Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

A Robust Big Data Analytic Framework for IoMT Leveraged by Federated Learning

Dr.S.Dhivya[1], Harshini R[2], Nivetha N[3], Ali Bostani[4], Ponmurugan Panneerselvam[5], R. Naveenkumar[6], Sai Krishna Edpuganti[7]

[1]Assistant Professor, Department of commerce, SRM Institute of Science and Technology, Ramapuram ,Bharathi salai Chennai -89, Email: dhivyas1@srmist.edu.in, Orchid ID- 0009-0003-1555-7952
[2]Assistant Professor, Department of Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Email: harshinir@maher.ac.in
[3]Assistant Professor, Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Email: nivethan@maher.ac.in
[4]Associate Professor, College of Engineering and Applied Sciences, American University of Kuwait, Salmiya, Kuwait, Email: abostani@auk.edu.kw, 0000-0002-7922-9857
[5]Professor & Dean-Doctoral Studies & IPR, Department of Research, Meenakshi Academy of Higher Education and Research, Chennai, Tamilnadu, India, Email: ponmurugan@maher.ac.in, 0000-0003-2212-8219
[6]Dept of CSE, School of Engineering and Technology, CGC University Mohali-140307, Punjab India, drnk1983@gmail.com, 0000-0001-9033-9400
[7]Assistant Professor,Department of Computer Science & Engineering,Koneru Lakshmaiah Education Foundation,Vaddeswaram, India, Email: saikrishnaedpuganti@gmail.com

Abstract

With the fast development of the Internet of Medical Things, continuous health monitoring using interconnected smart-medical devices became possible, which produces large amounts of heterogeneous data. Nevertheless, the centralization of the processing of such data also presents significant issues concerning the scalability, latency, and privacy of data. This paper will offer a new model to overcome these shortcomings by proposing a new model known as FL-BDA-IoMT (Federated Learning-driven Big Data Analytics IoMT). The suggested algorithm combines distributed Federated Learning with scalable big data analytics that would allow privacy-sensitive, decentralized model training on various healthcare nodes. With the FL-BDA-IoMT, IoMT devices process and train models on sensitive medical data locally, and only model parameters are exchanged with a central aggregator, this way removing the sharing of raw data. The hierarchical structure with edge (fog) computing and cloud layers is used to decrease the overhead and latency in the communication. Adaptive aggregation, resource-conscious client selection are also proposed as part of the suggested approach to address the issue of data heterogeneity and client device limitations. The findings of the experiment demonstrate that the suggested solution improves predictive performance, minimizes the use of network bandwidth and achieves privacy requirements in contrast to conventional centralized machine learning methods. Moreover, with the incorporation of big data analytics, it is now possible to effectively manage the high-velocity and high-volume medical streams, which are critical to real-time decision-making within the healthcare system. The FL-BDA-IoMT algorithm offers an efficient, scalable and privacy-conscious system to next-generation intelligent healthcare analytics, which covers key issues in the IoMT setting.

Keywords: Internet of Medical Things (IoMT), Big Data Analytics, Federated Learning, Healthcare IoT, Privacy-Preserving Machine Learning, Edge and Fog Computing, Distributed Learning Systems, Medical Data Mining, Smart Healthcare Systems, Clinical Decision Support

1. Introduction

Internet of Things (IoT) has numerous physical objects that are connected through network whereby the objects makes decision by perceiving the situation of the atmosphere [1]. The objects reside in the IoT network acts as intelligent objects and the every process is automated. The main intent of IoT is to automate the work without intervention of humans. IoT is widely used in the applications namely industry, business, education, transportation and healthcare [2]. IoT technology is prominently used in medical and healthcare analytics also emerged due to the progression of Internet of Medical Things (IoMT) [3].

