iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.82478

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

A Survey on Privacy-Preserving Learning using Vertical Federated Learning with Representation Synthesis

Asst. Prof. Vyshnavi M, Dhanusha Naik, Bhoomika H, Harshitha B, Jahanavi P Gowda
Dept. of CSE Sapthagiri College of Engineering

***Abstract*—In an era where data is one of the most valuable resourceswehave,protectingithasbecomejustasimportant as using it. Traditional machine learning has long relied on centralizingdatabypullingeverythingintooneplacesoa model can learn from it. However, as datasets become larger and more sensitive, this approach introduces serious risks including privacyviolations,databreaches,andreducedwillingnessamong organizations to share confidential information.**
ThispaperpresentsaVerticalFederatedLearning(VFL)frame- work designed to allow multiple organizations to collaboratively trainamachinelearningmodelwithoutexchangingrawdata. In this approach, each participating organization holds different featuresofthesameusersandtrainsalocalmodelindependently. Onlyintermediaterepresentationsaresharedwithacentralserver, whichcombinesthemintoaunifiedglobalmodelwhilepreserving privacy.
The proposed framework significantly reduces security vul- nerabilities and enables privacy-preserving collaboration among industries such as healthcare, banking, and government sectors wheredataconfidentialityiscritical.Thestudydemonstratesthat privacyandperformancecancoexisteffectivelythroughcarefully designed distributed learning architectures.*

I. INTRODUCTION

The rapid growth of digital technologies has led to an enormous increase in data generation across industries such as healthcare, finance, and e-commerce. Machine learning systems relyheavilyonsuchdatatogenerateaccuratepredictions and intelligent services. Traditional machine learning methods require centralized data collection, where all information is gathered and processed in a single location. While effective, this approach raises significant concerns regarding privacy, security, and regulatory compliance.

Organizations handling sensitive information are often un- willing or legally restricted from sharing raw datasets with externalparties.RegulationssuchasGDPRandHIPAAimpose strict rules on how personal information should be stored and processed. As a result, there is a growing demand for machine learning frameworks that can support collaborative learning without compromising data confidentiality.

Federated Learning addresses this challenge by allowing machine learning models to be trained directly where the data resides. Instead of transferring raw data, only model updatesor intermediate outputs are shared during training. Vertical FederatedLearning(VFL)extendsthisconceptfurtherby enabling organizations with different feature sets of the same users to collaboratively train models securely.

Thispaperexploresaprivacy-preservingVerticalFederated Learning framework combined with representation synthesis techniques. The proposed system enables organizations to contribute their local knowledge without exposing sensitive data, thereby creating a secure and efficient collaborative learning environment.

II. BACKGROUND AND MOTIVATION

Modern organizations possess massive volumes of valuable data.Hospitalsmaintainpatienthealthrecords,banksstore transactionhistories,andtechnologycompaniescollectbe- havioral information. Individually, each dataset provides only partial insights. Combining these datasets could significantly improvepredictivecapabilitiesanddecision-makingprocesses. However, directly sharing sensitive information introduces major risks including data breaches, unauthorized access, and regulatoryviolations.Organizationsarethereforeforcedtochoosebetweencollaborationandprivacyprotection.Thischallengemotivatedthede velopmentofprivacy-preservingdistributedlearningsystems.

VerticalFederatedLearningoffersapromisingsolutionbyal- lowingmultiplepartiestocollaborativelytrainmachinelearning models without exchanging raw data.

Each participant retains complete ownership of its data while contributing privacy-safe intermediate representations to the learning process.

The motivation behind this work is to design a framework where organizations can benefit from collective intelligence without compromising confidentiality, trust, or legal compli- ance.

III. LITERATURE SURVEY

A. *Personalized Federated Learning for Privacy-Preserving and Scalable IoT-Driven Smart Healthcare*

- AuthorandYear:DilipKumarJangBahadurSaini,Nilesh Shelke,AmitPimpalkar,PrajwalasimhaSN,RanjimaP,Vinitha V (2025)
- Methodology:Thisstudyproposesapersonalizedfederated learning framework for smart healthcare applications. The systemintegratesfederatedlearningwithIoT-basedmedicaldevices to support privacy-preserving healthcare analytics. Meta-learning techniques and lightweight homomorphic en- cryption are used to improve personalization and securitywhile maintaining scalability across distributed healthcare environments.
- Limitation: The framework introduces increased computa- tionaloverheadduetopersonalizationmechanisms.Scalability beyond large client networks remains a challenge, and multi- modalhealthcaredataintegrationrequiresfurtherimprovement.

