Addressing missing values and fragmented data to improve core and management processes in hospitals: a federated machine learning approach

Elion Shala, Jens O. Brunner, Jörg J. Vehreschild, Axel R. Heller, Christina C. Bartenschlager

Angaben zur Veröffentlichung / Publication details:

Shala, Elion, Jens O. Brunner, Jörg J. Vehreschild, Axel R. Heller, and Christina C. Bartenschlager. 2026. "Addressing missing values and fragmented data to improve core and management processes in hospitals: a federated machine learning approach." *Flexible Services and Manufacturing Journal*. https://doi.org/10.1007/s10696-025-09646-1.

Flexible Services and Manufacturing Journal
https://doi.org/10.1007/s10696-025-09646-1

Addressing missing values and fragmented data to improve core and management processes in hospitals: a federated machine learning approach

Elion Shala[1] · Jens O. Brunner[2,3] · Jörg J. Vehreschild[4] · Axel R. Heller[5] · Christina C. Bartenschlager[5,6]

Received: 8 October 2024 / Accepted: 21 November 2025

Abstract
Despite the increasing digitization of healthcare data, which offers opportunities to enhance core and management processes, hospitals frequently encounter challenges related to missing data and data fragmentation across institutions. In healthcare, missing data arise from both human and technological errors. Additionally, fragmented health data are often difficult to transfer due to their highly sensitive nature, which is subject to strict ethical and legal regulations. This study builds on the state-of-the-art Federated Machine Learning technique, Federated Averaging. We introduce two novel Federated Learning approaches: The Flexible and the Modified Federated Stochastic Variance Reduced Gradient (F-FSVRG and M-FSVRGS). These methods are designed to address missing and fragmented data with increased flexibility and robustness. We evaluate the performance of F-FSVRG and M-FSVRGS through a computational study using real-world multicenter COVID-19 diagnosis and Intensive Care Unit (ICU) data. M-FSVRGS generally achieves equal or higher accuracy than F-FSVRG, mostly independent of dataset size and structure, with statistically significant improvements observed for the ICU dataset when the proportion of missing values is high. Our findings demonstrate that M-FSVRGS effectively mitigates the challenges of fragmented and missing data, facilitating the integration of AI-driven approaches. Furthermore, this advancement supports sustainable improvements in healthcare processes and operations.

Keywords Federated machine learning · Missing data · Sparse data · Decision making in hospitals

Extended author information available on the last page of the article

Published online: 08 January 2026

1 Introduction

In hospitals, a multitude of decisions are made daily, impacting both the core process of patient care and management processes. For example, diagnosing a SARS-CoV-2 infection not only affects the patient's treatment but also necessitates capacity management due to required isolation and increased staffing demands. Similarly, reserving Intensive Care Unit (ICU) capacity for post-operative elective patients can influence surgical scheduling and ICU resource allocation. The breadth, intricacy, and volume of these decisions highlight the crucial need for digital, data-driven decision support systems for healthcare professionals and administrators in hospitals. Machine Learning (ML) algorithms can play a pivotal role in predicting outcomes such as ICU admissions or COVID-19 diagnoses. Digital COVID-19 diagnosis, for instance, enables a significantly expedited categorization of patients in emergency departments, thereby promoting the efficient and effective utilization of scarce resources (Bartenschlager et al. 2022).

Despite the facilitated digitization of healthcare data (Wang and Preininger 2019), which offers a multitude of opportunities to leverage patient outcomes (Wani and Malhotra 2018) and medical processes (Bartenschlager et al. 2023; Schiele et al. 2020), healthcare institutions often face the problems of (a) missing data (Carter and Busby 2023; Ismail et al. 2022; Luo 2022; Wells et al. 2013) and of (b) data being fragmented across different institutions (The Lancet Regional Health – Europe 2025; Xu et al. 2021). In the healthcare domain, missing data exists "due to potential reasons such as human errors, instrumental failures, and adverse death" (Ismail et al. 2022, p. 143). Fragmented health data is often difficult to transfer because health data in particular contains highly sensitive information that is subject to tight ethical and legal regulations in order to preserve data privacy and security (Rieke et al. 2020; Sarma et al. 2021; Van Panhuis et al. 2014; Xu et al. 2021). Data fragmentation creates a barrier for applying ML in healthcare (Xu et al. 2021), as both the performance and generalizability of ML models are highly dependent on a large volume of data (Obermeyer and Emanuel 2016; Oleynik et al. 2019; Sarma et al. 2021).

To break this barrier, Federated Machine Learning (FL), a framework of different local institutions collaborating to build a global ML model while keeping the sensitive data at their local institutions rather than sharing it, provides a remedy. FL enables different owners of fragmented healthcare data, such as hospitals, to interact with each other in a distributed manner (Xu et al. 2021) and simultaneously protect the privacy and security of their confidential patient data (Rajendran et al. 2021). Thereby, instead of exchanging data, institutions interact by periodically sending updates of their locally generated ML models to a central server, which aggregates them to a global model and then sends back the updated parameters of the global model to the institutions (Xu et al. 2021). This technology enables decisions to be made using a globally trained generalized model, rather than relying solely on locally trained models (see Fig. 1). While FL holds the potential to tackle the challenge of fragmented data, the issue of missing data in state-of-the-art FL approaches and existing literature is frequently neglected.

The purpose of this work is to expand upon the existing state-of-the-art FL technique, commonly known as Federated Averaging (FA), and to present an innova-

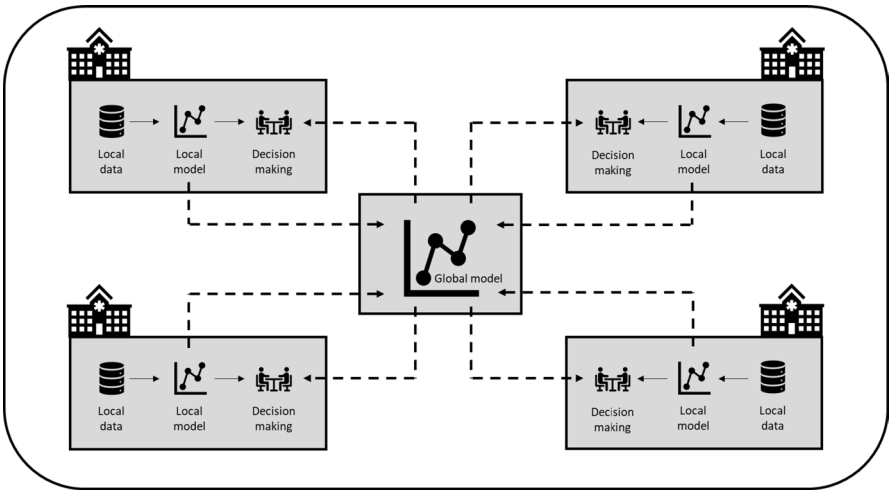

Fig. 1 Decision making based on local versus global models for an example with four hospitals

tive Federated Learning approach designed to address missing data robustly. This new approach is a modification based on the Federated Stochastic Variance Reduced Gradient (FSVRG) algorithm initially proposed by Konečný et al. (2016). FSVRG uses full inclusion and zero imputation of missing values which is known for biased outcomes, shortcomings regarding multicollinearity and robustness. Hence, the contributions of this study are:

- We introduce an extension of FSVRG, referred to as the Flexible Federated Stochastic Variance Reduced Gradient (F-FSVRG) that can replace missing values flexibly, rather than with the value 0.
- Building upon this extension, we propose the Modified Federated Stochastic Variance Reduced Gradient (M-FSVRGS) for partially including missing observations regarding parameter updates.
- We assess the performance of F-FSVRG and M-FSVRGS in decision-making through a computational study employing real-world multicenter COVID-19 diagnosis data and ICU data.
- We illustrate how addressing the simultaneous challenges of fragmented and missing data not only facilitates the integration of AI approaches but also fosters sustainable enhancements in processes and operations within the healthcare sector.

We find that M-FSVRGS guarantees highly accurate AI-based predictions in healthcare settings. Especially with a flexible imputation parameter and a high proportion of missing values, M-FSVRG demonstrates significantly higher accuracies. This independency guarantees medical decision makers flexibility in the choice of the inputation parameter, while saving scarce medical resources. In addition, our findings hold for different healthcare dataset structures, including sample size and the proportion of categorical and numerical features.

The structure of this work is outlined as follows: Sect. 2 offers insights into FL, FL algorithms and missing data as well as application areas of FL. Moving to Sect. 3, we introduce the F-FSVRG and the M-FSVRGS algorithms. Subsequently, Sect. 4 presents the findings derived from our computational study utilizing real-world multicenter COVID-19 diagnosis data and ICU data. Section 4 also encapsulates a discussion on medical and managerial insights. Finally, Sect. 5 concludes and discusses potential avenues for further research.

2 Federated machine learning and missing data

In this section, we first introduce the concept of FL in Sect. 2.1 and present existing FL algorithms, including FA and FSVRG. We then focus on the handling of missing data in the context of FSVRG, highlighting the potential drawbacks and research gaps of the zero-imputation approach and parameter updates employed in FSVRG. Furthermore, in Sect. 2.2, we provide an overview of FL applications, highlighting the potential of FL for healthcare applications.

2.1 Concept of federated machine learning

In ML, it is assumed that available data is centrally located and therefore not distributed among different organizations (the interested reader is referred to Appendix 1 for an introduction to the general ML concept). However, data often contains sensitive information such as facial images or health information (Lyu et al. 2020a), which is why its access is limited by legal regulations in order to protect the privacy of the respective owners of the data. The data, given by the set of input-output pairs $\{\boldsymbol{x}_i,\ y_i\}_{i=1}^{n}$ with $(\boldsymbol{x}_i,\ y_i)$ representing one sample associated with a particular patient i, $\boldsymbol{y} = (y_1, \ldots,\ y_n)^T$ representing the outputs as well as $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots,\ \boldsymbol{x}_n)^T$ representing the features, is therefore decentralized and distributed across different local organizations. As a result of such legal restrictions, data can often no longer be aggregated centrally in order to address ML problems (Liu et al. 2022). At this point, FL initially proposed by Google in 2016 (Xu et al. 2021) provides a remedy and is defined and categorized in the following.

FL represents a learning system that allows K local owners of the data, e.g., the four hospitals in Fig. 1, hereafter referred to as clients and indexed by $k = 1, \ldots, K$, to exploit the distributed data by collaboratively training a global ML model (Yang et al. 2019a, b). At the same time, FL systems comply with legal regulations on data security and data privacy (Liu et al. 2022), as clients do not share their local data with other clients (Yang et al. 2019a, b). For distinguishing these different local datasets, the data held by client k is denoted by P_k. Accordingly, $n_k = |P_k|$ is the corresponding local sample size, i.e., the number of samples located at client k. Considering the locally distributed data collectively, it follows that $P_1 \cup P_2 \cup \cdots \cup P_K = \{\boldsymbol{x}_i,\ y_i\}_{i=1}^{n}$ and $\sum_{k=1}^{K} n_k = n$ (Konečný et al. 2016; McMahan et al. 2017).

Depending on how the data is distributed among the clients with respect to the feature and sample space, FL systems can be divided into different categories. A

distinction between these concepts is depicted in Yang et al. (2019a, b). As we study a healthcare environment, we focus on Horizontal Federated Learning (HFL) in the following. The HFL system used in this work relies on a central parameter server for aggregating the model parameters of multiple clients (e.g., four clients, see Fig. 1). Accordingly, a centralized aggregation algorithm is required to aggregate the locally fitted parameters to a global set of model parameters. To this end, a series of such algorithms have been proposed (Liu et al. 2022). As we concentrate on fragmented and missing data in this work, we will narrow our focus to the state-of-the-art FA algorithm and the FSVRG algorithm. The interested reader is referred to the appendix (Appendix 2) for an overview of FL algorithms.