With the fast development of lightweight communication protocols, smart sensors, and embedded medical devices, the interconnection of healthcare systems has become very smooth, which has led to the Internet of Medical Things. This paradigm enables continuous observation of the biomedical signals with the help of interrelated smart medical objects that provide an opportunity to diagnose diseases and administer health independently of the medical specialist without direct intervention. IoMT is changing the traditional healthcare to an intelligent, real-time, and patient-centric system through the combination of diagnosis, treatment, and preventive care. This kind of automation has a major role to play in enhancing the individual health outcomes and general quality of life [4].

One of the main benefits of the IoMT-based healthcare system is that it allows conducting constant monitoring of the patient. It is possible to remotely monitor patients in real time and avoid making them too dependent on visiting hospitals and healthcare expenses are minimized. Moreover, IoMT systems allow sending timely alerts and notifications to patients and healthcare providers, which allow medical intervention to be proactive and early-stage. Nevertheless, the ongoing sensing and monitoring activities produce huge amounts of heterogeneous medical data, which are sent and stored on network infrastructures. This proliferation of data is an exponential one that requires sophisticated computational structures to process and analyse efficient data.

In order to manage such difficulties, Big Data Analytics has become an essential element of IoMT-based medical systems. The Big Data Analytics (BDA) aids in storage, processing as well as analysis of high-dimensional and high-velocity medical data. BDA identifies meaningful patterns and insights with the help of advanced analytical models and algorithms and can be employed to enhance the process of clinical decision-making. Moreover, IoMT systems must be able to operate in extremely heterogeneous settings, which are supported by different communication protocols and distributed architecture, which also explains the need to have scalable and efficient data analytics solutions.

Despite its effectiveness, big data-based healthcare analytics has several challenges, including the heterogeneity of data, storage complexity, complexity of computer processing, and, most of all, the problem of privacy and security [5]. The traditional machine learning approaches are premised on the notion of centrally collected data, according to which the data of different devices or institutions are initially processed in one central server or cloud to be trained into predictive models. This kind of centralized paradigm contributes to the overhead of communication not to mention that sensitive medical information is at risk of security breaches and an invasion of privacy.

To minimize the impact of these issues in the context of the IoMT-based healthcare analytics, Federated Learning has been suggested as a decentralized learning paradigm. Federated Learning (FL) allows the joint training of models on many distributed devices without the need to pass raw data. Instead, each device trains a local model on the data local to it and transfers model parameters or gradients to an aggregate server located centrally. This will ensure that the sensitive information on patients is localized and therefore ensuring privacy and complying with the regulatory stipulations.

In contrast to classic machine learning techniques, federated learning separates the storage of data and the training of the models so that the computation could be done at the network periphery. It minimizes the cost of communication, latency and improves scalability of the system. Google has greatly advanced the concept by creating a secure cloud-based platform of decentralized model training. Within this context, devices involved in communication all update a global model in a cycle of submitting locally trained parameters which are combined to enhance the overall model performance.

Moreover, federated learning solves the problems of non-homogeneous data distributions and capabilities of devices, which can be typical of the IoMT setting. It assists in distributed and adaptive learning processes that are capable of functioning effectively in a variety of medical data and resource-limited devices. The absence of centralized data storage will mean that federated learning will greatly lessen the possibility of data leakage and increase the credibility of healthcare systems.

To conclude, IoMT, Big Data Analytics and Federated Learning integration forms a strong and scalable model of smart healthcare systems. IoMT can be used to achieve continuous data collection, Big Data Analytics can be used to perform efficient data processing and knowledge extraction, and Federated Learning can be used to train models that can operate without privacy violations in a distributed fashion. All these technologies are solutions to key issues of data volume, heterogeneity, scalability, and security [6].

The rest of this paper follows the following structure. Section 2 addresses the importance and the ideals of federated learning in the context of healthcare analytics based on the IoMT. Section 3 will introduce some applications and the current works concerning medical IoT technologies. Section 4 outlines the suggested big data analytic system combined with federated learning to operate in an IoMT setting. Lastly, Section 5 summarizes the paper and provides the opportunities of further research.

