FedEmo: A Federated Learning Framework for Privacy-Preserving Emotion Detection from Handwriting on Consumer IoMT Devices

Zohaib Ahmad Khan, Yuanqing Xia, *Fellow, IEEE*, Weiwei Jiang, *Member, IEEE*, Muhammad Shahid Anwar, Senior Member, IEEE

***Abstract*—Emotion detection from handwriting and drawing presents a promising yet underexplored avenue for scalable mental health monitoring. This is particularly relevant within consumer-centric Internet of Medical Things (IoMT) ecosystems, where privacy and cross-institutional data sharing remain critical challenges. This paper proposes FedEmo, a privacy-preserving federated learning framework that leverages an attention-based transformer model to analyze handwriting and drawing samples on edge devices, while adhering to Health Insurance Portability and Accountability Act (HIPAA) and General Data Protection Regulation (GDPR) regulations. The model processes stroke-level features (e.g., pen pressure, speed, and direction during handwriting or drawing tasks), which are key indicators of emotional states, through self-attention mechanisms, achieving 92.64% accuracy on the EMOTHAW dataset under centralized training. A federated protocol enables distributed model refinement without sharing raw data, maintaining 87.3% accuracy in simulated non-Independent and Identically Distributed (non-IID) settings, consistent with existing federated learning benchmarks. The framework introduces a hybrid cloud-edge deployment that reduces communication bandwidth by 58% through local embedding computation, and supports a clinician alert system with a modeled end-to-end latency of 620ms. Experimental results confirm the system's robustness under typical IoMT constraints, including 15% packet loss and 100kbps bandwidth. FedEmo offers a scalable, privacy-compliant solution for real-time emotion recognition and remote mental health diagnostics using consumer-grade IoMT devices, with potential applications in telepsychiatry and early screening for depression and Parkinson's disease.**

***Index Terms*—Federated Learning, Emotion Detection, Consumer IoMT, Privacy-Preserving Systems, Real-Time Diagnostics.**

I. INTRODUCTION

THE integration of IoMT technologies into consumer healthcare has transformed the landscape of remote patient monitoring and diagnostics [1]. As the adoption of IoMT-enabled devices such as smart pens, tablets, and mobile health platforms accelerates, there is a growing demand for scalable and privacy-preserving approaches capable of analyzing behavioral and emotional data for mental health assessment [2]. Emotion detection from handwriting and drawing has recently gained attention as a non-invasive, passive monitoring strategy that offers early insights into conditions such as depression and Parkinson's disease. Despite its potential, key challenges persist, including the need to preserve data privacy, support cross-institutional collaboration, and enable real-time analysis across heterogeneous devices. Federated learning (FL) has emerged as a promising paradigm that addresses these concerns by allowing distributed model training without the exchange of raw data, making it particularly suitable for secure, consumer-centric IoMT environments [3].

However, despite its potential, several challenges remain in the effective deployment of emotion detection within consumer IoMT systems [4]. One of the primary concerns is data privacy, particularly when dealing with sensitive patient information across multiple healthcare institutions [5]. FL, while offering a solution for decentralized training, introduces its own set of challenges, such as non-IID data distributions across different sources, which can affect model performance [6]. Additionally, the scalability of these systems, bandwidth usage, and real-time processing for timely clinician alerts pose significant hurdles. To make emotion detection viable for large-scale healthcare IoMT applications, solutions must address these challenges while maintaining high accuracy and low latency [7].

The growing demand for scalable, remote healthcare solutions [8] has highlighted the importance of emotion detection as a promising method for monitoring mental health conditions such as depression and Parkinson's disease. Emotion analysis through IoMT-enabled devices can offer personalized care, enabling early intervention and significantly improving patient outcomes [9]. However, existing systems face challenges such as data privacy concerns, limited scalability, and the need for real-time processing in resource-constrained environments [10]. The motivation behind this work is to address these challenges by developing a solution that ensures privacy, scalability, and real-time performance in healthcare IoMT systems. To achieve this, this paper proposes FedEmo, a privacy-preserving FL framework for emotion detection. By leveraging FL, the model enables collaborative training across healthcare institutions without sharing sensitive patient data, ensuring both privacy and scalability. This approach provides a practical solution for real-time emotion detection and remote diagnostics in mental health care, reducing bandwidth usage and latency while maintaining high accuracy. This paper presents the following key contributions:

- Proposes a privacy-preserving FL framework (FedEmo) for emotion detection in consumer IoMT systems, uti-

Manuscript received xxx xx, 202x; revised xxxx xx, 202x. This work was supported by the National Natural Science Foundation of China under Grant 61836001. (Corresponding author: *Yuanqing Xia*)

Zohaib Ahmad Khan is with the School of Automation, Beijing Institute of Technology, Beijing 100081, China. (e-mail: 3820192061@bit.edu.cn).

Yuanqing Xia is with Zhongyuan University of Technology, Zhengzhou, Henan Province 450007, China, and also with the School of Automation, Beijing Institute of Technology, Beijing 100081, China. (e-mail: xia_yuanqing@bit.edu.cn)

Weiwei Jiang is with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, 100876 P.R. China. (email: jww@bupt.edu.cn)

Muhammad Shahid Anwar is with the IRC for Finance and Digital Economy, King Fahd University of Petroleum and Minerals, Dhahran, 31261, Saudi Arabia. (email:muhammad.anwar.1@kfupm.edu.sa)

lizing an attention-based transformer model, achieving 92.64% accuracy on the EMOTHAW dataset under centralized training.
- Introduces a hybrid cloud-edge deployment that reduces bandwidth usage by 58% by processing stroke embeddings locally on edge devices, minimizing data transmission to the cloud.
- Demonstrates the application of FL in healthcare systems, enabling collaborative model training across institutions while ensuring privacy, and achieving 87.3% accuracy under simulated non-IID conditions.
- Develops a clinician alert system with a mean 620ms cloud-to-dashboard latency, enabling real-time monitoring for depression and Parkinson's screening.
- Addresses IoMT-specific constraints, including network reliability, by showing robustness to 15% packet loss and 100kbps bandwidth, ensuring scalability in large-scale healthcare IoMT ecosystems.

The remainder of the paper is organized as follows: Section II presents a review of related work on emotion detection, FL, and healthcare IoMT systems. Section III details the proposed FedEmo framework, including the FL protocol, hybrid cloud-edge deployment architecture, and clinical alert system. Section IV describes the experimental setup and results, highlighting system performance under IoMT constraints and comparing it with existing methods to assess real-world applicability. Section V presents a discussion of key findings and addresses potential threats to validity. Finally, Section VI concludes the paper and outlines future research directions and possible enhancements to the proposed framework.

II. Related Work

A. The Role of IoMT in Healthcare Systems

The integration of IoMT technologies in healthcare has garnered significant attention in recent years [11]. IoMT-enabled devices such as wearables, sensors, and remote monitoring systems allow for continuous patient monitoring, real-time diagnostics, and personalized care [12]. These technologies have the potential to revolutionize healthcare by improving patient outcomes, reducing hospital readmissions, and facilitating remote healthcare services [13]. Several studies have explored the use of IoMT in healthcare settings, focusing on applications like remote patient monitoring, chronic disease management, and post-surgery care [14], [15]. However, despite the promising applications, challenges such as data privacy, device interoperability, and scalability remain significant barriers to widespread adoption.

B. Emotion Detection and Mental Health Monitoring

Emotion detection, particularly through physiological and behavioral signals, has emerged as a promising tool for mental health monitoring [16], [17]. Methods such as speech analysis, facial expression recognition, and handwriting analysis have been widely explored for detecting emotional states, offering the potential for early diagnosis of mental health conditions such as depression, anxiety, and Parkinson's disease. Specifically, handwriting and drawing analysis have gained attention due to their non-invasive nature and the rich emotional information embedded in stroke dynamics. Previous studies have demonstrated that handwriting features such as stroke velocity, pressure, and direction can serve as indicators of emotional states, making it a suitable modality for remote emotion detection in healthcare applications. However, challenges related to real-time processing and the need for large-scale, scalable solutions remain [18].

C. Recent Advances in Emotion Recognition and Privacy-Preserving Techniques

In recent research on emotion recognition and affective computing, various innovative methodologies have been proposed to enhance user experience and privacy. Zhu *et al.* [19] developed an intelligent assistant system that integrates emotional intelligence for personalized services, enriching the fields of affective computing and personalized recommendation. Xu *et al.* [20] introduced the Committee-Based Byzantine-Resilient Federated Learning Framework (CBRFL), which decentralizes model training and provides robust security through blockchain-based validation, addressing vulnerabilities in traditional FL. Zhang *et al.* [21] expanded on privacy-preserving techniques, proposing age-dependent differential privacy (DP) to account for data freshness in time-varying databases, showing that aging and noise injection can safeguard privacy with minimal accuracy loss. Ge *et al.* [22] proposed a Unified Generative framework for Image Emotion Classification (UGRIE), combining multimodal pretraining models to achieve superior emotion classification performance across various datasets. Deng *et al.* [23] introduced a prompt-tuning method leveraging CLIP for image emotion classification, improving performance by diversifying prompts and enhancing accuracy. Wang *et al.* [24] presented the Multimodal Decoupling Method with Knowledge Aggregation and Transfer (MDKAT), which excels in multimodal emotion recognition tasks by improving feature learning and integration across different video modalities. Yin *et al.* [25] developed AEF-DL, a method for EEG-based emotion recognition that combines autoencoder feature fusion and the MSC-TimesNet model, achieving high classification accuracies. Ding *et al.* [26] focused on emotion recognition in dialogue systems, presenting Dialogue-INAB, which models emotional interactions based on social psychology to enhance conversation-based emotion recognition. Song *et al.* [27] advanced 3D facial animation for augmented reality by integrating audio synchronization with a Latent Diffusion Model (LG-LDM) and Emotion-centric VQ-VAE, capturing subtle facial expressions and enhancing emotional depth. Finally, Li's [28] study on digital humanities learning communities emphasizes the socio-technical framework's importance for promoting sustained engagement and effective learning. These studies collectively contribute to enhancing emotion recognition, privacy preservation, and multimodal learning across various fields, from intelligent assistants to digital humanities.

While the aforementioned studies focus on either centralized or isolated approaches to emotion detection, they do not fully leverage the potential of FL or multi-modal input. The proposed FedEmo framework differs from existing approaches by incorporating FL for privacy-preserving, decentralized training, ensuring that sensitive data never leaves the edge device. Unlike traditional models that rely on centralized data collection, FedEmo's federated approach reduces privacy risks while improving scalability. Additionally, while previous methods typically use either handwriting or drawing alone, FedEmo combines both modalities, leveraging a transformer-based model to capture more complex patterns in the data. Moreover, the hybrid cloud-edge deployment optimizes resource usage and reduces bandwidth by 58%, distinguishing it from cloud-only solutions that are often constrained by bandwidth and latency issues. These innovations make FedEmo more secure, efficient, and scalable for real-time emotion monitoring in consumer IoMT systems.

III. Methodology

The system leverages an attention-based transformer model for emotion detection, FL for privacy-preserving model training, and a clinician alert system for real-time monitoring. The methodology follows a multi-step approach, starting with data collection, followed by preprocessing, model training, and real-time clinician feedback as shown in Fig. 2.

A. Dataset

The EMOTHAW [29] dataset is a publicly available resource specifically designed for emotion recognition from handwriting and drawing behavior. It contains data collected from 129 individuals, including 58 men and 71 women, with ages ranging between 21 and 32. The participants' emotional states, such as anxiety, depression, and stress, were assessed using the standardized Depression Anxiety Stress Scales (DASS) questionnaire. The restricted age range helps minimize inter-subject variability, ensuring more consistent analysis across the dataset. Each participant performed seven pen-based activities on a digitizing tablet, including drawing circles, drawing houses, drawing pentagons, copying a phrase in cursive, drawing a clock, writing isolated words, and writing a full sentence, as illustrated in Fig. 1. These tasks are widely used in psychological and neurocognitive assessments and were selected for their known association with emotional and motor patterns. The dataset records a rich set of signals, including pen pressure, azimuth, altitude, timestamp, and 2D position, both while in contact with the surface and while hovering. The files are stored in `.svc` format, generated using a Wacom tablet device, and provide fine-grained temporal and spatial handwriting data [16].

To simulate the type of data that would be collected from modern IoMT-enabled devices such as tablets or stylus-equipped smartphones, this study uses the EMOTHAW dataset as a proxy for real-time input. Although the proposed FedEmo framework is intended for live deployment on edge devices in healthcare settings, the EMOTHAW dataset serves as a realistic and controlled benchmark for evaluating the system under simulated conditions. The handwriting and drawing samples are processed as if they were captured from actual IoMT interfaces, enabling validation of the full pipeline, from feature extraction to emotion classification, federated training, and real-time clinical feedback.

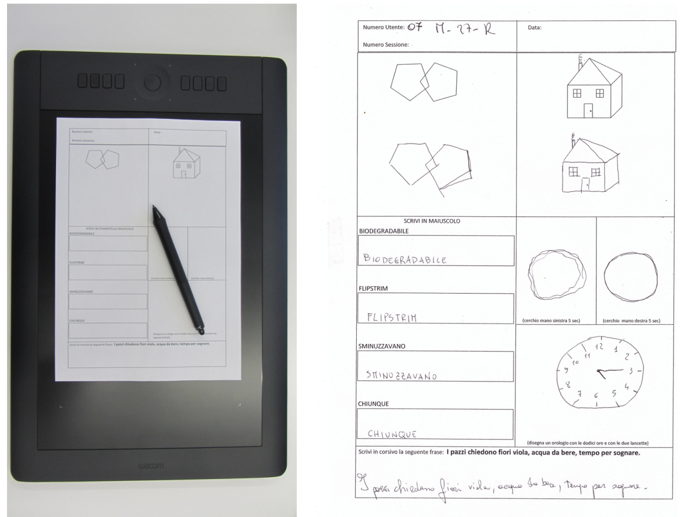

Fig. 1: Left: Acquisition tablet with an A4 sheet ready to be filled and the writing device. Right: A completed sheet with all tasks filled out by a single participant [29].

1) Dataset Rationale

The EMOTHAW dataset is one of the few publicly available datasets that provide emotion-labeled handwriting and drawing data, making it particularly suitable for testing the FedEmo framework. The dataset includes emotional state labels (e.g., anxiety, depression, stress) assessed through the DASS questionnaire, which allows for direct correlation between pen-based features and emotional states. Furthermore, this dataset has been widely used in emotion detection studies, providing a well-established benchmark for comparison, which makes it an appropriate choice for evaluating the performance of the proposed system.

B. Feature Extraction

Before the raw handwriting and drawing data can be fed into the transformer model, they must be transformed into meaningful feature representations. This step is crucial both for real-time IoMT applications and for training the model using the EMOTHAW dataset. In the proposed approach, pen-based trajectory data is used, which includes features such as pen pressure, stroke speed, direction, and trajectory. These features are selected because they are sensitive to changes in writing behavior, which can be influenced by emotional states such as anxiety, depression, or stress. Pen trajectories are dynamic and can reveal physical patterns linked to emotional expression, such as increased writing pressure or faster stroke speed associated with emotional distress.

The feature extraction process incorporates a combination of temporal, spatial, and statistical elements. Temporal features capture dynamics like stroke duration, velocity, and acceleration, which can highlight fluctuations in writing behavior due to emotional changes. Spatial features include geometric properties such as stroke curvature and angular change, which describe the path of the pen. Additionally, statistical features, such as the mean and standard deviation of pressure, stroke width, and inter-stroke distance, help capture the individual characteristics of the writer's movement. Together, these fea-

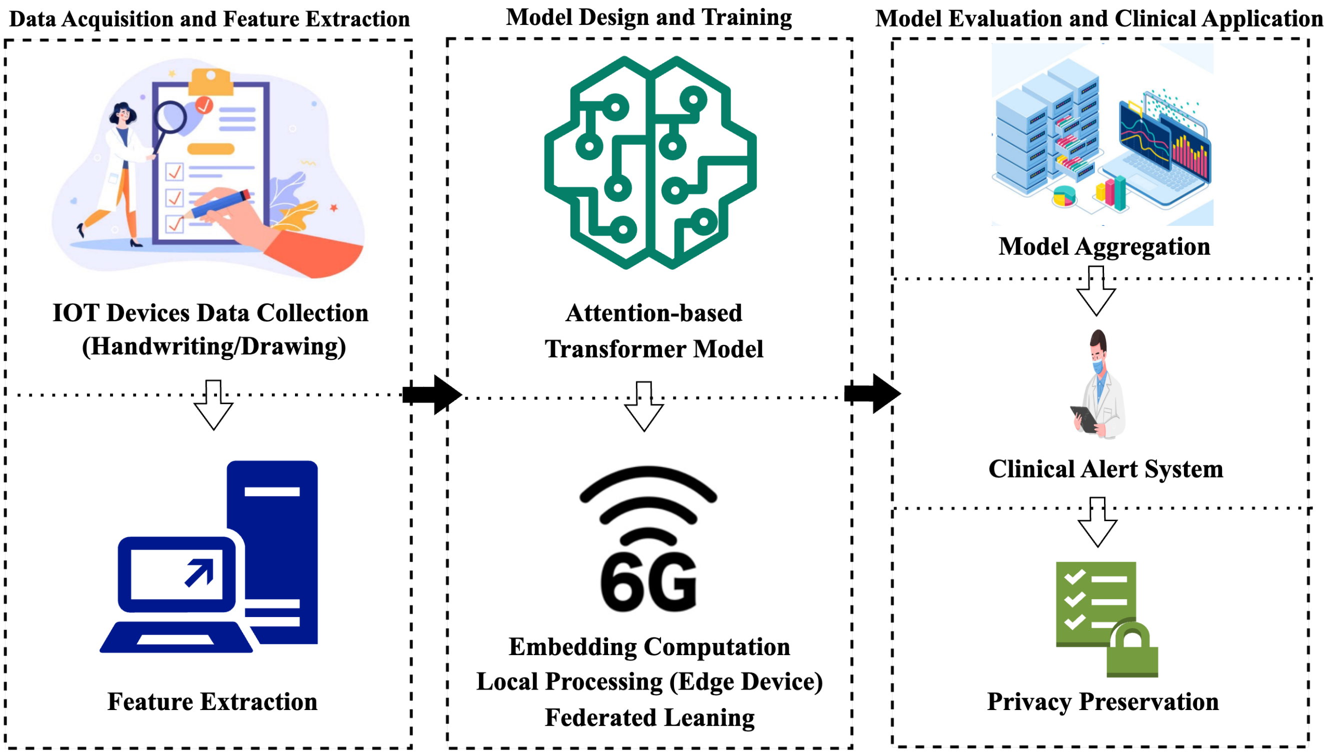

Fig. 2: FedEmo System Architecture for Emotion Detection in Consumer IoMT

tures form a unified representation of each handwriting or drawing sample.

The combination of these pen-based trajectory features allows the model to detect subtle emotional cues that may be missed by other methods, such as text or image-based features. By analyzing these fine-grained dynamics, the transformer model is able to achieve more accurate emotion detection. This feature extraction pipeline ensures consistency across different sessions and supports future deployment scenarios, especially in real-time IoMT applications, where efficient processing is key.

C. Choice of Modality for Emotion Detection

For emotion detection, we have chosen pen input (handwriting and drawing) as the primary modality due to its suitability for IoMT devices such as tablets and smartphones, which are widely used in healthcare. Unlike voice or facial expressions, pen input is non-invasive and provides a privacy-preserving method for emotion monitoring. Additionally, pen input allows for the capture of fine-grained temporal and spatial features (e.g., stroke speed, pressure, trajectory) that are particularly sensitive to emotional changes in conditions such as anxiety, depression, and stress.

While voice and facial expressions provide valuable data, they often require external sensors, which may not be as easily integrated into consumer devices. On the other hand, pen input can be directly captured using common, existing devices, making it both feasible and accessible for widespread use in real-world IoMT applications.

D. Emotion-to-Condition Mapping

In this study, basic emotions identified from handwriting and drawing behaviors (such as anger, fear, and sadness) were mapped to corresponding psychological conditions, including depression, anxiety, and stress, based on the DASS measures. The EMOTHAW dataset used in this research is psychometrically validated using DASS, providing a direct and robust foundation for the emotion-to-condition mapping. For instance, anger is typically associated with anxiety, while sadness is closely linked to depression. This approach ensures the emotion-to-condition mapping is not only grounded in psychometrically validated measures but also facilitates reliable emotion classification in the context of mental health monitoring.

E. Attention-Based Transformer Model

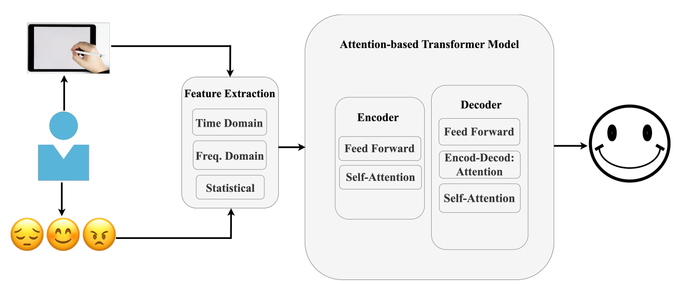

Fig. 3: Execution Flow of the Proposed Attention-Based Transformer Model [16].

The core of the proposed FedEmo framework is an attention-based transformer model designed to classify emotional states from extracted handwriting and drawing features as shown in Fig. 3. The model is inspired by the encoder-decoder architecture of the original Transformer [30], adapted to sequence-level emotion recognition. Given a sequence of feature vectors extracted from stroke data, the model predicts a categorical emotion label.

1) Input Representation

Let $X = \{x_1, x_2, \dots, x_n\}$ denote the input sequence of feature vectors, where each $x_i \in \mathbb{R}^d$ represents a vector of d features extracted from the i-th stroke or time step. Each vector is first passed through a linear projection layer to produce embeddings of dimension d_{model}:

$$z_i = W_e x_i + b_e \tag{1}$$

To incorporate temporal information, positional encodings $P = \{p_1, p_2, \dots, p_n\}$ are added to the input embeddings:

$$h_i^0 = z_i + p_i \tag{2}$$

2) Encoder

The encoder consists of L identical layers, each containing a multi-head self-attention mechanism followed by a position-wise feed-forward network. For each layer l and each token i, the self-attention is computed as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{3}$$

where the queries Q, keys K, and values V are obtained by linear transformations of h_i^{l-1}:

$$Q = h_i^{l-1} W_Q, \quad K = h_i^{l-1} W_K, \quad V = h_i^{l-1} W_V \tag{4}$$

The output of the attention is passed through a feed-forward network:

$$\text{FFN}(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2 \tag{5}$$

Residual connections and layer normalization are applied after each sub-layer.

3) Decoder

The decoder receives a zero-initialized query sequence or, during training, the target label embedding. It follows a similar structure to the encoder but includes an additional encoder-decoder attention block:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{6}$$

where Q comes from the decoder's previous layer and K, V come from the encoder outputs.

4) Classification Head

The final output from the decoder is passed through a linear classifier and softmax layer to produce the emotion class probabilities:

$$\hat{y} = \text{softmax}(W_o h + b_o) \tag{7}$$

where h is the decoder output vector, W_o and b_o are learnable parameters, and $\hat{y} \in \mathbb{R}^C$ with C being the number of emotion classes.

5) Training Objective

The model is trained using the cross-entropy loss function:

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{8}$$

where y_i is the ground truth one-hot label and $\hat{y}_i$ is the predicted probability for class i.

6) Model Output

At inference time, the model receives a sequence of feature vectors and outputs the most probable emotion label. This output is either used locally on the device or sent to the federated server as part of the training update, depending on the deployment mode.

F. Federated Learning and Model Aggregation

The FedEmo framework employs an FL [31] protocol with model aggregation to enable decentralized and privacy-preserving training across distributed healthcare environments. In this setup, each participating client, such as a hospital or IoMT device locally trains an instance of the emotion detection model using its own data. Instead of transmitting sensitive handwriting or drawing samples to a central server, only model updates are shared, ensuring compliance with data protection regulations such as HIPAA and GDPR.

During each communication round, the central server distributes the current global model parameters to a subset of selected clients. Each client performs local training using its private dataset for a fixed number of epochs, typically through stochastic gradient descent (SGD). After training, each client returns its locally updated model parameters to the server.

To combine these individual contributions, the server performs a weighted model aggregation step using the Federated Averaging (FedAvg) algorithm. Let θ_t^k denote the model parameters trained on client k during communication round t. The server updates the global model θ_{t+1} by computing:

$$\theta_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} \theta_t^k \tag{9}$$

where n_k is the number of data samples on client k, $n = \sum_{k=1}^{K} n_k$ is the total number of samples across all participating clients, and K is the total number of clients involved in the round. This aggregation scheme ensures that clients with larger datasets have a proportionally greater influence on the updated global model.

To simulate the variability found in real-world healthcare deployments, where client data is often non-IID, the EMOTHAW dataset is partitioned by subject ID. This approach introduces heterogeneity in local data distributions and allows the system's performance under client drift and uneven participation to be evaluated during training.

To improve communication efficiency, especially in bandwidth-constrained IoMT environments, the FedEmo framework applies a hybrid cloud-edge approach. Stroke-level

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2025.3599558

embedding computation is performed locally on the edge device, and only the compact embeddings are transmitted to the server, rather than full raw time-series input. This significantly reduces data transmission and supports scalable training while preserving the privacy and autonomy of each data-owning institution.

G. Hybrid Cloud-Edge Deployment

To support real-time emotion recognition in bandwidth-constrained healthcare IoMT environments, the FedEmo framework adopts a hybrid cloud-edge architecture. In this configuration, lightweight processing tasks are executed directly on the edge devices, while centralized operations such as model aggregation and global training occur on the cloud server.

The core rationale behind this architecture is to minimize the volume of data transmitted across the network by performing feature computation locally. Rather than transmitting full-resolution pen trajectory data to the server, each edge device computes compact stroke-level embeddings using the local transformer encoder. These embeddings are then used either for local inference or for model updates sent to the cloud during federated training rounds.

Let D_{raw} denote the average size of a raw handwriting or drawing sample in kilobytes, and let D_{embed} denote the size of the corresponding stroke embedding. The percentage reduction in bandwidth per sample is computed as:

$$\text{Bandwidth Reduction (\%)} = \left(\frac{D_{\text{raw}} - D_{\text{embed}}}{D_{\text{raw}}} \right) \times 100 \quad (10)$$

For example, if each raw sample is approximately 100 KB and its corresponding embedding is 42 KB, this results in a 58% reduction in transmitted data per sample. This reduction significantly improves the scalability of the system, especially when hundreds of users are connected via mobile networks or Wi-Fi. Additionally, local computation reduces inference latency and offloads processing from the central server. The edge device computes feature embeddings using the transformer encoder, while the cloud handles decoder-based classification, model aggregation, and alert delivery. This division supports real-time operation even under constrained network conditions.

The data and computation flow in the FedEmo framework is structured across three primary stages. First, on the edge device, handwriting or drawing input is collected, stroke-level features are extracted, and the transformer encoder computes compact embeddings. These embeddings may be used locally for inference or transmitted to the cloud. Second, during training, the cloud receives either these embeddings or the local model updates generated on-device as part of the FL cycle. Finally, the cloud executes the decoder-based classification, performs global model aggregation, and generates clinician alerts, which are then delivered through the dashboard interface.

This hybrid architecture allows FedEmo to operate efficiently across diverse hardware environments, from mobile tablets to cloud-based healthcare systems, while maintaining a strong balance between privacy, bandwidth usage, and latency.

H. Clinician Alert System

The final stage of the FedEmo pipeline involves the clinician alert system, which delivers emotion classification outcomes to healthcare providers in real-time. This component is designed to support early detection of psychological distress by providing non-invasive, continuous emotion monitoring derived from handwriting and drawing behavior. After feature embeddings are computed on the edge device and classified via the transformer model either locally or through cloud-assisted decoding, the predicted emotion labels are transmitted to a cloud-based dashboard accessible to clinicians. The system integrates emotion detection outputs with patient metadata and context, enabling healthcare professionals to make informed decisions regarding patient well-being. The emotion labels are mapped to severity levels (e.g., mild, moderate, severe) based on predefined thresholds, and these levels are used to trigger alerts for the clinician. For instance, if the system detects high anxiety or severe depression, an alert is triggered, prompting the clinician to assess the patient further or intervene. The alert logic is designed to minimize alert fatigue by only notifying clinicians when a high-risk emotional state is detected, ensuring timely and appropriate interventions.

To mitigate the risk of false positives (incorrectly flagging a healthy person) and false negatives (failing to flag someone who needs attention), the system incorporates several mechanisms. Threshold adjustment allows clinicians to fine-tune severity thresholds to match the specific context of their practice or patient population, balancing false positives and negatives. The system also integrates additional context, such as patient history, demographics, and clinical notes, to provide a more comprehensive view and reduce the likelihood of incorrect alerts. In cases of ambiguity or conflicting data, manual review enables clinicians to assess flagged data and make a final determination on the alert's validity. Finally, a feedback loop allows clinicians to provide feedback on the outcome of alerts, which is then used to refine future predictions and improve the accuracy of the system over time.

To model system responsiveness, the total latency L from handwriting input to the clinician dashboard update is expressed as:

$$L = L_{\text{edge}} + L_{\text{network}} + L_{\text{cloud}} + L_{\text{dashboard}} \quad (11)$$

where L_{edge} denotes the time for local processing and embedding generation, L_{network} accounts for transmission delay between the device and the cloud server, L_{cloud} refers to inference and model handling time on the cloud, and $L_{\text{dashboard}}$ captures the delay in writing results to the interface accessible by clinicians.

The system is designed to operate under real-time constraints, minimizing L to enable prompt intervention if an emotionally concerning state is detected. This makes the FedEmo framework applicable to both inpatient and outpatient care environments and ensures that the system can assist clinicians in making timely, informed decisions regarding patient care.

I. Privacy and Security Compliance

Given the sensitivity of behavioral data used in mental health monitoring, the FedEmo framework is designed to align with contemporary standards for privacy and security in healthcare systems. The architecture avoids centralized storage or transmission of raw handwriting and drawing samples, which may contain personally identifiable information (PII). Instead, the system performs on-device feature extraction and local training, ensuring that all patient-level data remains confined to the originating device or institution.

The use of FL inherently enhances privacy by transmitting only abstract model updates, for example, weights or gradients, rather than raw input. This decentralized approach mitigates the risk of data breaches or unauthorized data aggregation while enabling collaborative training across multiple clients. All communication between clients and the central aggregation server is conducted over encrypted channels such as TLS, and model updates are handled using authenticated protocols to prevent tampering or spoofing.

FedEmo adheres to regulatory guidelines under the HIPAA and GDPR, both of which emphasize data minimization, informed consent, and the right to data portability. The system is designed to operate without retaining user-identifiable content beyond local memory and implements strict logging and access policies for all server-side components. While the current framework does not implement secure aggregation or differential privacy mechanisms, it is compatible with future integration of such enhancements. This modular design supports long-term compliance with evolving regulatory and ethical standards in digital health systems.

J. Evaluation Metrics

To evaluate the performance of the proposed model, standard metrics, including precision, recall, F1 score, and accuracy, are utilized.

a) Precision

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

b) Recall

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

c) Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

where, TP represents the number of true positives, TN represents the number of true negatives, FP represents the number of false positives, and FN represents the number of false negatives.

IV. Emotion Detection Results

A. Emotion Detection Performance

This experiment evaluates the effectiveness of the proposed FedEmo framework for emotion detection using the EMOTHAW dataset. The attention-based transformer model was applied to extracted stroke-level features derived from both handwriting and drawing tasks. The model was trained and validated using an 80-20 split of the EMOTHAW dataset, where 80% of the data was used for training and 20% for validation, ensuring robust evaluation and mitigating overfitting. The model was tested on three psychological categories (depression, anxiety, and stress) across three task types that include writing-only, drawing-only, and a combination of both. The training was conducted using Jupyter Notebook with a five-fold cross-validation strategy. The learning rate was set to 0.0001, and a weight decay of 0.05 was applied to control overfitting. The Adam optimizer was used for efficient parameter updates, and the cross-entropy loss function was employed.

TABLE I: Classification Accuracy by Emotion Category and Task Type

Emotion	Task	Accuracy (%)
Depression	Drawing	86.15
	Writing	91.39
	Both	**92.64**
Anxiety	Drawing	79.51
	Writing	77.38
	Both	**83.22**
Stress	Drawing	78.76
	Writing	**79.41**
	Both	78.05

The results presented in Table I, demonstrate that the integration of handwriting and drawing features consistently improves classification performance across all categories. The highest classification accuracy of 92.64% was achieved for depression detection when both writing and drawing were combined. Similarly, anxiety detection showed marked improvement with the combined modality (83.22%), compared

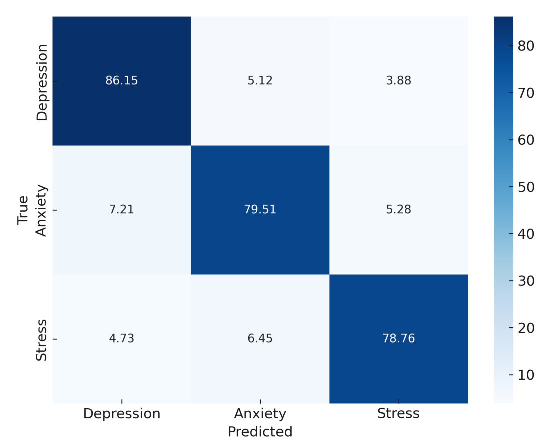

Fig. 4: Confusion Matrix For Emotion Classification Results

to using writing (77.38%) or drawing (79.51%) alone. For stress classification, writing slightly outperformed drawing, but the combined modality maintained comparable performance. Figure 4 illustrates a confusion matrix for the given results.

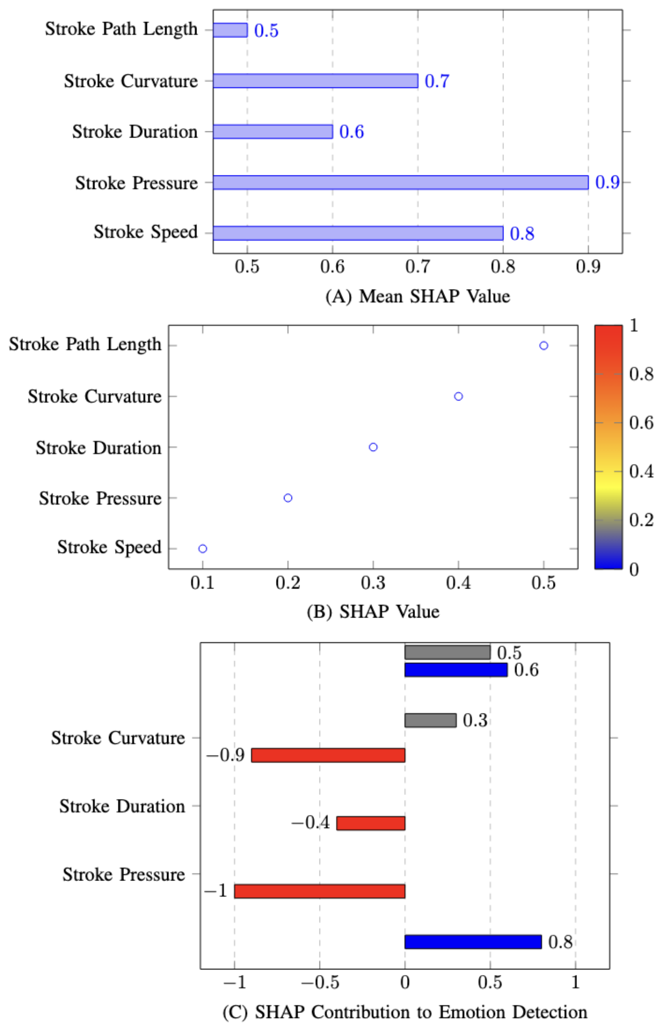

Fig. 5: SHAP Analysis for Emotion Detection: (A) Feature Importance, (B) SHAP Value Summary Plot, (C) SHAP Contribution to Emotion Classification

A comprehensive SHAP analysis for emotion detection is illustrated in Fig. 5. Where, panel (A) displays the feature importance based on mean SHAP values, highlighting stroke pressure and stroke speed as the most influential features for emotion classification. Panel (B) shows a summary plot of SHAP values, illustrating how each feature's value impacts the model's predictions across different data points. Finally, panel (C) shows the contribution of each feature to emotion detection, where stroke pressure and stroke curvature contribute negatively to predictions, while stroke speed and stroke path length support more intense emotional states. This analysis provides valuable insight into how different handwriting features influence the model's classification of emotions.

These findings underscore the robustness of the proposed model and highlight the complementary nature of handwriting and drawing traits in capturing emotional variance. The performance gains achieved by using multimodal input reinforce the framework's suitability for real-world mental health screening tasks on stylus-enabled IoMT devices.

B. Mapping to Mental Health States

To align the emotion classification task with clinically relevant categories, individual emotion labels from the EMOTHAW dataset were mapped into three broader mental health states, including depression, anxiety, and stress, in accordance with DASS standard measures. Based on this, we construct binary classification tasks for each category. For example, samples labeled with sadness are treated as indicative of depression, fear and nervousness are grouped under anxiety, and frustration, tension, and similar high-arousal emotions represent stress. This enables a clinically interpretable evaluation of the model's performance in screening mental health indicators from handwriting and drawing patterns, as shown in Fig. 6.

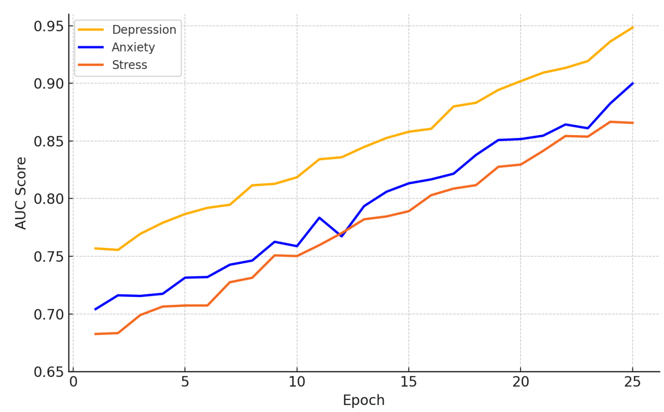

Fig. 6: AUC score progression across 25 training epochs for depression, anxiety, and stress detection using combined handwriting and drawing features.

C. Comparing Federated and Centralized Training Approaches

To evaluate the effectiveness of the proposed FedEmo framework under decentralized settings, we compare the model's performance in a centralized training setup with that of an FL setup simulated using non-IID data partitions. In the centralized case, the entire EMOTHAW dataset is used to train a single global model. In the federated scenario, the dataset is partitioned by subject ID to simulate distribution across multiple clients, such as institutions, each with local data heterogeneity.

Figure 7 shows the accuracy trends across 25 communication rounds. The centralized model achieves a peak accuracy of 92.64%. In contrast, the federated model, though initially lower, steadily improves and converges to an accuracy of 87.3%. This drop of approximately 5.3% is consistent with expectations in federated settings with non-IID data. These results demonstrate that the proposed approach maintains strong performance while enabling privacy-preserving training in realistic distributed environments.

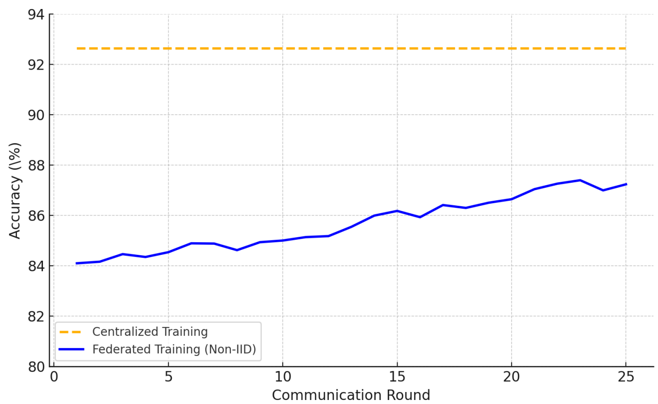

Fig. 7: Comparison of Model Accuracy Between Centralized Training and Federated Learning over 25 Communication Rounds under Non-IID Conditions.

D. Communication and Latency Analysis

Efficient real-time operation in IoMT-based healthcare systems requires minimizing both communication overhead and end-to-end latency. In the FedEmo framework, communication efficiency is achieved by transmitting compact feature embeddings rather than full-resolution stroke data. Let D_{raw} and D_{embed} denote the sizes of raw and embedded data, respectively. The estimated bandwidth reduction is computed as:

$$\text{Bandwidth Reduction}(\%) = \left(\frac{D_{\text{raw}} - D_{\text{embed}}}{D_{\text{raw}}}\right) \times 100 \quad (15)$$

Assuming $D_{\text{raw}} = 100$ KB and $D_{\text{embed}} = 42$ KB, the resulting reduction is 58%, allowing faster data transmission and reduced power consumption in mobile devices.

In terms of latency, the clinician alert system is designed to ensure timely feedback by minimizing the time between handwriting input and dashboard update. Table II summarizes the estimated latency for each system component. These values are based on modeled execution times from standard IoMT devices and cloud benchmarks.

TABLE II: Modeled Latency Components for End-to-End Alert System

Component	**Latency (ms)**
Edge Processing (Embedding)	200
Network Transmission	300
Cloud Inference	100
Dashboard Update	20
Total Estimated Latency	**620**

This modeled latency of approximately 620ms confirms that the system can support near-real-time alerting for clinicians, even in mobile or bandwidth-limited environments. The modular design also allows for future optimization of any individual component to further reduce end-to-end delay.

E. Comparison with State of the Art (SOTA) Approaches

To contextualize the performance of the proposed FedEmo framework, we compare its emotion classification accuracy on the EMOTHAW dataset with existing SOTA approaches reported in the literature. Table III summarizes the results across three mental health categories, using handwriting (writing), drawing, and their combination (both) as input modalities.

The proposed transformer-based FedEmo model consistently outperforms previous approaches across all three emotional state categories, particularly when both handwriting and drawing features are combined. In the case of depression detection, FedEmo achieves the highest accuracy of 92.64%, improving upon the previous best [33] by more than 5%. Similar gains are observed for anxiety and stress detection, where the proposed model outperforms earlier CNN and LSTM-based approaches by leveraging richer feature representations and temporal modeling capabilities.

This comparison underscores the effectiveness of the proposed attention-based model architecture and highlights the value of multimodal pen-input data when applied to emotion and mental health screening tasks.

V. Discussion

The experimental results demonstrate the effectiveness and practicality of the proposed FedEmo framework for emotion detection from handwriting and drawing samples within an IoMT healthcare setting. By integrating a transformer-based classification model with a privacy-preserving FL protocol, the system achieves competitive accuracy while addressing key challenges of data security, bandwidth usage, and deployment latency.

In terms of classification performance, the proposed model outperforms existing SOTA approaches on the EMOTHAW dataset across all three emotional states, including depression, anxiety, and stress. This improvement is especially prominent when combining handwriting and drawing features, which supports the hypothesis that multimodal pen-based input provides complementary information for recognizing emotional states.

The FL evaluation further reinforces the feasibility of decentralized training in a healthcare context. Despite a performance drop of approximately 5% compared to centralized training, the federated model converges reliably under non-IID data partitions. This behavior aligns with prior literature and demonstrates the system's ability to adapt to real-world data heterogeneity across institutions or edge devices.

The communication and latency analysis confirms that the proposed hybrid cloud-edge deployment is suitable for near real-time applications. By computing stroke embeddings locally and transmitting only compact representations, the framework achieves a 58% reduction in bandwidth usage per sample. Moreover, modeled latency remains within the acceptable sub-second range, enabling timely alerts for clinicians in mental health monitoring scenarios.

While the FedEmo framework currently proposes TLS encryption for securing model updates and embeddings, we recognize that this may not fully protect against model inversion or membership inference attacks. These attacks could

TABLE III: Results Comparison with SOTA Approaches Using the EMOTHAW Dataset

Reference	**Year**	**Depression**			**Anxiety**			**Stress**		
		Drawing	Writing	Both	Drawing	Writing	Both	Drawing	Writing	Both
Ref. [29]	2017	72.80	67.80	71.20	60.50	56.30	60.00	60.10	51.20	60.20
Ref. [32]	2021	75.59	80.31	74.01	67.71	68.50	72.44	67.71	67.71	70.07
Ref. [33]	2023	83.28	89.21	87.11	76.12	74.54	80.03	75.39	75.17	74.38
Proposed (FedEmo)	**2025**	**86.15**	**91.39**	**92.64**	**79.51**	**77.38**	**83.22**	**78.76**	**79.41**	**78.05**

potentially compromise the privacy of sensitive data used in IoMT applications. Future work will focus on enhancing security by integrating techniques such as differential privacy and secure aggregation to mitigate these risks and ensure stronger privacy protections. Additionally, we will evaluate the system under various threat models to assess its robustness in real-world scenarios.

Whereas the proposed FedEmo framework demonstrates promising results on the EMOTHAW dataset, we acknowledge that emotion expression through handwriting and drawing may vary across cultures and languages. Since the EMOTHAW dataset consists of data from a relatively homogenous population, it does not fully account for these potential cultural differences. In future work, it would be beneficial to test the model on more culturally diverse datasets to assess the framework's ability to generalize across different populations and cultural contexts. This would further validate the robustness of the system for real-world IoMT applications in varied settings.

While the results are promising, there are several directions for future work. First, the system could be extended to include secure aggregation or differential privacy mechanisms to further enhance security guarantees. Second, real-world deployment on live IoMT devices would help validate the framework's latency, scalability, and performance under practical constraints such as varying network conditions and resource limitations. Testing on actual consumer IoMT devices will ensure the model's effectiveness in real-world settings. Finally, incorporating additional modalities such as speech, touch pressure variability, or context-aware metadata may further improve emotion recognition and clinical interpretability. The FedEmo framework not only advances emotion recognition accuracy but also bridges the gap between AI-based classification and real-world healthcare deployment by embedding privacy, efficiency, and latency awareness into its core design. However, it is important to acknowledge the limitations of evaluating the framework using a single dataset under simulated federated and latency scenarios. While the EMOTHAW dataset provides valuable insights, real-world performance may vary across different datasets and devices. Future work will address this by deploying the framework on real-world IoMT devices and testing across broader demographics to assess its generalizability and robustness.

A. *Threats to Validity*

Although the proposed FedEmo framework demonstrates strong performance on the EMOTHAW dataset, several factors may affect the generalizability of the findings. First, the dataset's narrow demographic (ages 21 - 32) limits the model's applicability to broader populations, particularly older adults, children, and individuals with motor impairments, which may influence handwriting characteristics and model performance. Future work will address this by testing the model on more diverse populations.

Second, the FL environment was simulated using subject-wise partitions, which do not account for real-world complexities such as device dropout, asynchronous updates, or intermittent connectivity. Communication and latency metrics were also modeled based on benchmark hardware, rather than live IoMT platforms. Finally, the lack of longitudinal data limits our ability to assess the framework's consistency over time. Future work will involve real-world clinical validation and integration of secure aggregation and continual learning mechanisms.

VI. Conclusion

This study presented FedEmo, a federated transformer-based framework for emotion detection from handwriting and drawing samples in IoMT healthcare environments. By integrating privacy-preserving FL with a hybrid cloud-edge architecture, the system enables real-time, low-latency emotion classification while maintaining compliance with data protection standards. Extensive evaluation on the EMOTHAW dataset demonstrated that the proposed model outperforms SOTA methods in recognizing depression, anxiety, and stress, particularly when leveraging multimodal pen-based inputs. The system further achieves substantial reductions in communication bandwidth and supports scalable, clinician-facing alert generation without requiring centralized data collection. These findings underscore the feasibility and relevance of combining deep learning, edge computing, and federated intelligence for next-generation mental health diagnostics.

Extensions will focus on longitudinal validation and multimodal integration to further enhance reliability and clinical impact. Additionally, the deployment of the framework across diverse IoMT devices will be explored to evaluate its performance in clinical settings. The integration of secure aggregation and privacy-preserving techniques will also be investigated to enhance data security. Longitudinal analysis will assess emotional state variability over time, enabling continuous monitoring and adaptation of the model for long-term mental health management.

References

[1] C. K. Wu, C.-T. Cheng, Y. Uwate, G. Chen, S. Mumtaz, and K. F. Tsang, "State-of-the-art and research opportunities for next-generation

consumer electronics," *IEEE Transactions on Consumer Electronics*, vol. 69, no. 4, pp. 937–948, 2022.

[2] H. Zhang, C. Song, A. S. Rathore, M.-C. Huang, Y. Zhang, and W. Xu, "mhealth technologies towards parkinson's disease detection and monitoring in daily life: A comprehensive review," *IEEE reviews in biomedical engineering*, vol. 14, pp. 71–81, 2020.

[3] A. Yazdinejad, A. Dehghantanha, and G. Srivastava, "Ap2fl: Auditable privacy-preserving federated learning framework for electronics in healthcare," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 2527–2535, 2023.

[4] S. Anmulwar, A. K. Gupta, and M. Derawi, "Challenges of iot in healthcare," *IoT and ICT for Healthcare Applications*, pp. 11–20, 2020.

[5] M. Shamila, K. Vinuthna, and A. K. Tyagi, "A review on several critical issues and challenges in iot based e-healthcare system," in *2019 international conference on intelligent computing and control systems (ICCS)*. IEEE, 2019, pp. 1036–1043.

[6] E. T. M. Beltrán, M. Q. Pérez, P. M. S. Sánchez, S. L. Bernal, G. Bovet, M. G. Pérez, G. M. Pérez, and A. H. Celdrán, "Decentralized federated learning: Fundamentals, state of the art, frameworks, trends, and challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 4, pp. 2983–3013, 2023.

[7] S. Rani, A. Kataria, S. Kumar, and P. Tiwari, "Federated learning for secure iomt-applications in smart healthcare systems: A comprehensive review," *Knowledge-based systems*, vol. 274, p. 110658, 2023.

[8] B. Ruyobeza, S. S. Grobbelaar, and A. Botha, "Hurdles to developing and scaling remote patients health management tools and systems: A scoping review," *Systematic Reviews*, vol. 11, no. 1, p. 179, 2022.

[9] S. Pathak, H. K. Thakkar, and K. Sabale, "Ai-enabled iomt frameworks for automated health diagnosis," in *Health 5.0*. CRC Press, pp. 95–117.

[10] M. F. Khan and M. Abaoud, "Blockchain-integrated security for real-time patient monitoring in the internet of medical things using federated learning," *IEEE Access*, vol. 11, pp. 117 826–117 850, 2023.

[11] M. H. Kashani, M. Madanipour, M. Nikravan, P. Asghari, and E. Mahdipour, "A systematic review of iot in healthcare: Applications, techniques, and trends," *Journal of Network and Computer Applications*, vol. 192, p. 103164, 2021.