B. *Federated Learning: Next-Gen Privacy-Preserving AI Framework for Consumer and Industrial Applications*

- Author and Year: Ravinder Singh, Smriti Mahajan, Sofia Singh (2025)
- Methodology: This paper proposes a multi-layer federated learning architecture integrating Differential Privacy, Homo morphicEncryption,SecureMulti-PartyComputation(SMPC), andTrustedExecutionEnvironments.Theframeworksupports secure collaborative learning across industrial and consumer applications using non-IID distributed datasets.
- Limitation: The study relies mainly on synthetic datasets and simulated security mechanisms. Real-world deployment challenges and adversarial attack resistance are not fully addressed.

C. *Vertical Federated Representation Synthesis for Non- Aligned Samples in the Active Party*

- Author and Year: Jintao Liang, Sen Su, Zhenya Wang (2025)
- Methodology: This paper introduces a Vertical Federated Learning framework called VFedRS that handles non-aligned samples through representation synthesis. Vertical Federated PCA is used to generate missing feature representations securely while improving collaborative learning performance without exposing raw data.
- Limitation: The approach increases communication and computational overhead during representation synthesis. Model accuracy also depends heavily on the availability of aligned samples.

D. *Secured Cost-Effective Anonymous Federated Learning With Proxied Privacy Enhancement for Personal Devices*

- Author and Year: Muhammad Senoyodha Brennaf, Po Yang, and Vitaveska Lanfranchi (2025)
- Methodology: This paper introduces a proxy-based anony- mous federated learning framework for personal devices. The proxy removes client identity information before forwarding en- cryptedmodelupdatestotheserver,improvinganonymityand privacyprotection.Twocommunicationprotocols,namelyTwo- Stage Communication (2SC) and Three-Stage Communication (3SC), are proposed along with AES encryption and LZ-string compressiontoreducebandwidthusagewhilepreservingmodel accuracy.
- Limitation:Collusionbetweentheproxyandtheservercan still expose client identities. The framework also lacks formal Differential Privacy guarantees and faces tradeoffs between bandwidth efficiency, memory usage, and inference speed.

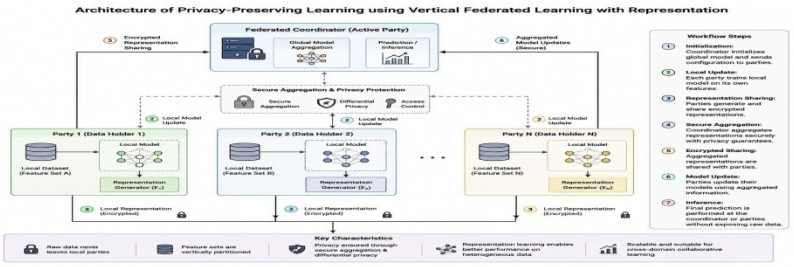


Fig.1.Artitecture

E. FederatedLearningforDistributedIoTSecurity:APrivacy- Preserving Approach to Intrusion Detection

- Author and Year:Gutti Chandu, Thumula Karthik, and Balbudhe Parag (2025)
- Methodology: This paper proposes a federated learning- based intrusion detection framework for distributed IoT net- works. IoT devices locally train security models and share encryptedupdateswiththeserverinsteadoftransmitting raw network traffic data. The framework improves privacy preservation, reduces centralized attack risks, and supports collaborative cyberattack detection across heterogeneous IoT devices.
- Limitation: The framework mainly focuses on intrusion detection and is not evaluated for broader IoT analytics applications. Communication overhead and synchronization delaysmayincreasesignificantlyinlarge-scaledeployments.

