International Journal of Emerging Science and Engineering (IJESE)
ISSN: 2319–6378 (Online), Volume-14 Issue-4, March 2026

Scaffold Extensions for Client Drift Mitigation in Federated Learning: A Synthesis of Approaches, Limitations, and Future Directions

James Mburu Muthii, Stephen Kahara Wanjau, Stephen Njenga

***Abstract:** Client drift arising from non-independent and identically distributed (non-IID) data across participating clients remains one of the most critical obstacles to effective Federated Learning. The Scaffold algorithm, which introduces control variates to correct local gradient updates, has emerged as one of the most prominent variance reduction methods for mitigating this drift. Although numerous extensions to Scaffold have been proposed, no systematic review has exclusively examined the Scaffold algorithm and the control variate mechanism for client drift mitigation, leaving the research community without a consolidated understanding of how Scaffold has been extended, what limitations persist, and which characteristics remain underexplored. This study addresses that gap through a systematic literature review guided by PRISMA 2020 guidelines. Seven electronic databases were searched for publications from 2016 to 2026, yielding 1,847 records, from which 33 studies were included after duplicate removal, screening, and full-text eligibility assessment based on criteria requiring each study to address Scaffold or control variates for client drift in FL and cover at least two performance metrics. Data were synthesized thematically using frequency counts and tabular summaries. The review reveals nine distinct extension approaches: variance reduction via gradient estimation techniques was the most prevalent (11 studies, 34%), followed by integration with advanced optimization algorithms (8 studies, 25%), together accounting for 59% of the reviewed work. Twelve Scaffold characteristics were targeted for extension, with variance reduction the most commonly modified (37%, rising to 50% with combined categories), while communication mechanism, privacy budget allocation, and similarity-based approaches remained significantly underexplored. Recurring limitations across all approaches included communication and computational overhead, hyperparameter sensitivity, restrictive theoretical assumptions, performance degradation under extreme data heterogeneity, and limited large-scale empirical validation. A notable finding is that similarity-based approaches for client drift mitigation are largely absent from the literature, with only one study employing a similarity measure. The review, therefore, recommends future investigation of similarity-based methods as adaptive control variates within the Scaffold protocol, alongside prioritization of communication-efficient, privacy-preserving designs validated at scale. This research was self-sponsored with no external funding.*

***Keywords:** Client Drift, Federated Learning, Scaffold, Variance Reduction.*

Nomenclature:
FL: Federated Learning
Non-IID: Non-Independent and Identically Distributed
Scaffold: Stochastic Controlled Averaging with Variance Reduction
SVRG: Stochastic Variance Reduced Gradient
SARAH: Stochastic Recursive Gradient Algorithm
SAM: Sharpness-Aware Minimization
PID: Proportional-Integral-Derivative
ADMM: Alternating Direction Method of Multipliers
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

I. INTRODUCTION

Federated Learning (FL) is a distributed machine learning paradigm that enables multiple decentralized clients to collaboratively train a shared global model without exchanging their raw data. [1], [2], [3]. Originally proposed by McMahan *et al.* [4] through the Federated Averaging (FedAvg) algorithm, FL keeps training data localized on each participating client, transmitting only model updates to a central server for aggregation [2]. This architectural distinction makes FL inherently amenable to privacy preservation, and it has consequently found extensive application in healthcare, mobile computing, autonomous driving, financial services, and Internet of Things (IoT) ecosystems [1] [5] [6] [7] [8].

However, FL faces several fundamental challenges, including systems heterogeneity, communication overhead, privacy and security threats, and limited client resources [3] [9] [10]. Among these, statistical heterogeneity, manifested as non-independent and identically distributed (non-IID) data across clients, is widely regarded as one of the most critical obstacles to effective federated optimization [11]. When client data distributions diverge significantly, local model updates can point in conflicting directions, a phenomenon termed client drift [12] [13]. Client drift degrades the convergence rate of the global model, increases the communication rounds required to reach target accuracy, and in severe cases can cause the global model to diverge entirely [11] [14].

To address client drift, federated optimisation algorithms have been extensively studied and can be broadly categorised into aggregation-based, regularisation-based, variance-reduction, and personalisation-based methods [15]. Evidence in the

Manuscript received on 26 February 2026 | First Revised Manuscript received on 04 March 2026 | Second Revised Manuscript received on 10 March 2026 | Manuscript Accepted on 15 March 2026 | Manuscript published on 30 March 2026.
*Correspondence Author(s)

James Mburu Muthii*, Department of Computer Science, School of Computing and Information Technology, Murang'a University of Technology, Murang'a, Kenya. Email ID: muthiimburu@gmail.com. ORCID: 0009-0008-1150-4817

Dr. Stephen Kahara Wanjau, Department of Computer Science, Murang'a University of Technology, Murang'a, Kenya. Email ID: skahara@mut.ac.ke. ORCID: 0000-0002-6567-6912

Dr. Stephen Njenga, Department of Computer Science, Murang'a University of Technology, Murang'a, Kenya. Email ID: njengast@gmail.com. ORCID: 0000-0001-7795-6287

Retrieval Number:100.1/ijese.D264114040426
DOI:10.35940/ijese.D2641.14040326
Journal Website: www.ijese.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

literature demonstrates that variance reduction methods offer stronger theoretical convergence guarantees under heterogeneous settings and can achieve target accuracy in fewer communication rounds [12] [16] [17]. This makes variance reduction a particularly compelling strategy, and, in particular, the Stochastic Controlled Averaging with Variance Reduction (Scaffold) algorithm, proposed by Karimireddy et al. [12], is one of the most prominent methods in this category.

Scaffold addresses client drift by introducing control variates to correct the direction of local updates on each client [12] [18]. The algorithm maintains a server-level control variate that estimates the global update direction and a client-level control variate for each participant that estimates the local update direction. During local training, each client's gradient step is corrected by the difference between the server and client control variates, effectively aligning local updates with the global optimization direction and reducing accumulated drift [12]. Unlike FedAvg [4], which applies no correction and is susceptible to arbitrary drift, and FedProx [11], which constrains but does not correct local updates, Scaffold directly estimates and counteracts the source of drift. Theoretically, Scaffold achieves linear speedup in the number of clients under heterogeneous settings, with a convergence rate independent of the degree of data heterogeneity [12], [16].

Nevertheless, Scaffold introduces additional communication overhead and memory requirements through the maintenance and transmission of control variates [18] [19]. Its effectiveness can further be affected by partial client participation, variations in local training steps, and the choice of control variate update strategies [20] [21]. These considerations have motivated a substantial body of research aimed at extending, enhancing, or modifying Scaffold to improve its practical performance and broaden its applicability.

The rest of the paper is organised as follows: Section 2 presents a discussion of related work. Section three discusses the research method. Section four focuses on the study discussion, presenting findings related to the three research questions: the approaches employed to extend Scaffold for client drift mitigation in FL, the limitations of these approaches, and the core characteristics of Scaffold targeted for extension. Section five focuses on the study conclusion and future work.

II. RELATED WORKS

There exist several surveys and reviews in the literature on federated learning optimization, data heterogeneity, and aggregation algorithms that address client drift. However, none of them exclusively and systematically reviews the Scaffold algorithm and the control variates mechanism for client-drift mitigation. The relevant existing reviews are discussed below.

Kairouz *et al*. [1] produced the most comprehensive FL survey to date, spanning 210 pages in Foundations and Trends in Machine Learning (2021), covering optimization challenges including client drift and variance reduction. While it positions Scaffold as a state-of-the-art method, the treatment is embedded within a broader discussion and does not systematically compare Scaffold variants or their empirical performance. Wang *et al*. [22] authored "A Field Guide to Federated Optimization" (2021), a practical guide that discusses Scaffold's control variates and highlights a critical limitation: Scaffold's incompatibility with cross-device FL due to statefulness requirements. However, this work is not structured as a systematic review, lacks a formal methodology, and does not provide a comprehensive account of the evolution of control variate-based methods. Li, *et al*. [23] published a systems-oriented survey in IEEE Transactions on Knowledge and Data Engineering (2023) that describes Scaffold as applying variance reduction to correct local updates, but treats it briefly alongside many other algorithms without a focused analysis of its mechanism, variants, or conditions under which it excels or fails.