FA. The current state-of-the-art algorithm for implementing a FL system is FA (Nguyen et al. 2021). FA relies on a centralized server that aggregates multiple local ML models from local clients by averaging, i.e., by taking a weighted sum of the respective local model parameters. The weights $\frac{n_k}{n}$ are defined based on the proportion of data held by client k. The aggregated parameters then compose a global model, which is redistributed to the clients via the server, and the training process continues iteratively (federated rounds). The FA algorithm relies on the well-known Stochastic Gradient Decent (SGD) algorithm as a nested baseline algorithm where the local model parameters are updated using a stochastic gradient $\nabla f_{i_t}(\boldsymbol{\omega}_t)$ resulting in the following update of the local model parameters $\boldsymbol{\omega}$: $\boldsymbol{\omega}_{t+1} = \boldsymbol{\omega}_t - \eta_t \cdot \nabla f_{i_t}(\boldsymbol{\omega}_t)$. Note that η_t denotes the learning rate in update step t. Pseudo code of the FA algorithm is provided in Algorithm 1. While FA represents a straightforward technique for implementing a FL system (Liu et al. 2022), Li et al. (2020a) state that FA can lead to biased results under certain conditions, such as when dealing with heterogeneous data. In addition, using FA is associated with slower convergence arising from the variance inherent in the process of gradient estimation and does not at all address the problem of missing data (Konečný et al. 2016).

Algorithm 1: Federated Averaging (FA)

Parameters: η = learning rate
Initialize $\boldsymbol{\omega}_0$
For $t = 0, 1, 2, \ldots$ (federated rounds)
  For $k = 1, 2, \ldots, K$
    Select $i_t \in P_k$ uniformly at random
    SGD-based local updates: $\boldsymbol{\omega}_{t+1}^k = \boldsymbol{\omega}_t - \eta \cdot \nabla F_{k,i_t}(\boldsymbol{\omega}_t)(\boldsymbol{\omega}_t)$
  End for
  Global aggregation step: $\boldsymbol{\omega}_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} \cdot \boldsymbol{\omega}_{t+1}^k$
End for

FSVRG. FSVRG addresses the problem of reducing the variance associated with the estimation of gradients (Konečný et al. 2016). The implementation of Stochastic Variance Reduced Gradient (SVRG, see Algorithm 2) rather than SGD explicitly

aims at reducing variance. The core aspect of SVRG is the use of a benchmark, which standardizes the magnitude of directional changes associated with parameter updates (Johnson & Zhang 2013).

Besides the purpose of variance reduction, the design of FSVRG is driven by the need to account for missing data (Konečný et al. 2016) since the missing data problem is a common issue in real-world ML applications. A ML model is based on a set of features $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$. If some of these features do not occur for some observations, the data is said to be sparse. To this end, a variety of methods, known as imputation methods, for handling missing data have been developed over the past decades. These methods include simplistic approaches such as *Mean Imputation* (Little and Rubin 2019) but also ML techniques including *Support Vector Regression* (Wang et al. 2006) or *MissForest* (Stekhoven and Bühlmann 2012) for imputing missing data (Bertsimas et al. 2018) and are applied in the data preprocessing step.

Algorithm 2: Stochastic Variance Reduced Gradient (SVRG)

Parameters: η = learning rate, T = number of stochastic updates per epoch
Initialize $\tilde{\boldsymbol{\omega}}_0$
For $u = 1, 2, \ldots$ (epochs)
  $\tilde{\boldsymbol{\omega}} = \tilde{\boldsymbol{\omega}}_{u-1}$
  $\nabla f(\tilde{\boldsymbol{\omega}}) = \frac{1}{n} \cdot \sum_{i=1}^{n} \nabla f_i(\tilde{\boldsymbol{\omega}})$
  For $t = 1, 2, \ldots, T$ (stochastic updates)
    Select $i_t \in \{1, \ldots, n\}$ uniformly at random
    $\boldsymbol{\omega}_t = \boldsymbol{\omega}_{t-1} - \eta \cdot \left(\nabla f_{i_t}(\boldsymbol{\omega}_{t-1}) - \nabla f_{i_t}(\tilde{\boldsymbol{\omega}}) + \nabla f(\tilde{\boldsymbol{\omega}}) \right)$
  End for
  $\tilde{\boldsymbol{\omega}}_u = \boldsymbol{\omega}_T$
End for

In the FSVRG algorithm, sparsity patterns in the data are implemented in the algorithm itself. Thereby, Konečný et al. (2016) label the missing data points, i.e., the missing values for certain features, with zero. Based on this labeling, the quantities from Table 1 as well as the underlying SVRG algorithm, the authors establish the FSVRG algorithm, which is presented in Algorithm 3. Please find a detailed description of Algorithm 3 in Appendix 3.

Table 1 Quantities for FSVRG (Konečný et al. 2016)

Notation FSVRG	
n	Total number of samples (patients), i.e., the sample size
$N = \{1, \ldots, n\}$	Set of indices
P_k	Set of indices belonging to samples located at client k
$n_k = \lvert P_k \rvert$	Number of samples located at client k
$n^j = \left\lvert i \in N : \boldsymbol{x}_i^T \boldsymbol{e}_j \neq 0 \right\rvert$	Number of samples with non-zero j^{th} feature
$n_k^j = \left\lvert i \in P_k : \boldsymbol{x}_i^T \boldsymbol{e}_j \neq 0 \right\rvert$	Number of samples located at client k with non-zero j^{th} feature
$\varphi^j = n^j / n$	Share of samples with non-zero j^{th} feature
$\varphi_k^j = n_k^j / n_k$	Share of samples located at client k with non-zero j^{th} feature
$s_k^j = \varphi^j / \varphi_k^j$	Ratio of local and global presence of feature j for client k
$\boldsymbol{S}_k = \text{Diag}\left(s_k^j\right)$	Diagonal matrix with s_k^j being the j^{th} diagonal element
$w^j = \left\lvert \left\{ P_k : n_k^j \neq 0 \right\} \right\rvert$	Number of clients with at least 1 sample with j^{th} feature $\neq 0$
$a^j = K / w^j$	Standardization parameter for feature j
$\boldsymbol{A} = \text{Diag}\left(a^j\right)$	Diagonal matrix with a^j being the j^{th} diagonal element

Algorithm 3: Federated SVRG (FSVRG)

Parameters: learning rate η, data partition $\{P_k\}_{k=1}^K$, $\boldsymbol{A}, \boldsymbol{S}_k \in \mathbb{R}^{d \times d}\ \forall\ k \in \{1, \ldots, K\}$
Initialize the global parameter vector $\tilde{\boldsymbol{\omega}}_0$
For $t = 1, 2, \ldots$ (federated rounds)
  $\tilde{\boldsymbol{\omega}} = \tilde{\boldsymbol{\omega}}_{t-1}$
  Global gradient computation: $\nabla f(\tilde{\boldsymbol{\omega}}) = \frac{1}{K}\sum_{k=1}^K \frac{1}{n_k} \sum_{i \in P_k} \nabla f_i(\tilde{\boldsymbol{\omega}}) = \frac{1}{K}\sum_{k=1}^K \nabla F_k(\tilde{\boldsymbol{\omega}})$
  For $k = 1, 2, \ldots, K$ (local subproblems)
    $\boldsymbol{\omega}_0^k = \tilde{\boldsymbol{\omega}}$ and local learning rate $\eta_k = \eta / n_k$
    Set $\left\{i_q\right\}_{q=1}^{n_k}$ as random permutation of indices stored in P_k
    For $q = 1, 2, \ldots, n_k$ (stochastic updates)
      $\boldsymbol{\omega}_q^k = \boldsymbol{\omega}_{q-1}^k - \eta_k \cdot \left(\boldsymbol{S}_k \cdot \left[\nabla f_{i_q}\left(\boldsymbol{\omega}_{q-1}^k\right) - \nabla f_{i_q}(\tilde{\boldsymbol{\omega}}) \right] + \nabla f(\tilde{\boldsymbol{\omega}}) \right)$
    End for
  End for
  Global aggregation step: $\tilde{\boldsymbol{\omega}}_t = \tilde{\boldsymbol{\omega}} + \boldsymbol{A} \cdot \sum_{k=1}^K \frac{n_k}{n} \left(\boldsymbol{\omega}_q^k - \tilde{\boldsymbol{\omega}} \right)$
End for

*Zero imputation***.** With FSRVG, missing data points are replaced by zeros. However, in their study, Fraser et al. (2009) find that replacing missing values with zeros can lead to suboptimal, biased predictions in certain settings where the proportion of missing data is larger than five to ten%. Consequently, the degree of bias increases the more this threshold is exceeded, i.e., the more data points are missing. This bias is directly related to the presence of so-called true zeros, i.e., data points that are equal to zero. If true zeros do occur, replacing missing values with zero assumes an equivalence between the true zeros and the missing observations.

In terms of the classification problem of COVID-19 diagnosis, let the concentration of C-reactive protein (cCRP) be a particular feature $(x_{i,j})_{i=1}^{n}$ indexed by j, with the cCRP of healthy patients usually being close to zero (measured in mg/L or µg/L). For example, the cCRP values of three patients are shown in Table 2.

As can be seen, the cCRP observation for patient $i = 1$ represents a value close to zero, whereas for patient $i = 3$ the value is missing. If, for instance, robust data scaling is applied, the metrically scaled cCRP values $(x_{i,j})_{i=1}^{n}$ are scaled based on their median $\widetilde{x}_{0.5,j}$ as well as their interquartile range $QA_j = \widetilde{x}_{0.75,j} - \widetilde{x}_{0.25,j}$ with (Brownlee 2020):

$$\bar{x}_{i,j} = \frac{x_{i,j} - \widetilde{x}_{0.5,j}}{QA_j}. \tag{1}$$

Note that $\widetilde{x}_{0.25,j}$ and $\widetilde{x}_{0.75,j}$ denote the first and the third quartile of the j^{th} feature, respectively. The median of the cCRP data available to us (which, in the first step, is filtered for outliers, as explained later in Sect. 4.2) is 3.67. The third quartile $\widetilde{x}_{0.75,j}$ is 9.3350, the first quartile $\widetilde{x}_{0.25,j}$ is 1.1625, and thus the interquartile range QA_j is 8.1725. According to (1), the scaled cCRP values for the three patients are as follows:

- Scaled value for patient 1: $\bar{x}_{1,j} = \frac{0.01-3.67}{8.1725} \approx -0.4475$
- Scaled value for patient 2: $\bar{x}_{2,j} = \frac{1.0-3.67}{8.1725} \approx -0.3263$
- Scaled value for patient 3: $\bar{x}_{3,j} = \frac{0-3.67}{8.1725} \approx -0.4497.$

If the imputation technique used in FSVRG (zero imputation) is applied, patients $i = 1$ and $i = 3$ are assumed to have virtually identical cCRP values after robust data scaling, although the true cCRP value of the third patient remains unknown (Fraser et al. 2009). Since small cCRP values, i.e., values close to 0, are typically observed in healthy patients, incorrect conclusions regarding the health status of patient 3 could be drawn, which could lead to serious consequences in the context of a COVID-19 diagnosis. This illustrates how the zero imputation approach can incorrectly equate missing values with actual zero values or when dealing with very small quantities that might be difficult to measure with standard equipment (e.g., values that are close to zero but not exactly zero, often referred to as 'rounded zeros') (Martín-Fernández et al. 2003).

Table 2 Zero imputation for missing data

Patient	$x_{i,j}$: cCRP [mg/l]	$\bar{x}_{i,j}$ scaled (robust scaling)
$i = 1$	0.01	−0.4475
$i = 2$	1.0	−0.3263
$i = 3$	not available	−0.4497 (after zero imputation)

For instance, the cCRP observation for patient $i = 1$ could be considered a rounded zero, where the actual concentration is very small (e.g., 0.01 mg/L) but measurable techniques may not detect such small values. While changing the unit of measurement or scaling the data – for example, through robust data scaling – may reduce the influence of very small values, it does not entirely resolve the issues with zero imputation, particularly when missing values are wrongly assumed to be true zeros. This becomes even more problematic when dealing with large amounts of missing data, as is often the case in healthcare settings. Scaling techniques bring all data points to a common scale, but they do not address the core problem: missing values are treated as if they are equal to zero, even when they represent an unknown value that could differ from zero due to measurement limitations or the accuracy of the equipment used (Fraser et al. 2009). Even when applying robust scaling, the issue persists: missing values imputed as 0 are still treated as equivalent to small non-zero cCRP values, potentially leading to biased results. This demonstrates that while robust data scaling can reduce the influence of outliers, the underlying problem of zero imputation remains, which is why we propose the M-FSVRGS method to handle such issues in a more robust way.