2. Federated Learning in IoT

The field of machine learning has transformed healthcare with the use of big medical data and the best methods of analysis, and the Internet of Things is one of the primary sources of data creation. The IoT-based systems are used in applications like real-time monitoring, fitness tracking, elderly care, and chronic disease management, which always allow the collection of biomedical signals and allow personalized healthcare. Nonetheless, the conventional centralized methods of storing and processing IoT data pose serious privacy and safety issues because the information on the medical sphere is sensitive. In order to address these shortcomings, Federated Learning is embraced, that is, local model training is allowed on devices, only parameters are exchanged between the devices and the server, supporting privacy-sensitive and safe healthcare analytics.

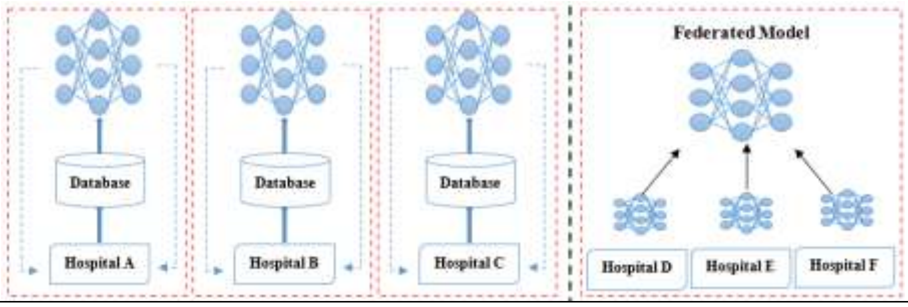

Figure 1. ML vs FL Architecture in IoMT.

Federated Learning is not exactly the same as conventional distributed learning in a number of fundamental ways. On the one hand, federated learning is concerned with user data privacy, and the sensitive information is kept locally on the client machines, unlike distributed learning, which is mainly concerned with how to enhance computational efficiency and convergence rate. Second, in comparison with distributed learning, federated learning does not reveal or describe the allocation of data to client devices: in a distributed learning, the dataset can be divided and distributed between the nodes. This is a natural privacy-saving design, which makes federated learning especially appropriate in sensitive fields of activity, including healthcare.

Federated learning is not as simple and dynamic as distributed learning. It encompasses a huge population of non-homogeneous clients, including mobile devices and edge nodes, and can be characterized by different computational capacity, amount of data, and intermittent connectivity. Such gadgets can often malfunction causing asynchronous involvement in the training process. Moreover, unlike distributed learning, federated learning does not presuppose the existence of a full world data set and it is difficult to describe data distribution in a homogenous way.

Federated learning involves the training process which is iterative and collaborative in nature. A global model is firstly established at the central server and is sent to the client devices involved. The model is then trained with its own data in a few epochs by each client. Upon local training, instead of raw data clients only transfer the updated model parameters or gradients to the central server. Because of the heterogeneity of the devices and limitations in communication, client updates may come in asynchronous.

The server aggregates these updates—either synchronously or asynchronously—to refine the global model. This process of local training and global aggregation is repeated iteratively until the model converges. Overall, federated learning enables scalable, privacy-preserving, and distributed model training, addressing key limitations of traditional centralized and distributed learning approaches.

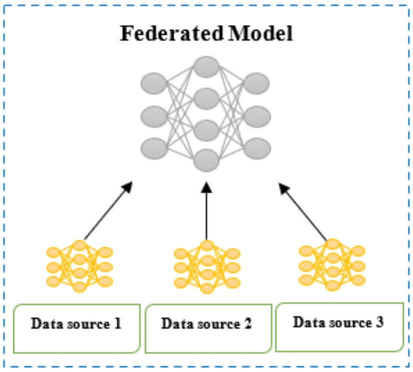

Figure 2. Overall Process of Federated Learning

A mathematical formulation of Federated Learning can be constructed to highlight its distinction from traditional distributed learning while formalizing its optimization process. Consider a set of Mclients (data owners) denoted as $\{1,2,\ldots,M\}$, where each client $j$possesses a local dataset $L_j$that remains private and is not shared with any central entity. Unlike distributed learning, where a global dataset is partitioned and distributed across nodes, federated learning assumes that data is inherently decentralized and non-identically distributed (non-IID), i.e., $L_i \neq L_j$for $i \neq j$, and the global data distribution is unknown.