Several reviews focus specifically on the non-IID data problem that Scaffold was designed to address. Jiménez G., *et al*. [24] Conducted a systematic review of 235 papers on non-IID data in FL for IEEE Communications Surveys & Tutorials (2024), providing a detailed taxonomy of data heterogeneity types. Scaffold is discussed as one of many solutions, but without an in-depth examination of its variants or convergence properties. Ye, *et al*. [25] published a comprehensive survey on heterogeneous FL in ACM Computing Surveys (2024) covering data, model, and system heterogeneity, discussing gradient correction and variance reduction as a core category. However, its broad scope means that the treatment of Scaffold-specific extensions, such as BN-SCAFFOLD, partial variance reduction, or FedDC, is absent. Zhu *et al*. [26] surveyed methods for non-IID data in Future Generation Computer Systems (2022), classifying Scaffold within variance reduction approaches, but the early publication date means many subsequent extensions are not covered. Chen *et al*. [27] surveyed robust FL with heterogeneity considerations, reviewing optimization methods including variance reduction, but Scaffold receives limited focused attention within the paper's broad five-dimensional heterogeneity framework.

From the reviewed literature, it is evident that while numerous surveys address federated learning optimization and data heterogeneity, none provides a dedicated, systematic review specifically focused on the Scaffold algorithm and the control variate mechanism for client drift mitigation. Existing reviews treat Scaffold as one of many algorithms within broader discussions, without systematically cataloguing its variants and extensions, analysing the specific conditions under which control variates provide advantages or exhibit limitations, or employing a formal review methodology that targets this mechanism. This systematic review, therefore, aims to fill this gap.

III. RESEARCH METHOD

This study sought to conduct a systematic literature review to expose the approaches employed to extend the Scaffold and control variates techniques for client-drift mitigation in Federated Learning (FL), their limitations, and the characteristics are extended. Scaffold, by its nature, is a control variates technique that

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

corrects for client drift by maintaining control variates at both the server and client levels. Consequently, this review encompasses not only extensions to Scaffold itself but also enhancements to the broader family of control variate methods used in FL to address client drift. The research was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [28] [29], since PRISMA provides a comprehensive, transparent, and reproducible framework for conducting and reporting systematic reviews. The study is categorized as a secondary study since it aims to review primary studies on approaches employed in extending Scaffold and control variates techniques for client drift mitigation in FL. Therefore, the focus is on uncovering the approaches in the literature that extend or modify the Scaffold and control variates techniques to mitigate client drift in FL, understanding their limitations, and pinpointing the characteristics extended in these approaches. This will expose practical and relevant technological approaches and best practices in Scaffold and control variates techniques aimed at client drift mitigation. The PRISMA process involved identification, screening, eligibility assessment, and inclusion of studies.

A. Planning for the Review

In the planning phase, the researcher established the research aim, research questions, search strategy, and inclusion and exclusion criteria. Further, the online databases to be utilized were selected.

B. Aim of the Systematic Literature Review

The systematic literature review sought to expose the approaches employed to extend Scaffold and control variates techniques for client drift mitigation in Federated Learning, their limitations, and the characteristics extended for this purpose.

C. Research Questions

The following research questions were crafted to help achieve the research objective.

i. What are the approaches to extending Scaffold and control variates techniques for client drift mitigation in Federated Learning?

ii. What are the limitations of the approaches employed in extending the Scaffold and control variates techniques for client drift mitigation in Federated Learning?

iii. What characteristics of the Scaffold and control variates techniques are extended for client drift mitigation in Federated Learning?

D. Search Strategy

A search strategy was developed using keywords and Boolean operators, as outlined below, to retrieve relevant papers from digital databases. The use of multiple keywords ensured that no relevant studies were missed. Since Scaffold is fundamentally a control variates technique, the search string was designed to capture both Scaffold-specific extensions and broader enhancements to control variates methods for addressing client drift in Federated Learning.

(SCAFFOLD OR SCAFFOLD Enhancement OR SCAFFOLD Optimization OR SCAFFOLD Modification OR SCAFFOLD Extension OR Control Variates OR Variance Reduction) AND (Client Drift OR Client Drift Mitigation OR Client Drift Correction OR Client Drift Reduction OR Data Heterogeneity OR Non-IID) AND (Federated Learning OR Federated Optimization)

The search was conducted across the following digital libraries: IEEE Xplore, Science Direct, Springer, ACM Digital Library, Semantic Scholar, arXiv, and Google Scholar. These databases were selected because they contain extensive literature on the knowledge domain under study, including peer-reviewed journal articles, conference proceedings, and preprints relevant to Federated Learning, control variates techniques, and variance reduction algorithms. A search was conducted in each digital library in turn. The crafted search strategy was employed in the search process to assist in retrieving journal articles and conference proceedings of interest, coupled with the year of publication between 2016 and 2026. The ten-year window was selected to capture the foundational works on control variates methods in distributed optimization as well as contemporary advancements following the introduction of Scaffold by Karimireddy *et al.* in 2020 [12]. The retrieved studies were written in English. The retrieved journal articles and conference proceedings were subsequently analysed and reported.

E. Inclusion and Exclusion Criteria

The criteria for including or excluding studies were established as follows:

- **Inclusion Criteria**
  - *i.* Peer-reviewed journal publications, conference proceedings, or established preprints (e.g., arXiv with subsequent peer-reviewed versions or significant citations).
  - *ii.* Studies published between the years 2016 and 2026.
  - *iii.* Studies about Scaffold, Scaffold extension or modification or enhancement, control variates techniques, variance reduction methods, and client drift mitigation in Federated Learning.
  - *iv.* Comparative studies about extended, enhanced, optimized, or modified Scaffold and other control variates or variance reduction techniques for client drift mitigation in Federated Learning.
  - *v.* Studies addressing at least two of the performance metrics under consideration, namely convergence rate, communication efficiency, model accuracy, client drift reduction, and computational overhead.
- **Exclusion Criteria**
  - *i.* Journal publications not peer-reviewed and unpublished conference proceedings (except established preprints with significant citations).
  - *ii.* Non-English publications.
  - *iii.* Duplicate studies.
  - *iv.* Publications not addressing Scaffold, control variates techniques, or variance reduction for client drift in Federated Learning.
  - *v.* Publications not about Federated Learning.
  - *vi.* Studies older than ten years (published before 2016).

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

vii. Studies addressing one or none of the performance metrics are under consideration.

F. Conducting the Review

This phase involves conducting the search, selecting relevant studies, assessing study quality, extracting data, and synthesising data. The PRISMA 2020 flow diagram was used to document and visualise the systematic process of study identification, screening, eligibility assessment, and inclusion.

G. Search Process

The study considered the following digital libraries: IEEE Xplore, Science Direct, Springer, ACM Digital Library, Semantic Scholar, arXiv, and Google Scholar. These databases were selected because they contained extensive literature on the knowledge domain under study, including peer-reviewed journals, top-tier machine learning conferences (such as NeurIPS, ICML, ICLR, and AAAI), and established preprint repositories. A search was conducted in each digital library in turn. The crafted search strategy was employed in the search process to assist in retrieving journal articles, conference proceedings, and preprints of interest, coupled with the year of publication between 2016 and 2026. The retrieved studies were written in English. A total of 1,847 records were identified across all seven databases. The retrieved journal articles, conference proceedings, and preprints were subsequently analysed and reported.

H. Selection of Relevant Studies

The retrieved journal articles, conference proceedings, and preprints were screened using PRISMA. In the identification phase, 412 duplicate records were removed before screening, leaving 1,435 unique records. These records were then screened by examining their titles and abstracts, and subjected to the following criteria:

i. Publications not addressing Scaffold, control variates techniques, or variance reduction for client drift in Federated Learning were excluded.

ii. Publications not about Federated Learning were excluded.

iii. Studies addressing one or none of the performance metrics under consideration were excluded.

iv. Studies not about Scaffold, Scaffold extension or modification or enhancement, control variates techniques, or variance reduction and client drift mitigation in Federated Learning were excluded.

After this screening phase, 1,240 records were excluded, leaving 195 reports for full-text retrieval. Of these, 7 reports could not be retrieved due to restricted access or unavailability, leaving 188 reports for eligibility assessment.

I. Study Quality Assessment

The 188 reports assessed for eligibility were subjected to full-text review and quality assessment. During this eligibility phase, 155 reports were excluded for the following reasons: not addressing Scaffold or control variates techniques (n = 47), not primarily about client drift mitigation (n = 39), addressing fewer than two performance metrics (n = 28), not about Federated Learning (n = 24), published outside the 2016–2026 range (n = 11), and non-English publications (n = 6). The quality threshold for the study was set as the ability of the selected journal article, conference proceeding, or preprint to answer all the research questions. Any study which did not meet the quality threshold criteria was excluded. To determine how well each study satisfied the quality criteria, the researcher used the following approach:

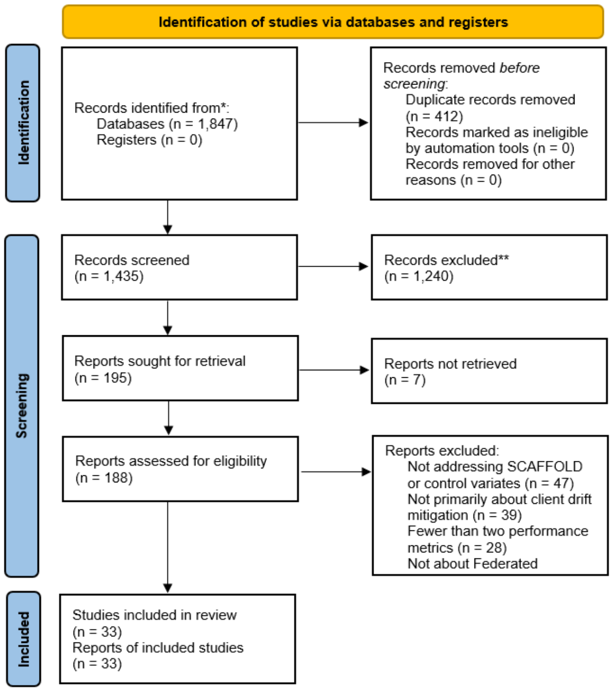
i. To satisfy research question 1 (RQ1), the researcher read the abstract, methodology, results, and discussion sections documenting the approaches used in extending Scaffold and control variates techniques for client drift mitigation in Federated Learning.

ii. To satisfy research question 2 (RQ2), the researcher read the results, discussion, conclusion, and future works sections documenting the limitations of the approaches.

iii. To satisfy research question 3 (RQ3), the researcher read the methodology, results, and discussion sections documenting the characteristics of Scaffold and control variates techniques that are extended for client drift mitigation in Federated Learning.

After the quality assessment, thirty-three (33) studies were included in the final review.

Fig. 1 shows the PRISMA 2020 flow diagram for identifying, screening, assessing eligibility, and including the relevant studies in this systematic literature review.



[Fig.1.PRISMA 2020 Flow Diagram for Selection and Evaluation of Studies]

J. Data Extraction and Synthesis

After the study quality assessment, data were extracted from the chosen thirty-three (33) studies using spreadsheets. The following fields were extracted: database, paper code, title, authors, publication year, research methods, results, conclusions, and future work. A thematic analysis approach was employed to identify patterns and themes within the data. This process involved

Retrieval Number:100.1/ijese.D264114040426
DOI:10.35940/ijese.D2641.14040326
Journal Website: www.ijese.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

classifying data based on the approaches employed to extend Scaffold and control variates techniques for client-drift mitigation in Federated Learning, the limitations of these approaches, the performance impact, and the characteristics of the extended techniques. The extracted data were finally synthesised to answer the research questions and achieve the research objectives.

K. Reporting and Review

This section presents the findings from the studies considered in the research in line with the research questions.

L. RQ1: What are the Approaches in Extending SCAFFOLD for Client Drift Mitigation in Federated Learning?

The first research question sought to uncover the approaches used to extend Scaffold for client-drift mitigation in Federated Learning. After analysis of the thirty-two (32) studies included in the review, nine (9) distinct approaches were identified:

The first approach was variance reduction via gradient estimation techniques, the most prevalent, and was addressed by eleven (11) of the thirty-two (32) studies reviewed. This approach extends Scaffold's control variates by incorporating advanced stochastic gradient estimation methods, primarily from the SVRG, SARAH, and SAGA families of variance-reduction methods. The core idea is to reduce the variance introduced by stochastic gradient sampling and client heterogeneity by maintaining and recursively updating gradient estimates. For instance, [19] computes a global optimizer state that remains frozen during all local client updates and then applies an SVRG-style correction locally based on this global state [30]. extends control variates by maintaining an aggregated control variate that incorporates both fresh and stale local model updates from all clients, including inactive ones, thereby leveraging stale client updates without incurring extra communication cost [31]. combines historical gradient mappings from proximal operators with inertial Douglas-Rachford splitting to reduce variance from both stochastic sampling and client heterogeneity [32]. adds SARAH-style recursive variance reduction and random model selection to the control variate approach, enabling it to break the communication barrier that standard Scaffold cannot overcome [33]. eliminates partial participation variance by storing and reusing each client's most recent update at the server using SAGA-style variance reduction, avoiding the extra client computation and communication overhead required by standard control variate methods. Similarly, [34] combines proximal regularization with SVRG/SARAH gradient estimators in a nested loop structure to enable faster convergence for non-convex objectives [35]. maintains persistent gradient statistics for all clients and uses the global aggregate to make episodic clipping decisions, enabling convergence under client subsampling, data heterogeneity, and unbounded smoothness simultaneously [36]. hybridizes SAGA's gradient storage with SARAH's recursive updates to achieve near-optimal complexity without periodic full gradients [37]. introduces a novel analysis technique that tracks variances and covariances between parameters and control variates to prove that control variate methods can simultaneously achieve logarithmic communication complexity and linear speed-up [38]. combines SAGA's single-loop structure with SARAH's accurate recursive estimation to build a variance reduction method that avoids periodic full gradient computation. Finally, [39] applies the REINFORCE Leave-One-Out (RLOO) estimator on top of basic control variates, dynamically constructing control variates by excluding each sample in turn and using the average of remaining gradients as the baseline.

The second approach was integration with advanced optimization algorithms, addressed by eight (8) out of thirty-two (32) studies. This approach enhances Scaffold by combining its control variates with other established optimization techniques to achieve improved convergence, generalization, or robustness [40]. replaces Scaffold's correction term with a PID (Proportional-Integral-Derivative) controller that includes proportional, integral, and derivative components, better to regulate the client's drift from the global model [41]. performs drift correction after a Sharpness-Aware Minimization (SAM) gradient ascent step, combining control variates with global flatness seeking to improve model generalization beyond what standard control variates achieve alone [42]. enhances Scaffold by adding one Anderson Acceleration step after local gradient descent iterations, which mixes history points to approximate Newton-GMRES directions and achieve second-order-like convergence without accessing the Hessian [43]. combines control variates with Gradient Norm-Aware Minimization (GAM), guiding each client to train models in a globally flat direction, directly smoothing the global model's loss landscape [44]. adds momentum to Scaffold's local update steps, providing anchoring that reduces client drift by using lower-variance gradient estimates [45]. reveals that federated ADMM is inherently a client-variance-reduction scheme and proposes an extension that introduces adaptive step sizes to control the level of variance reduction and global model bias, enabling better adaptation to heterogeneous clients [46]. combines control variates with client-side momentum and adaptive gradient normalization, creating a parameter-free federated learning algorithm that handles arbitrary data heterogeneity without requiring problem-specific hyperparameters [47]. combines Scaffold with an accelerated primal-dual (APD) framework to solve distributionally robust optimization (DRO) problems in federated learning, rather than just minimizing average client loss.

The third approach was partial client participation handling, addressed by three (3) out of thirty-two (32) studies. Scaffold's original design assumes full client participation, which is unrealistic in practical federated learning deployments. This approach specifically addresses the challenges of stale control variates, cyclic availability, and selective participation [48]. scales both client gradients and client control variates by a factor matched to the client's participation rate to reduce the impact of stale control variates and balance global correction in settings with partial participation [49]. extends Scaffold to handle periodic or cyclic client participation by using amplified updates and long-range control variates across participation windows,

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

achieving significantly improved communication complexity with resilience to data heterogeneity [50]. applies control variates only to the final classification layers, not to the entire model, achieving better communication efficiency and faster convergence while maintaining diverse feature representations.

The fourth approach was hybrid framework integration and domain adaptation, also addressed by three (3) out of thirty-two (32) studies. This approach integrates Scaffold's control variates into broader unified frameworks or adapts them to specialized problem domains beyond standard supervised federated learning [51]. uses model contrastive learning with adaptive control variates to handle skewness among Human Activity Recognition (HAR) clients, combining representation-level alignment with drift correction [52]. introduces a unified aggregation technique that integrates Scaffold's control variates with model averaging from FedAvg and a proximal term from FedProx, utilizing accuracy-weighted aggregation to improve stability and performance on non-IID datasets [53]. Adapts Scaffold's control variates technique from supervised federated learning to federated fuzzy C-means clustering, using gradient-based client optimisation and a co-initialisation module to manage both statistical heterogeneity and system heterogeneity.

The fifth approach was privacy and security enhancements, addressed by two (2) out of thirty-two (32) studies. Since Scaffold transmits both model updates and control variates, it introduces additional privacy risks. This approach extends Scaffold with formal privacy or security guarantees [54]. adds two time-varying noise allocation schemes into the Scaffold technique to improve convergence while satisfying total Differential Privacy requirements, balancing the trade-off between privacy protection and drift control [55]. applies multi-key homomorphic encryption (EMK-BFV) to protect both model parameters and control variates during transmission, while adding an anti-offline mechanism that generates supplementary ciphertexts and reweights aggregation to preserve accuracy under non-IID data when users disconnect.

The sixth approach was communication efficiency and channel-aware methods, also addressed by two (2) out of thirty-two (32) studies. Scaffold's requirement to transmit both model updates and control variates doubles the communication overhead compared to FedAvg. This approach seeks to reduce the communication burden or account for transmission-channel effects [56]. adds Minimum Mean Square Error (MMSE) estimation, which corrects for client drift caused by heterogeneous data while simultaneously mitigating channel noise effects in wireless federated learning [57]. proposes compressed federated learning algorithms built on a reformulated Scaffold that halves uplink communication by transmitting a single increment variable, which is then compressed using unbiased or biased compressors with local momentum, achieving state-of-the-art convergence under arbitrary data heterogeneity and partial participation.

The seventh approach was client similarity-based grouping, addressed by one (1) out of thirty-two (32) studies [58]. uses the Pearson correlation coefficient as a similarity measure to group nodes with a high similarity, modifying the variance control mechanism. By clustering clients with correlated model updates, the approach aims to reduce the divergence between local and global models before applying drift correction.

The eighth approach was enhanced bias and drift correction mechanisms, addressed by one (1) out of thirty-two (32) studies [59]. uses dual variables as control variates, accumulating historical gradient information to correct for systematic bias in heterogeneous federated learning. Unlike standard control variates that track gradient differences, the dual-variable formulation provides an adaptive bias-correction mechanism that adjusts based on the accumulated history of gradient updates.

The ninth category was theoretical analysis of existing control variates, addressed by one (1) out of thirty-two (32) studies. Unlike the other approaches that propose algorithmic extensions, [60] provides a novel theoretical analysis of the existing Scaffold algorithm. The study shows that Scaffold's global parameters and control variates jointly form a Markov chain that converges geometrically to a unique stationary distribution in Wasserstein distance. While not proposing a new enhancement, this theoretical contribution establishes previously unknown results about Scaffold's convergence behaviour, including that Scaffold still suffers from a higher-order bias due to stochastic gradient noise.

M. RQ2: What are the Limitations of the Approaches Employed in Extending SCAFFOLD for Client Drift Mitigation in Federated Learning?

The second research question sought to uncover the limitations of the approaches used to extend Scaffold for client-drift mitigation in Federated Learning. The discussion of these limitations, organized by approach, is presented below.

Variance reduction via gradient estimation techniques**.** The most prevalent approach in the reviewed literature exhibits several recurring limitations. A fundamental challenge is that many of these methods impose significant computational overhead. For instance, [34] requires full batch gradient evaluations at each global iteration, which becomes prohibitively expensive with large local datasets and undermines the communication-computation trade-off that federated learning aims to optimize. Similarly, [36] demands large local minibatch sizes, which may be infeasible when clients have limited local data, whereas [38] requires each client to store auxiliary variables, trading increased per-client storage for communication efficiency. Communication overhead is another recurrent concern; [19] reports that the approach incurs communication overhead and computational requirements, and that it performs poorly on sparse optimisation problems, limiting its applicability to dense, smooth settings with moderate heterogeneity. A critical limitation observed across several studies is performance degradation under high data heterogeneity [32]. notes that the technique only works well when client data heterogeneity is small, with performance degrading to match non-local methods as heterogeneity increases. Likewise, [38] reports that all communication benefits

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

degrade to the centralised rate as Hessian heterogeneity approaches the smoothness constant. The issue of stale updates is highlighted by [30], where the technique fails when client participation becomes extremely sparse because stale updates lose correlation with current gradients. Furthermore, [33] deliberately does not address client drift error, which persists at O(1/T) in their bounds, because it only focuses on eliminating the partial participation variance term. Scalability concerns are also evident: [35] requires increasingly conservative learning rates as the number of clients or rounds grows, making it impractical for large-scale federated learning. At the same time, [37] is restricted to linear stochastic approximation problems rather than to general non-convex objectives, such as neural network training [31], which lacks explicit regularization to pull diverging client models back toward the global optimum. Finally, [39] acknowledges that its variance reduction estimator is anchored on a single technique (RLOO) without exploring integration with diverse control variate methods, and its scalability evaluation is restricted to a single dataset and model architecture.

Integration with advanced optimization algorithms**.** The most prominent limitation of this approach is its hyperparameter sensitivity and the complexity of tuning [40] demonstrates that the PID controller's performance depends heavily on precise tuning of gain parameters and can lead to performance degradation rather than improvement, without providing clear guidelines for optimal parameter selection. Both [44] and [45] report that achieving theoretical convergence guarantees requires carefully tuning multiple interdependent hyperparameters with complex scaling relationships that depend on problem-specific constants, with [45] providing only heuristic guidance rather than principled automatic selection methods. Convergence limitations under non-convexity represent another critical concern [42]. reports that the Anderson Acceleration approach fails in non-convex settings, where it gets trapped at critical points due to rapidly decreasing gradient norms, as demonstrated in MNIST experiments [47] similarly notes that convergence guarantees rely on local objective functions being strongly convex, limiting direct applicability to deep neural networks. Communication and computational overhead are recurring issues as well [41] reports doubled communication cost from transmitting both model parameters and control variates each round, while [43] suffers from a doubled computational cost due to the requirement of two gradient computations per step, combined with a lack of theoretical convergence guarantees [46] highlights that gradient normalization, while enabling parameter-free operation, discards magnitude information, making the approach critically dependent on momentum to maintain effective descent directions.

Partial client participation handling. The limitations of this approach reflect the inherent tension between addressing partial participation and maintaining practical efficiency [48] reports that the additional communication overhead from transmitting both scaled and unscaled gradients directly contradicts federated learning's fundamental goal of communication efficiency [49]. requires well-defined, known participation windows with equal expected client participation and consequently cannot manage dynamic client populations, unknown participation patterns, or clients who freely join and leave during training. This represents a significant limitation for real-world deployments where client availability is inherently unpredictable [50]. requires manual, architecture-specific selection of which layers to apply variance reduction to, with no principled or adaptive method provided for determining the optimal split point for arbitrary neural network architectures.

Hybrid framework integration and domain adaptation**.** The limitations of this approach centre on performance under distribution skew, scalability, and complexity [51]. reports severe performance degradation under label distribution skew (up to a 10.75% drop in F1 score), because both its contrastive learning and control variates depend on a global model that becomes heavily biased when clients have different label sets. [52] is constrained by evaluation on simple benchmarks (MNIST/CIFAR-10) with only 100 clients, without validating whether the approach scales to real-world scenarios with thousands of heterogeneous clients and varying computational and network conditions [53]. reports increased algorithmic complexity and hyperparameter sensitivity, as the gradient-based optimization with control variates requires careful tuning of both local and global step sizes compared to simpler federated methods.

Privacy and security enhancements**.** The limitations of this approach highlight the inherent tension between privacy guarantees and model performance [54]. acknowledges that the technique balances privacy and drift control but does not resolve the inherent conflict between them, constituting a compromise that requires stopping training early and accepting lower accuracy for privacy guarantees [55]. relies on Scaffold's existing control variate mechanism without improving it; the approach only encrypts and transmits control variates securely, but does not address Scaffold's fundamental limitation of requiring full participation or the large communication overhead from transmitting both model updates and control variates each round. This indicates that while security can be layered atop Scaffold, the underlying algorithmic limitations persist.

Communication efficiency and channel-aware methods**.** The limitations of this approach relate to restrictive assumptions and narrow experimental validation [56]. relies on a Gaussian modelling assumption for local weights and gradients, which becomes suboptimal for real heterogeneous data with non-Gaussian distributions, potentially degrading the drift mitigation performance that Scaffold's control variates are designed to provide [57] has experiments limited to small-scale datasets (MNIST, Fashion-MNIST) with a simple two-hidden-layer neural network, leaving unclear whether the advantages over baselines hold for large-scale, real-world federated learning settings with modern deep architectures and more complex data heterogeneity patterns.

Client similarity-based grouping [58] reports that the Pearson correlation coefficient used as the similarity measure cannot distinguish between healthy similarity arising from models trained on related data and malicious similarity arising from models subjected to the same attack. This creates a security vulnerability that allows adversarial clients to exploit the grouping mechanism, potentially undermining the

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

integrity of the drift mitigation process.

Enhanced bias and drift correction mechanisms [59]. reports that the adaptive bias correction mechanism, while effective in moderate heterogeneity scenarios, requires complex hyperparameter tuning and imposes computational overhead on clients. Furthermore, the approach has fundamental limits when facing extreme non-IID data distributions, concept drift, or adversarial conditions, where historical gradient accumulation becomes unreliable or insufficient for effective drift correction.

Theoretical analysis of existing control variates [60]. reveals that while Scaffold successfully eliminates heterogeneity bias (client drift), it still suffers from a higher-order bias due to stochastic gradient noise that scales with the step size divided by the number of clients. This bias does not decrease beyond a certain threshold, leaving the design of truly unbiased stochastic federated algorithms as an open problem. This finding has significant implications for all Scaffold extensions, as the fundamental stochastic-noise bias identified applies to the base mechanism on which all other approaches build.

N. RQ3: What Characteristics of SCAFFOLD are Extended for Client Drift Mitigation in Federated Learning?

The third research question sought to identify the characteristics of the Scaffold protocol that are extended or modified to mitigate client drift in Federated Learning. The study findings from the thirty (30) studies that explicitly extended a Scaffold characteristic are presented below. Twelve (12) distinct characteristics, including two combined characteristic categories, were identified:

i. ***Variance Reduction:*** Scaffold's core mechanism uses control variates to reduce the variance introduced by heterogeneous client gradients. This was the most commonly extended characteristic, addressed by eleven (11) out of thirty (30) studies. The extensions modify how control variates are estimated and correct gradient variance by incorporating advanced stochastic variance-reduction techniques such as SAGA, SARAH, and SVRG, applying similarity-based client grouping, amplifying updates for cyclic participation, and dynamically constructing control variates. Studies extending this characteristic include [58], [48], [40], [51], [31], [35], [49], [36], [37], [38], [39].

ii. ***Drift Correction:*** Scaffold corrects client drift by maintaining control variates that track the difference between client and server gradient directions, pulling local updates toward the global optimum. Three (3) out of thirty (30) studies extended this characteristic. These extensions enhance how the drift correction operates by incorporating channel-aware estimation [56], combining drift correction with global flatness seeking via Sharpness-Aware Minimization [41], and guiding clients to train in globally flat directions through Gradient Norm-Aware Minimization [43].

iii. ***Gradient Correction:*** The gradient correction characteristic relates to how Scaffold adjusts client gradient updates to account for the discrepancy between local and global optimization objectives. Three (3) out of thirty (30) studies extended this characteristic. The extensions introduce momentum-based anchoring to the local update steps [44], adaptive gradient normalization combined with client-side momentum for parameter-free operation [46], and adaptation of gradient-based correction to federated fuzzy C-means clustering [53].

iv. ***Bias Correction:*** The bias correction characteristic addresses the systematic bias that arises in heterogeneous federated learning due to non-IID data distributions. Two (2) out of thirty (30) studies extended this characteristic [59] uses dual variables as control variates that accumulate historical gradient information to correct for systematic bias, while [47] combines Scaffold with an accelerated primal-dual framework to address bias in the context of distributionally robust optimization problems.

v. ***Server Aggregation Strategy:*** Scaffold's server aggregation determines how client updates and control variates are combined at the server to form the global model. Two (2) out of thirty (30) studies extended this characteristic [52] integrates Scaffold's control variates into a unified framework with model averaging and proximal regularization, using accuracy-weighted aggregation [33] stores and reuses each client's most recent update on the server using SAGA-style variance reduction, modifying the server-side aggregation to eliminate partial-participation variance without requiring additional client-side computation.

vi. *Variance Reduction and Drift Correction:* Two (2) out of thirty (30) studies extended variance reduction and drift correction as a combined unit [19]. computes a global optimizer state that remains frozen during all local updates and applies SVRG-style correction locally, simultaneously reducing gradient variance and correcting drift [50]. applies control variates selectively to the final classification layers rather than the entire model, achieving both variance reduction and drift correction with improved communication efficiency.

vii. ***Variance Reduction and Bias Correction:*** Two (2) out of thirty (30) studies extended variance reduction and bias correction as a combined unit [45] introduces adaptive step sizes that control both the variance-reduction level and the global model bias, revealing that federated ADMM inherently serves as a variance-reduction scheme [32] adds SARAH-style recursive variance reduction and random model selection to the control variates approach, addressing both gradient variance and systematic bias, thereby breaking the communication barrier that standard Scaffold cannot overcome.

viii. ***Gradient Estimation:*** The gradient estimation characteristic relates to how Scaffold estimates gradients during local updates. One (1) out of thirty (30) studies extended this characteristic [34]. combines proximal regularization with SVRG/SARAH gradient estimators in a nested loop structure, modifying the gradient estimation process to enable faster convergence for non-convex objectives under heterogeneous data.

ix. *Local Update Mechanism:* The local update mechanism governs how clients perform their local optimization steps

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

International Journal of Emerging Science and Engineering (IJESE)
ISSN: 2319–6378 (Online), Volume-14 Issue-4, March 2026

between communication rounds. One (1) out of thirty (30) studies extended this characteristic [42], which enhances the Scaffold by adding one Anderson Acceleration step after the local gradient descent iterations, mixing history points to approximate second-order convergence directions without computing the Hessian.

x. *Communication Mechanism:* Scaffold's communication mechanism defines what information is transmitted between clients and the server, including both model updates and control variates. One (1) out of thirty (30) studies extended this characteristic [57] reformulates Scaffold to halve uplink communication by transmitting a single increment variable instead of separate model updates and control variates, applying unbiased or biased compression with local momentum to achieve state-of-the-art convergence under arbitrary data heterogeneity.

xi. *Privacy Budget Allocation:* One (1) out of thirty (30) studies extended Scaffold's control variates with a privacy budget allocation mechanism [54] adds two time-varying noise allocation schemes into the Scaffold technique, modifying how differential privacy noise is distributed over training rounds to improve convergence while satisfying total differential privacy constraints.

Correlation Term: The correlation term in Scaffold captures the relationship between local and global gradient directions through control variates. One (1) out of thirty (30) studies extended this characteristic [30]. maintains a control variate that aggregates both fresh and stale local model updates from all clients, including inactive ones, modifying the correlation term to leverage historical update information without incurring extra communication cost.

IV. DISCUSSION

The study aimed to review the available scholarly work on approaches to extending Scaffold for client-drift mitigation in Federated Learning, their limitations, and the characteristics of Scaffold commonly extended. Data were gathered from published studies, and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were adopted to guide the research. After consideration of the inclusion and exclusion criteria and study quality assessment, thirty-two (32) papers were chosen for final review. The study discussions are summarized as follows:

RQ1. The first research question was to review approaches employed to extend Scaffold for client-drift mitigation in Federated Learning. The study results revealed nine (9) distinct approaches. Variance reduction via gradient estimation techniques was the most prevalent approach, addressed by 34% of studies, followed by integration with advanced optimization algorithms at 25%. Together, these two dominant approaches account for 59% of all studies, indicating that the research community has predominantly focused on enhancing the mathematical and algorithmic foundations of Scaffold's control variate mechanism. Partial client participation handling and hybrid framework integration each accounted for 9%, while privacy and security enhancements and communication efficiency methods each accounted for 6%. The remaining approaches, client similarity-based grouping, enhanced bias and drift correction, and theoretical analysis, were each addressed by a single study (3%). Approaches such as privacy and security enhancements, communication efficiency, and client similarity-based grouping have been least studied yet possess the potential to improve drift mitigation and could therefore be investigated in future research.

RQ2. The second research question was to review the limitations of the approaches employed. Several recurring themes emerged. Communication and computational overhead were the most prominent limitation, reported by [19], [48], [41], [43], [34], [55], which is particularly concerning given that Scaffold already doubles communication costs relative to FedAvg. Hyperparameter sensitivity and tuning complexity were reported by [40], [44], [59], [45], [53], limiting practical applicability. Restrictive assumptions such as strong convexity, Gaussian distributions, or linear problem formulations were imposed by [56], [42], [37], [47], limiting generalizability to deep neural network training. Performance degradation under high data heterogeneity [51], [32], [38] and limited experimental validation at scale [52], [57], [39] further constrain the practical adoption of these extensions. Collectively, these limitations suggest that future approaches should prioritise communication-efficient designs, adaptive hyperparameter mechanisms, relaxed theoretical assumptions, and rigorous large-scale empirical validation.

RQ3. The third research question was reviewing the Scaffold characteristics extended for client drift mitigation. The study revealed twelve (12) distinct characteristics. Variance reduction was addressed in 37% of studies. When including combined categories (variance reduction and drift correction at 7%, and variance reduction and bias correction at 7%), the proportion rises to 50%, underscoring the centrality of this core mechanism. Drift correction and gradient correction were each addressed by 10%, while bias correction, server aggregation strategy, and the two combined categories accounted for 7%. The remaining characteristics, gradient estimation, local update mechanism, communication mechanism, privacy budget allocation, and correlation term, were each extended by only one study (3%), indicating that these aspects of Scaffold remain underexplored. In particular, the communication mechanism and privacy budget allocation represent characteristics where further research could yield significant practical benefits, given that communication efficiency and privacy are among the most pressing challenges in real-world federated learning deployments.

V. CONCLUSION AND FUTURE WORK

This paper has uncovered approaches from the literature for extending or modifying Scaffold to mitigate client drift in Federated Learning, and their limitations. Further, the study revealed the characteristics of the Scaffold protocol commonly used for client drift mitigation. Nine (9) approaches were identified from the thirty-two (32) studies reviewed. The two most frequently employed approaches were variance reduction via gradient estimation techniques (11 studies) and integration with advanced optimisation algorithms (8 studies),

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

together accounting for 59% of all reviewed work. Despite these approaches extending Scaffold to improve drift mitigation, recurring limitations persist, notably communication and computational overhead, hyperparameter sensitivity, restrictive theoretical assumptions, and performance degradation under extreme data heterogeneity. The variance reduction characteristic was the most addressed, appearing in 37% of studies, rising to 50% when combined categories are included.

A notable finding of this review is that similarity-based approaches for client drift mitigation remain significantly underexplored. Only one (1) out of thirty-two (32) studies [58] employed a similarity measure, specifically the Pearson correlation coefficient, to group clients for variance control. However, that study's own limitation reveals that Pearson correlation cannot distinguish between healthy similarity from well-trained models and malicious similarity from compromised ones. This narrow adoption is surprising given that similarity measures offer a fundamentally different and potentially more powerful mechanism for drift mitigation: rather than correcting drift after it occurs through variance reduction or bias correction, similarity-based approaches can proactively detect and quantify directional divergence between client and global gradients before corrections are applied. A similarity measure incorporated as a third Scaffold control variate, alongside the existing server and client control variates, could provide a real-time, per-round signal of client-to-global gradient alignment that dynamically modulates the intensity of drift correction. Clients whose gradients are closely aligned with the global direction would receive minimal correction, preserving their useful local information, while clients exhibiting significant directional divergence would receive stronger correction. Such an adaptive mechanism has the potential to improve upon Scaffold's uniform correction strategy, reduce unnecessary overhead, and enhance robustness to heterogeneous data distributions, thereby addressing several key limitations identified in this review.

Therefore, the study recommends that future work investigate the use of similarity measures as a third control variate in the Scaffold protocol to mitigate client drift. Several candidate similarity measures warrant consideration. Cosine similarity, which measures directional alignment between vectors irrespective of magnitude, aligns naturally with the directional nature of gradient drift. Euclidean distance could quantify the absolute magnitude of gradient divergence between clients and the global model. Centred Kernel Alignment (CKA) offers a representation-level similarity measure that could capture structural differences in learned features across clients. Jaccard similarity could be adapted to compare the sparsity patterns of gradient updates in large-scale models. Each of these measures presents distinct trade-offs in terms of computational cost, sensitivity to gradient scale, and suitability for high-dimensional parameter spaces; therefore, comparative studies evaluating their effectiveness as control variates would be a valuable contribution to the field.

DECLARATION STATEMENT

As the article's author, I must verify the accuracy of the following information after aggregating input from all authors.

- **Conflicts of Interest/ Competing Interests:** Based on my understanding, this article has no conflicts of interest.
- **Funding Support:** This article has not been funded by any organizations or agencies. This independence ensures that the research is conducted objectively and without external influence.
- **Ethical Approval and Consent to Participate:** The content of this article does not necessitate ethical approval or consent to participate with supporting documentation.
- **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
- **Author's Contributions:** The authorship of this article is contributed equally to all participating individuals.

REFERENCES

1. "Advances and Open Problems in Federated Learning | now eBooks | IEEE Xplore." Accessed: Feb. 13, 2026. [Online]. Available: https://ieeexplore.ieee.org/book/9464278
2. "Federated Machine Learning: Concept and Applications: ACM Transactions on Intelligent Systems and Technology: Vol 10, No 2," *ACM Trans. Intell. Syst. Technol. TIST*, Accessed: Feb. 13, 2026. [Online]. Available: https://dl.acm.org/doi/10.1145/3298981
3. T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated Learning: Challenges, Methods, and Future Directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, May 2020, DOI: https://doi.org/10.1109/MSP.2020.2975749
4. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, PMLR, Apr. 2017, pp. 1273–1282. Accessed: Feb. 13, 2026. [Online]. Available: https://proceedings.mlr.press/v54/mcmahan17a.html
5. N. Rieke *et al.*, "The future of digital health with federated learning," *NPJ Digit. Med.*, vol. 3, p. 119, 2020, DOI: https://doi.org/10.1038/s41746-020-00323-1.
6. T. Yang *et al.*, "Applied Federated Learning: Improving Google Keyboard Query Suggestions," Dec. 07, 2018, *arXiv*: arXiv:1812.02903. DOI: https://doi.org/10.48550/arXiv.1812.02903.
7. S. R. Pokhrel and J. Choi, "Federated Learning With Blockchain for Autonomous Vehicles: Analysis and Design Challenges," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4734–4746, Aug. 2020, DOI: https://doi.org/10.1109/TCOMM.2020.2990686.
8. W. Y. B. Lim *et al.*, "Federated Learning in Mobile Edge Networks: A Comprehensive Survey," *IEEE Commun. Surv. Tutor*, vol. 22, no. 3, pp. 2031–2063, 2020, DOI: https://doi.org/10.1109/COMST.2020.2986024.
9. L. Zhu, Z. Liu, and S. Han, "Deep Leakage from Gradients," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2019. Accessed: Feb. 13, 2026. [Online]. Available: https://proceedings.neurips.cc/paper/2019/hash/60a6c4002cc7b29142def8871531281a-Abstract.html
10. V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, "A survey on security and privacy of federated learning," *Future Gener. Comput. Syst.*, vol. 115, pp. 619–640, Feb. 2021, DOI: https://doi.org/10.1016/j.future.2020.10.007.
11. T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated Optimization in Heterogeneous Networks," *Proc. Mach. Learn. Syst.*, vol. 2, pp. 429–450, Mar. 2020. https://proceedings.mlsys.org/paper_files/paper/2020/hash/1f5fe83998a09396ebe6477d9475ba0c-Abstract.html
12. S. P. Karimireddy, S. Kale, M. Mohri, S. J. Reddi, S. U. Stich, and A. T. Suresh, "SCAFFOLD: stochastic controlled averaging for federated learning," in *Proceedings of the 37th International Conference on Machine Learning*, in ICML'20, vol. 119. JMLR.org, Jul. 2020, pp. 5132–5143. Accessed: Feb. 13, 2026.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

[Online]. Available: https://dl.acm.org/doi/10.5555/3524938.3525414
13. J. Wang, Q. Liu, H. Liang, G. Joshi, and H. V. Poor, "Tackling the Objective Inconsistency Problem in Heterogeneous Federated Optimization," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2020, pp. 7611–7623. Accessed: Feb. 13, 2026. [Online]. Available: https://proceedings.neurips.cc/paper/2020/hash/564127c03caab942e503ee6f810f54fd-Abstract.html
14. X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, "On the Convergence of FedAvg on Non-IID Data," presented at the Eighth International Conference on Learning Representations, Apr. 2020. Accessed: Feb. 13, 2026. [Online]. Available: https://iclr.cc/virtual_2020/poster_HJxNAnVtDS.html
15. S. J. Reddi et al., "ADAPTIVE FEDERATED OPTIMIZATION," ICLR, 2021, [Online]. Available: https://iclr.cc/virtual/2021/poster/2691
16. K. Mishchenko, G. Malinovsky, S. Stich, and P. Richtarik, "ProxSkip: Yes! Local Gradient Steps Provably Lead to Communication Acceleration! Finally!," in *Proceedings of the 39th International Conference on Machine Learning*, PMLR, Jun. 2022, pp. 15750–15769. Accessed: Feb. 13, 2026. [Online]. Available: https://proceedings.mlr.press/v162/mishchenko22b.html
17. S. U. Stich, "Local SGD Converges Fast and Communicates Little," presented at the International Conference on Learning Representations, Sep. 2018. Accessed: Feb. 13, 2026. [Online]. Available: https://openreview.net/forum?id=S1g2JnRcFX
18. G. Malinovskiy, D. Kovalev, E. Gasanov, L. Condat, and P. Richtarik, "From Local SGD to Local Fixed-Point Methods for Federated Learning," in *Proceedings of the 37th International Conference on Machine Learning*, PMLR, Nov. 2020, pp. 6692–6701. Accessed: Feb. 13, 2026. [Online]. Available: https://proceedings.mlr.press/v119/malinovskiy20a.html
19. S. P. Karimireddy *et al.*, "Breaking the centralized barrier for cross-device federated learning," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2021, pp. 28663–28676. Accessed: Feb. 03, 2026. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2021/hash/f0e6be4ce76ccfa73c5a540d992d0756-Abstract.html
20. Y. J. Cho, J. Wang, and G. Joshi, "Towards Understanding Biased Client Selection in Federated Learning," in *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, PMLR, May 2022, pp. 10351–10375. Accessed: Feb. 13, 2026. [Online]. Available: https://proceedings.mlr.press/v151/jee-cho22a.html
21. D. A. E. Acar, Y. Zhao, R. Matas, M. Mattina, P. Whatmough, and V. Saligrama, "Federated Learning Based on Dynamic Regularization," presented at the International Conference on Learning Representations, Oct. 2020. Accessed: Feb. 13, 2026. [Online]. Available: https://openreview.net/forum?id=B7v4QMR6Z9w
22. "[2107.06917] A Field Guide to Federated Optimization." Accessed: Feb. 13, 2026. [Online]. Available: https://arxiv.org/abs/2107.06917
23. Q. Li *et al.*, "A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 3347–3366, Apr. 2023, DOI: https://doi.org/10.1109/TKDE.2021.3124599.
24. "Non-IID data in Federated Learning: A Survey with Taxonomy, Metrics..." Accessed: Feb. 14, 2026. [Online]. Available: https://openreview.net/forum?id=7tSnVgHG7F
25. M. Ye, X. Fang, B. Du, P. C. Yuen, and D. Tao, "Heterogeneous Federated Learning: State-of-the-art and Research Challenges," *ACM Comput Surv*, vol. 56, no. 3, p. 79:1-79:44, Oct. 2023, DOI: https://doi.org/10.1145/3625558.
26. H. Zhu, J. Xu, S. Liu, and Y. Jin, "Federated learning on non-IID data: A survey," *Neurocomputing*, vol. 465, pp. 371–390, Nov. 2021, DOI: https://doi.org/10.1016/j.neucom.2021.07.098.
27. C. Chen, T. Liao, X. Deng, Z. Wu, S. Huang, and Z. Zheng, "Advances in Robust Federated Learning: A Survey With Heterogeneity Considerations," *IEEE Trans. Big Data*, vol. 11, no. 3, pp. 1548–1567, Jun. 2025, DOI: https://doi.org/10.1109/TBDATA.2025.3527202.
28. M. J. Page *et al.*, "PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews," *BMJ*, vol. 372, p. n160, Mar. 2021, DOI: https://doi.org/10.1136/bmj.n160.
29. M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, vol. 372, p. n71, Mar. 2021, DOI: https://doi.org/10.1136/bmj.n71.
30. H. Zhang, C. Li, W. Dai, J. Zou, and H. Xiong, "Federated Learning Based on Model Discrepancy and Variance Reduction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 6, pp. 10407–10421, Jun. 2025, DOI: https://doi.org/10.1109/TNNLS.2024.3517658.
31. J. Xue and C. Wang, "FedSAGA: Composite Federated Learning with Inertial Douglas-Rachford Splitting and Variance Reduction Method," in *2025 IEEE 31st International Conference on Parallel and Distributed Systems (ICPADS)*, Dec. 2025, pp. 1–9. DOI: https://doi.org/10.1109/ICPADS67057.2025.11322895.
32. T. Murata and T. Suzuki, "Bias-Variance Reduced Local SGD for Less Heterogeneous Federated Learning," in *Proceedings of the 38th International Conference on Machine Learning*, PMLR, Jul. 2021, pp. 7872–7881. Accessed: Feb. 07, 2026. [Online]. Available: https://proceedings.mlr.press/v139/murata21a.html
33. D. Jhunjhunwala, P. Sharma, A. Nagarkatti, and G. Joshi, "FedVARP: Tackling the Variance Due to Partial Client Participation in Federated Learning," presented at the 38th Conference on Uncertainty in Artificial Intelligence, May 2022. Accessed: Feb. 07, 2026. [Online]. Available: https://openreview.net/forum?id=HlWLLdUocx5
34. C. T. Dinh, N. H. Tran, T. D. Nguyen, W. Bao, A. Y. Zomaya, and B. B. Zhou, "Federated Learning with Proximal Stochastic Variance Reduced Gradient Algorithms," in *Proceedings of the 49th International Conference on Parallel Processing*, in ICPP '20. New York, NY, USA: Association for Computing Machinery, Aug. 2020, pp. 1–11. DOI: https://doi.org/10.1145/3404397.3404457.
35. M. Crawshaw, Y. Bao, and M. Liu, "Federated Learning with Client Subsampling, Data Heterogeneity, and Unbounded Smoothness: A New Algorithm and Lower Bounds," *Adv. Neural Inf. Process. Syst.*, vol. 36, pp. 6467–6508, Dec. 2023. https://proceedings.neurips.cc/paper_files/paper/2023/hash/14ecbfb2216bab76195b60bfac7efb1f-Abstract-Conference.html
36. K. Oko, S. Akiyama, T. Murata, and T. Suzuki, "Reducing Communication in Nonconvex Federated Learning with a Novel Single-Loop Variance Reduction Method," presented at the OPT 2022: Optimization for Machine Learning (NeurIPS 2022 Workshop), Nov. 2022. Accessed: Feb. 12, 2026. [Online]. Available: https://openreview.net/forum?id=pYBZZzbJtE
37. P. Mangold *et al.*, "SCAFFLSA: taming heterogeneity in federated linear stochastic approximation and TD learning," in *Proceedings of the 38th International Conference on Neural Information Processing Systems*, in NIPS '24, vol. 37. Red Hook, NY, USA: Curran Associates Inc., Dec. 2024, pp. 13927–13981. DOI: https://dl.acm.org/doi/10.5555/3737916.3738363
38. K. Oko, S. Akiyama, D. Wu, T. Murata, and T. Suzuki, "SILVER: Single-loop variance reduction and application to federated learning," in *Proceedings of the 41st International Conference on Machine Learning*, PMLR, Jul. 2024, pp. 38683–38739. Accessed: Feb. 12, 2026. [Online]. Available: https://proceedings.mlr.press/v235/oko24a.html
39. Y. Liu, X. Yang, X. Chen, H. Du, and L. Xu, "FedNCV: Optimizing Federated Learning With Networked Control Variates," *Trans. Emerg. Telecommun. Technol.*, vol. 36, no. 11, p. e70287, 2025, DOI: https://doi.org/10.1002/ett.70287.
40. O. Mashaal, S. Baadel, and H. AbouZied, "Extending Control Theory into Federated Learning Data Heterogeneity Problem," in *2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, Feb. 2024, pp. 1–4. DOI: https://doi.org/10.1109/AIMS61812.2024.10512401.
41. R. Dai *et al.*, "FedGAMMA: Federated Learning With Global Sharpness-Aware Minimization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 12, pp. 17479–17492, Dec. 2024, DOI: https://doi.org/10.1109/TNNLS.2023.3304453.
42. X. Feng, M. P. Laiu, and T. Strohmer, "FedOSAA: Improving Federated Learning with One-Step Anderson Acceleration," presented at the Second Conference on Parsimony and Learning (Proceedings Track), Mar. 2025. Accessed: Feb. 03, 2026. [Online]. Available: https://openreview.net/forum?id=OoYcaWhfwB#discussion
43. Y. Xu, W. Ma, C. Dai, Y. Wu, and H. Zhou, "Generalized Federated Learning via Gradient Norm-Aware Minimization and Control Variables," *Mathematics*, vol. 12, no. 17, p. 2644, Jan. 2024, DOI: https://doi.org/10.3390/math12172644.
44. Z. Cheng, X. Huang, P. Wu, and K. Yuan, "Momentum Benefits Non-iid Federated Learning Simply and Provably," *Int. Conf. Learn. Represent.*, vol. 2024, pp. 9815–9848, May 2024. https://proceedings.iclr.cc/paper_files/paper/2024/hash/291d92e9097276da23f9aba4a45a5041-Abstract-Conference.html
45. S. Wang, Y. Xu, Z. Wang, T.-H. Chang, T. Q. S. Quek, and D. Sun, "Beyond ADMM: A Unified Client-Variance-Reduced Adaptive Federated Learning Framework," *Proc. AAAI Conf. Artif. Intell.*, vol. 37, no. 8, pp. 10175–10183, Jun. 2023, DOI: https://doi.org/ 10.1609/aaai.v37i8.26212.
46. W. Yan, K. Zhang, X. Wang, and X. Cao, "Problem-Parameter-Free Federated Learning," presented at the Thirteenth International Conference on Learning