Furthermore, Caan et al. (1991) as well as Michels and Willett (2009) state in their works that zero imputation can often lead to a high correlation between the imputed part of the data and the original non-missing part, in some cases even above 90% (Fraser et al. 2009). Consequently, this can cause different features to be correlated with each other. This phenomenon is known as multicollinearity and refers to the problem that in case of correlated features, it is no longer possible to clearly identify those features that have an explanatory impact on the target variable (James et al. 2013).

Parameter updates. With FSVRG all initially missing data points, which are then replaced by zero, are used for parameter updates. This approach may lead to suboptimal results for datasets that do not end up being scaled to value ranges close to zero (as is the case with robust scaling in this work), where imputing zeros might result in much worse outcomes compared to using non-zero imputation values.

2.2 Applications

FL has already been implemented in various application areas. Below, we explore some well-known areas where FL is applied, specifically healthcare, industrial engineering, autonomous vehicles, and mobile devices (Li et al. 2020a, b, c; Shaheen et al. 2022), to highlight the potential of FL for healthcare processes.

Healthcare. Kim et al. (2017) were among the first to apply FL in the medical field, using tensor factorization models for phenotyping analysis without sharing individual patient data. Later, Pfohl et al. (2019) demonstrated that differentially private FL models for Electronic Health Records (EHRs) can perform similarly to centralized models, highlighting FL's potential in privacy-preserving healthcare analytics. Huang et al. (2019) used FL to predict heart disease mortality rates based on

decentralized EHRs from multiple hospitals, ensuring no data or parameters were exchanged between hospitals during the training. Similarly, Brisimi et al. (2018) employed FL to predict heart disease hospitalizations using a method called Cluster Primal Dual Splitting, which also avoided data leakage. FL has also been used for patient matching across hospitals, with Lee et al. (2018) proposing a federated framework to detect similar patients without sharing individual-level data. Furthermore, Huang et al. (2020) used a FL-based algorithm to predict mortality rates based on drug usage data, optimizing both computation and communication efficiency. In the field of natural language processing, Liu et al. (2019a) introduced the first FL approach for processing unstructured clinical notes. Their model used a two-stage federated training process to create patient representations and study disease phenotypes. FL has also made strides in biomedical imaging analysis, as demonstrated by Silva et al. (2019), who proposed Federated Principal Component Analysis to analyze magnetic resonance imaging data from multiple medical centers. Lastly, Gao et al. (2019) developed a hierarchical FL framework for electroencephalography classification, addressing challenges like limited labeled data and privacy concerns. Moreover, companies like Owkin and Intel are also exploring how FL might be employed to safeguard patient privacy while still enabling the valuable use of their data for research purposes (Owkin 2025; Intel 2020). Owkin is creating a platform that utilizes FL to protect patient data in research aimed at identifying drug toxicity, predicting disease progression, and estimating survival rates for rare cancers. Intel has showcased how FL can be applied in medical imaging. Their FL-based deep learning model achieves an accuracy that is 99% comparable to a model trained with traditional methods.

Industrial engineering. As FL matures, it holds great promise for widespread adoption in data-sensitive industrial sectors. For example, in environmental monitoring, Hu et al. (2018) developed a FL framework to aggregate sensor data across different locations, improving model performance while preserving data privacy. Similarly, FL has been applied in visual inspection tasks (Han et al. 2019), where it helps detect defects in production processes without sharing sensitive data. In the realm of image processing, FL has been used to enhance vision-and-language tasks by acquiring diverse representations from federated tasks (Liu et al. 2020a, b). It is also effective in detecting malicious attacks in communication systems, such as those involving Unmanned Aerial Vehicles (UAVs), where data challenges like unbalanced distribution and unreliable communication align well with FL's strengths (Mowla et al. 2020). FL is also proving valuable in energy prediction, as demonstrated by Saputra et al. (2019), who developed an FL model to predict energy demand across charging stations for electric vehicles, helping to prevent energy congestion.

Autonomous vehicles. Regarding self-driving cars, FL enhances user data privacy by allowing the algorithms to process data locally on vehicles rather than sending sensitive information to a central server. This way, journey logs and travel data remain private while still improving the system's performance. On the other hand, FL

can reduce latency. As the number of autonomous vehicles on the roads grows, they must respond quickly to safety situations. Traditional cloud-based learning involves large data transfers, which can slow down the learning process. In contrast, FL allows vehicles to make faster, more accurate decisions, potentially reducing accidents and improving safety (Shaheen et al. 2022). Research, such as that by Elbir et al. (2022), highlights FL's transformative potential for autonomous vehicles and the Internet of Things (IoT). Pokhrel and Choi (2020) further emphasize the need for rapid response in real-world scenarios and point out that FL can address security concerns inherent in cloud-based systems by minimizing data transfer and speeding up learning processes. Moreover, Liu et al. (2020a, b) introduced a FL-based data fusion method for robotic imitation learning in networked robots, which could also be applied to autonomous vehicles to develop guidance models and predict potential emergencies.

Mobile devices. In addition to autonomous vehicles, FL is being applied to other domains of IoT, such as in smart homes. Devices in these settings often upload user data to cloud servers, which can lead to privacy concerns. To address this, Aïvodji et al. (2019) proposed a secure federated architecture to build joint models, while Yu et al. (2020) developed a federated multi-task learning framework for smart home IoT systems to learn user behavior patterns and detect potential physical hazards. Beyond that, smart devices make use of FL to predict human trajectories (Feng et al. 2020) and behavior (Sozinov et al. 2018), expanding its range of potential uses. Despite rapid advancements in mobile device storage and computing power, mobile subscribers' increasing demand for high-quality services is often constrained by communication bandwidth limitations. To alleviate network congestion, many service providers now prefer to deploy services at the edge of the cellular network, closer to customers, instead of relying on centralized cloud computing and storage. This approach, known as mobile edge computing (MEC), does, however, increase the risk of information leakage. One potential solution is combining MEC with FL. Wang et al. (2019) explored a framework that integrates FL with deep reinforcement learning to optimize resource allocation in MEC systems. Similarly, Qian et al. (2019) developed a privacy-aware service placement scheme using FL on MEC to provide high-quality service by caching desired services on edge servers.

Other application fields. In finance, FL has been used to detect credit card fraud by analyzing transactions across different banks (Yang et al. 2019b). For example, FL can help detect multi-party borrowing, a major risk in banking where users borrow from one bank to pay off loans from another. This fraudulent activity can destabilize the financial system. Using FL, banks can detect these borrowers without sharing sensitive user data. By encrypting the user lists at each bank and intersecting the encrypted data, they can identify multi-party borrowers while keeping the identities of other legitimate users private (Yang et al. 2019a). Lastly, in text mining, Wang et al. (2020a, b) proposed an industrial federated framework for spam filtering and sentiment analysis based on Latent Dirichlet Allocation, showing its effectiveness with real-world data.

3 F-FSVRG and M-FSVRGS

This section addresses the challenge of handling missing data in Federated Learning (FL) and presents two extended versions of the FL algorithm FSVRG: F-FSVRG and M-FSVRGS. At first, we introduce the concept of flexible imputation for handling missing data in FSVRG. It describes the modification from zero imputation to a flexible parameter ϵ and explains how this change affects the algorithm's notation and behavior. Then, we show how our proposed method M-FSVRGS performs gradient and parameter updates on a per-feature basis. It highlights the differences between FSVRG and M-FSVRGS in terms of handling missing data and introduces the use of an indicator function for excluding missing values during updates. Finally, we provide an example to illustrate how incomplete observations are handled in practice. This section demonstrates the per-feature approach and how it is implemented in the update process using a small and simplified dataset.

F-FSVRG. Building on the discussion of zero imputation (see Sect. 2.1), the F-FSVRG algorithm replaces missing data points with a parameter ϵ instead of zero. This allows for a more precise handling of missing values in the learning process. The notations in Table 1 have been modified to reflect this flexible approach (see Table 3). The remaining notations can be derived analogously to those in Table 1 .

M-FSVRGS. The M-FSVRGS algorithm extends the F-FSVRG approach by performing gradient and parameter updates on a per-feature basis. This means that only the non-missing values for a given feature are considered when computing the gradients. Unlike FSVRG, where all initially missing data points are replaced by zero, M-FSVRGS only includes initially available values. This is achieved by using the following indicator function $I\left(\boldsymbol{x}_i^T \boldsymbol{e}_j \neq \epsilon\right)$, which equals 1 if the j-th feature of sample i is not missing and 0 otherwise:

$$I\left(\boldsymbol{x}_i^T \boldsymbol{e}_j \neq \epsilon\right) = \begin{cases} 1, & \text{if } \boldsymbol{x}_i^T \boldsymbol{e}_j \neq \epsilon \\ 0, & \text{otherwise} \end{cases} . \tag{2}$$

This function is used for both the global gradient update (Algorithm 4, Line 6) and the stochastic parameter updates (Line 13).

Table 3 Modified quantities (own table based on Konečný et al. (2016)

Notation F-FSVRG & M-FSVRGS	
$n^j = \left\lvert i \in N : \boldsymbol{x}_i^T \boldsymbol{e}_j \neq \epsilon \right\rvert$	Number of samples with j^{th} feature $\neq \epsilon$
$n_k^j = \left\lvert i \in P_k : \boldsymbol{x}_i^T \boldsymbol{e}_j \neq \epsilon \right\rvert$	Number of samples located at client k with j^{th} feature $\neq \epsilon$
$w^j = \left\lvert \left\{ P_k : n_k^j \neq \epsilon \right\} \right\rvert$	Number of clients storing at least 1 sample with j^{th} feature $\neq \epsilon$

Algorithm 4: Modified Federated SVRG (M-FSVRGS)

1: Parameters: learning rate η, data partition $\{P_k\}_{k=1}^{K}$, $\boldsymbol{A}, \boldsymbol{S}_k \in \mathbb{R}^{d\times d}\ \forall\ k \in \{1,\ldots,K\}$
2: Initialize the global parameter vector $\widetilde{\boldsymbol{\omega}}_0$
3: **For** $t = 1, 2, \ldots$ (federated rounds)
4: $\quad\widetilde{\boldsymbol{\omega}} = \widetilde{\boldsymbol{\omega}}_{t-1}$
5: $\quad$**For** $j = 0, 1, 2, \ldots, m$ (features)
6: $\quad\quad$Global FODs: $\nabla f\left(\widetilde{\omega}_j\right) = \frac{1}{K}\cdot\sum_{k=1}^{K}\frac{1}{n_k^j}\cdot\sum_{i\in P_k} I\left(\boldsymbol{x}_i^T\boldsymbol{e}_j \neq \varepsilon\right)\cdot\nabla f_i\left(\widetilde{\omega}_j\right)$
7: $\quad$**End for**
8: $\quad$**For** $k = 1, 2, \ldots, K$ (local subproblems)
9: $\quad\quad\boldsymbol{\omega}_0^k = \widetilde{\boldsymbol{\omega}}$ and $\eta_k = \eta/n_k$
10: $\quad\quad$Set $\{i_q\}_{q=1}^{n_k}$ as random permutation of indices stored in P_k
11: $\quad\quad$**For** $q = 1, 2, \ldots, n_k$ (stochastic updates)
12: $\quad\quad\quad$**For** $j = 0, 1, 2, \ldots, m$ (features)
13: $\quad\quad\quad\quad\omega_{j,q}^k = \omega_{j,q-1}^k - I\left(\boldsymbol{x}_{i_q}^T\boldsymbol{e}_j \neq \varepsilon\right)\ \eta_k\left(\boldsymbol{S}_k\left[\nabla f_{i_q}\left(\omega_{j,s-1}^k\right) - \nabla f_{i_q}\left(\widetilde{\omega}_j\right)\right] + \nabla f\left(\widetilde{\omega}_j\right)\right)$
14: $\quad\quad\quad$**End for**
15: $\quad\quad$**End for**
16: $\quad$**End for**
17: $\quad$Global aggregation step: $\widetilde{\boldsymbol{\omega}}_t = \widetilde{\boldsymbol{\omega}} + \boldsymbol{A}\cdot\sum_{k=1}^{K}\frac{n_k}{n}\left(\boldsymbol{\omega}_q^k - \widetilde{\boldsymbol{\omega}}\right)$
18: **End for**

The following example, shown in Table 4, illustrates how incomplete observations are handled in practice. Assume that only one client is involved (i.e., $K = 1$ and $n_k^j = n^j$) for simplicity. Note that the data in Table 4 is randomly generated, dimensionless, with $n = 2$ samples and $m = 4$ features, for illustrative purposes. With M-FSVRGS (Algorithm 4, lines 5–7), the global gradient is updated on a per-feature basis, and the Fist Order Derivatives (FODs) $\nabla f\left(\widetilde{\omega}_j\right)\ \forall\ j = 0,\ 1, \ldots, m$, are determined as components of the global gradient $\nabla f\left(\widetilde{\boldsymbol{\omega}}\right) = \left(\nabla f\left(\widetilde{\omega}_0\right), \nabla f\left(\widetilde{\omega}_1\right), \ldots, \nabla f\left(\widetilde{\omega}_m\right)\right)^T$ following the per-feature approach. Looking at the example from Table 4, both samples $i = 1$ and $i = 2$ are fully included in the update of the second and fourth component of the gradient. In contrast, only the first sample $i = 1$ is included in the update of the first and third gradient components, since the second sample originally has missing values for features

Table 4 Illustrative example for the indicator function

Sample	Feature/Indicator	$j = 1$	$j = 2$	$j = 3$	$j = 4$
$i = 1$	x_{ij}	2.89	8.31	3.34	0.06
	$\boldsymbol{x}_i^T\boldsymbol{e}_j$	2.89	8.31	3.34	0.06
	$I\left(\boldsymbol{x}_i^T\boldsymbol{e}_j \neq \epsilon\right)$	1	1	1	1
$i = 2$	x_{ij}	ϵ	1.57	ϵ	4.06
	$\boldsymbol{x}_i^T\boldsymbol{e}_j$	ϵ	1.57	ϵ	4.06
	$I\left(\boldsymbol{x}_i^T\boldsymbol{e}_j \neq \epsilon\right)$	0	1	0	1

$x_{2,1}$ and $x_{2,3}$, which are replaced by ϵ. Consequently, the first two components of the gradient can be calculated as follows:

$$\nabla f\left(\widetilde{\omega}_1\right) = \frac{1}{2} \cdot \sum_{i=1}^{2} I\left(\boldsymbol{x}_i^T \boldsymbol{e}_1 \neq \epsilon\right) \cdot \nabla f_i\left(\widetilde{\omega}_1\right) = \frac{1}{2} \cdot \left(1 \cdot \nabla f_1\left(\widetilde{\omega}_1\right) + 0 \cdot \nabla f_2\left(\widetilde{\omega}_1\right)\right),$$

$$\nabla f\left(\widetilde{\omega}_2\right) = \frac{1}{2} \cdot \sum_{i=1}^{2} I\left(\boldsymbol{x}_i^T \boldsymbol{e}_2 \neq \epsilon\right) \cdot \nabla f_i\left(\widetilde{\omega}_2\right) = \frac{1}{2} \cdot \left(1 \cdot \nabla f_1\left(\widetilde{\omega}_2\right) + 1 \cdot \nabla f_2\left(\widetilde{\omega}_2\right)\right).$$

The third and fourth gradient components are computed in a similar manner. Moreover, it is important to note that the contribution of client k to the j^{th} component of the global gradient (i.e., the FOD with respect to the j^{th} parameter) is no longer averaged over the number of samples n_k assigned to this client (Algorithm 3, line 5). Instead, averaging over n_k^j gradient contributions is applied, as the indicator function is set to one exactly n_k^j times when passing through the samples $i \in P_k$ of client k (Algorithm 4, line 6).

The per-feature approach is also embedded in the stochastic updates. The parameter for the j^{th} feature is updated only if the randomly selected sample (Algorithm 4, line 10) has a non-missing value for that feature (Algorithm 4, line 13). In the example (see Table 4), selecting the first sample $i_q = 1$ updates all $m = 4$ parameters. However, if the second sample $i_q = 2$ is chosen, only $\omega_{2,q}^k$ and $\omega_{4,q}^k$ are updated, as $I\left(\boldsymbol{x}_2^T \boldsymbol{e}_j \neq \epsilon\right) = 0$ for $j \in \{1, 3\}$. Thus, $\omega_{j,q}^k = \omega_{j,q-1}^k$ holds for $j \in \{1, 3\}$.

4 Computational study

The goal of our computational analysis is to study the proposed federated algorithms in the context of various problem dimensions. On one hand, we aim to investigate the influence of the magnitude of the sparsity patterns, i.e., the share of missing data, on the performance of the algorithms. Furthermore, the performance will be evaluated concerning the imputation parameter ϵ. The comparison of the algorithms will be also based on different dataset structures such as sample size and feature structure to address the impact of different dataset structures on the performance of the algorithms. Before the algorithms are compared with each other, the difference between FSVRG and the flexible FSVRG (F-FSVRG) needs to be clarified. The F-FSVRG algorithm refers to the variant of FSVRG parameterized by the imputation parameter ϵ. Thus, FSVRG with $\epsilon = 0$ represents a special case of F-FSVRG. Consequently, the following discussion will refer to F-FSVRG with $\epsilon = 0$ in the context of FSVRG, and F-FSVRG will be compared with M-FSVRGS for different values of ϵ, where we refer to each specific value of ϵ as an experimental setting. For each setting, the value of ϵ is set prior to the experiment and remains fixed throughout each experimental setting. Based on the multidimensional analysis of the algorithms, we derive medical and managerial insights.

4.1 Datasets

The computational study conducted hereafter is based on data from 3,670 patients collected in the period of the first and second wave of the pandemic (COVID-19 diagnosis) as well as ICU data containing patient records of 26,676 surgeries carried out between 2017 and 2021. For the COVID-19 diagnosis, the patient records are provided by the University Hospital of Augsburg (UHA), Germany, the Hospital Göppingen (HG), Germany, as well as the *Lean European Open Survey on SARS-COV-2 Infected Patients (LEOSS)* registry. Besides the COVID-19 diagnosis denoted by $\boldsymbol{y} \in \{0, 1\}^n$, the dataset contains information on the age, sex as well as laboratory values of patients tested positive and negative for COVID-19. This additional information represents the features underlying the classification problem. In total, $m = 15$ features, which are selected by expert consensus and specified in Table 5, are considered for the COVID-19 diagnosis. For the ICU treatment problem, we examine the classification problem a surgeon faces when planning a few weeks before an elective surgery. The surgeon must decide whether ICU capacity needs to be reserved for a patient after the procedure and, consequently, on which day the patient can be scheduled, considering the available ICU capacity. The patient data is provided by the UHA and comprises 13 features, the definition of which is also listed in Table 5. The features are selected by expert consensus as for the COVID-19 diagnosis data. Every patient record of the ICU data is labeled based on whether a patient has been sent to the ICU after an elective surgery treatment. The features in Table 5 marked with a star represent categorical features, while the other features are numerical. Note that we focus on the COVID-19 diagnosis data in the following. The ICU data, which come from a single hospital with two locations, are primarily intended to evaluate the impact of dataset size and structure on our results.

Before the federated algorithms are studied based on the available datasets, this data first needs to be prepared in a preprocessing step. Then, the preprocessed data is used to compare the different FL algorithms with each other.

Table 5 Features used for COVID-19 diagnosis and the ICU treatment problem in alphabetical order

Index j	Description (COVID-19 diagnosis)	Description (ICU treatment)
1	Age	Age
2	C-reactive Protein	ASA-Score
3	D-Dimer	Charlson Comorbidity Index
4	Direct Bilirubin	Cutting-suture time
5	Erythroblasts	Estimated anesthesia duration
6	Gamma Glutamyltransferase	Estimated surgery duration
7	Hemoglobin	Height
8	Platelet (Thrombocyte) Count	Main diagnosis for a hospital stay*
9	Partial Thromboplastin Time	Medical specialty*
10	Serum Alanin Aminotransferase	Number of comorbidity diagnoses
11	Serum Creatinin	Sex*
12	Serum Lactatdehydrogenase	Type of anesthesia*
13	Serum Urea	Weight
14	Sex*	–
15	White blood cells	–

4.2 Data preprocessing

The preprocessing step consists of several sub-steps, the sequence of which may differ depending on the algorithm under consideration. The processing is illustrated in a flowchart (see Fig. 2) using the COVID-19 diagnosis use case as an example.

Data preparation starts by identifying outliers and incorrect values, which are labeled as missing values. Incorrect values are impossible or unrealistic values that occur due to measurement errors and transmission errors, among other sources of errors. For example, in the dataset at hand, such an incorrect value is given by the age $x_{i,2}$, i.e., the second feature, of a particular patient i, which is stated to be 120 years. As this age appears to be unrealistic, it can be interpreted as an incorrect value, which is why it is labeled as a missing value in the course of this work. In addition, it is important to identify outliers, which are extreme values, as they can distort the results of the data analysis. Outliers can either represent rare extreme observations that do not fit the rest of the data, or incorrect values, and are identified according to the rules of thumb of John Tukey (Tukey 1977). According to these rules, a data point $x_{i,j}$ represents an outlier if

$$x_{i,j} < \widetilde{x}_{0.25,j} - 1.5 \cdot QA_j \text{ or } x_{i,j} > \widetilde{x}_{0.75,j} + 1.5 \cdot QA_j \tag{3}$$

holds. Here, $\widetilde{x}_{0.25,j}$ and $\widetilde{x}_{0.75,j}$ denote the first and the third quartile of the j^{th} feature, respectively, and $QA_j = \widetilde{x}_{0.75,j} - \widetilde{x}_{0.25,j}$ is referred to as the corresponding interquartile range. As with the incorrect values, the identified outliers are labeled as missing values (Kaptein and van den Heuvel 2022).

The next sub-step is to encode the categorical features in such a way that differences between individual classes can be examined. The dataset at hand includes one such categorical feature, that is, the patient sex $x_{i,1}$. Subsequently, the data imbalance, which refers to the problem of unequally distributed classes of the output variable or categorical features, needs to be addressed. For instance, regarding the COVID-19 diagnosis problem at hand, class imbalance can be attributed to the output variable $\boldsymbol{y}$, as there are much more patients tested positive (2627 patients) for COVID-19 than negative (1,043 patients). This raises the problem that, with imbalanced data, ML algorithms predict accurately for samples belonging to the majority class but

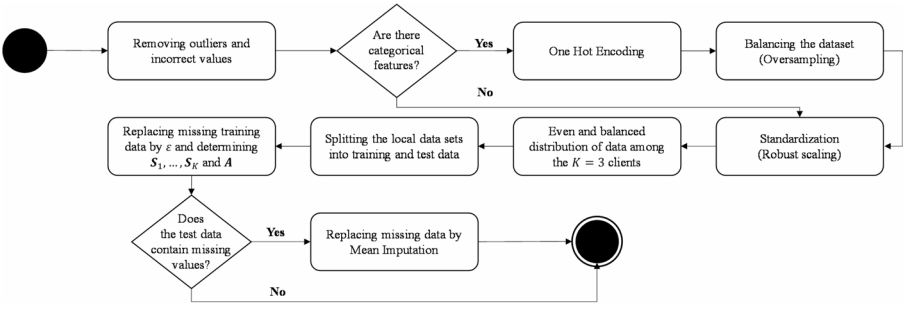

Fig. 2 Preprocessing

perform poorly for samples of the minority class. Following the recommendations of Mohammed et al. (2020), who found that oversampling yields better results than undersampling, we apply oversampling to both the COVID-19 diagnosis data and the ICU data. Specifically, we use the Synthetic Minority Oversampling Technique (SMOTE), initially proposed by Chawla et al. (2002), a widely used oversampling method that utilizes interpolation to synthesize new samples, which are then assigned to the minority class. We choose SMOTE over random oversampling, which involves simply duplicating samples in the minority class, as synthesized samples contribute new information to the model, unlike random oversampling (Brownlee 2021). For the COVID-19 diagnosis (ICU admission) problem at hand, existing patient records are used to synthesize new samples until the minority class contains as many patient records as the majority class. The augmented dataset then contains a total of 5,254 (47,794) samples, with 2617 (23,897) patients tested positive (admitted to the ICU) and 2617 (23,897) patients tested negative for COVID-19 (not admitted to the ICU).

Next, the need for scaling the data arises because features with comparatively large values can dominate the training process, giving these features even more weight when estimating the corresponding parameters. For this purpose, robust standardization, also known as robust data scaling, is applied according to (1). The reason for applying robust data scaling rather than regular standardization based on the mean and standard deviation is that robust scaling is recommended especially when the underlying dataset contains many outliers, which applies to the underlying dataset (Brownlee 2020).