Let $w \in \mathbb{R}^d$represent the model parameter vector. The objective of federated learning is to minimize a global loss function defined as an aggregation of local objective functions:

$$\min_{w} F(w) = \sum_{j=1}^{M} \frac{| L_j |}{\sum_{k=1}^{M} | L_k |} \cdot F_j(w)$$

where $F_j(w)$denotes the local loss function computed on dataset L_j, typically expressed as:

$$F_j(w) = \frac{1}{| L_j |} \sum_{x_i \in L_j} \ell(w; x_i)$$

Here, $\ell(w; x_i)$is the loss incurred on sample x_i. This formulation reflects horizontal federated learning, where all clients share the same feature space but differ in sample distributions.

The federated optimization process is iterative and proceeds in communication rounds indexed by t. At each round, the central server broadcasts the global model $w^{(t)}$to a subset of available clients $S_t \subseteq \{1,\ldots,M\}$. Each selected client performs local training using stochastic gradient descent (SGD) for Elocal epochs:

$$w_j^{(t+1)} = w^{(t)} - \eta \nabla F_j(w^{(t)})$$

where $\eta$is the learning rate. Due to heterogeneity in computation and connectivity, not all clients participate in every round, leading to asynchronous and partial updates.

After local training, each participating client sends only its updated model parameters $w_j^{(t+1)}$(or gradients) to the central server, preserving data privacy. The server then performs aggregation, commonly using the Federated Averaging (FedAvg) algorithm:

$$w^{(t+1)} = \sum_{j \in S_t} \frac{|L_j|}{\sum_{k \in S_t} |L_k|} \cdot w_j^{(t+1)}$$

This weighted aggregation ensures that clients with larger datasets have a proportionally greater influence on the global model. The process repeats until convergence, i.e., $\| w^{(t+1)} - w^{(t)} \| \leq \epsilon$.

In contrast to traditional distributed learning, where gradients are computed on partitioned datasets and synchronized frequently, federated learning reduces communication overhead by performing multiple local updates before aggregation. Additionally, the absence of raw data transmission ensures privacy preservation, which can be further enhanced using techniques such as differential privacy and secure aggregation.

The optimization problem can also be expressed in a compact form as:

$$\min_{w} \sum_{j=1}^{M} k_j \left(w \mid L_j \right)$$

where $k_j(w \mid L_j)$represents the empirical risk over client j. The goal is to find an optimal parameter w^*such that:

$$w^* = \arg \min_{w} \sum_{j=1}^{M} k_j \left(w \mid L_j \right)$$

This formulation emphasizes that federated learning constructs a global model by integrating locally trained models without centralizing data, thereby ensuring privacy, scalability, and robustness in heterogeneous environments such as IoT-based healthcare systems.

3. Medical IoT Applications

The elevated progression of federated learning technology is believed to enhance smart medicine, a future healthcare development. In the past, there were insufficient instances for machine learning due to the independence of hospitals and the confidentiality of patient information.Patient data is typically safeguarded in regional silos, thanks to the Personalized Medicine Initiative and the creation of a vast volume of general health electronic information. This renders establishing valid medical assistance in diagnosing that is challenging. It's critical to have a dependable medical assistant in diagnosing system. The performance of a model can be considerably improved by establishing a horizontal federated learning model. This enables tiny hospitals to provide a higher degree of diagnosis, lowering direct patient expenditures. There is, meanwhile, a growing demand to combine datasets from various medical services.Because the limitations of locally developing a calibration system may restrict the level of effectiveness.

