiangsu Province People's Hospital. His current research interests include the application of artificial intelligence in the medical domain.

Changwei Zhang (Member, IEEE) received his M.S. degree in electronics and communication engineering from the Nanjing University of Posts and Telecommunications (NJUPT), Nanjing, China, in 2018, the first Ph.D. degree in communication and information system from NJUPT, in 2021, and the second Ph.D. degree in electrical and computer engineering from the Western University, London, Canada, in 2023. He is currently a Postdoctoral Researcher with the Purple Mountain Laboratories. His research interests include the Internet-of-Things, random access, machine learning, and mmWave communications.

Chen Qiu received the Ph.D. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2020 and the B.E. degree from Beijing University of Posts and Telecommunications, Beijing, China and Queen Mary University of London, London, England in 2013. He was a visiting scholar in the Department of Electrical and Computer Engineering, University of California at Davis (2016-2017) and Department of Electrical and Computer Engineering, Texas A&M University (2015-2016). He is currently a postdoctoral researcher in Peng Cheng Laboratory, Shenzhen, China. His research interests include UAV-assisted wireless network, integrated sensing and communication and machine learning in wireless systems.

Mingkai Chen received his Ph.D. degree in information and communication engineering from Nanjing University of Posts and Telecommunications, China in 2019. He is currently an Associate Professor at Nanjing University of Posts and Telecommunications, China. His research interests are in the area of multimedia communication, semantic communication, resource allocation, and signal processing in wireless communications.