F. FederatedLearningforPrivacy-PreservingDataMining

- AuthorandYear:DattatrayRaghunathKale,TusharMane, Amar Buchade, Prashant Bansilal Patel, et al. (2024)
- Methodology: This paper presents a federated learning frameworkforprivacy-preservingdataminingacrosshealthcare, finance,andIoTdomains.Theframeworkcombinesencryption methodsandDifferentialPrivacytoenablecollaborativemodel training without sharing sensitive raw data. It also addresses communication efficiency and heterogeneous non-IID data distributionswhilemaintainingperformanceclosetocentralized systems.
- Limitation:Theuseofstrongprivacy-preservingtechniques increases computational complexity and communication over- head. The framework also faces challenges related to secure aggregation and protection against malicious participants.

IV. METHODOLOGY

The proposed system follows a Vertical Federated Learning architecture where multiple organizations collaboratively train a machine learning model without exchanging raw datasets. Each organization stores different attributes of the same users locally and independently trains a local model.

After local training, privacy-preserving embeddings are generatedandsecurelysharedwithacentralaggregationserver. The server combines these embeddings into a unified repre- sentation and trains a final global model capable of generating accurate predictions while maintaining data confidentiality.

A. DataAcquisitionandCollection

Theframeworkcollectsdistributedfeaturesetsfrommultiple organizations participating in the Vertical Federated Learning process. Each organization maintains different information about the same group of users.

Thecollecteddatamayinclude:

- StructuredData:Featurevectors,transactionrecords, demographic information, and numerical datasets.
- Semi-StructuredData:JSONfiles,metadata,system logs, and communication parameters.
- Unstructured Data: Medical reports, textual documents, and multimedia files.

Alldataremainsstoredlocallywithintheparticipating organizations.

B. DataValidationandPreprocessing

Before training begins, each organization preprocesses and validates its local datasets.

Thepreprocessingstageincludes:

- Datacleaninganddduplicateremoval
- Handlingmissingandinconsistentvalues
- Featurenormalizationandscaling
- Validation againstpredefined formats
- Privacy screening to remove identifiable information Thesestepsensurethatonlycleanandreliablerepresenta- tionsparticipateinthefederatedlearningprocess.

C. ApplicationLayerandUserInteraction

The application layer provides interfaces for participating organizations to interact with the system securely.

Keyfunctionalitiesinclude:

- User-friendlyclientinterfaces

- Secureauthenticationandauthorization
- Role-basedaccesscontrol
- Monitoringtrainingprogress
- Visualization of model results and performance metrics Thislayersimplifiesinteractionbetweenusersandthe federatedlearninginfrastructure.

D. BlockchainNetworkandDistributedLedger

Theblockchainlayermaintainssecureandimmutablerecords of federated learning activities. Transactions related to model updatesandclientparticipationarerecordedwithindistributed ledger blocks.

Cryptographic hashing techniques ensure data integrity, transparency,andtamperresistancethroughoutthecollaborative learning environment.

E. SecurityandPrivacyMechanisms

Tostrengthensystemsecurity,theframeworkintegrates several privacy-preserving mechanisms.

- Encryption:Securecommunicationchannelsprotect transmitted embeddings.
- SecureAggregation:Localrepresentationsareaggregated without exposing private data.
- Access Monitoring: Blockchain records maintain trans- parent audit logs.
- PrivacyPreservation:Rawdatasetsneverleavelocal organizational environments.

Thesemechanismscollectivelyimprovetrust,confidentiality, and system reliability.

