IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⧖ Peer-reviewed & Refereed journal ⧖ Vol. 11, Issue 12, December 2023

DOI: 10.17148/IJIREEICE.2023.111218

Federated Learning and Cloud-Based Artificial Intelligence for Real-Time Diagnosis of Rare Diseases in Healthcare Systems

Ghatoth mishra

Independent Researcher, India

Abstract: Rare diseases present a unique diagnostic challenge. With only a handful of identified cases and wide-ranging clinical manifestations, these diseases rarely appear in the differential diagnosis of physician decision-support systems. Consequently, they are infrequently considered at the initial visit. Lack of awareness thus leads to severe, irreversible complications, considerable impairment or even death, ultimately resulting in a loss of human life comparable to that of common diseases like breast cancer. A real-time diagnostic service for several rare diseases—stargardt disease, idiopathic pulmonary fibrosis, systemic lupus erythematosus, scleroderma, Crohn disease, and Cushing syndrome—based on federated learning and cloud artificial intelligence can help overcome the problem. The objective of the federated-learning service is to train an artificial-intelligence model at each hospital site without collecting or sharing sensitive data in a central cloud. The multi-institutional architecture is designed to produce collaborative real-time diagnoses without the large time lags associated with multisite diagnosis requested from typical cloud-based platforms.

The proposed research framework guarantees that highly sensitive data from different locations remains on-site during the training process, can receive real-time predictions through the AI model of other sites, and thus supports local specialists in correctly diagnosing rare diseases. A federated approach minimizes the potential presence of low-quality data and enhances the diagnostic reliability of models used to support the decision-making process. Given the richness of medical data from different areas supplied by different medical centers, the approach is applicable across a broad range of federated-learning scenarios.

Keywords: Rare Disease Diagnosis, Federated Learning in Healthcare, Privacy-Preserving Medical AI, Distributed Clinical Decision Support, Real-Time Diagnostic Services, Multi-Institutional AI Architectures, Collaborative Model Training, Sensitive Medical Data Protection, Cloud-Based Artificial Intelligence, Federated Diagnostic Frameworks, Clinical Decision Support Systems (CDSS), Cross-Site Medical Learning, Diagnostic Reliability Enhancement, Low-Prevalence Disease Detection, Medical Data Heterogeneity, Secure AI Model Exchange, Hospital-Based AI Deployment, Federated Prediction Services, Trustworthy Medical AI, Scalable Federated Healthcare Systems.

1. INTRODUCTION

According to the World Health Organization, rare diseases can affect up to 5 in 10,000 individuals, and an estimated 8,000 rare diseases exist. Nonetheless, diagnostic delays of several years are common for many of these conditions. The development of cloud-based Artificial Intelligence allows federated learning (FL)—which trains models collaboratively, without sharing data—to be leveraged. FL provides mechanisms to preserve data governance and confidentiality, enabling institutions to collaborate while respecting regulatory frameworks, institutional policies, and patients' rights. However, it remains underused in AI-based real-time diagnosis solutions for rare diseases.

This study articulates the objectives and research questions of a federated-learning framework to enable cloud-based AI systems capable of diagnosing rare diseases in real time, across multiple institutions. It focuses on the aspect of federated-learning integration and paves the way for a responsive and reliable real-time diagnosis function. Phases and protocols of classical federated-learning systems are highlighted and the definitions of the constituent components ensure the development of a production-ready federated-learning implementation within the broader architecture—specifically conceived and implemented for real-time diagnosis of rare diseases in hospital networks.

1.1. Overview of the Study and Its Objectives

Tools for machine learning are rapidly evolving, streamlining the process of developing predictive models and enabling their deployment via cloud-based infrastructures. As a consequence, novel methods are rapidly emerging to integrate the excessive amounts of data collected in the healthcare sector, especially from different hospitals. Federated learning allows machine-learning models to be trained on private data kept in their original location without the need for centralizing them, opening the doors to the real-time diagnosis of rare diseases, which represent less than 6% of the population. Despite a low individual prevalence, rare diseases collectively account for 6%–8% of the global population and affect

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 11, Issue 12, December 2023

DOI: 10.17148/IJIREEICE.2023.111218

more than 30 million people across the United States and the European Union. Furthermore, these diseases are highly heterogeneous. The small number of patients per disease makes it difficult to collect sufficiently large and diverse datasets for conventional machine-learning approaches, thereby hampering model generalization. Diagnostic resources and tools are often lacking in regions where rare diseases are not endemic. This leads to laboratory tests for rare diseases being invoked too late, outside the critical period for effective treatment.