As mentioned at the beginning of this section, patient data is provided by UHA, HG as well as LEOSS, hence there are $K = 3$ local clients to consider in total. The distribution of the oversampled data among the clients is both uniform and balanced, i.e., each client is assigned as many samples with a positive COVID-19 test result ($y_i = 1$) as samples with a negative test result ($y_i = 0$). Then, the locally distributed datasets are each split into training and test data. Note that each client uses $75\,\%$ of their assigned samples to fit local models. The test data is distributed locally to the clients in the same way as the training data, and each client evaluates the performance based on local test data. To evaluate the overall performance of the algorithms, the average of the local accuracy values, which are derived from the local test datasets, is used as the performance metric.

Finally, as F-FSVRG and M-FSVRGS explicitly do account for sparsity patterns in the data, missing values in the training dataset are replaced by a parameter ϵ. To evaluate the performance of these two algorithms, each missing value x_{ij} of a particular feature $\boldsymbol{x}_j = (x_{1j}, \ldots, x_{jn})^T$ is imputed by using the mean, the first and the third quartile of the non-missing observations of that feature, and zero, as values for ϵ. It should be mentioned that when imputing epsilon using the mean or the first or third quartile, these statistics are always computed from the scaled, available values of that feature, ensuring that the imputed values fit within the range of the scaled data used to train the algorithms. The usage of zero as an imputation value can be seen as benchmark corresponding to the initial proposal of Konečný et al. (2016).

Table 6 Final accuracy results after 50 federated rounds (COVID-19 diagnosis data)

Accuracy [%]	F-FSVRG				M-FSVRGS			
	τ =0%	τ =5%	τ =10%	τ =25%	τ =0%	τ =5%	τ =10%	τ =25%
ϵ = mean	89.51	87.98	87.53	85.10	89.51	87.98	**87.53**	**85.17**
ϵ =0		88.14	87.83	85.17		**88.21**	**87.83**	**85.25**
ϵ =1st quar.		**88.29**	**86.92**	**83.96**		87.91	86.77	83.42
ϵ =3rd quar.		87.45	86.31	82.66		**87.45**	**86.46**	**82.74**

Table 7 Final accuracy results after 50 federated rounds (ICU data)

Accuracy [%]	F-FSVRG				M-FSVRGS			
	τ =0%	τ =5%	τ =10%	τ =25%	τ =0%	τ =5%	τ =10%	τ =25%
ϵ = mean	83.24	82.92	82.51	80.64	83.24	**82.93**	82.51	**80.95**
ϵ =0		82.85	82.40	**80.50**		**82.85**	**82.41**	80.49
ϵ =5		82.77	**82.16**	78.90		**82.79**	82.12	**79.13**
ϵ =10		82.73	**82.44**	80.31		**82.74**	82.35	**80.39**

4.3 Comparison of FSVRG, F-FSVRG and M-FSVRGS

As described before, zero, the mean, the first and the third quartile of the non-missing observations of the feature under consideration are used as flexible values for ϵ for both F-FSVRG and M-FSVRGS. Please note that the imputation parameter is not a tunable parameter, but flexibility refers to the fact that, with F-FSVRG and M-FSVRGS, we are able to vary the imputation parameter other than for FSVRG. Furthermore, the two algorithms are also compared in terms of the share of missing data to examine the robustness of the algorithms under different magnitudes of sparsity patterns. To implement this, for each of the aforementioned values of ϵ , 5%, 10%, and 25% of the training data points are randomly removed for both F-FSVRG and M-FSVRGS. This procedure is carried out for both the COVID-19 data and the ICU data. Let the share of this missing data be denoted as $\tau \in \{0.05, 0.10, 0.25\}$. It should be mentioned that a default regularization parameter $\lambda = 1$ is used for both F-FSVRG and M-FSVRGS. The learning rate $\eta = 0.35$ is chosen for both the COVID-19 diagnosis problem and the ICU treatment problem, because these choices led to the best results. The performance of F-FSVRG and M-FSVRGS is evaluated subject to the choice of the imputation parameter ϵ as well as with respect to the magnitude of the sparsity patterns τ (varying share of missing data points). Therefore, accuracy based on test data is used as a performance measure for comparison and the corresponding results can be seen in Tables 6 and 7. These tables show the accuracies depending on ϵ and τ after 50 federated rounds for both F-FSVRG and M-FSVRGS. In the comparison between F-FSVRG and M-FSVRGS regarding various settings, where each setting is characterized by the choice of ϵ and τ , the better accuracy value for the given setting is highlighted in bold in Tables 6 and 7. If the accuracy values of both algorithms are equivalent, M-FSVRGS is marked as superior (bold), because the influence of missing values and biased outcomes is minimized while allowing for a flexible imputation achieving same accuracies. For the ICU data, with a considerable amount of testing data, we, in addition, tested the differences of accuracies for statistical significance with the McNemar test and a 10% significance

level. Note that $\tau = 0\%$ corresponds to the setting with full data, i.e., with no data points missing. Therefore, the accuracy with 0% missing data constitutes an upper bound on the accuracy associated with settings where $\tau > 0\%$ for both F-FSVRG and M-FSVRGS. Consequently, the accuracy associated with settings where $\tau = 0\%$ does not depend on ϵ as there are no values to be imputed.

For $\epsilon = 0$, in which FSVRG can be compared to M-FSVRGS, M-FSVRGS shows higher accuracies in terms of the COVID-19 diagnosis data, in terms of the ICU dataset the accuracies are comparable. For other values of ϵ, i.e., mean, 25% quantile and 75% quantile, M-FSVRGS shows, in the majority of settings, superior performance. The inferential statistical analysis shows significant results for the setting with a high amount of missing values, i.e. $\tau = 25\%$ (p-values: 0 for $\epsilon =$ mean and 0.06 for $\epsilon =$ 1st quar), which supports the findings of Fraser et al. (2009). The results indicate comparable performance and suggest that our findings are mostly independent of dataset size and structure. However, it should be noted that, for the COVID-19 diagnosis data, the limited size of the dataset did not allow us to statistically demonstrate significant superiority of M-FSVRGS. This limitation should be considered when interpreting the comparative performance on COVID-19 data.

4.4 Medical and managerial implications

By tackling the concurrent issues of fragmented and missing data through F-FSVRG and M-FSVRGS, we not only streamline the integration of AI methodologies in hospitals but also drive improvements in core and management processes, as evidenced by our two use cases:

Especially at the onset of the COVID-19 pandemic, there was a dual challenge: a shortage of physical testing capacities (Calabrese and Demers 2022) and a lack of comprehensive data on COVID-19 patients in a single hospital, hindering the expansion of testing capabilities using digital testing. Moreover, e.g., due to a lack of initial treatment experience, medical data was often inconsistently collected (Binkheder et al. 2021), leading to numerous incomplete datasets. Leveraging the existing but frequently incomplete data from various hospitals with M-FSVRGS could have facilitated early and robust training of accurate digital solutions like digital COVID-19 diagnosis during the pandemic. With increased testing capacity and accurate predictions, patients can receive prompt treatment, while unnecessary isolation of suspected cases - requiring significant additional space and personnel - can be minimized. Consequently, using M-FSVRGS considerably reduces the risk of making incorrect decisions such as regarding the allocation of resources. Likewise, in the preoperative classification of elective patients regarding postoperative ICU admission, collaboration among different hospitals can enhance the data foundation and address the challenge of missing and incomplete data. In this example, the wrong decision would cause the scarce resource, i.e., a bed in the ICU, not to be used for a patient who actually requires intensive care, thereby increasing the risk of consequential damages. Reliable forecasts with accuracies over 80% generated by M-FSVRGS can reduce last-minute surgery cancellations due to insufficient ICU capacity, thereby easing constraints on personnel and space and potentially enhancing medical outcomes for patients.

As we show the independency of our results of the given dataset structure and size, we moreover provide, especially with M-FSVRGS, decision makers in hospitals with a flexible and robust AI-based tool at the same time. On the one hand, the proposed way of dealing with missing data enables institutions to flexibly adapt the choice of the imputation parameter towards the particularities of their local data. M-FSVRGS offers medical decision makers an increased degree of flexibility regarding the specific choice of the imputation parameter. On the other hand, our proposed method provides a robust decision basis as the dependency of the results on the choice of the imputation value and therefore the risk for deriving wrong decisions resulting from the imputation technique can be substantially reduced. Furthermore, our approach fosters the exploitation of incomplete information, especially in settings with a high amount of missing data. In this regard, we can make use of the available features of incomplete data points while simultaneously reducing the impact of imputed, initially missing feature values on the predicted outcome.

5 Conclusion

This study presents innovative Federated Learning approaches, F-FSVRG and M-FSVRGS, designed to address fragmented and missing data robustly. The new approaches are modifications based on the Federated Stochastic Variance Reduced Gradient (FSVRG) algorithm initially proposed by Konečný et al. (2016). Compared to FSVRG, our proposed M-FSVRGS algorithm offers the advantage of not being limited to using a fixed value of zero as the imputation parameter (FSVRG). Instead, M-FSVRGS allows for imputing missing values with a parameter ϵ, which can differ from zero. Furthermore, our approach enables feature-specific imputation by utilizing descriptive statistics such as the mean, as well as the first and third quartiles, for the respective numerical features. This means that not all missing values across features need to be replaced with the same imputation value. Additionally, the use of the indicator function allows us to reduce the impact of imputed values by excluding originally missing observations from gradient and parameter updates. We assess the performance of F-FSVRG and M-FSVRGS in decision-making in hospitals through a computational study employing real-world healthcare datasets. Especially with a flexible imputation parameter and a high proportion of missing values, M-FSVRG demonstrates significantly higher accuracies. This behavior is independent of dataset size and structure. In conclusion, we illustrate how addressing the simultaneous challenges of fragmented and missing data not only facilitates the integration of AI approaches in hospitals but also fosters sustainable enhancements in processes and operations within the healthcare sector.

The findings of our study guide future research avenues. The impact of initially missing observations could not be fully eliminated. To further reduce this impact, the explicit consideration of sparsity patterns in the data could be combined with the concept of regularization for the purpose of controlling the impact of rarely occurring features. Accordingly, modified local loss functions could be defined such that the penalizing effect of the regularization term on the feature parameters is larger for features with a higher number of missing data points. Furthermore, future research might

consider further dataset structures that substantially differ from the data structures associated with the use cases considered in this work, that is, COVID-19 diagnosis and ICU treatment. These differences might be, for example, differently scaled data which enforces different scaling methods and therefore results in different impacts of imputation values on the performance of algorithms. In terms of the imputation parameter, a further research avenue could involve investigating datasets that do not end up being scaled to value ranges close to zero (as is the case with robust scaling in this work), where imputing zeros might result in much worse outcomes compared to using non-zero imputation values. In addition, simulation studies and mathematical modelling techniques could be applied to quantify the effect of the application of our new approaches for specific use cases and decision-making processes in hospitals. Finally, it should be noted that, for the COVID-19 diagnosis dataset, the limited size of the dataset did not allow us to statistically demonstrate significant superiority of M-FSVRGS, as we could for the ICU dataset. This limitation should be considered when generalizing our findings.

Moreover, the proposed FL methodology could be extended to other sectors where privacy-preserving, decentralized learning is crucial. Potential applications include industrial engineering, for defect detection and energy demand prediction (Hu et al. 2018; Saputra et al. 2019); autonomous vehicles, for real-time decision making and data fusion (Elbir et al. 2022; Pokhrel and Choi 2020); mobile devices and IoT systems, for behavior prediction and secure resource allocation (Aïvodji et al. 2019; Yu et al. 2020; Wang et al. 2019); as well as finance, for fraud detection and secure multi-party analysis (Yang et al. 2019a, b). Highlighting these possibilities demonstrates the broader relevance of our approach and encourages future research in diverse application domains.

Appendix

Appendix 1: Machine learning concept

The idea behind ML is to transfer the ability of learning associated with human intelligence to computational algorithms (Goodfellow et al. 2016), enabling them to extract patterns experimentally and automatically from data (Bini 2018; Naylor 2018). These algorithms rely on multidimensional input data that is usually large in scale and is processed based on an underlying model consisting of a set of model parameters in order to generate an output, e.g., prediction or classification results. The process of generating the model, i.e., determining the model parameters, is called the training process, and the data used for this is correspondingly referred to as training data (He et al. 2020a, b). Therefore, the model parameters are iteratively modified using a training algorithm such that the performance of the selected ML model is continuously improved. This performance is measured, for example, by the accuracy of the prediction results or the generalizability (Liu et al. 2022).

Throughout this work, we focus on supervised learning problems, namely binary classification problems with the COVID-19 diagnosis and ICU treatment after elective surgery of patients as use cases. Therefore, the classification tasks consist in classifying patients into those with(out) COVID-19 and those (not) requiring intensive care treatment after elective surgery based on laboratory parameters and other patient characteristics. The patients $i = 1, \ldots, n$ are therefore the objects of observation and the corresponding laboratory parameters as well as other patient characteristics are referred to as features or inputs $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{im})$ with m denoting the number of features. An enumeration and description of the features $j = 1, \ldots, m$ underlying the classification problems are given in Sect. 4. Furthermore, each patient i is known to belong to one of the two classes, COVID-19-positive (ICU treatment after elective surgery) or COVID-19-negative (no ICU treatment after elective surgery). This categorization is represented by the binary target variable y_i, which is also referred to as the output variable:

$$y_i = \begin{cases} 1, & \text{if patient } i \text{ is diagnosed as COVID - 19 - positive/ treated on the ICU} \\ 0, & \text{if patient } i \text{ is diagnosed as COVID - 19 - negative/not treated on the ICU.} \end{cases}$$

Hence, the available data set is given by the set of input-output pairs $\{\boldsymbol{x}_i, y_i\}_{i=1}^{n}$ with $(\boldsymbol{x}_i, y_i)$ representing one sample associated with a particular patient i. Furthermore, the outputs $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ as well as the features $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$ of all observations, i.e., all patients, can be summarized using vector and matrix notation, respectively (Hastie et al. 2009). A model commonly used for classification tasks, such as the COVID-19 diagnosis of patients, is the Logistic Regression (LR) model (Feng et al. 2014) which is based on the following hypothesis:

$$h_\omega(\boldsymbol{x}_i) = \mathcal{P}(y_i = 1 | \boldsymbol{x}_i) = \frac{1}{1 + e^{-\boldsymbol{\omega}^T \boldsymbol{x}_i}},$$

with the $\boldsymbol{\omega}^T = (\omega_0, \omega_1, \omega_2, \ldots, \omega_m)$ representing the vector of weights indicating the impact of the features on the output. Note that the bias parameter ω_0 is introduced to account for any fixed offset in the data. In this regard, from a mathematical point of view, it is convenient to introduce an auxiliary feature $x_{i0} = 1 \, \forall \, i = 1, \ldots, n$, which corresponds with the bias (Bishop 2006; Hastie et al. 2009; Joshi 2020). The hypothesis specifies the probability that patient i belongs to the class of patients diagnosed as COVID-19 positive / treated on the ICU, i.e., $y_i = 1$, given the information $\boldsymbol{x}_i$ on this patient. Therefore, patient i is predicted to be COVID-19-positive / treated on the ICU ($\widehat{y}_i = 1$) if the corresponding probability $h_\omega(\boldsymbol{x}_i)$ is at least 50%, which is equivalent to $\boldsymbol{\omega}^T \boldsymbol{x}_i \geq 0$, otherwise the patient is predicted to be COVID-19-negative / not to be treated on the ICU ($\widehat{y}_i = 0$) (Géron 2019). Relying on the specified LR model, the loss function $f(\boldsymbol{\omega}) = \frac{1}{n} \sum_{i=1}^{n} f_i(\boldsymbol{\omega})$, which is used to measure the accuracy of the predicted probabilities $h_\omega(\boldsymbol{x}_i)$, can be specified based on individual, patient-specific losses f_i:

$$f_i(\boldsymbol{\omega}) = -\left[y_i \log\left(h_\omega(\boldsymbol{x}_i)\right) + (1-y_i)\log\left(1-h_\omega(\boldsymbol{x}_i)\right)\right] + \left[\frac{\lambda}{2n}\sum_{j=1}^{m}\omega_j^2\right].$$

Equation (2.3) denotes the regularized binary cross-entropy loss (Ho and Wookey 2020) for patient i with $\frac{\lambda}{2n}\sum_{j=1}^{m}\omega_j^2$ representing the regularization term, which is used to prevent overfitting and hence to ensure the generalizability of the ML model. Accordingly, to determine the optimal parameters $\boldsymbol{\omega}^* = \underset{\boldsymbol{\omega}}{\operatorname{argmin}} f(\boldsymbol{\omega})$, the resulting optimization problem is to minimize the loss function $f(\boldsymbol{\omega})$ (Algamal and Lee 2015):

$$\min_{\boldsymbol{\omega}\in\mathbb{R}^d} f(\boldsymbol{\omega}) \quad \text{with} \quad f(\boldsymbol{\omega}) = \frac{1}{n}\sum_{i=1}^{n} f_i(\boldsymbol{\omega}).$$

Note that d represents the number of parameters, including the bias ω_0, so $d = m + 1$ holds (Konečný et al. 2016).

Appendix 2: Overview of federated learning algorithms

See Table 8.

Table 8 Centralized aggregation algorithms based on Liu et al. (2022)

Algorithm name	References
Adaptive personalized Federated Learning (APFL)	Deng et al. (2020)
Attentive Federated Aggregation (FedAttOpt)	Jiang et al. (2020)
Federated Averaging (FA)	McMahan et al. (2017)
Federated Learning with added proximal term (FedProx)	Li et al. (2020b)
Federated Learning with Personalization Layers (FedPer)	Arivazhagan et al. (2019)
Federated Matched Averaging (FedMA)	Wang et al. (2020a, b)
Federated Multiple Gradient Descent Algorithm (FedMGDA+)	Hu et al. (2022)
Federated Stochastic Block Coordinate Descent (FedBCD)	Liu et al. (2019b)
Federated Stochastic Variance Reduced Gradient (FSVRG)	Konečný et al. (2016)
Personalized Federated Learning using Moreau Envelope (pFedMe)	Dinh et al. (2020)
Stochastic Agnostic Federated Learning (SAFL)	Mohri et al. (2019)
Stochastic Controlled Averaging for Federated Learning (SCAFFOLD)	Karimireddy et al. (2020)

The current state-of-the-art algorithm for implementing a FL system is FA (Nguyen et al. 2021) introduced by Google. FA relies on a centralized server that aggregates multiple local ML models from local clients by averaging, i.e., by taking a weighted sum of the respective local model parameters. The aggregated parameters then compose a global model, which is redistributed to the clients via the server, and the training process continues iteratively. However, while FA represents a straightforward technique for implementing a FL system (Liu et al. 2022), Li et al. (2020a) state that FA can lead to biased results under certain conditions, such as when dealing with heterogenous data. Accordingly, other concepts such as FedBCD have recently

been introduced to address limitations associated with FA. Specifically, the FedBCD approach focuses on the communication effort among the clients, i.e., the number of federated rounds, as a bottleneck of the collaborative training process (Li et al. 2020a).

The purpose of FedBCD is to reduce the number of federated rounds, by performing multiple local updates of local model parameters prior to communication between the clients and the centralized server, i.e., prior to the aggregation step (Liu et al. 2022). Another approach that also focuses on communication efficiency, similar to FedBCD, is provided by FSVRG.

Moreover, FSVRG also addresses the problem of heterogeneous data by reducing the variance associated with the estimation of gradients (Konečný et al. 2016). As already indicated, another challenge that arises with FL problems is the heterogeneity inherent in federated systems of multiple clients collaborating with each other, distinguishing between statistical heterogeneity and system heterogeneity. While statistical heterogeneity represents the situation of data being non-identically distributed across the clients, system heterogeneity refers to the substantially differing characteristics among the systems of the devices used by the clients. More precisely, system heterogeneity appears when the clients have to face different resource constraints with regard to their devices, such as different hardware, network connections, or battery levels. One approach to deal with heterogeneity is given by FedProx, which can be seen as both a generalization and a reparameterization of FA that enables the clients to contribute a variable amount of local work to the collaboration, i.e., running a different number of local iterations (Li et al. 2020b). Besides, there is another approach that addresses heterogeneity, SCAFFOLD, where statistical heterogeneity is tackled and in addition, similar to FedBCD, the communication effort is also reduced. For reducing the variance between clients, which is also referred to as client drift resulting from the non-identically distributed data across clients, control variables are used in SCAFFOLD to counteract the client drift (Karimireddy et al. 2020). For completeness, it is worth mentioning that with APFL, FedPer, and pFedMe, further concepts are available that focus on tackling statistical heterogeneity (Liu et al. 2022).

In addition to reducing the communication effort and dealing with heterogeneity, the fairness among the collaborating clients is also an important aspect in FL, which is addressed via SAFL and FedMGDA+. In general, fairness can be understood as the possibility to equally consider the distribution of data among the collaborating clients without irrelevant factors being involved (Liu et al. 2022). Taking this understanding of fairness even further, the entitlement of a client to a collaboratively trained model can be determined based on the contribution to the collaborative training process in terms of data provided by the respective client. Using this FL system, as opposed to the common FL paradigm where each client receives the same final model regardless its contribution, each client is provided with a different version of the final global model, which depends on its individual contribution (Lyu et al. 2020b). Another interpretation of fairness is provided by Li et al. (2019), where fairness is defined as uniformity with respect to the performance, i.e., a FL system that achieves an even distribution of the performance across the clients. Similar to the interpretation of fairness according to Li et al. (2019), Jiang et al. (2020) propose FedAttOpt, an approach that uses a so-called attention-augmented mechanism aiming to generate as much

benefit as possible for all collaborating clients. A more detailed discussion of these fairness concepts is beyond the scope of this work, so the reader is referred to Lyu et al. (2020b); Li et al. (2019); Jiang et al. (2020), respectively.

So far, it is assumed that the listed FL concepts can be generally applied to any ML model and are not exclusively designed for a particular type of model, such as neural networks (NNs). However, for the latter, including convolutional NNs (CNNs), an approach called FedMA exists that explicitly exploits the architecture of NNs to fit a collaboratively trained model in a layer-wise manner. As indicated, these models consist of so-called hidden layers, each composed of hidden elements. With FedMA, those hidden elements of different local networks that are similar to each other are matched and averaged (Wang et al. 2020a, b).

Appendix 3: Description of algorithm 3

At first, a pre-nested computation of the global gradient using a central server is necessary. Since it is assumed that the clients only use local data and do not make this data available for other clients, the clients first compute their local gradients $\nabla F_1\left(\widetilde{\boldsymbol{\omega}}\right), \ldots, \nabla F_K\left(\widetilde{\boldsymbol{\omega}}\right)$, which are then sent to a centralized server for aggregation (Algorithm 3, line 5). The aggregated global gradient $\nabla f\left(\widetilde{\boldsymbol{\omega}}\right)$ is then distributed to all local clients and used as part of local updates, where each local client k individually carries out SVRG-updates of the parameter vector (Algorithm 3, lines 6–11).

Finally, the server is also used for aggregating the local results $\boldsymbol{\omega}_q^1, \ldots, \boldsymbol{\omega}_q^K$ obtained from the last stochastic update steps $q = n_1, \ldots, n_K$ by weighting them according to the respective different local data sizes (Algorithm 3, line 13). Thereby, the diagonal matrix $\boldsymbol{A}$ is used to assign a higher weight to the scarce information of features available at only a few clients than to the information of a feature occurring at more clients, according to the degree of information scarcity captured by a^j. In other words, the scarcer the availability of a particular feature, the more valuable the associated information is compared to features that are available to a less limited extent. Another aspect that can be attributed to the sparsity patterns in the data, is the use of another diagonal matrix, $\boldsymbol{S}_k$, during the stochastic update step (Algorithm 3, line 10). This serves the purpose of balancing the contribution of those features, which are unequally distributed among the clients, to the stochastic gradient computation. This is necessary as sampling a non-missing observation of a feature (Algorithm 3, line 8), which mainly occurs at a specific client k, is much more likely for this client k than for those clients who lack non-missing observations of that feature. This may cause the FSVRG algorithm failing to converge and diverging instead. To avoid this, the local stochastic gradients used for the approximation of $\nabla f\left(\widetilde{\boldsymbol{\omega}}\right)$ must be scaled on a per-feature basis such that these approximations are of the same order of magnitude.

Acknowledgements We express our deep gratitude to all study teams supporting the LEOSS study. The LEOSS study group contributed at least 5 per mille to the analyses of this study: University Hospital Regensburg (Frank Hanses), Technical University of Munich (Christoph Spinner), Hospital Ingolstadt (Stefan Borgmann), University Hospital Freiburg (Siegbert Rieg), Klinikum Dortmund gGmbH, Hospital of University Witten/Herdecke (Martin Hower), University Hospital Frankfurt (Maria Vehreschild), University Hospital Jena (Maria Madeleine Ruethrich), Practice at Ebertplatz Cologne (Christoph Wyen), Hospital Bremen-Center (Christian Hohmann), Hospital Passau (Julia Lanznaster), University Hospital Augsburg (Christoph Roemmele), Johannes Wesling Hospital Minden Ruhr University Bochum (Kai Wille), Hospital Ernst von Bergmann (Lukas Tometten), University Hospital Essen (Sebastian Dolff), University Hospital Munich/ LMU (Michael von Bergwelt-Baildon), University Hospital Heidelberg (Uta Merle), Robert-Bosch-Hospital Stuttgart (Katja Rothfuss), University Hospital Wuerzburg (Nora Isberner), University Hospital Cologne (Norma Jung), University Hospital Tuebingen (Siri Göpel), Hospital Maria Hilf GmbH Moenchengladbach (Juergen vom Dahl), Municipal Hospital Karlsruhe (Christian Degenhardt), University Hospital Erlangen (Richard Strauss), University Hospital Ulm (Beate Gruener), Hospital Leverkusen (Lukas Eberwein), Catholic Hospital Bochum (St. Josef Hospital) Ruhr University Bochum (Kerstin Hellwig), Bundeswehr Hospital Koblenz (Dominic Rauschning), Evangelisches Hospital Saarbruecken (Mark Neufang), Marien Hospital Herne Ruhr University Bochum (Timm Westhoff), Tropical Clinic Paul-Lechler Hospital Tuebingen (Claudia Raichle), Hacettepe University (Murat Akova), University Hospital Duesseldorf (Bjoern-Erik Jensen), Elbland Hospital Riesa (Joerg Schubert), Center for Infectiology Prenzlauer Berg Berlin (Stephan Grunwald), University Hospital Schleswig-Holstein Kiel (Anette Friedrichs), University Hospital of Giessen and Marburg (Janina Trauth), University Hospital Dresden (Katja de With), Clinic Munich (Wolfgang Guggemos), Hospital Braunschweig (Jan Kielstein), Agaplesion Diakonie Hospital Rotenburg (David Heigener), Hospital Fulda (Philipp Markart), University Hospital Saarland (Robert Bals), Petrus Hospital Wuppertal (Sven Stieglitz), Elisabeth Hospital Essen (Ingo Voigt), Richmond Research Institute (Jorg Taubel), Malteser Hospital St. Franziskus Flensburg (Milena Milovanovic).
The LEOSS study infrastructure group: Jörg Janne Vehreschild (Goethe University Frankfurt), Susana M. Nunes de Miranda (University Hospital of Cologne), Margarete Scherer (University Hospital of Cologne), Carolin E. M. Koll (University Hospital of Cologne), Melanie Stecher (University Hospital of Cologne), Lisa Pilgram (Goethe University Frankfurt), Nick Schulze (University Hospital of Cologne), Annika Claßen (University Hospital of Cologne), Bernd Franke (University Hospital of Cologne), Sandra Fuhrmann (University Hospital of Cologne), und Fabian Praßer (Charité, Universitätsmedizin Berlin).

Funding Open Access funding enabled and organized by Projekt DEAL. Not applicable.

Data availability Data will be made available upon reasonable request due to the sensitive nature of the data. Please contact the corresponding author.

Declarations

Conflict of interest Not applicable.

Generative AI and AI-assisted technologies in the writing process During the preparation of this work the author(s) used ChatGPT in order to improve language and readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

Aïvodji UM, Gambs S, Martin A (2019) IOTFLA: a secured and privacy-preserving smart home architecture implementing federated learning. In: Proceedings of the 2019 IEEE security and privacy workshops, pp 175–180. https://doi.org/10.1109/SPW.2019.00041

Algamal ZY, Lee MH (2015) Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification. Expert Syst Appl 42(23):9326–9332. https://doi.org/10.1016/j.eswa.2015.08.016

Arivazhagan MG, Aggarwal V, Singh AK, Choudhary S (2019) Federated learning with personalization Layers. Preprint at https://doi.org/10.48550/arXiv.1912.00818

Bartenschlager CC, Ebel SS, Kling S, Vehreschild J, Zabel LT, Spinner CD et al (2022) COVIDAL: A machine learning classifier for digital COVID-19 diagnosis in German hospitals. ACM Trans Manage Inform Syst 14(2):1–16. https://doi.org/10.1145/3567431

Bartenschlager CC, Grieger M, Erber J, Neidel T, Borgmann S, Vehreschild JJ et al (2023) Covid-19 triage in the emergency department 2.0: how analytics and AI transform a human-made algorithm for the prediction of clinical pathways. Health Care Manag Sci 26:412–429. https://doi.org/10.1007/s10729-023-09647-2

Bertsimas D, Pawlowski C, Zhuo YD (2018) From predictive methods to missing data imputation: an optimization approach. J Mach Learn Res 18:7133–7171

Bini SA (2018) Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? J Arthroplast 33(8):2358–2361. https://doi.org/10.1016/j.arth.2018.02.067

Binkheder S, Asiri MM, Altowayan KW, Alshehri TM, Alzarie MF, Aldekhyyel RN et al (2021) Real-world evidence of COVID-19 Patients' data quality in the electronic health records. Healthcare 9(12):1648

Bishop CM (2006) Pattern recognition and machine learning. Springer, New York, USA

Brisimi TS, Chen R, Mela T, Olshevsky A, Paschalidis IC, Shi W (2018) Federated learning of predictive models from federated electronic health records. Int J Med Informatics 112:59–67. https://doi.org/10.1016/j.ijmedinf.2018.01.007

Brownlee J (2020) Data Preparation for machine learning: data cleaning, feature selection, and data transforms in python. Machine Learning Mastery, Vermont, Australia

Brownlee J (2021) SMOTE for imbalanced classification with Python. https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification . Accessed on 14 Mar 2025

Caan B, Hiatt RA, Owen AM (1991) Mailed dietary surveys: response Rates, error Rates, and the effect of omitted food items on nutrient values. Epidemiology 2(6):430–436

Calabrese JM, Demers J (2022) How optimal allocation of limited testing capacity changes epidemic dynamics. J Theor Biol 538:111017

Carter MW, Busby CR (2023) How can operational research make a real difference in healthcare? Challenges of implementation. Eur J Oper Res 306(3):1059–1068. https://doi.org/10.1016/j.ejor.2022.04.022

Deng Y, Kamani MM, Mahdavi M 2020 Adaptive Personalized Federated Learning. Preprint at https://doi.org/10.48550/arXiv.2003.13461

Dinh CT, Tran NH, Nguyen TD (2020) Personalized federated learning with Moreau envelopes. Adv Neural Inf Process Syst 33:21394–21405

ElbirAM, Soner B, Cöleri S, Gündüz D, Bennis M (2022) Federated learning in vehicular networks. In: Proceedings of the 2022 IEEE international Mediterranean conference on communications and networking, pp 72–77. https://doi.org/10.1109/MeditCom55741.2022.9928621

Feng J, Huan X, Mannor S, Yan S (2014) Robust logistic regression and classification. Adv Neural Inf Process Syst 27:253–261

Feng J, Rong C, Sun F, Guo D, Li Y (2020) PMF: A Privacy-preserving human mobility prediction framework via federated learning. Proc ACM Interact Mob Wearable Ubiquitous Technol 4(1):1–21. https://doi.org/10.1145/3381006

Fraser GE, Yan R, Butler TL, Jaceldo-Siegl K, Beeson WL, Chan J (2009) Missing data in a long food frequency questionnaire: are imputed zeros correct? Epidemiology 20(2):289–294. https://doi.org/10.1097/EDE.0b013e31819642c4

Gao D, Ju C, Wei X, Liu Y, Chen T, Yang Q (2019) HHHFL: Hierarchical heterogeneous horizontal federated learning for electroencephalography. Preprint at https://doi.org/10.48550/arXiv.1909.05784

Géron A (2019) Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems, 2nd edition. O'Reilly Media, Sebastopol, USA
Goodfellow I, Bengio Y, Courville A (2016) Deep learning. The MIT Press, Cambridge, USA
HanX, Yu H, Gu H (2019) Visual inspection with federated learning. In: Proceedings of the 16th international conference on image analysis and recognition, pp 52–64. https://doi.org/10.1007/978-3-030-27272-2_5
Hastie T, Tibshirani R, Friedman J (2009) The Elements of statistical learning–data mining, inference, and prediction, 2nd edn. Springer, New York, USA
He C, Annavaram M, Avestimehr S (2020a) Towards Non-I.I.D. and invisible data with FedNAS: federated deep learning via neural architecture search. Preprint at https://doi.org/10.48550/arXiv.2004.08546
He C, Ye H, Shen L, Zhang T (2020b) MiLeNAS: efficient neural architecture search via mixed-level reformulation. In: Proceedings of the 2020 IEEE/CVF conference on computer vision and pattern recognition, pp 11993–12002. https://doi.org/10.1109/CVPR42600.2020.01201
Ho Y, Wookey S (2020) The real-world-weight cross-entropy loss function: modeling the costs of mislabeling. IEEE Access 8:4806–4813. https://doi.org/10.1109/ACCESS.2019.2962617
HuB, Gao Y, Liu L, Ma H (2018) Federated region-learning: An edge computing based framework for urban environment sensing. In: Proceedings of the 2018 IEEE global communications conference, pp 1–7. https://doi.org/10.1109/GLOCOM.2018.8647649
Hu Z, Shaloudegi K, Zhang G, Yu Y (2022) Federated learning meets multi-objective optimization. IEEE Trans Netw Sci Eng 9(4):2039–2051. https://doi.org/10.1109/TNSE.2022.3169117
Huang L, Shea AL, Qian H, Masurkar A, Deng H, Liu D (2019) Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records. J Biomed Inform 99:103291. https://doi.org/10.1016/j.jbi.2019.103291
Huang L, Yin Y, Fu Z, Zhang S, Deng H, Liu D (2020) LoAdaBoost: Loss-Based adaboost federated machine learning with reduced computational complexity on IID and non-IID intensive care data. PLoS ONE 15(4):e0230706. https://doi.org/10.1371/journal.pone.0230706
Intel (2020) Intel works with University of Pennsylvania in using privacy-preserving AI to identify brain tumors. https://www.intc.com/news-events/press-releases/detail/5/intel-works-with-university-of-pennsylvania-in-using . Accessed on 13 Mar 2025
Ismail AR, Abidin NZ, Maen MK (2022) Systematic review on missing data imputation techniques with machine learning algorithms for healthcare. J Rob Control 3(2):143–152. https://doi.org/10.18196/jrc.v3i2.13133
James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning-with applications in R. Springer, New York, USA
Jiang J, Ji S, Long G (2020) Decentralized knowledge acquisition for mobile internet applications. World Wide Web 23(5):2653–2669. https://doi.org/10.1007/s11280-019-00775-w
Johnson R, Zhang T (2013) Accelerating stochastic gradient descent using predictive variance reduction. Adv Neural Inf Process Syst 26:315–323
Joshi AV (2020) Machine learning and artificial intelligence. Springer International Publishing, Cham, Switzerland
Kaptein M, van den Heuvel E (2022) Statistics for data scientists: an introduction to Probability, Statistics, and data analysis. Springer International Publishing, Cham, Switzerland
Karimireddy SP, Kale S, Mohri M, Reddi S, Stich S, Suresh AT (2020) SCAFFOLD: stochastic controlled averaging for federated learning. Proc Mach Learn Res 119:5132–5143
Kim Y, Sun J, Yu H, Jiang X (2017) Federated tensor factorization for computational phenotyping. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, pp 887–895
Konečný J, McMahan HB, Ramage D, Richtárik P (2016) Federated optimization: distributed machine learning for on-device intelligence. Preprint at https://doi.org/10.48550/arXiv.1610.02527
Lee J, Sun J, Wang F, Wang S, Jun C-H, Jiang X (2018) Privacy-preserving patient similarity learning in a federated environment: development and analysis. JMIR Med Inf 6(2):e20. https://doi.org/10.2196/medinform.7744
Li T, Sanjabi M, Beirami A, Smith V (2019) Fair resource allocation in federated learning. Preprint at https://doi.org/10.48550/arXiv.1905.10497
Li T, Sahu AK, Talwalkar A, Smith V (2020a) Federated learning: challenges, methods, and future directions. IEEE Signal Process Mag 37(3):50–60. https://doi.org/10.1109/MSP.2020.2975749
Li T, Sahu AK, Zaheer M, Sanjabi M, Talwalkar A, Smith V (2020b) Federated optimization in heterogenous networks. In: Proceedings of machine learning and systems, vol 2, pp 429–450

Li L, Fan Y, Tse M, Lin K-Y (2020c) A review of applications in federated learning. Comput Ind Eng 149:106854. https://doi.org/10.1016/j.cie.2020.106854
Little RJ, Rubin DB (2019) Statistical analysis with missing data, 3rd edn. Wiley, Hoboken, USA
Liu D, Dligach D, Miller T (2019a) Two-stage federated phenotyping and patient representation learning. In: Proceedings of the 18th BioNLP Workshop and shared task, pp 283–291. https://doi.org/10.18653/v1/W19-5030
LiuY, Kang Y, Zhang X, Li L, Cheng Y, Chen T et al (2019b) A communication-efficient collaborative learning framework for distributed features. Preprint at https://doi.org/10.48550/arXiv.1912.11187
Liu B, Wang L, Liu M, Xu C-Z (2020a) Federated imitation learning: a novel framework for cloud robotic systems with heterogeneous sensor data. IEEE Rob Autom Lett 5(2):3509–3516. https://doi.org/10.1109/LRA.2020.2976321
Liu F, Wu X, Ge S, Fan W, Zou Y (2020b) Federated learning for Vision-and-Language grounding problems. Proc AAAI Conf Artif Intell 34(7):11572–11579. https://doi.org/10.1609/aaai.v34i07.6824
Liu J, Huang J, Zhou Y, Li X, Ji S, Xiong H et al (2022) From distributed machine learning to federated learning: a survey. Knowl Inf Syst 64(4):885–917. https://doi.org/10.1007/s10115-022-01664-x
Luo Y (2022) Evaluating the state of the Art in missing data imputation for clinical data. Brief Bioinform 23(1):1–9. https://doi.org/10.1093/bib/bbab489
Lyu L, Yu H, Yang Q (2020a) Threats to federated learning: a survey. Preprint at https://doi.org/10.48550/arXiv.2003.02133
Lyu L, Yu J, Nandakumar K, Li Y, Ma X, Jin J et al (2020b) Towards fair and Privacy-Preserving federated deep models. IEEE Trans Parallel Distrib Syst 31(11):2524–2541. https://doi.org/10.1109/TPDS.2020.2996273
Martín-Fernández JA, Barceló-Vidal C, Pawlowsky-Glahn V (2003) Dealing with zeros and missing values in compositional data sets using nonparametric imputation. Math Geol 35(3):253–278. https://doi.org/10.1023/A:1023866030544
McMahan B, Moore E, Ramage D, Hampson S, Agüera y Arcas B (2017) Communication-Efficient learning of deep networks from decentralized data. Proc 20th Int Conf Artif Intell Stat 54:1273–1282
Michels KB, Willett WC (2009) Self-Administered semiquantitative food frequency questionnaires: patterns, predictors, and interpretation of omitted items. Epidemiology 20(2):295–301. https://doi.org/10.1097/EDE.0b013e3181931515
Mohammed R, Rawashdeh J, Abdullah M (2020) Machine learning with oversampling and undersampling techniques: overview study and experimental results. In: Proceedings of 11th International conference on information and communication systems 243–248. https://doi.org/10.1109/ICICS49469.2020.239556
Mohri M, Sivek G, T Suresh A (2019) Agnostic federated learning. Proc Mach Learn Res 97:4615–4625
Mowla NI, Tran NH, Doh I, Chae K (2020) Federated learning-based cognitive detection of jamming attack in flying Ad-Hoc network. IEEE Access 8:4338–4350. https://doi.org/10.1109/ACCESS.2019.2962873
Naylor CD (2018) On the prospects for a (Deep) learning health care system. JAMA 320(11):1099–1100. https://doi.org/10.1001/jama.2018.11103
Nguyen VD, Sharma SK, Vu TX, Chatzinotas S, Ottersen B (2021) Efficient federated learning algorithm for resource allocation in wireless IoT networks. IEEE Internet Things J 8(5):3394–3409. https://doi.org/10.1109/JIOT.2020.3022534
Obermeyer Z, Emanuel EJ (2016) Predicting the future—big data, machine learning, and clinical medicine. N Engl J Med 375(13):1216–1219. https://doi.org/10.1056/NEJMp1606181
Oleynik M, Kugic A, Kasáč Z, Kreuzthaler M (2019) Evaluating shallow and deep learning strategies for the 2018 n2c2 shared task on clinical text classification. J Am Med Inform Assoc 26(11):1247–1254. https://doi.org/10.1093/jamia/ocz149
Owkin (2025) Federated learning—The future of federated learning in health care starts here. https://owkin.com/federated-learning/ . Accessed on 13 Mar 2025
PfohlSR, Dai AM, Heller K (2019) Federated and differentially private learning for electronic health records. Preprint at https://doi.org/10.48550/arXiv.1911.05861
Pokhrel SR, Choi J (2020) Federated learning with blockchain for autonomous vehicles: analysis and design challenges. IEEE Trans Communication 68(8):4734–4746. https://doi.org/10.1109/TCOMM.2020.2990686
Qian Y, Hu L, Chen J, Guan X, Hassan MM, Alelaiwi A (2019) Privacy-aware service placement for mobile edge computing via federated learning. Inf Sci 505:562–570. https://doi.org/10.1016/j.ins.2019.07.069

Rajendran S, Obeid JS, Binol H, D`Agostino R, Foley K, Zhang W et al (2021) Cloud-Based federated learning implementation across medical centers. JCO Clin Cancer Inf 5:1–11. https://doi.org/10.1200/CCI.20.00060

Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S et al (2020) The future of digital health with federated learning. NPJ Digit Med 3:119. https://doi.org/10.1038/s41746-020-00323-1

Saputra YM, Hoang DT, Nguyen DN, Dutkiewicz E, Mueck MD, Srikanteswara S (2019) Energy demand prediction with federated learning for electric vehicle networks. In: Proceedings of the 2019 IEEE global communications conference, pp 1–6. https://doi.org/10.1109/GLOBECOM38437.2019.9013587

Sarma KV, Harmon S, Sanford T, Roth HR, Xu Z, Tetreault J et al (2021) Federated learning improves site performance in multicenter deep learning without data sharing. J Am Med Inform Assoc 28(6):1259–1264. https://doi.org/10.1093/jamia/ocaa341

Schiele J, Koperna T, Brunner JO (2020) Predicting intensive care unit bed occupancy for integrated operating room scheduling via neural networks. Nav Res Logist 68(1):65–88. https://doi.org/10.1002/nav.21929

Shaheen M, Farooq MS, Umer T, Kim BS (2022) Applications of federated Learning; Taxonomy, Challenges, and research trends. Electronics 11(4):670. https://doi.org/10.3390/electronics11040670

Silva S, Gutman BA, Romero E, Thompson PM, Altmann A, Lorenzi M (2019) Federated learning in distributed medical databases: meta-analysis of large-scale subcortical brain data. In: Proceedings of the 16th IEEE international symposium on biomedical imaging, pp 270–274. https://doi.org/10.1109/ISBI.2019.8759317

Sozinov K, Vlassov V, Girdzijauskas S (2018) Human activity recognition using federated learning. In: Proceedings 2018 IEEE International Conference on parallel & distributed processing with applications, ubiquitous computing & communications, big data & cloud computing, social computing & sustainable computing, pp 1103–1111. https://doi.org/10.1109/BDCloud.2018.00164

Stekhoven DJ, Bühlmann P (2012) MissForest – non-parametric missing value imputation for mixed-type data. Bioinformatics 28(1):112–118. https://doi.org/10.1093/bioinformatics/btr597

The Lancet Regional Health – Europe (2025) The Italian health data system is broken. Lancet Reg Health – Europe 48:101206. https://doi.org/10.1016/j.lanepe.2024.101206

Tukey JW (1977) Exploratory data analysis. Addison-Wesley Pub Co 688:581–582

Van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ et al (2014) A systematic review of barriers to data sharing in public health. BMC Public Health 14:1144. https://doi.org/10.1186/1471-2458-14-1144

Wang X, Li A, Jiang Z, Feng H (2006) Missing value estimation for DNA microarray gene expression data by Support Vector Regression imputation and orthogonal coding scheme. BMC Bioinformatics 7(32). https://doi.org/10.1186/1471-2105-7-32

Wang F, Preininger A (2019) AI in health: state of the Art, Challenges, and future directions. IMIA Yearbook Med Inf 28(1):16–26. https://doi.org/10.1055/s-0039-1677908

Wang X, Han Y, Wang C, Zhao Q, Chen X, Chen M (2019) In-Edge AI: intelligentizing mobile edge Computing, caching and communication by federated learning. IEEE Netw 33(5):156–165. https://doi.org/10.1109/MNET.2019.1800286

Wang H, Yurochkin M, Sun Y, Papailiopoulos D, Khazaeni Y (2020a) Federated learning with matched averaging. Preprint at https://doi.org/10.48550/arXiv.2002.06440

Wang Y, Tong Y, Shi D (2020b) Federated latent dirichlet allocation: A local differential privacy based framework. Proc 2020 AAAI Conf Artif Intell 34(4):6283–6290. https://doi.org/10.1609/aaai.v34i04.6096

Wani D, Malhotra M (2018) Does the meaningful use of electronic health records improve patient outcomes? J Oper Manag 60(1):1–18. https://doi.org/10.1016/j.jom.2018.06.003

Wells BJ, Nowacki AS, Chagin K, Kattan MW (2013) Strategies for handling missing data in electronic health record derived data. J Electron Health Data Methods 1(3):7. https://doi.org/10.13063/2327-9214.1035

Xu J, Glicksberg BS, Su C, Walker P, Bian J, Wang F (2021) Federated learning for healthcare informatics. J Healthc Inf Res 5(1):1–19. https://doi.org/10.1007/s41666-020-00082-4

Yang Q, Liu Y, Chen T, Tong Y (2019a) Federated machine learning: concept and applications. ACM Trans Intell Syst Technol 10(2):1–19. https://doi.org/10.1145/3298981

YangW, Zhang Y, Ye K, Li L, Xu C-Z (2019b) FFD: a federated learning based method for credit card fraud detection. In: Proceedings of the 8th international congress, held as part of the services conference federation, pp 18–32. https://doi.org/10.1007/978-3-030-23551-2_2

Yu T, Li T, Sun Y, Nanda S, Smith V, Sekar V et al (2020) Learning context-aware policies from multiple smart homes via federated multi-task learning. In: Proceedings of the 5th IEEE/ ACM international conference on internet-of-things design and implementation, pp 104–115. https://doi.org/10.1109/IoTDI49375.2020.00017

Authors and Affiliations

Elion Shala[1] · Jens O. Brunner[2,3] · Jörg J. Vehreschild[4] · Axel R. Heller[5] · Christina C. Bartenschlager[5,6]

✉ Christina C. Bartenschlager
christina.bartenschlager@th-nuernberg.de

1 Chair for Analytics and Optimization, Faculty of Business and Economics, University of Augsburg, Augsburg, Germany

2 Health Care Operations/Health Information Management, Faculty of Business and Economics, Faculty of Medicine, University of Augsburg, Augsburg, Germany

3 Department of Technology, Management, and Economics, Technical University of Denmark, Kopenhagen, Denmark

4 Faculty of Medicine, Institute for Digital Medicine and Clinical Data Science, Goethe University Frankfurt, Frankfurt, Germany

5 Anaesthesiology and Operative Intensive Care, University Hospital of Augsburg, Augsburg, Germany

6 Applied Data Science in Health Care, Nürnberg School of Health, Ohm University of Applied Sciences Nuremberg, Nuremberg, Germany

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH ("Springer Nature").

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users ("Users"), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use ("Terms"). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com