V. CONCLUSION

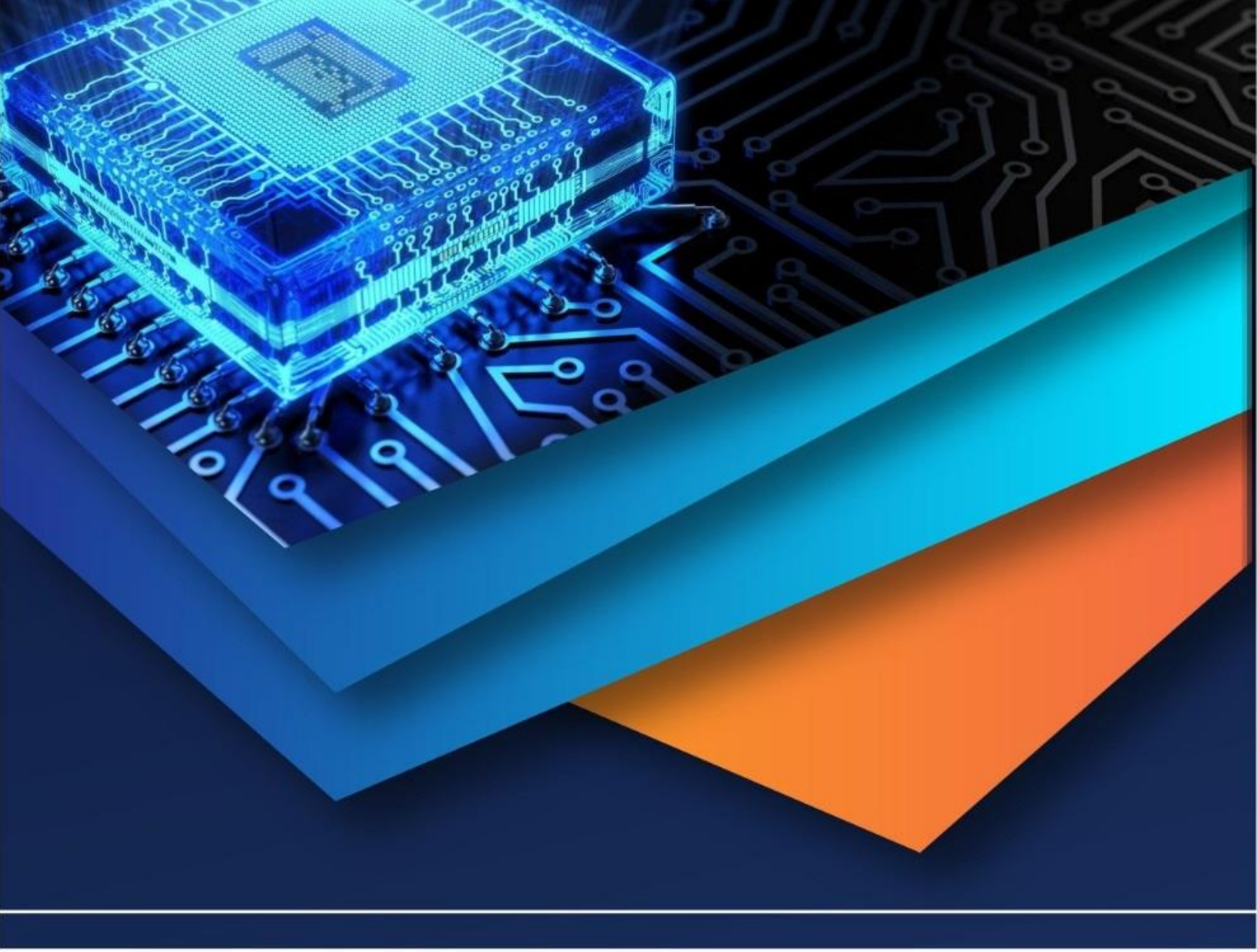
This paper presented a privacy-preserving learning frame- work using Vertical Federated Learning with Representation Synthesis. The proposed architecture enables multiple organi- zations to collaboratively train machine learning models while ensuring that sensitive raw data remains protected within local environments.

By integrating secure aggregation, encryption techniques, and distributed learning mechanisms, the framework enhances privacy, transparency, and collaboration among participating organizations. The system demonstrates that organizationscan achieve high-quality collaborative intelligence without sacrificing confidentiality or regulatory compliance.

Future work may focus on reducing communication over- head, improving scalability for large-scale deployments, and integrating advanced privacy-preserving technologies such as fully homomorphic encryption and differential privacy.

REFERENCES

[1] DilipKumarJangBahadurSainietal.,“PersonalizedFederatedLearningfor Privacy-Preserving and Scalable IoT-Driven Smart Healthcare,” 2025.
[2] Ravinder Singh, Smriti Mahajan, Sofia Singh, “Federated Learning:Next-GenPrivacy-PreservingAIFrameworkforConsumerandIndustrialApplications,” 2025.
[3] JintaoLiang,SenSu,ZhenyaWang,“VerticalFederatedRepresentationSynthesis for Non-Aligned Samples in the Active Party,” 2025.
[4] Zhang et al., “Privacy-Preserving Federated Learning using SecureAggregation,” 2024.
[5] Chenetal.,“DifferentialPrivacyinFederatedLearningSystems,”2023.
[6] GuttiChandu,ThumulaKarthik,andBalbudheParag(2025)



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY