Federated Fuzzy Clustering for Decentralized Incomplete Longitudinal Behavioral Data

Hieu Ngo, Hua Fang, *Senior Member, IEEE*, Joshua Rumbut, *Member, IEEE*, and Honggang Wang, *Fellow, IEEE*

***Abstract*—The use of medical data for machine learning, including unsupervised methods, such as clustering, is often restricted by privacy regulations, such as the health insurance portability and accountability act (HIPAA). Medical data is sensitive and highly regulated and anonymization is often insufficient to protect a patient's identity. Traditional clustering algorithms are also unsuitable for longitudinal behavioral health trials, which often have missing data and observe individual behaviors over varying time periods. In this work, we develop a new decentralized federated multiple imputation-based fuzzy clustering algorithm for complex longitudinal behavioral trial data collected from multisite randomized controlled trials over different time periods. Federated learning (FL) preserves privacy by aggregating model parameters instead of data. Unlike previous FL methods, this proposed algorithm requires only two rounds of communication and handles clients with varying numbers of time points for incomplete longitudinal data. The model is evaluated on both empirical longitudinal dietary health data and simulated clusters with different numbers of clients, effect sizes, correlations, and sample sizes. The proposed algorithm converges rapidly and achieves desirable performance on multiple clustering metrics. This new method allows for targeted treatments for various patient groups while preserving their data privacy and enables the potential for broader applications in the Internet of Medical Things.**

***Index Terms*—Behavior, decentralized computing, diet, federated learning (FL), fuzzy clustering, Internet of Medical Things (IoMT), longitudinal trial, missing data.**

I. Introduction

INCREASING reliance on big data has become an emerging trend in health research. Machine learning algorithms are trained across disciplines to mine large data sets for patterns; examples include detecting infectious disease outbreaks, predicting mortality, and diagnosing health conditions [1]. For longitudinal behavioral health data, clustering algorithms are especially useful, as they can identify groups of similar patients to gain insight into their behaviors and associated health outcomes [2], [3], [4], [5], [6], [7].

Employing machine learning methods, such as clustering, often may require the aggregation and harmonization of large data repositories [8], [9], [10]. While modern healthcare institutions collect a vast array of data on their patients, its use is often restricted due to data privacy regulations, such as the health insurance portability and accountability act (HIPAA) and other laws [11], [12], [13]. Medical data is sensitive and highly regulated; anonymization is often insufficient to protect a patient's identity [14], and other technical challenges, such as data security during delivery, create difficulties in preserving privacy as well [15], [16]. Because of this, healthcare institutions are, justifiably, often unable to share data with researchers and machine learning engineers [17], [18].

Much research has been devoted to solving the data privacy problem in machine learning using federated learning (FL). To allow a decentralized network of clients to collaboratively train a machine learning model without sharing data, the FL paradigm has been developed. Rather than training a model using a set of aggregated data, model parameters are distributed to individual data sources for updates, and results are averaged (Fig. 1) [17], [18], [19]. This allows models to be trained while circumventing the need for healthcare institutions to share data with researchers.

Longitudinal behavioral health trials often generate high-dimensional data with complex intercorrelation [6], [20]. In longitudinal behavioral health trials, clustering has the potential to identify groups that share distinct trajectories related to outcomes [7]. However, current FL methods are often inadequate for clustering on longitudinal data. FL can require excessive rounds of communication, slowing down the algorithm, and cannot always handle data sets containing missing values, different numbers of time points, or features, a common occurrence in longitudinal behavioral health studies. Furthermore, previous FL research has often treated data from different clients as representing a single participant rather than recognizing the possibility of pooling data at the institution or trial level, which may be more useful from a policy perspective.

Manuscript received 16 August 2023; revised 13 November 2023; accepted 12 December 2023. Date of publication 18 December 2023; date of current version 9 April 2024. The work of Hua Fang was supported in part by NIH under Grant 1R56DK114514-01A1 and Grant R01DK129432, and in part by NSF-IIS under Grant 2140729. The work of Honggang Wang was supported in part by NSF-IIS under Grant 2140729. *(Corresponding author: Hua Fang.)*

Hieu Ngo is with the College of Engineering, University of Massachusetts Dartmouth, North Dartmouth, MA 02747 USA (e-mail: hngo1@umassd.edu).

Hua Fang is with the Department of Computer and Information Science, University of Massachusetts Dartmouth, North Dartmouth, MA 02747 USA, and also with the Department of Population and Quantitative Health Science, University of Massachusetts Chan Medical School, Worcester, MA 01655 USA (e-mail: hfang2@umassd.edu).

Joshua Rumbut is with the College of Engineering, University of Massachusetts Dartmouth, North Dartmouth, MA 02747 USA, and also with the Department of Population and Quantitative Health Science, University of Massachusetts Chan Medical School, Worcester, MA 01655 USA (e-mail: jrumbut@umassd.edu).

Honggang Wang is with the Department of Graduate Computer Science and Engineering, Katz School of Science and Health, Yeshiva University, New York City, NY 10033 USA (e-mail: Honggang.wang@yu.edu).

Digital Object Identifier 10.1109/JIOT.2023.3343719

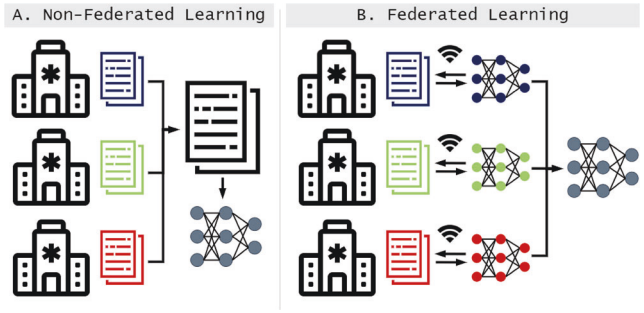

Fig. 1. Diagram depicting the difference between federated and non-FL. In A, data from healthcare institutions are aggregated to train a model. In B, model parameters are distributed to each institution for training and then aggregated.

Additionally, current FL cannot handle the complexity of longitudinal behavioral health trials, such as missing data and varying numbers of observations and time points between different data collection sites. Missing data are inevitable in longitudinal trials, and this missingness could be non-ignorable (informative drop-out) resulting from intermittent missingness (i.e., occasional missing and can relapse) and drop-out missingness (i.e., premature withdrawal and never relapse). Using only cases with no missing values typically ignores many cases, leading to a loss of power in outcome tests. At the same time, single-imputation-based clustering (e.g., mean, regression, and hot deck) misleads the clustering accuracy due to the unaccounted uncertainty in imputation [21], [22], [23], [24].

In this work, we aim to develop and evaluate federated fuzzy clustering for applications in the Internet of Medical Things (IoMT) that maintains the privacy-preserving quality of FL while overcoming the aforementioned limitations for longitudinal behavioral health studies. We formulate a federated multiple imputation-based fuzzy (FeMIFuzzy) algorithm based on a decentralized FL framework and our previously developed MIFuzzy method [2], [3], [4], [5], [23]. FeMIFuzzy incorporates information from multiple clients, allowing incomplete longitudinal data from numerous healthcare institutions to be aggregated. FeMIFuzzy is built on our MIFuzzy algorithm, which is robust to a range of missing data mechanisms (missing completely at random (MCAR), missing at random (MAR), and non-MAR) [2]. With a fairly wide range of missing rates in our tested observational studies (OSs) and randomized controlled trials (RCTs) and simulation (real: 8%–42.5%; simulation: 5%–40%), our studies consistently found that MIFuzzy outperforms these comparators in OS, RCT, and simulation with average clustering accuracy of 97%, an inconsistency rate of 3% across real and simulation studies, and at least 14% accuracy and 18% inconsistency gap between these comparators and MIFuzzy [2].

After data collection, each individual client runs the MIFuzzy algorithm in parallel. The optimal number of clusters is voted by clients, and the global cluster model is created by integrating results from each client using Sammon Mapping and weighted averaging. This method preserves privacy, requires minimal communication, effectively aggregates all studies, and identifies clusters on incomplete longitudinal data that may collect varying numbers of time points, thereby representing a novel contribution to the FL literature.

We evaluate FeMIFuzzy on an empirical health scenario using harmonized data from multiple longitudinal dietary studies in Massachusetts [25], [26], [27], [28] and simulated data with various effect sizes, correlations, and sample sizes. In the empirical scenario, each study represents a different client in the distributed framework. The data collected are incomplete. Hence, the MIFuzzy algorithm is needed to effectively process the missing data while handling high-dimensional data and overlapped clusters. Thus, we illustrate the novel use of federated MIfuzzy clustering in real-world longitudinal trial settings. For numerical analysis, we simulate data using various effect sizes, correlations, sample sizes, numbers of clients, and numbers of clusters. The simulated data are labeled into clusters to assess the performance of the proposed FeMIFuzzy clustering method. Empirical and simulation tests demonstrate the efficacy of our method for decentralized fuzzy clustering based on several clustering evaluation metrics.

This article is structured as follows. First, in Section II, we review the literature and background information related to health data privacy and fuzzy clustering. Next, in Section III, we describe the FeMIFuzzy model and the empirical simulation, which evaluates its performance on harmonized dietary data from [10]. Section IV discussed the evaluation methods and results. Finally, we conclude with a discussion in Section V that explores potential future work in this area.

II. Background and Literature Review

As our capacity to collect medical data expands, longitudinal behavioral health studies are becoming more prevalent and collecting larger quantities of observations and variables. Consequently, current and future longitudinal behavioral health studies need to employ methods for analyzing big data, including trajectory pattern recognition and data mining [21], [22], [23], [29]. These methods allow the development of adaptive treatment plans for individual patients. By relying on health data for pattern recognition in the IoMT, healthcare institutions can provide higher quality care. In this section, we survey the challenges of analyzing large longitudinal health study data and explain the foundational computational techniques used in this study.

A. Data Privacy

In the United States, the privacy of personal health data is protected under federal law by the *HIPAA of 1996*, or HIPAA. This law limits healthcare providers from disclosing protected health information (PHI) to external entities. This includes *personally identifiable information* relating to physical or mental health, provision of care, or payment linked to an individual [11], [12]. Other countries similarly regulate health data privacy [13].

Although HIPAA allows de-identified information to be disclosed, this requires either a formal determination by an investigator or the removal of all identifiers that could provide identifying knowledge [11], [12]. However, full de-identification may be difficult or even impossible, as removing basic personal information, such as name or address, is often insufficient to preserve privacy due to modern identity reconstruction techniques [14]. It may even reduce data usefulness - for example, if a removed street address would be necessary to perform spatial data analysis. Since healthcare institutions must legally protect privacy, they have minimal incentive to share data with outsiders.

This poses a problem because future advances in biomedical and health research will require large data sets. As surveyed by Mooney and Pejaver, public health research has begun to leverage machine learning algorithms that rely on big data to mine for patterns that predict patient behaviors and outcomes [1]. These include geospatial, omics, electronic health record, personal monitoring (i.e., FitBit), and effluent (i.e., online search activity) data. Even in the field of nutrition alone, machine learning has been applied for predicting high-blood pressure and body fat, weight loss success during interventions, and risk of metabolic syndrome [30], as well as studying gut microbiome features associated with diabetes [31] and integrating genomics data into nutrition studies [32].

While previous biomedical studies using machine learning have relied on large aggregated databases - for example, the Human Connectome [33], the Cancer Imaging Archive [8], or the U.K. Biobank [34] - such databases are not easily compiled due to privacy and other concerns. However, as discussed by Tresp et al. [15] and Chen et al. [16], the digitization of health data and cloud computing technologies have not only allowed additional information on health and fitness to be collected, such as via wearable sensors but has also promoted sharing, mobilizing, and advanced analysis of data across of healthcare institutions. Sharing medical data is easier than ever, but doing so requires developing privacy-protecting mechanisms such as federated analysis.

FL was first proposed in 2016 by Konečný et al. [19], who discuss a variety of methods for optimization across distributed nodes. In FL, rather than updating model parameters based on a single aggregate data set, model parameters are distributed to individual clients, who update the parameters based on their local data. Then, they transmit these model parameters and aggregate each client's parameters using *federated averaging* to obtain a global model. This process repeats iteratively. Performing federated averaging preserves privacy but requires a round of communication between clients at each iteration, which can be time-consuming.

Applications and challenges associated with FL for healthcare are surveyed by Rieke et al. [17] and Xu et al. [18]. These works discuss how FL can be used to preserve privacy in healthcare. By distributing model parameters, FL circumvents the data aggregation necessary for machine learning, a process called *data-private collaborative learning* [35]. Case studies of FL applied to medical areas have been conducted - for instance, Sheller et al. [35] used FL to detect cancer in brain tissue, finding comparable performance. Hence, FL allows multiple healthcare institutions to train different types of models, including clustering models, without sharing patient data.

B. Problems With Existing Federated Clustering

As mentioned previously, behavioral health data is often collected in longitudinal trials. This means that some health measure is collected at two or more separate points in time. For example, a study might measure the quality of a person's diet at the start of a study and one year after some intervention [41]. For clustering on longitudinal studies, current FL methods are inadequate for several reasons.

First, longitudinal behavioral health data from different studies (or clients) could have different numbers of timepoints. FL traditionally requires the same number of attributes [17], [42]. Thus, performing federated clustering on these data sets would be difficult.

The second problem is that FL can be slow and inconsistent. FL typically requires a round of communication between each client at each iteration of the algorithm [17], [42]. Machine learning algorithms can take hundreds of iterations to converge, with each round introducing substantial network communications overhead. Each communication can also pose an opportunity for a cyberattack. Additionally, because FL requires incorporating information from all other clients, depending on the implementation of the algorithm, technical problems or delays in communications in just a single client could halt progress for all others. Therefore, methods that reduce the amount of necessary communication are preferable.

The third problem is that federated clustering treats data from different clients as one aggregated set of data rather than as individual trials [37], [42]. This may be undesirable when clustering data from longitudinal behavioral health studies, as some resulting clusters could only be present in a single study or client. This means that simple federated averaging will not suffice since some clusters may form from observations in just a single client. For applications that require every cluster to incorporate information or observations from all studies, researchers may prefer a method that performs clustering on individual clients first and then aggregates the results rather than aggregating results within the training process. Our novel algorithm incorporates FL and fuzzy clustering to overcome these issues in federated clustering for longitudinal behavioral health studies.

The final problem is the incomplete data existing in longitudinal studies. In longitudinal clinical behavioral data, it is very common for patients to miss one or multiple observations over the study process. Using only the complete data may introduce bias, such as excluding groups, that are more likely to miss appointments. Therefore, it is essential to have federated clustering cope with missing values.

C. Previous Research in Federated Fuzzy Clustering

Recently, researchers have attempted to develop federated clustering algorithms to address the challenges of different domains. The method F-FCM proposed in [10] preservers privacy but requires many rounds of communication to converge

TABLE I
FEATURES COVERED IN EXISTING FEDERATED LEARNING LITERATURE AND PROPOSED FEMIFUZZY

Method	Fuzzy Clustering	Federated	Missing Data	Decentralized	Non-IID Data
FLSC [36]	Y	Y	N	N	Y
F-FCM [37]	Y	Y	N	N	N
kfed [38]	N	Y	N	Y	N
FL+HC [39]	Y	Y	N	N	Y
FFCM [40]	Y	Y	N	Y	Y
FeMIFuzzy (Our proposed method)	Y	Y	Y	Y	Y

and does not address missing data. Another federated clustering method called FCFLA was proposed by Yoo et al. [43] for solar power generation forecasting, where each local solar generator can be included in more than one set of local generators (cluster). However, this method does not work with longitudinal data and does not focus on preserving privacy, which is important for longitudinal health data. Additionally, Zhu et al. [44] proposed a rule-based horizontal federated fuzzy clustering, which is distinct from FCM or MIFuzzy. However, even though this rule-based method maintains user privacy, it does not address the problem of the same data being characterized in different feature spaces. For comparison, in Table I, we show a clear gap in the current federated clustering literature.

For medical data, CIT2FR-FL-NAS is a FL method developed in the neuro-fuzzy architecture search domain [45]. However, this method does not focus on clustering.

D. MIFuzzy for Longitudinal Behavioral Health Studies

Longitudinal behavioral health studies often employ *fuzzy clustering* to identify distinct groups in a data set. First developed in the 1960s and '70s, fuzzy clustering is a soft computing technique that partitions data into groups based on some criteria, usually related to distance or dissimilarity [46]. Unlike "hard" clustering, which places each subject into one discrete group, fuzzy clustering measures the *degree of belonging*, or *membership*, of each participant to each cluster. This effectively allows each subject to belong to more than one cluster at a time.

Fuzzy clustering is highly useful for health studies that group patients depending on behavioral intervention responses, as it captures the complex relationship between a patient, their behavior, and their outcomes [2], [23], [47]. In health studies, fuzzy clustering can be used to, for example, group patients depending on their responses to behavioral interventions, thereby informing researchers how different types of intervention may yield varying outcomes for different patient types [2], [6], [7], [21], [48]. Allowing observations to belong to more than one cluster is important in health domains since patients may share characteristics across different intervention responses. Here, fuzzy logic captures the uncertainty and imprecision in our understanding of health conditions [49].

Additionally, multiple imputation (MI) allows us to deal with missing data that is common in OSs and RCTs in longitudinal health studies, in favor of single imputation [24], [50], [51], [52], [53]. MI is a method for generating multiple data sets with replacement values for the missing data [54]. For longitudinal health studies, integrating the MI approach to clustering will help reduce the uncertainty of imputation, hence improving the accuracy of the clustering [2]. The MI-based fuzzy clustering (MIFuzzy) Clustering algorithm has been developed in OSs and random controlled trials (RCTs) to cope with real-world longitudinal data that are error-prone, nonnormal, high dimensional, and contain missing and zero-inflated values. It has evolved from its stepwise concept to a current iterative integrated MIFuzzy clustering model by learning the features and data structure in real OS and RCT, and comparing to other major pattern recognition methods in real data and simulation. MIFuzzy represents a full theoretical integration of MI, fuzzy-logic-based clustering, and visualization-aided validation for trajectory pattern analysis in longitudinal studies. It addresses: the extent to which an individual's behaviors partially involves them in more than one cluster, e.g., due to food-intake changes over time (technically, clusters touch or overlap); nonnormal, high-dimensional longitudinal data with missing values and zero-inflation; and the need to visualize and validate patterns. MIFuzzy embedded visualization-aided pattern-validation process is replicable in contrast to most clustering models that generate clusters without or unclear verification. MIFuzzy uses observed scores to capture individual behavioral change over time and identify latent clusters in populations that describe distinct behavioral trajectories. MIFuzzy approach to testing outcomes remains novel in that it relates identified behavioral trajectory patterns that account for individual behavior changes and variations influenced by different contexts over time, to other important risk factors (e.g., demographics and psychosocial variables) and outcomes (e.g., obesity, diabetes, and CVD).

Hence, this research aims to develop a decentralized version of MIFuzzy for longitudinal data - specifically, dietary data - that preserves privacy as effectively as FL while having the benefit of MIFuzzy. A few papers have explored federated clustering. For example, Kumar et al. [42] have introduced a federated k-means clustering algorithm and evaluated its performance on classical ML problems. In this algorithm, each node transmits the means aggregated using federated averaging. Pedrycz developed a federated FCM algorithm [37] that functions similarly. However, these methods did not address the problem of incomplete data in their approaches. Thus, we propose the Federated MIFuzzy clustering, or FeMIFuzzy for short, a method that overcomes the existing problems for federated fuzzy clustering on longitudinal behavioral health data.

This method can cluster longitudinal behavioral health data from multiple decentralized clients, such as healthcare institutions or medical research centers, including those with varying time points and incomplete data, with only two rounds of communication.

III. Decentralized Federated MIFuzzy Clustering

This section outlines our proposed FeMIFuzzy algorithm which clusters data from multiple decentralized clients, such as healthcare institutions or medical research centers. Clustering input takes the form of a matrix or data frame; rows represent observations (patients or other individuals) and columns represent attributes (features) at given time points in the longitudinal study.

At each client, FeMIFuzzy involves three modules, including Sammon mapping, MI, and MIFuzzy clustering. The MI module handles the arbitrary missing patterns and is embedded in the FeMIFuzzy clustering procedure. Sammon Mapping is also integrated into this procedure to map high-dimensional data to a space of lower dimensionality. It preserves the inherent high-dimensional data structure in the lower dimension projection. By preserving the data structure, Sammon Mapping allows the construction of a global cluster model on longitudinal data sets containing different numbers of time points.

At the global stage, the clients' results are aggregated with respect to clients' sample sizes to find the optimal number of clusters and global centroids.

We assume that all clients possess an adequate number of sample observations - a number larger than the potential number of clusters in the data - compared to other clients. In addition, each client requires the capability to communicate model parameters with every other client for the model to function.

MI: The first module in FeMIFuzzy is MI. MI works by: generating multiple replacement values ("imputations") for missing data, resulting in many data sets with replaced missing information; and analyzing and integrating the results of the imputed sets. Let Q be the data distribution of the population. The MI module can be expressed as

$$Q_{\mathrm{MI}} = E_{Y_M|Y_O} E[Q(Y_O, Y_M)] \tag{1}$$

where Y_O is the observed data and Y_M is the missing data. Detailed discussions and evaluation are described for MIFuzzy in [2] regarding the three classical missing mechanisms, MCAR, MAR, and missing not at random (MNAR) [52], [53].

MIFuzzy Clustering: The second module in FeMIFuzzy is MIFuzzy. The first step for MIFuzzy is identifying attributes. In longitudinal multiple-component behavioral studies, the type and number of components are likely to differ, with varying numbers of time points. The MIFuzzy's objective function is as follows:

$$f'_m(X, U, V, \lambda) = \sum_{k=1}^{C}\sum_{i=1}^{N}(\mu_{ik})^w |x_i - v_k|_A^2 + \sum_{i=1}^{N}\lambda_k\left(\sum_{k=1}^{C}\mu_{ik} - 1\right) \tag{2}$$

where X denotes the attributes, U is the degree of membership for each subject ($i = 1, 2, \ldots, n$) in the respective cluster ($k = 1, 2, \ldots, C$) and V are the cluster centroids.

The advantage of MIfuzzy clustering is that rather than treating each point as a member of just one single cluster at each iteration, each observation possesses a degree of belonging to each cluster [2], [3], [4], [5], [23] which is stored in a matrix U. Equation (3) shows an example of how each imputed set U is structured - each row represents an observation, and each column represents a centroid, with the rows summing to 1

$$U_{\mathrm{MI}} = \begin{array}{c} \\ x_1 \\ \cdots \\ x_p \end{array} \begin{array}{c} \begin{array}{ccc} c_1 & c_2 & c_3 \end{array} \\ \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ \cdots & \cdots & \cdots \\ 0.2 & 0.3 & 0.5 \end{bmatrix} \end{array}. \tag{3}$$

U_{MI} is calculated using the two equations below. The degree of belonging $(u_{\mathrm{MI}})_{ij,t}$ between observation i and cluster j is calculated at each iteration as in (4), where m represents the fuzzifier. $(u_{\mathrm{MI}})_{ij,t}$ corresponds to the degree that a given observation belongs to a given cluster and represents the ijth entry of the partition matrix U

$$(u_{\mathrm{MI}})_{ij,t} = \frac{1}{\sum_{k=1}^{C}\frac{\|x_i - V_{j,t}\|^2}{\|x_i - V_{k,t}\|^2}^{\frac{2}{m-1}}}. \tag{4}$$

Equation (5) provides each updated centroid V_j at iteration $t+1$

$$(V_{\mathrm{MI}})_{j,t+1} = \frac{\sum_{i=1}^{N}(u_{\mathrm{MI}})_{ij,t}^m x_i}{\sum_{i=1}^{N}(u_{\mathrm{MI}})_{ij,t}^m}. \tag{5}$$

Following MIFuzzy validation procedure, the clustering results from each imputed data set are first analyzed at each client, involving averaging validation indices across imputed data sets, e.g., 10 imputed sets for our included longitudinal studies, matching centroids and cluster labels across the imputed data sets. To validate the clustering results and identify the optimal number of clusters, several validation indices were revised in the framework of MI-based clustering validation [5]: Xie-Beni index, which is well known for fuzzy clustering, the lower the better [55], Silhouette Score, a popular score measuring the goodness of a clustering technique, ranging from -1 to 1, the higher the better [56], and Inertia, which is the target function for MIFuzzy clustering. For cluster label and centroid matching, multiple criteria are used, including the number of observations, mean values, minimum values, maximum values, standard deviations, and median values in each cluster. The identified centroids at each client are the average values of the matched centroids across the imputed data sets.

Sammon Mapping The third module of FeMIFuzzy is Sammon Mapping. It is the module that allows FeMIFuzzy to construct a global cluster model on longitudinal data sets containing different numbers of time points. Sammon mapping projects the longitudinal data to a lower dimensional space while preserving the original structure of interpoint distances in high-dimensional space. The Sammon Mapping's

target function, also known as Sammon's Stress, is as follows:

$$E_{\mathrm{MI}} = \frac{1}{\sum_{i<j}(d_{\mathrm{MI}})^*_{ij}} \sum_{i<j} \frac{\left((d_{\mathrm{MI}})^*_{ij} - (d_{\mathrm{MI}})_{ij}\right)^2}{(d_{\mathrm{MI}})^*_{ij}} \tag{6}$$

where $(d_{\mathrm{MI}})^*_{ij}$ denotes the distance between the ith and jth objects in the original space and $(d_{\mathrm{MI}})_{ij}$ the distance in the projections.

Federated MIFuzzy Global Model: At the global stage, the clients' results are aggregated with respect to clients' sample sizes to find the optimal number of clusters and the global centroids. To generate the global model, three primary factors are considered: 1) the clients need to agree on the optimal number of clusters; 2) the cluster labels across the clients are correctly matched together; and 3) all clients contain the same features. To identify the optimal number of clusters, we developed an observation-weighted majority voting, where the client with the most observations, i.e., with the largest number of subjects, has the most votes. To match the cluster labels and centroids across the clients, we also use the mean values, minimum values, maximum values, standard deviations, and median values in each cluster to match them across the clients. To handle varying numbers of time points, each client performs MIFuzzy and Sammon Mapping modules on its own side before communicating with other clients. The centroids are projected on the Sammon space along with calculating the mean, minimum, maximum, standard deviations, and median of the cluster centroids. After that, we can calculate the global model.

First, to identify the global optimal number of clusters, we use

$$c_{\mathrm{global}} = \sum_{k=1}^{M} \frac{n_k}{N} c_k \tag{7}$$

where n_k is the number of observations from client k, N is the total number of observations across all the clients, and c_k is the vote for the optimal number of clusters from client k.

After identify the global optimal number of cluster c_global, we can generate the global centroids. Let $V_{j,k}$ represent a final given centroid j for the client k after the convergence of FeMIFuzzy and M is the number of clients. Then, (8) calculates each final centroid V^*_j for the global cluster model by weighted-averaging the centroid j across all clients where the weight is the ratio between the client's number of observations and the total number of observations. This average is weighted by the number of samples in each client, so clients with more observations have a greater influence on the global model

$$V^*_j = \sum_{k=1}^{M} \frac{n_k}{N} V_{j,k}. \tag{8}$$

Hence, (8) describes the global centroids produced by FeMIFuzzy. These steps are included in Algorithm 1 and depicted graphically in Fig. 2.

FeMIFuzzy Communication Architecture: The process of calculating the global model using data from across many clients is modeled as a decentralized parallel system. In this architecture, each client runs the FeMIFuzzy algorithm simultaneously in parallel until a round of communication is required. There are two rounds of communication for our proposed FeMIFuzzy. In the first round, clients communicate the data dictionary, including the number of time points, the number of attributes, and their descriptions. The identified attributes used for clustering must use all intervention or treatment components and include all repeated measures of components. In the second, each client transmits their calculated centroids and their vote for the optimal number of clusters to every other client. This communication process between clients is depicted graphically in Fig. 3. From the communicated votes of the optimal number of clusters and centroids, FeMIFuzzy uses (7) and (8) to generate the global model(as explained in Algorithm 1).

Algorithm 1 Federated MIFuzzy

1) **Communication Round 1:** The clients communicate the data dictionary to identify the attributes that:
   a) use all intervention components
   b) include repeated measures of components.
2) For each client:
   a) Perform Sammon mapping for the data to reduce dimension to 2-D space.
   b) for each number of clusters for 1 to M: perform MIFuzzy Clustering
   c) Calculate the Xie-Beni Index for each number of clusters
   d) Calculate the Silhouette Score for each number of clusters
   e) For each number of clusters, match the centroids and clusters across the imputed data sets using the minimum Euclidean distance between imputed centroids
   f) For each number of clusters, calculate the final client's centroids by averaging the values of the imputed data sets' centroids
   g) Each imputed set selects the optimal number of clusters based on a clustering quality index such as the XB_{MI} index and some criterion - for example, the "elbow method" which selects the first k clusters at which the clustering index ceases to improve.
   h) Vote for the clients' cluster number using majority voting across the imputed sets
3) **Communication Round 2:** Each client transmits their Sammon-projected centroids for each number of clusters along with the vote for the optimal number of clusters to all other clients.
4) Use Equation (7) to decide the global optimal number of clusters
5) Match the centroids and clusters across the clients using the minimum Euclidean distance between clients' centroids
6) Calculate the global centroids using Equation (8).

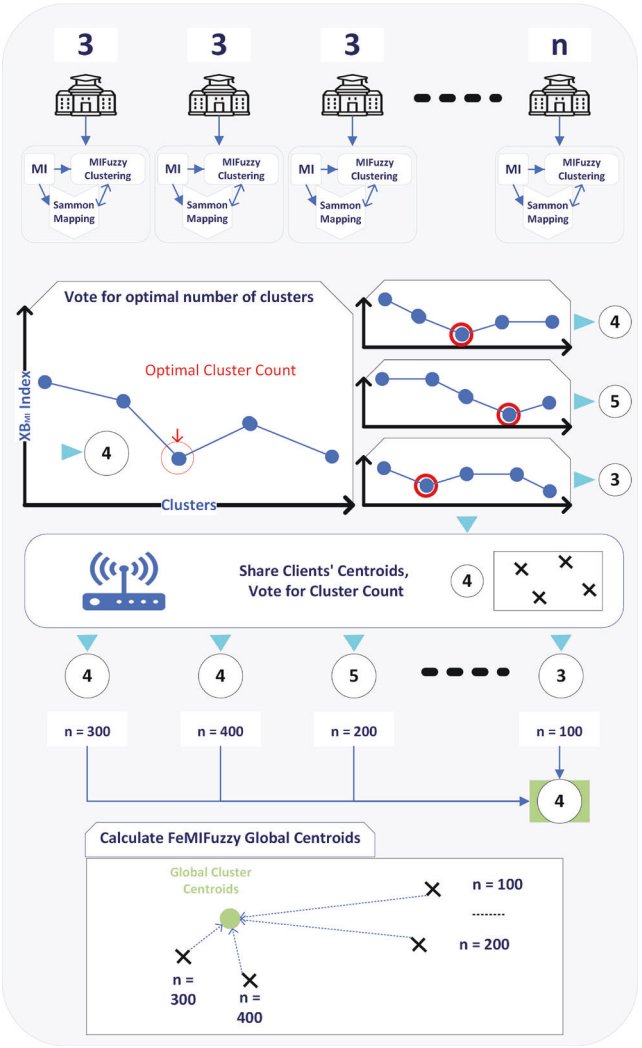

Fig. 2. Graphical depiction of the FeMIFuzzy process. At each client, FeMIFuzzy involves three modules, including Sammon mapping, MI, and MIFuzzy clustering. Validation indices, such as XB_{MI}, identify the optimal number of clusters at each client. At the global stage, the clients' results are aggregated with respect to clients' sample sizes to find the optimal number of clusters and the global centroids.

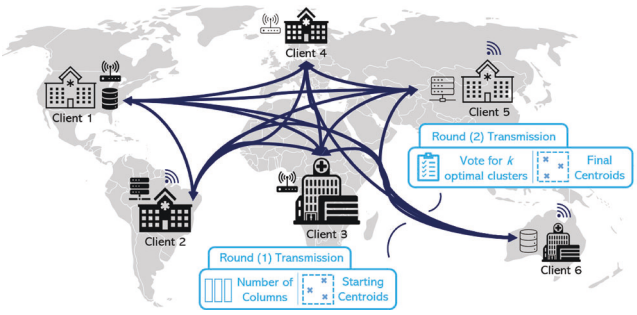

Fig. 3. Communication process in the proposed FeMIFuzzy method. Each client transmits parameters, such as the model centroids $V_{j,k}$, at each round of communication.

IV. Evaluation

The FeMIFuzzy algorithm was implemented in MATLAB and empirically evaluated on a desktop with an Intel Core i7-6700 CPU. FeMIFuzzy was evaluated using both real data and simulation. First, the algorithm was applied to a set of harmonized dietary data collected from four different clients. This harmonized data set contains sets of observations provided by multiple different healthcare institutions that may require clustering while maintaining privacy. Then, FeMIFuzzy was applied to simulated data with varying effect sizes, correlations, number of clients, and sample sizes to understand its performance under different conditions better.

TABLE II
Number of Time Points and Samples in Each Empirical Client

Empirical Client	Time Points	n Samples
1	4	191
2	4	240
3	3	252
4	2	570

Dietary Study Data: First, this study performed clustering on a harmonized data set containing observations from four longitudinal dietary health studies, as described in [10]. These studies survey participants about their diet habits, recording the types of foods that they consume over 24 h three times in two weeks over a varied number of study years. For each participant, eight variables were recorded, each describing a different quantity of food (vegetables, meat, etc.) that the participant consumed in the study period. These were then calculated to obtain individual dietary quality indices; in this study, we focus on the Alternate Healthy Eating Index (AHEI-2005) [57].

Simulated Data: In our simulation, each dietary study represents a separate client in the distributed algorithm. Each client data set contains 2, 3, or 4 attributes, each representing an AHEI-2005 score at a different time point in the longitudinal study. Table II displays the number of samples and time points for each client. To simulate the data using parameters generated from our longitudinal dietary studies, we randomly generated cluster centroids a minimum distance apart, with distances of 0.2 (small effect size, typical of dietary health studies), 0.5 (medium effect), and 0.8 (large effect) [58]. Then, we sampled new observations from a multivariate distribution centered on each cluster centroid to generate each new cluster. We also use different number settings of correlation for each distribution generated, including small, medium, and large correlation [58].

Synthetic data sets containing 4, 8, and 100 clients with different total sample sizes were simulated. In each data set, different numbers of clusters (3, 4, 5, 6) are simulated and labeled. Observations were drawn from a multivariate Gaussian distribution with degrees of freedom equal to the number of samples for each effect size. The FeMIFuzzy algorithm was then evaluated on each data set.

A. FeMIFuzzy Evaluation Methods

The performance of FeMIFuzzy was evaluated using several metrics. The algorithm was repeated across 10 trials for each client. Each client also has multiple imputed sets. The metric means and standard errors are reported in Table III. The metrics recorded are as follows.

TABLE III
FEMIFUZZY PERFORMANCE ACROSS CLIENTS ON EMPIRICAL DATA

Client	mean (±SD) 1		2		3		4		Global Model	
XB_{MI} Index	0.332	(±0.0014)	0.275	(±0.0009)	0.419	(±0.0097)	0.388	(±0.0000)	0.3641	(±0.0007)
$Silhouette_{MI}$ Score	0.257	(±0.0085)	0.299	(±0.0008)	0.3376	(±0.0044)	0.4643	(±0.00)	0.3755	(±0.0008)
$Inertia_{MI}$	1727	(±4.9992)	2375	(±0.5819)	1980	(±3.2923)	4078	(±0.0135)	8886	(±4.93)
Largest Cluster	0.26	(±0.0099)	0.231	(±0.0079)	0.28	(±0.0116)	0.245	(±0.00)	0.25	(±0.05)
Iterations	114	(±8.6894)	146	(±17.6278)	130	(±13.1918)	92	(±8.4704)		
Time (s)	6.1442	(±0.134)	8.256	(±0.1524)	8.95	(±0.8433)	9.6367	(±0.3676)		

1) *XB_{MI} Index* The averaged Xie-Beni index value across imputed data sets, representing the ratio of the compactness of each cluster to their separation [55], taking into account degrees of membership in each cluster. Smaller values are preferable since they indicate more compact and separate clusters. The $\mathrm{XB_{MI}}$ index is computed as in (9). This metric is standard for fuzzy clustering evaluation and has been used previously [2]

$$\mathrm{XB_{MI}} = \sum_{\mathrm{mi}=1}^{10} \frac{\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m}||V_i - X_j||^2}{n\left(\min_{i\neq j}||V_i - V_j||^2\right)} \Big/ 10. \quad (9)$$

2) *$Silhouette_{MI}$ Score:* The averaged Silhouette score imputed data sets, measuring how well entries fit their own cluster compared to others, on average, and ignores fuzzy degrees of belonging. A higher score is preferable, and a positive score indicates that the algorithm has functioned effectively [56].
3) *$Inertia_{MI}$*, or the averaged within-cluster sum of squares across imputed data sets, is the measure that FeMIFuzzy seeks to minimize, but is only useful for examining individual trials since scenarios with more observations will always feature higher inertia. Lower inertia is preferable.

Finally, we report the total number of iterations for convergence and the time FeMIFuzzy taken in seconds to indicate whether the algorithm converges at a reasonable speed. For FeMIFuzzy, we used a fuzzifier of m = 2.7 using a tolerance of $1e-4$. The fuzzifier was selected based on the optimal choice from a range of fuzzifiers for dietary data from previous studies [2], [59].

With empirical dietary study data, we also report the proportion of observations in the largest cluster to demonstrate that the algorithm does not simply place every observation into one cluster, which would not be useful.

B. FeMIFuzzy Evaluation Methods

1) Dietary Study Results Using FeMIFuzzy: After applying the FeMIFuzzy clustering algorithm to the empirical data set with four clients, we found that the output matched expectations. The algorithm consistently selected 5 clusters as the optimal using the majority vote method. Fig. 4 represents how the algorithm uses MIFuzzy and Sammon mapping to cluster and visualize data. The final global centroids are calculated using a weighted average from the clients' centroids.

FeMIFuzzy results across all four clients for 5 clusters are displayed in Table III. Though metrics vary across clients, we

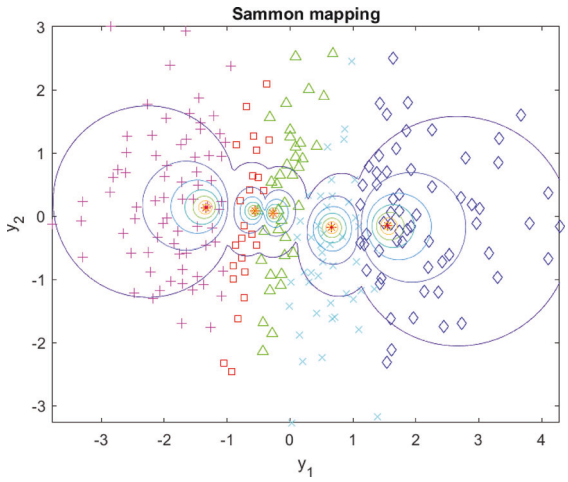

Fig. 4. Sammon mapping of the dietary study data set with clustering results from FeMIFuzzy.

observe consistently low-$\mathrm{XB_{MI}}$ values of roughly 0.27 to 0.42 and positive $Silhouette_{\mathrm{MI}}$ scores of 0.25 to 0.47. These metrics also displayed low-standard error, indicating that performance was consistent across repeated trials or imputed data sets. The largest cluster size only made up 23-28% of observations, confirming the method's functionality for further analysis. Finally, the algorithm converged relatively quickly, averaging between 92 to 146 iterations per client and only requiring 8-9 s on average across clients to run in MATLAB (MATLAB R2022b [60]).

For most evaluation metrics, the global model yielded similar or improved performance compared to individual clients. With an average $Silhouette_{\mathrm{MI}}$ score of 0.33 and $\mathrm{XB_{MI}}$ index of 0.35, we can be assured quality clusters are still produced. Furthermore, while the $inertia_{\mathrm{MI}}$ was larger than that of any individual client, it accounts for observations across all clients. The global model inertia of 8886 was lower than the summed average inertia across individual clients, which totaled about 10 160. Hence, regarding inertia, the global model actually *improved* over individual cluster models.

2) Simulation Results Using FeMIFuzzy (Distributions): After empirical evaluation, we applied FeMIFuzzy to a variety of simulated data sets with different numbers of clusters and clients, different distances between centroids (representing effect sizes), and correlation. For distribution, we use the Gaussian distribution. Fig. 5 displays examples of how data were distributed for a 2-D client across distributions and effect sizes. As depicted, larger effect sizes yielded more easily

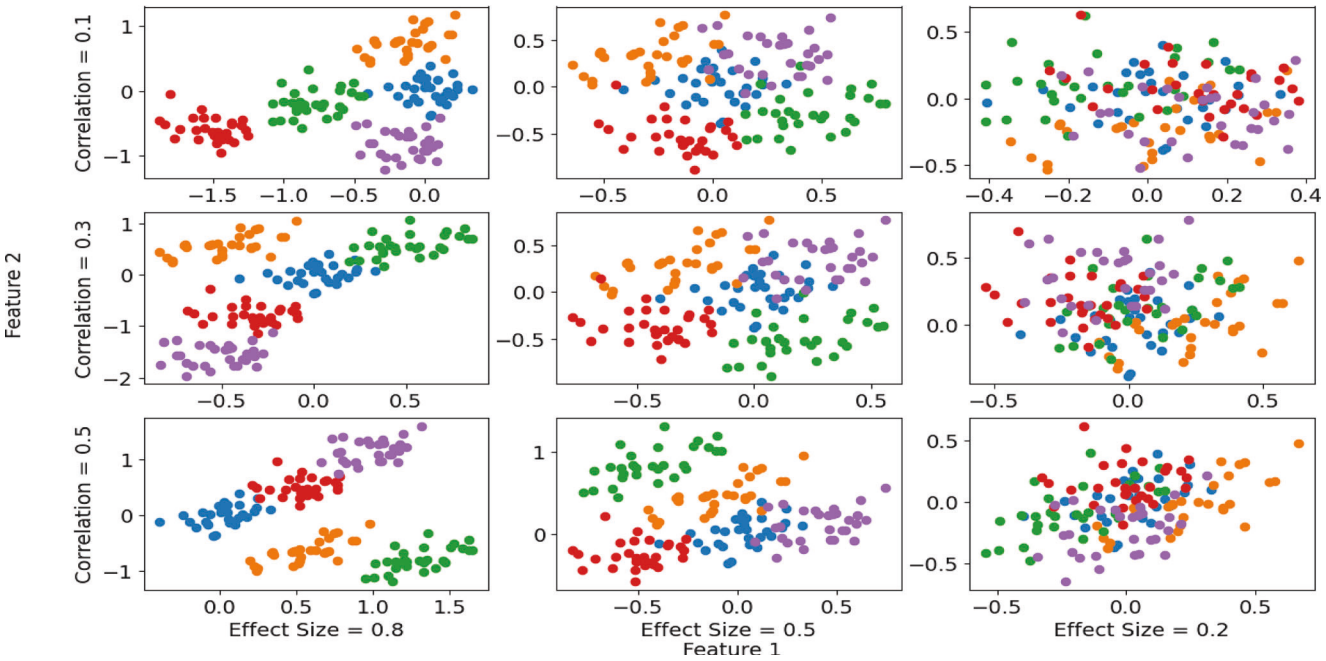

Fig. 5. Examples of clustering problems on 2-feature client, comparing different effect sizes and distributions for five simulated clusters.

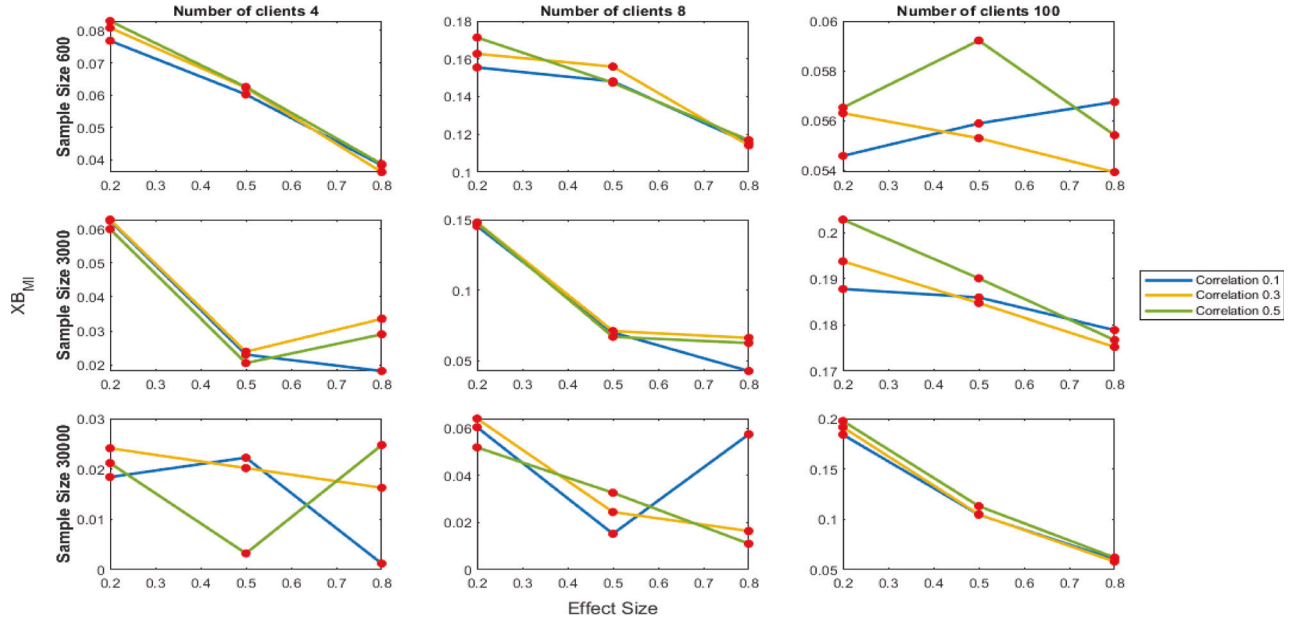

Fig. 6. Median XB_{MI} index attained by FeMIFuzzy global cluster solution on simulated data ranging from 4 to 100 clients. Graph compares scenarios with different numbers of clusters, effect sizes, and sample sizes.

separable clusters, while small effect sizes resulted in clusters that were difficult to separate by the human eye. Subtly different spreads of data points also occurred between small, medium, and large correlations.

Clustering Index: We started by evaluating scenarios with different numbers of clients, including 4, 8, and 100 clients. Fig. 6 displays the average global XB_{MI} index across different effect sizes from the global cluster solution for each given number of clusters with the total sample sizes at 600, 3000, and 30 000 across different effect sizes and correlations. An average XB_{MI} index in the range of [0,1] indicates that the clustering solutions were of high quality, as this indicates that the average cluster separation was larger than the average cluster variation - in other words, clusters did not overlap. We can observe, on average that the larger the sample size, the smaller the global XB_{MI} and clustering quality. Across most clients, correlations, and effect sizes, as measured by the average XB_{MI} index, FeMIFuzzy consistently produced high-quality clusters. Better XB indices tended to be produced for larger effect sizes and larger sample sizes.

Across the majority of scenarios with reasonable numbers of clients, varied correlation, and effect sizes, as measured by the

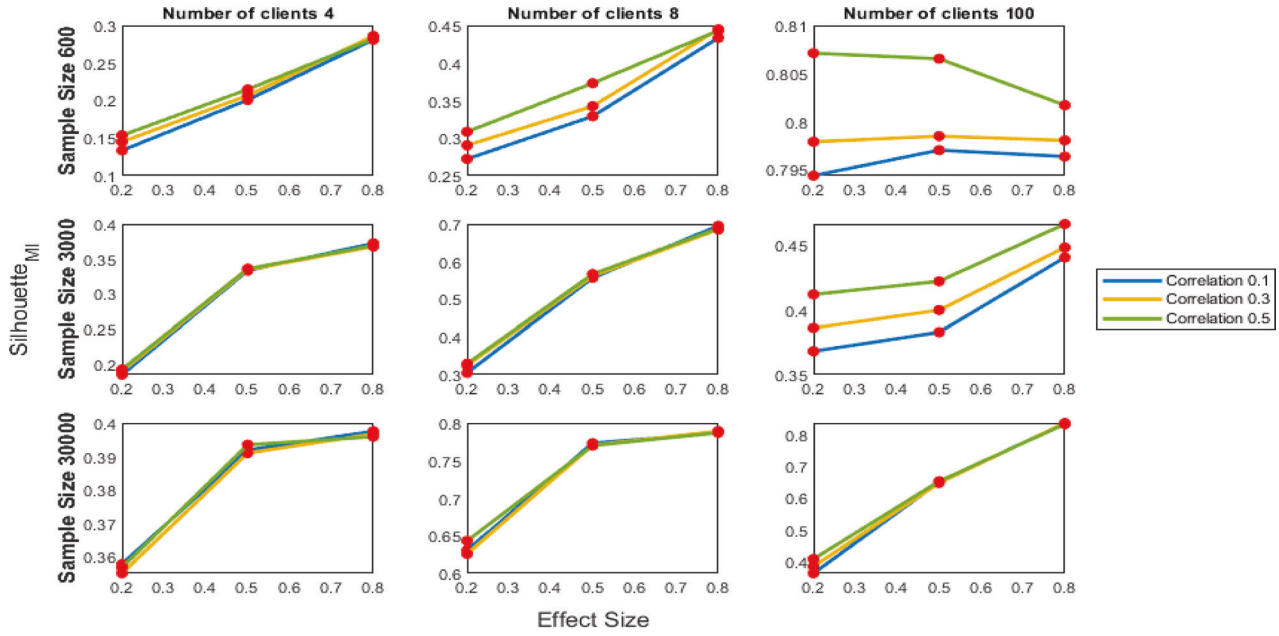

Fig. 7. Average silhouette score attained by FeMIFuzzy global cluster solution on simulated data ranging between 4, 8, and 100 clients. Graph compares scenarios with different numbers of clusters, effect sizes, and sample sizes.

average XB_{MI} index, FeMIFuzzy consistently produced high-quality clusters. Better XB_{MI} indices tended to be produced for larger effect sizes and larger sample sizes. However, for smaller sample sizes, FeMIFuzzy sometimes produced poor outlier solutions with much larger XB_{MI} indices. Hence, it is important to have a sufficient sample size for each client, at least 5 samples per cluster to ensure that the algorithm does not select a suboptimal clustering result.

We can also measure the cluster quality using the average $Silhouette_{MI}$ score, which does not factor in the "degree of membership" feature of the fuzzy clustering solution. Fig. 7 displays the average $Silhouette_{MI}$ score across 4, 8, and 100 clients for each scenario for each given number of clusters with sample sizes of 600, 3000, and 30 000. A positive $Silhouette_{MI}$ score indicates a quality clustering solution, informing us that, on average, observations fit better into their assigned cluster than into other clusters.

Based on the silhouette scores, FeMIFuzzy always produced quality clustering solutions across client counts. Like the XB_{MI} index, the algorithm produced better solutions at larger effect sizes, with the highest average silhouette scores tending to occur at the larger effect size of 0.8. However, similar to XB_{MI}, for a large number of clients at 100 clients and a small sample size of 600, the average $Silhouette_{MI}$ decreases at the effect size 0.8 with medium and high correlation. Regardless, the overall results show that the algorithm mostly produced average $Silhouette_{MI}$ scores in the positive range, with $Silhouette_{MI}$ scores up to 0.8 in the case of effect size 0.8.

Clustering Performance: Aside from cluster indices, we also evaluate FeMIFuzzy performance based on the clustering accuracy and the number of iterations. To calculate the clustering accuracy, we keep labels of each cluster when generating the simulated data. The global labels are calculated using Euclidean distance from the global centroids on the 2-D Sammon Mapping space. The final accuracy is then calculated from the confusion matrix between the predicted and assigned labels.

Fig. 8 displays the accuracy score across 4, 8, and 100 clients for each scenario for each given number of clusters with sample sizes of 600, 3000, and 30 000. The method shows better accuracy at a larger effect size compared to a lower effect size as expected. Additionally, the accuracy increases as the sample size increases. At a sample size of 30 000, the accuracy goes above 90% accuracy for both effect sizes of 0.5 and 0.8 across difference correlations, with numbers of clients 4 and 8. At 100 clients, FeMIFuzzy still shows good accuracy at around 80% accuracy. Furthermore, for comparison, we also run a decentralized federated FCM method (FFCM) on the same simulated data sets [40]. This comparison method also uses 2 rounds of communication. However, the comparison method does not use MI to deal with the missing data (missing rate = 20%). To deal with varying timepoints, FFCM used PCA instead of Sammon Mapping. The results show significant improvements using FeMIFuzzy compared to the FFCM in almost all cases with a sufficient sample size. It shows that with a sufficient sample size per client and an effect size of 0.5 or 0.8, FeMIFuzzy can achieve high accuracy ranging from 60-80%.

Finally, FeMIFuzzy converged rapidly. Fig. 9 displays the algorithm's convergence speed, measured both in the number of iterations and computational time. Convergence speed was consistent across clients and distributions, generally falling in the range of about 10-80 iterations. The algorithm FeMIFuzzy converges much more quickly on data sets with larger effect sizes - a sensible result, given that larger effect sizes result in more clearly separable clusters. Additionally, on average, FeMIFuzzy needs fewer iterations when the number of samples per client is sufficiently large. The reduction in iterations,

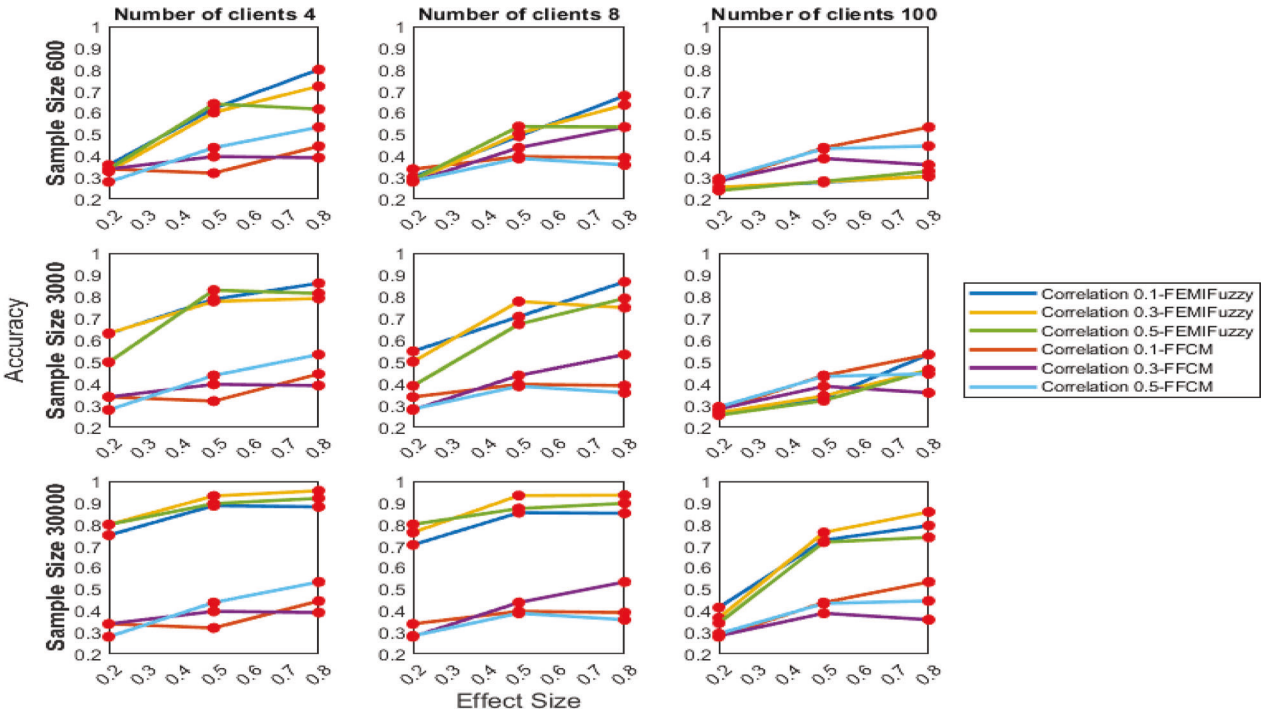

Fig. 8. Accuracy rate of FeMIFuzzy algorithm on simulated data ranging from 4 to 100 clients. Accuracy is computed using the predicted label from FeMIFuzzy and the assigned label from the simulated data; the graph compares scenarios with different numbers of clusters, effect sizes, and sample sizes.

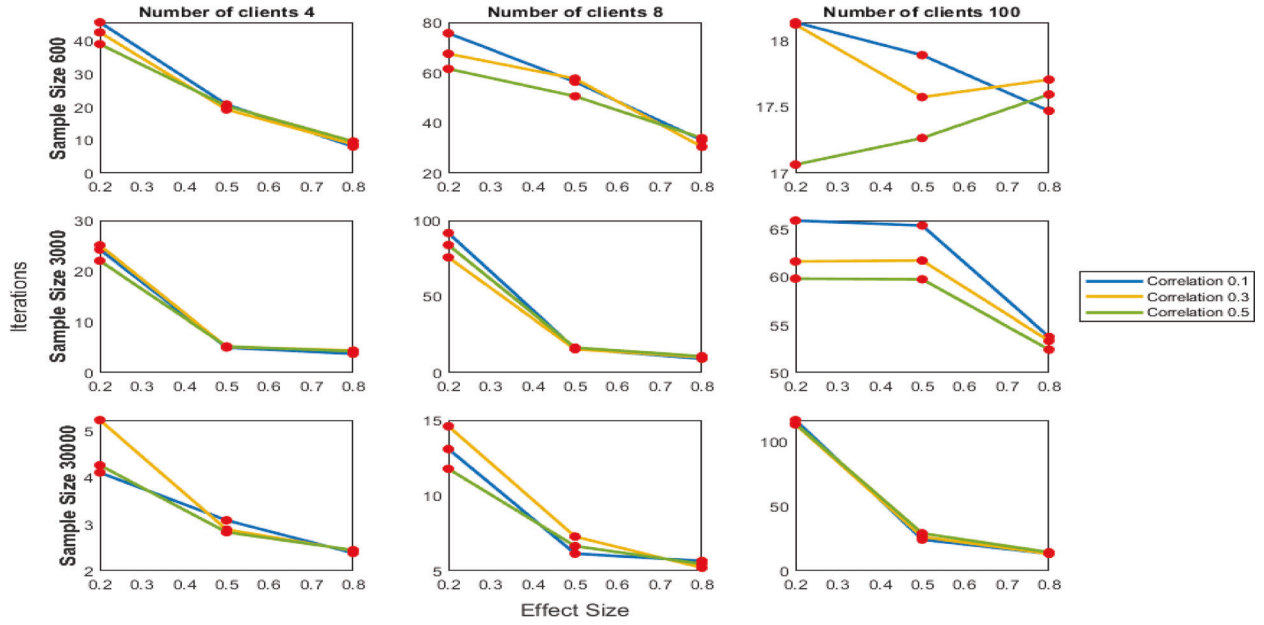

Fig. 9. Convergence rate of FeMIFuzzy algorithm on simulated data ranging from 4 to 100 clients. Convergence is measured in time (in seconds) and iterations; the graph compares scenarios with different numbers of clusters, effect sizes, and sample sizes.

however, does not reduce the execution speed significantly as the execution speed correlates more with the sample size.

Voting Method Evaluation: While FeMIFuzzy produced high-quality solutions in our simulations, it is also worth considering whether its selection of the optimal *number* of clusters produced results that matched the actual number of clusters generated. FeMIFuzzy performed better in finding the optimal number of clusters when effect sizes were large (0.8), especially in the 4-cluster scenario where it voted correctly in every instance for both distributions. When effect sizes were small (0.2), FeMIFuzzy tended to select higher numbers of clusters. Given a small effect size, minuscule trivial clusters could appear. This would happen in empirical longitudinal behavioral studies where behaviors fluctuate and clusters may overlap. When analyzed more closely, the trivial clusters (those with few cases) could be merged with larger nearby clusters.

In Table IV, we compare the mode vote for the 5 cluster scenario across all clients, rounding up in the presence of ties.

TABLE IV
MODE NUMBER OF CLUSTERS SELECTED BY FEMIFUZZY (COMPARED TO FIVE TRUE CLUSTERS)

Correlation	Effect Size / Number of Clients	0.3	0.5	0.8
0.1	4	5	4	5
	8	6	4	5
	100	5	4	4
0.3	4	5	4	5
	8	6	5	5
	100	5	5	4
0.5	4	5	5	5
	8	6	4	5
	100	5	5	4

Results did not particularly vary across different client counts, and FeMIFuzzy always either identified the correct number of clusters or was one cluster off (meaning that it merged two very similar clusters or identified a small trivial cluster in a larger existing cluster). As such, the results were relatively consistent.

V. DISCUSSION

The FeMIFuzzy algorithm described in this article supports fuzzy clustering among a decentralized network of clients. This allows many healthcare institutions to cooperatively construct global cluster models while preserving data privacy. FeMIFuzzy also overcomes longitudinal data issues, permitting clustering on data sets that are incomplete, may not share the same number of attributes (features or columns), incorporating observations from all clients in each cluster, and requiring only two round of communication.

On empirical dietary data, FeMIFuzzy converged rapidly and achieved desirable clustering performance on various metrics, including XB_{MI} index and $Silhouette_{MI}$ score. On simulated data, FeMIFuzzy also demonstrated consistent high-$Silhouette_{MI}$ scores and low XB_{MI} indices across different numbers of clients, effect sizes, numbers of clusters, correlations, and sample sizes. It converged faster regarding the number of iterations for clusters of larger effect sizes - a logical outcome given that such clusters appear more distinctly partitioned. Additionally, using labels from simulated data, we show that FeMIFuzzy can achieve very high accuracy in clustering, especially with sufficient sample size ($\tilde{3}0$ per client) and large effect size. Finally, FeMIFuzzy voted for the optimal number of clusters accurately. Based on these results, our method appears suitable for use on longitudinal dietary health data and similar studies.

While this study used the elbow method heuristic with weighted voting to select the number of clusters based on the XB index, our proposed approach can easily be extended to many other methods to find the optimal number of clusters. For example, this method could be adapted to methods like the gap statistic [61]. Other voting schemes could be proposed as well; for instance, in some scenarios, it may be more appropriate to perform a simple majority vote rather than a weighted average of votes. In addition, we rounded cluster count votes down, which may have caused FeMIFuzzy's underestimation of the true number of clusters observed; changing this assumption may result in higher accuracy in cluster count estimation, as our evaluation approach was more conservative.

This method does contain limitations. Similar FL methods often assume that data is identically and independently distributed (IID) among clients [19]. Though the dietary data used in this study is non-IID, more extreme data distributions, such as large imbalances in the sample sizes across clients, might result in poor performance. Hence, it may be worth adopting strategies for non-IID data from existing FL research [62], [63].

In future work, we plan to integrate FeMIFuzzy into existing work in pattern recognition for longitudinal behavioral health trials, and broader application in the IoMT. Specifically, we also can extend the model to our neuro-fuzzy classification model [64] for predicting clinical outcomes of interventions depending on patient characteristics.

Our proposed FeMIFuzzy clustering approach can identify clusters in incomplete longitudinal behavioral health data distributed across several disparate healthcare institutions or other clients. By protecting individual patient privacy and allowing the clustering of longitudinal incomplete data containing different numbers of time points, the algorithm can improve health research and treatment plans.

Further research testing the performance of federated fuzzy clustering on a real-world platform would provide better performance insights, yielding a valuable tool for improving targeted patient treatments while preserving privacy.

REFERENCES

[1] S. J. Mooney and V. Pejaver, "Big data in public health: Terminology, machine learning, and privacy," *Annu. Rev. Public Health*, vol. 39, no. 1, pp. 95–112, Apr. 2018, doi: 10.1146/annurev-publhealth-040617-014208.
[2] H. Fang, "MIFuzzy clustering for incomplete longitudinal data in smart health," *Smart Health*, vols. 1–2, pp. 50–65, Jun. 2017, doi: 10.1016/j.smhl.2017.04.002.
[3] Z. Zhang, H. Fang, and H. Wang, "A new MI-based visualization aided validation index for mining big longitudinal Web trial data," *IEEE Access*, vol. 4, pp. 2272–2280, 2016, doi: 10.1109/access.2016.2569074.
[4] Z. Zhang and H. Fang, "Multiple-vs non-or single-imputation based fuzzy clustering for incomplete longitudinal behavioral intervention data," in *Proc. IEEE 1st Int. Conf. Connect. Health Appl., Syst. Eng. Technol. (CHASE)*, 2016, pp. 219–228, doi: 10.1109/chase.2016.19.
[5] Z. Zhang, H. Fang, and H. Wang, "Multiple imputation based clustering validation (MIV) for big longitudinal trial data with missing values in eHealth," *J. Med. Syst.*, vol. 40, no. 6, p. 146, Apr. 2016, doi: 10.1007/s10916-016-0499-0.
[6] S. S. Kim et al., "Acculturation, depression, and smoking cessation: A trajectory pattern recognition approach," *Tob. Induc. Dis.*, vol. 15, no. 1, p. 33, Jul. 2017, doi: 10.1186/s12971-017-0135-x.
[7] H. Fang, C. Johnson, C. Stopp, and K. A. Espy, "A new look at quantifying tobacco exposure during pregnancy using fuzzy clustering," *Neurotoxicol. Teratol.*, vol. 33, no. 1, pp. 155–165, Jan. 2011, doi: 10.1016/j.ntt.2010.08.003.
[8] K. Clark et al., "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *J. Digit. Imag.*, vol. 26, no. 6, pp. 1045–1057, Jul. 2013, doi: 10.1007/s10278-013-9622-7.
[9] J. Borovec et al., "ANHIR: Automatic non-rigid histological image registration challenge," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3042–3052, Oct. 2020, doi: 10.1109/tmi.2020.2986331.
[10] V. S. Gurugubelli et al., "A review of harmonization methods for studying dietary patterns," *Smart Health*, vol. 23, Mar. 2022, Art. no. 100263. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352648321000763