The distance correlation in between original information and the intermediate presentation is minimised. While retaining model correctness can minimizes the leaking of important patterns in raw data during client contact. It minimises transmission of original data and payload leakage in medical information. A FL framework for multi-site functional imaging analysis with privacy protection is ensured [9, 10]. The medical applications based on FL is give in Table 1.

Table 1. Medical Applications Leveraging Federated Learning

Category	Significant Inference	Framework
Medical Imaging	FL paradigm evaluated across five models; ResNet18 shows superior performance	FL
Medical Imaging	Variation-aware FL (VAFL) improves prostate cancer classification	FL
Medical Information	FL evaluated on two datasets; does not significantly improve logistic regression accuracy	FL
Medical Information	Multi-hospital data used to evaluate models for death prediction	FL
Medical Information	FL demonstrates robustness under skewed and heterogeneous data distributions	FL
Medical Information	Privacy-preserving medical Named Entity Recognition (NER) developed using FL	FL

Processing of Medical Information	Dynamic model fusion based on client training time; applied to COVID-19 diagnostic image datasets	FL
Processing of Medical Information	Knowledge extraction approach reduces communication bottleneck; validated on three datasets	FL

4. Federated Learning-Enabled Big Data Analytics Framework for IoMT

Big Data Analytics plays a critical role in modern healthcare by enabling disease prediction, epidemic management, reduction of preventable mortality, and overall improvement in quality of life. In IoMT-enabled environments, large volumes of heterogeneous data—including structured, semi-structured, and unstructured data—are continuously generated from wearable sensors, environmental sources, medical records, and contextual parameters such as location and climate. These data are processed, visualized and aggregated with the help of cloud-based structures where preprocessing stages including data retrieval, data cleaning and data statistical analysis are conducted and results of the entire process are delivered to healthcare providers, remote users and emergency services.

The wearable IoT gadgets constantly measure physiological values, and the resulting large volumes of data become the subject of big data. Nonetheless, it is still a complicated process to pull out meaningful information on such high volume and high velocity information especially when it needs to be used in real time decision-making. Despite the ability to have scalability in the storage of data and computing power, cloud computing presents the issue of latency caused by the transmission of data over a distance. Mog computing is incorporated to overcome this weakness to have an intermediate between the IoT devices and the cloud to provide localized data processing, reduce latency, and response time. Processed data is then sent to the cloud to be stored and create sophisticated analytics in the long run. Continuous remote monitoring is essential for chronic disease management, including conditions such as heart disease, hypertension, and diabetes. Overall, the integration of IoMT with big data analytics and fog-cloud architectures provides a scalable and efficient framework for intelligent, real-time healthcare delivery, while also addressing challenges related to data management and system responsiveness.

As the number of patients grows, so does the problem of scalability, or deciding who to service first is also increased. They take population ageing and disaster into mind when providing adequate telemedicine services. Prioritization of patients takes place in three areas namely transplantation, surgery, and accessing the operating room. To tackle the problem, triage is utilised to determine the order in which patients should be treated based on their severity. Big Data analytics and IoT play a critical part in the smart health service in the emerging countries.