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Representations, Oct. 2024. Accessed: Feb. 07, 2026. [Online]. Available: https://openreview.net/forum?id=ZuazHmXTns
47. Y. Yu, S. P. Karimireddy, Y. Ma, and M. I. Jordan, "Scaff-PD: Communication-Efficient Fair and Robust Federated Learning," Jul. 2023, DOI: https://doi.org/10.48550/arXiv.2307.13381.
48. X. Wang, L. Tian, C. Yang, F. Lin, and M. Li, "Enhancing Federated Learning under Partial Participation via Proportional Variance Reduction," in *2025 IEEE 24th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, Nov. 2025, pp. 344–351. DOI: https://doi.org/10.1109/Trustcom66490.2025.00045.
49. M. Crawshaw and M. Liu, "Federated Learning under Periodic Client Participation and Heterogeneous Data: A New Communication-Efficient Algorithm and Analysis". https://dl.acm.org/doi/10.5555/3737916.3738181
50. B. Li, M. N. Schmidt, T. S. Alstrøm, and S. U. Stich, "On the Effectiveness of Partial Variance Reduction in Federated Learning with Heterogeneous Data," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2023, pp. 3964–3973. DOI: https://doi.org/10.1109/CVPR52729.2023.00386.
51. I. Iwan, B. N. Yahya, and S.-L. Lee, "Federated Model with Contrastive Learning and Adaptive Control Variates for Human Activity Recognition," *Front. Inf. Technol. Electron. Eng.*, vol. 26, no. 6, pp. 896–911, Jun. 2025, DOI: https://doi.org/10.1631/FITEE.2400797.
52. N. Moneesh, M. S, D. G. Nair, and J. J. Nair, "FedHybrid: Unifying Aggregation Strategies to Optimise Federated Learning on Non-IID Datasets," *Procedia Comput. Sci.*, vol. 258, pp. 3126–3134, Jan. 2025, DOI: https://doi.org/10.1016/j.procs.2025.04.570
53. Q. Zhang *et al.*, "Robust Federated Fuzzy C-Means Algorithm in Heterogeneous Scenarios," *IEEE Trans. Fuzzy Syst.*, vol. 33, no. 9, pp. 3168–3181, Sep. 2025, DOI: https://doi.org/10.1109/TFUZZ.2025.3584697.
54. W.-T. Chang, M. Seif, and R. Tandon, "Differentially Private Federated Learning with Drift Control," in *2022 56th Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2022, pp. 240–245. DOI: https://doi.org/10.1109/CISS53076.2022.9751200.
55. W. Ding, H. Guo, Z. Yan, and M. Wang, "Secure Federated Learning Schemes Based on Multi-Key Homomorphic Encryption," in *2024 IEEE 23rd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, Dec. 2024, pp. 551–558. DOI: https://doi.org/10.1109/TrustCom63139.2024.00091.
56. T. Gafni, K. Cohen, and Y. C. Eldar, "Federated Learning from Heterogeneous Data via Controlled Air Aggregation with Bayesian Estimation," *IEEE Trans. Signal Process.*, vol. 72, pp. 1928–1943, 2024, DOI: https://doi.org/10.1109/TSP.2024.3351469.
57. X. Huang, P. Li, and X. Li, "Stochastic Controlled Averaging for Federated Learning with Communication Compression," presented at the Twelfth International Conference on Learning Representations, Oct. 2023. Accessed: Feb. 07, 2026. [Online]. Available: https://openreview.net/forum?id=jj5ZjZsWJe
58. M. G. Rahmat and M. Khalilian, "A Novel Pearson Correlation-Based Merging Algorithm for Robust Distributed Machine Learning with Heterogeneous Data," Jan. 24, 2025, *arXiv*: arXiv:2501.11112. DOI: https://doi.org/10.48550/arXiv.2501.11112.
59. H. Yang, Z. Liu, X. Zhang, and J. Liu, "SAGDA: achieving O(ε-2) communication complexity in federated min-max learning," Adv. Neural Inf. Process. Syst., vol. 35, pp. 7142–7154, Dec. 2022, Accessed: Feb. 03, 2026. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2022/hash/2f13806d6580db60d9d7d6f89ba529ca-Abstract-Conference.html
60. P. Mangold, A. O. Durmus, A. Dieuleveut, and E. Moulines, "Scaffold with Stochastic Gradients: New Analysis with Linear Speed-Up," presented at the Forty-second International Conference on Machine Learning, Jun. 2025. Accessed: Feb. 03, 2026. [Online]. Available: https://openreview.net/forum?id=2XvOJvUlKc

AUTHOR'S PROFILE

James Mburu Muthii is a student in the Computer Science Department of Murang'a University of Technology, School of Computing and Information Technology. He is currently pursuing his post-graduate degree. He holds a Bachelor of Science degree in Software Engineering. His interests include Data Science (AI), Computer Networking, Computer Security, Software Engineering, and Distributed Systems.

Dr. Stephen Kahara Wanjau holds a PhD in Computer Science from Murang'a University of Technology (MUT), an M.Sc. in Computer Systems (JKUAT), an M.A. in Organisational Development (USIU-A), and a B.Sc. in Information Sciences (Moi). Currently serving as the Director of Performance Management and Standards at MUT, he is a distinguished ICT strategist and certified QMS/ISMS Lead Auditor with over fifteen years of experience in operational excellence and information security. His research and professional practice focus on leveraging Artificial Intelligence, Machine Learning, Network Security, and Distributed Systems. A DAAD-UNILEAD alumnus and member of the Association of Computing Practitioners-Kenya (ACPK), Dr Kahara bridges the gap between rigorous international standards and impactful technological innovation to drive institutional development.

Dr. Stephen Njenga is a Lecturer in Computer Science at Murang'a University of Technology. He holds a PhD in Information Systems and an MSc in Computer Science from the University of Nairobi, Kenya. He graduated from Egerton University (Kenya) with a BSc in Computer Science. Dr Njenga has published extensively in the fields of Machine Learning, Intelligent Agents, and Distributed Ledger Technology.