[11] "Health insurance portability and accountability act of 1996," U.S. Department of Health & Human Services, Washington, DC, USA, Rep. CFR 164.502, 1996.
[12] (Office for Civil Rights, Dept. Health Hum Serv, Washington, DC, USA). *Summary of the HIPAA Privacy Rule*. Accessed: Jan. 3, 2022. [Online]. Available: https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html
[13] T. Schulte, "The protection of personal data in health information systems-principles and processes for public health," World Health Org. Reg. Office Europe, Copenhagen, Denmark, Rep. 2021–1994-41749-57154, 2020.
[14] L. Rocher, J. M. Hendrickx, and Y.-A. De Montjoye, "Estimating the success of re-identifications in incomplete datasets using generative models," *Nat. Commun.*, vol. 10, no. 1,p. 3069, Jul. 2019, doi: 10.1038/s41467-019-10933-3.
[15] V. Tresp, J. M. Overhage, M. Bundschus, S. Rabizadeh, P. A. Fasching, and S. Yu, "Going digital: A survey on digitalization and large-scale data analytics in healthcare," *Proc. IEEE*, vol. 104, no. 11, pp. 2180–2206, Nov. 2016, doi: 10.1109/jproc.2016.2615052.
[16] M. Chen, Y. Qian, J. Chen, K. Hwang, S. Mao, and L. Hu, "Privacy protection and intrusion avoidance for cloudlet-based medical data sharing," *IEEE Trans. Cloud Comput.*, vol. 8, no. 4, pp. 1274–1283, Oct. 2020, doi: 10.1109/tcc.2016.2617382.
[17] N. Rieke et al., "The future of digital health with federated learning," *NPJ Digit. Med.*, vol. 3, no. 1, Sep. 2020, doi: 10.1038/s41746-020-00323-1.
[18] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang, "Federated learning for healthcare informatics," *J. Healthc. Inform. Res.*, vol. 5, no. 1, pp. 1–19, Nov. 2020, doi: 10.1007/s41666-020-00082-4.
[19] J. Konečný, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated optimization: Distributed machine learning for on-device intelligence," 2016, *arXiv:1610.02527*.
[20] C. A. Mellins et al., "Sexual and drug use behavior in perinatally-HIV-infected youth: Mental health and family influences," *J. Amer. Acad. Child Adolesc. Psyc.*, vol. 48, no. 8, pp. 810–819, 2009.
[21] H. Fang, K. A. Espy, M. L. Rizzo, C. Stopp, S. A. Wiebe, and W. W. Stroup, "Pattern recognition of longitudinal trial data with nonignorable missingness: An empirical case study," *Int. J. Inf. Technol. Decis. Mak.*, vol. 08, no. 3, pp. 491–513, Sep. 2009, doi: 10.1142/s0219622009003508.
[22] H. Fang, Z. Zhang, C. J. Wang, M. Daneshmand, C. Wang, and H. Wang, "A survey of big data research," *IEEE Netw.*, vol. 29, no. 5, pp. 6–9, Sep. 2015, doi: 10.1109/mnet.2015.7293298.
[23] H. Fang and Z. Zhang, "An enhanced visualization method to aid behavioral trajectory pattern recognition infrastructure for big longitudinal data," *IEEE Trans. Big Data*, vol. 4, no. 2, pp. 289–298, Jun. 2018, doi: 10.1109/tbdata.2017.2653815.
[24] R. Little and D. Rubin, *Statistical Analysis with Missing Data* (Wiley Series in Probability and Statistics). Hoboken, NJ, USA: Wiley, 2014. [Online]. Available: https://books.google.com/books?id=AyVeBAAAQBAJ
[25] "A simple dietary message to improve dietary quality for metabolic syndrome (CANDO)." Accessed: Aug. 15, 2023. [Online]. Available: https://clinicaltrials.gov/study/NCT00911885
[26] "Diabetes management in low-income hispanic patients." Accessed: Aug. 15, 2023. [Online]. Available: https://clinicaltrials.gov/study/NCT00848315
[27] "Lawrence latino diabetes prevention project (LLDPP)." Accessed: Aug. 15, 2023. [Online]. Available: https://clinicaltrials.gov/study/NCT00810290
[28] "Effectiveness of behavioral treatments for obesity and major depression in women." Accessed: Aug. 15, 2023. [Online]. Available: https://clinicaltrials.gov/study/NCT00572520
[29] S. M. Schüssler-Fiorenza Rose et al., "A longitudinal big data approach for precision health," *Nat. Med.*, vol. 25, no. 5, pp. 792–804, May 2019, doi: 10.1038/s41591-019-0414-6.
[30] K. W. DeGregory et al., "A review of machine learning in obesity," *Obesity Rev.*, vol. 19, no. 5, pp. 668–685, Feb. 2018, doi: 10.1111/obr.12667.
[31] W. Gou et al., "Interpretable machine learning framework reveals robust gut microbiome features associated with type 2 diabetes," *Diabetes Care*, vol. 44, no. 2, pp. 358–366, Dec. 2020, doi: 10.2337/dc20-1536.
[32] L. Khorraminezhad, M. Leclercq, A. Droit, J.-F. Bilodeau, and I. Rudkowska, "Statistical and machine-learning analyses in nutritional genomics studies," *Nutrients*, vol. 12, no. 10, p. 3140, Oct. 2020, doi: 10.3390/nu12103140.
[33] O. Sporns, G. Tononi, and R. Kötter, "The human connectome: A structural description of the human brain," *PLoS Comput. Biol.*, vol. 1, no. 4, p. e42, 2005, doi: 10.1371/journal.pcbi.0010042.
[34] C. Sudlow et al., "U.K. Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age," *PLoS Med.*, vol. 12, no. 3, Mar. 2015, Art. no. e1001779, doi: 10.1371/journal.pmed.1001779.
[35] M. J. Sheller et al., "Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data," *Sci. Rep.*, vol. 10, no. 1, Jul. 2020, Art. no. 12598, doi: 10.1038/s41598-020-69250-1.
[36] C. Li, G. Li, and P. K. Varshney, "Federated learning with soft clustering," *IEEE Internet Things J.*, vol. 9, no. 10, pp. 7773–7782, May 2022.
[37] W. Pedrycz, "Federated FCM: Clustering under privacy requirements," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 8, pp. 3384–3388, Aug. 2022.
[38] D. K. Dennis, T. Li, and V. Smith, "Heterogeneity for the win: One-shot federated clustering," in *Proc. 38th Int. Conf. Mach. Learn.*, 2021, pp. 2611–2620. [Online]. Available: https://proceedings.mlr.press/v139/dennis21a.html
[39] C. Briggs, Z. Fan, and P. Andras, "Federated learning with hierarchical clustering of local updates to improve training on non-IID data," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2020, pp. 1–9.
[40] M. Stallmann and A. Wilbik, "Towards federated clustering: A federated fuzzy *c*-means algorithm (FFCM)," 2022, *arXiv:2201.07316*.
[41] M. Zhao et al., "Substantial increase in compliance with saturated fatty acid intake recommendations after one year following the American heart association diet," *Nutrients*, vol. 10, no. 10, p. 1486, Oct. 2018, doi: 10.3390/nu10101486.
[42] H. H. Kumar, V. R. Karthik, and M. K. Nair, "Federated K-means clustering: A novel edge AI based approach for privacy preservation," in *Proc. IEEE Int. Conf. Cloud Comput. Emerg. Markets (CCEM)*, 2020, pp. 52–56, doi: 10.1109/ccem50674.2020.00021.
[43] E. Yoo, H. Ko, and S. Pack, "Fuzzy clustered federated learning algorithm for solar power generation forecasting," *IEEE Trans. Emerg. Topics Comput.*, vol. 10, no. 4, pp. 2092–2098, Oct.–Dec. 2022, doi: 10.1109/tetc.2022.3142886.
[44] X. Zhu, D. Wang, W. Pedrycz, and Z. Li, "Horizontal federated learning of Takagi–Sugeno fuzzy rule-based models," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 9, pp. 3537–3547, Sep. 2022, doi: 10.1109/tfuzz.2021.3118733.
[45] X. Liu, J. Zhao, J. Li, B. Cao, and Z. Lv, "Federated neural architecture search for medical data security," *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5628–5636, Aug. 2022, doi: 10.1109/tii.2022.3144016.
[46] E. H. Ruspini, J. C. Bezdek, and J. M. Keller, "Fuzzy clustering: A historical perspective," *IEEE Comput. Intell. Mag.*, vol. 14, no. 1, pp. 45–55, Feb. 2019, doi: 10.1109/mci.2018.2881643.
[47] V. S. Gurugubelli, Z. Li, H. Wang, and H. Fang, "eFCM: An enhanced fuzzy C-means algorithm for longitudinal intervention data," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, 2018, pp. 912–916, doi: 10.1109/iccnc.2018.8390419.
[48] H. Fang, V. Dukic, K. E. Pickett, L. Wakschlag, and K. A. Espy, "Detecting graded exposure effects: A report on an east boston pregnancy cohort," *Nicot. Tob. Res.*, vol. 14, no. 9, pp. 1115–1120, Aug. 2012, doi: 10.1093/ntr/ntr272.
[49] A. Torres and J. J. Nieto, "Fuzzy logic in medicine and bioinformatics," *J. Biomed. Biotechnol.*, vol. 2006, pp. 1–7, Mar. 2006, doi: 10.1155/jbb/2006/91908.
[50] D. B. Rubin, "Multiple imputation after 18+ years," *J. Amer. Stat. Assoc.*, vol. 91, no. 434, pp. 473–489, 1996, doi: 10.1080/01621459.1996.10476908.
[51] J. L. Schafer, *Analysis of Incomplete Multivariate Data*. Boca Raton, FL, USA: CRC Press, 1997.
[52] J. L. Schafer, "Multiple imputation: A primer," *Stat. Methods Med. Res.*, vol. 8, no. 1, pp. 3–15, 1999.
[53] J. L. Schafer and J. W. Graham, "Missing data: Our view of the state of the art," *Psychol. Methods*, vol. 7, no. 2, p. 147, 2002.
[54] P. Li, E. A. Stuart, and D. B. Allison, "Multiple imputation: A flexible tool for handling missing data," *JAMA*, vol. 314, no. 18, pp. 1966–1967, Nov. 2015, doi: 10.1001/jama.2015.15281.
[55] X. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 841–847, Aug. 1991.
[56] K. R. Shahapure and C. Nicholas, "Cluster quality analysis using silhouette score," in *Proc. IEEE 7th Int. Conf. Data Sci. Adv. Anal. (DSAA)*, 2020, pp. 747–748, doi: 10.1109/dsaa49011.2020.00096.

[57] M. L. McCullough et al., "Diet quality and major chronic disease risk in men and women: Moving toward improved dietary guidance," *Amer. J. Clin. Nutrit.*, vol. 76, no. 6, pp. 1261–1271, Dec. 2002.

[58] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Cambridge, MA, USA: Academic, 2013.

[59] H. Ngo, H. Fang, and H. Wang, "Poster: Intelligent fuzzifier-based cluster validation for incomplete longitudinal digital trial data," in *Proc. IEEE/ACM Conf. Connect. Health Appl., Syst. Eng. Technol. (CHASE)*, 2022, pp. 146–147.

[60] (MathWorks Comput. Softw. Corp., Natick, MA, USA). *MATLAB version: 9.13.0 (R2022b)*. (2022). [Online]. Available: https://www.mathworks.com

[61] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *J. Roy. Stat. Soc. Ser. B, Stat. Methodol.*, vol. 63, no. 2, pp. 411–423, 2001, doi: 10.1111/1467-9868.00293.

[62] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-IID data," 2018, *arXiv:1806.00582*.

[63] H. Zhu, J. Xu, S. Liu, and Y. Jin, "Federated learning on non-IID data: A survey," *Neurocomputing*, vol. 465, pp. 371–390, Nov. 2021, doi: 10.1016/j.neucom.2021.07.098.

[64] V. S. Gurugubelli, H. Fang, and H. Wang, "Neuro-fuzzy classifier for longitudinal behavioral intervention data," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, 2019, pp. 385–389, doi: 10.1109/iccnc.2019.8685574.