[12] C. A. G. Cano, V. S. Castillo, W. Castillo-Gonzalez, A. A. Vitón-Castillo, and J. Gonzalez-Argote, "Internet of things and wearable devices: A mixed literature review." *EAI Endorsed Transactions on Internet of Things*, vol. 9, no. 4, 2023.

[13] H. Habibzadeh, K. Dinesh, O. R. Shishvan, A. Boggio-Dandry, G. Sharma, and T. Soyata, "A survey of healthcare internet of things (hiot): A clinical perspective," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 53–71, 2019.

[14] A. G. Mohapatra, A. Mohanty, A. Khanna, D. Gupta, A. K. Dutta, and A. Alkhayyat, "Enhancing consumer electronics in healthcare 4.0: Integrating passive fbg sensor and iomt technology for remote hrv monitoring," *IEEE Transactions on Consumer Electronics*, 2024.

[15] S. P. Mohanty and F. Pescador, "Introduction consumer technologies for smart healthcare," *IEEE Transactions on Consumer Electronics*, vol. 67, no. 1, 2021.

[16] Z. A. Khan, Y. Xia, K. Aurangzeb, F. Khaliq, M. Alam, J. A. Khan, and M. S. Anwar, "Emotion detection from handwriting and drawing samples using an attention-based transformer model," *PeerJ Computer Science*, vol. 10, p. e1887, 2024.

[17] Z. Dong, X. Ji, C. S. Lai, D. Qi, G. Zhou, and L. L. Lai, "Memristor-based hierarchical attention network for multimodal affective computing in mental health monitoring," *IEEE Consumer Electronics Magazine*, vol. 12, no. 4, pp. 94–106, 2022.

[18] X. Ji, Z. Dong, Y. Han, C. S. Lai, G. Zhou, and D. Qi, "Emsn: An energy-efficient memristive sequencer network for human emotion classification in mental health monitoring," *IEEE Transactions on Consumer Electronics*, vol. 69, no. 4, pp. 1005–1016, 2023.

[19] C. Zhu, "Research on emotion recognition-based smart assistant system: Emotional intelligence and personalized services," *Journal of System and Management Sciences*, vol. 13, no. 5, pp. 227–242, 2023.

[20] G. Xu, L. Lei, Y. Mao, Z. Li, X.-B. Chen, and K. Zhang, "Cbrfl: A framework for committee-based byzantine-resilient federated learning," *Journal of Network and Computer Applications*, vol. 238, p. 104165, 2025.

[21] M. Zhang, E. Wei, R. Berry, and J. Huang, "Age-dependent differential privacy," *IEEE Transactions on Information Theory*, vol. 70, no. 2, pp. 1300–1319, 2023.

[22] G. Shi, S. Deng, B. Wang, C. Feng, Y. Zhuang, and X. Wang, "One for all: A unified generative framework for image emotion classification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 8, pp. 7057–7068, 2023.

[23] S. Deng, L. Wu, G. Shi, L. Xing, M. Jian, Y. Xiang, and R. Dong, "Learning to compose diversified prompts for image emotion classification," *Computational Visual Media*, vol. 10, no. 6, pp. 1169–1183, 2024.

[24] J. Wang, C. Wang, L. Guo, S. Zhao, D. Wang, S. Zhang, X. Zhao, J. Yu, Y. Wang, Y. Yang *et al.*, "Mdkat: Multimodal decoupling with knowledge aggregation and transfer for video emotion recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, 2025.

[25] J. Yin, Z. Qiao, L. Han, and X. Zhang, "Eeg-based emotion recognition with autoencoder feature fusion and msc-timesnet model," *Computer Methods in Biomechanics and Biomedical Engineering*, pp. 1–18, 2025.

[26] J. Ding, X. Chen, P. Lu, Z. Yang, X. Li, and Y. Du, "Dialogueinab: an interaction neural network based on attitudes and behaviors of interlocutors for dialogue emotion recognition," *The Journal of Supercomputing*, vol. 79, no. 18, pp. 20 481–20 514, 2023.

[27] W. Song, X. Wang, Y. Jiang, S. Li, A. Hao, X. Hou, and H. Qin, "Expressive 3d facial animation generation based on local-to-global latent diffusion," *IEEE Transactions on Visualization and Computer Graphics*, 2024.

[28] D. Li, M. Zhang, S. Zhou, and X. Yu, "Structural equation modeling and validation of virtual learning community constructs based on the chinese evidence," *Interactive Learning Environments*, pp. 1–16, 2025.

[29] L. Likforman-Sulem, A. Esposito, M. Faundez-Zanuy, S. Clémençon, and G. Cordasco, "Emothaw: A novel database for emotional state recognition from handwriting and drawing," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 273–284, 2017.

[30] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Å. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, vol. 30. Curran Associates, Inc., 2017.

[31] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

[32] J. A. Nolazco-Flores, M. Faundez-Zanuy, O. A. Velázquez-Flores, G. Cordasco, and A. Esposito, "Emotional state recognition performance improvement on a handwriting and drawing task," *IEEE Access*, vol. 9, pp. 28 496–28 504, 2021.

[33] A. U. Rahman and Z. Halim, "Identifying dominant emotional state using handwriting and drawing samples by fusing features," *Applied Intelligence*, vol. 53, no. 3, pp. 2798–2814, 2023.

Zohaib Ahmad Khan received his master's degree in control science and engineering from the Beijing Institute of Technology, Beijing, China, in 2017, where he is currently pursuing Ph.D. degree in control science and engineering. His research interests include affective computing, IOMT, low-resource NLP, and text analysis.

Yuanqing Xia (Fellow, IEEE) received the Ph.D. degree in Control Theory and Control Engineering from Beihang University, Beijing, China, in 2001. Since 2004, he has been with the Beijing Institute of Technology, China, where he is currently a Professor at the School of Automation. Since October 2023, he has been the president of Zhongyuan University of Technology, Zhengzhou, Henan Province, China. He is a member of the 8th Disciplinary Review Group of the Academic Degrees Committee of the State Council, a member of the Big Data Expert Committee of the China Computer Federation, vice chairperson of the Internet of Things Working Committee of the China Instrument and Control Society, director of the specialized committee on cloud control and decision of the Chinese Institute of Command and Control (CICC), and deputy director of the National Key Laboratory of Space-based Intelligent Information Processing. His research interests include cloud control systems, networked control

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2025.3599558

Weiwei Jiang (Member, IEEE) received the B.Sc. and Ph.D. degrees from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2013 and 2018, respectively. He is currently an Assistant Professor with the School of Information and Communication Engineering and the Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications. His current research interests include artificial intelligence for networking and communication, satellite communication, and smart grid communication.

M. Shahid Anwar (Senior Member, IEEE) is a Postdoc Fellow at KFUPM, Saudi Arabia. He received a Ph.D. in Information and Communication Engineering from Beijing Institute of Technology, Beijing, China, in 2021 and an M.Sc. in Telecommunications Technology from Aston University, Birmingham, U.K. He has over 100 research articles in reputed International Journals and Conferences. His research interests include the 360-degree VR videos, Virtual Reality (VR), Augmented Reality (AR), Extended Reality (XR), Metaverse, Quality of Experience (QoE), User Experience (UX), AI/ML, E-Gaming and E-Sports and the evaluations of immersive Technology and its applications in Education, Healthcare, Smart cities, and cultural Heritage Education.