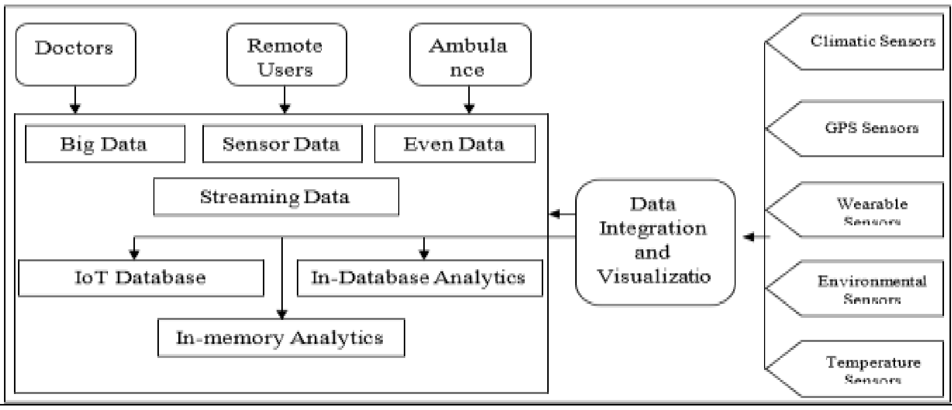

Figure 3. IoMT Architecture based on big data

Handling big data in healthcare is a major problem especially in real-time contexts when making decisions. Such aspects of telemedicine settings as patient prioritization must be dealt with efficiently to address such issues as the six Vs of big data (volume, velocity, variety, veracity, value, and variability).An IoT-based healthcare system is structured into three integrated layers that collectively enable efficient data acquisition, processing, and decision-making. The body sensor area network is the first layer that uses wearable or embedded sensors to constantly record physiological data on patients as well as environmental conditions which include temperature, humidity, and time. The second tier consists of a fog/edge-based smart gateway, which is used to conduct intermediate functionalities such as data preprocessing, protocol transformation, data mining, filtering as well as real time alert generation of patients. The third layer is the cloud-based big data infrastructure which is a repository of big sensor data, advanced analytics, prediction of clinical outcomes and emergency alerts to caregivers in case of need.

As the Electronic Health Records (EHR) have been expanded rapidly, the large-scale data analysis based on the Big Data Analytics is now a necessity in the healthcare decision-making process. Unless there is a solution to the rising patient volumes and enhanced clinical throughput, other models like Simultaneously Aided Diagnostic Model (SADM) have been developed. The model incorporates various steps such as data gathering, data preprocessing, data storage, feature extraction, performance measurement and machine learning based prediction and disease, which would help clinicians in diagnosis and would decrease workload. On the whole, the IoT-based healthcare systems offer a whole of real-time monitoring and smart decision support systems, as well as provide an insight into the system-level benefits and drawbacks.

Table 2. Advantages and Limitations of IoT-Based Healthcare Systems

Application Type	Advantages	Disadvantages
Continuous observation of ICU patients	Assists doctors in making accurate clinical decisions	Monitoring system is not fully utilized and accuracy remains unvalidated
Study about the healthcare observation system	Optimized energy usage reduces power gaps and meets energy requirements	IoT sensors consume high energy
Prevention and diagnosis system to regulate chikungunya virus	Enables real-time feedback system	Security aspects are imbalanced
Health observation system	Provides efficient healthcare services	Limited fault tolerance in data handling
Prevention and diagnosis of chikungunya virus	Rapid alert generation with high bandwidth support	IoT sensors consume high energy

In the data collection stage, different outpatient data such as medication records, treatment plans, costs, clinical outcomes, diagnostic reports, and medical images are collected and grouped together to facilitate a thorough analysis. Machine algorithms like Support Vector Machines and neural networks are used on historical data to perform activities like hyperlipidemia classification. Within this context, the Big Data Analytics is instrumental in the efficient management of large and diverse medical data through the ability to store, retrieve, update and manage medical data on-demand and their lifecycle.

5. Conclusion

This article identifies the importance of the Internet of Things in contemporary healthcare and especially in the development of the Internet of Medical Things where interconnected medical devices are constantly producing immense amounts of data. The sheer size and dissimilarity of this information require the use of Big Data Analytics that will provide an efficient storage, processing, and analysis. Nonetheless, the transfer of sensitive medical information through networks poses essential issues of data integrity, privacy and security.

Federated Learning is part of healthcare analytics based on big data to deal with these concerns. Federated learning, in contrast to conventional machine learning methods, makes it possible to train models in a decentralized manner, with individual devices constructing local models and transmitting only model parameters but not raw data. This is then combined into a global federated model composed of these local models which protects privacy and safeguards data. In general, the paper proves that the integration of IoMT, big data analytics, and federated learning is an effective and scalable system to achieve secure and intelligent healthcare analytics.

Reference

1. Ahad, A., Ahmed, K. I., Ullah, F., Sheikh, M. A., Tahir, M., Hayajneh, M., & Pires, I. M. (2026). Federated Learning and 5G/6G-Based Internet of Medical Things (IoMT): Applications, Key Enabling Technologies,

Open Issues and Future Research Directions. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 16(1), e70065.

2. Annappa, B., Hegde, S., Abhijit, C. S., &Ambesange, S. (2024). Fedcure: A heterogeneity-aware personalized federated learning framework for intelligent healthcare applications in iomt environments. IEEE Access, 12, 15867-15883.
3. Alamleh, A., Albahri, O. S., Zaidan, A. A., Albahri, A. S., Alamoodi, A. H., Zaidan, B. B., ... & Jasim, A. N. (2022). Federated learning for IoMT applications: A standardization and benchmarking framework of intrusion detection systems. IEEE Journal of Biomedical and Health Informatics, 27(2), 878-887.
4. Tharini, V. J., & Vijayarani, S. (2020). IoT in Healthcare: Ecosystem, Pillars, Design Challenges, Applications, Vulnerabilities, Privacy, and Security Concerns. In Incorporating the Internet of Things in Healthcare Applications and Wearable Devices (pp. 1-22). IGI Global.
5. Alkhalifa, A. K., Alanazi, M. H., Mahmood, K., Almukadi, W. S., Qurashi, M. A., Alshehri, A. H., ... & Mohamed, A. A. (2024). Harnessing Privacy-Preserving federated learning with blockchain for secure Iomt applications in smart healthcare systems. Fractals, 32(09n10), 2540020.
6. Almogadwy, B., &Alqarafi, A. (2025). Integrating IoMT and Federated Learning for Advanced Healthcare Monitoring in Healthcare 5.0. Current Pharmaceutical Biotechnology.
7. Ramani, R., Mary, A. R., Raja, S. E., &Shunmugam, D. A. (2024). Optimized data management and secured federated learning in the Internet of Medical Things (IoMT) with blockchain technology. Biomedical Signal Processing and Control, 93, 106213.
8. Islam, U., Ullah, H., Khan, N., Ahmad, I., & Saleem, K. (2025). Adaptive Federated Learning Framework for Privacy-Preserving Consumer-Centric IoMT: A Novel Secure Data Collaboration Model. IEEE Transactions on Consumer Electronics.
9. Olaitan, S. (2025). Federated Learning Models for Privacy-Preserving Patient Data Analytics in IoT-Based Healthcare Systems.
10. Misbah, A., Sebbar, A., & Hafidi, I. (2025). Securing Internet of Medical Things: An Advanced Federated Learning Approach. International Journal of Advanced Computer Science & Applications, 16(2).
11. Khan, A. A., Mahendran, R. K., Ullah, F., Ali, F., Alghamdi, N. S., AlZubi, A. A., & Kwak, D. (2025). Fed-iomt-block: A privacy-preserving framework for secure federated learning in consumer-centric internet of medical things. IEEE Transactions on Consumer Electronics.
12. Haitham M. Snousi, Fateh A. Aleej, M. F. Bara, Ahmed Alkilany. (2026). Design and Implementation of an Energy-Efficient AI Accelerator Architecture for Edge-Based Embedded VLSI Platforms. Progress in AI-Accelerated VLSI Systems, 22–31.
13. K.L Mrinh. (2026). Design of a Grid-Forming Battery Energy Storage System for Voltage and Frequency Stabilization. Transactions on Energy Storage Systems and Innovation , 30-36.
14. Sheela. V. (2026). Mathematical Modeling and Simulation of Nonlinear Wave Propagation in Complex Media. Frontiers in Mathematical and Computational Research, 28-38.