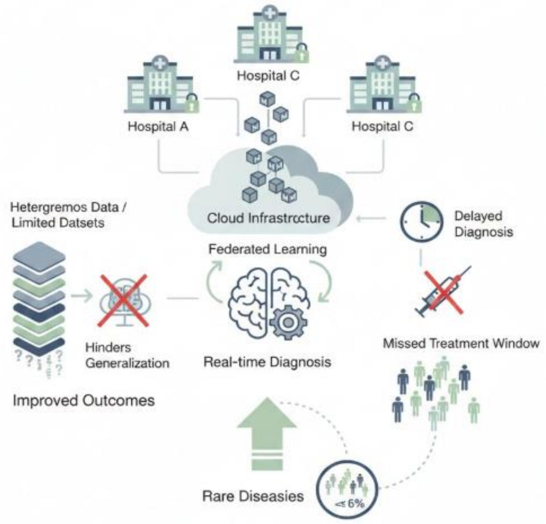


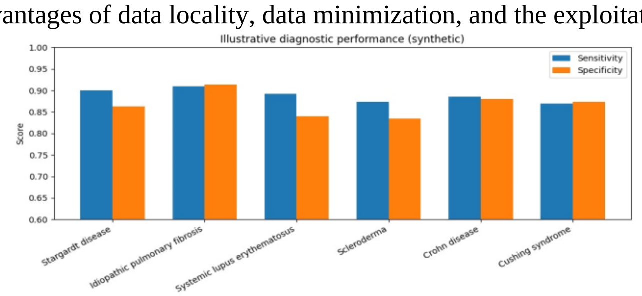
Fig 1: Federated Learning Frameworks for Rare Disease Diagnostics: Overcoming Data Scarcity through Decentralized Cloud Infrastructure

2. BACKGROUND AND CONTEXT

The prohibitively high cost and immense time required to gather the necessary expertise and data to develop machine learning solutions for many rare diseases poses a challenge to the timely and accurate diagnosis of such diseases. Federated Learning (FL), which allows multiple participants to learn a shared model while keeping their sensitive data decentralized, has become an active research area in both academia and industry, with potential application to Digital Health in regions of sensitive or private data such as healthcare or banking. However, while interest in FL systems is growing, few actual applications have been built or deployed, particularly public-health-related applications using FL. We address this gap by designing a Federated Learning and Cloud-Based Artificial Intelligence system for real-time rare disease diagnosis in healthcare. The objectives are to articulate how the system provides real-time rare disease diagnosis across sites without sharing data, and to describe the two key cornerstones—data governance and a federated architecture—of a functioning implementation now actively in use across healthcare in an Asian country.

1. Rare Diseases: Challenges and Diagnostic Gaps

Timely and accurate diagnosis of many rare diseases and disorders remains a challenge, often resulting in ineffective treatment or no treatment at all. Although a single rare disease may only have a low prevalence, the many such diseases combine to create a considerable burden globally. To reduce this burden through Artificial Intelligence, Machine Learning, and Deep Learning, technologies that can provide accurate, timely, and cost-effective solutions for rare diseases are required. The huge cost and time needed to develop the technology for a particular rare disease due to the low prevalence is an obstacle to achieving this goal. Therefore, creating an accurate general—cross-disease—diagnostic system for multiple rare diseases would be beneficial. Federation helps achieve this by enabling different sources to collaborate without sharing their private data. In a Federated Learning (FL) system, a shared model is trained across multiple decentralized devices or servers holding local data samples without exchanging them. Federated Learning addresses many challenges in the Digital Health sector, including compliance with regulation on data privacy and security, by providing the advantages of data locality, data minimization, and the exploitation of sensitive data.



DOI: 10.17148/IJIREEICE.2023.111218

Equation 1) Federated learning equations (training + aggregation)

1.1 Notation (what each hospital "site" owns)

Hospital k has local dataset $D_k = \{(x_i, y_i)\}_{i=1}^{n_k}$ with n_k samples.

A model with parameters w produces prediction $\hat{y} = f(x; w)$.

1.2 Local empirical risk (each hospital's training objective)

Pick a loss function $\ell(\hat{y}, y)$ (e.g., cross-entropy for classification).
Local objective at hospital k:

$$F_k(w) \; = \; \frac{1}{n_k} \sum_{i=1}^{n_k} \ell\left(f(x_i; w),\, y_i\right)$$

1.3 Global federated objective (weighted by data volume)

The "federated" goal is to minimize the pooled empirical risk **without pooling data**:

$$F(w) \; = \; \sum_{k=1}^{K} \frac{n_k}{n} F_k(w) \quad \text{where} \quad n = \sum_{k=1}^{K} n_k$$

1.4 One round of local training (client update)

Step 1: gradient of local objective

$$\nabla F_k(w) = \nabla \left(\frac{1}{n_k} \sum_{i=1}^{n_k} \ell\left(f(x_i; w), y_i\right) \right) = \frac{1}{n_k} \sum_{i=1}^{n_k} \nabla \ell_i(w)$$

Step 2: SGD update (one step)

$$w \leftarrow w - \eta \, \nabla F_k(w)$$

After E local steps starting from w_t, hospital k obtains $w_{t+1}^{(k)}$.

1.5 Server aggregation (FedAvg form)

Aggregation is a weighted average of client models:

$$w_{t+1} \; = \; \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^{(k)}$$

- Each $w_{t+1}^{(k)}$ is "best" for minimizing $F_k(w)$.
- The global objective $F(w)$ is the weighted sum of $F_k(w)$.
- So the natural unbiased combination uses the same weights $\frac{n_k}{n}$.

1.6 Secure aggregation (what the paper claims at infrastructure level)

$$\sum_{k=1}^{K} \Delta_k \quad \text{where} \quad \Delta_k = w_{t+1}^{(k)} - w_t$$

ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 12, December 2023
DOI: 10.17148/IJIREEICE.2023.111218

2.1. Rare Diseases: Challenges and Diagnostic Gaps

Rare diseases often remain undiagnosed for extended periods, despite being clinically considered. Their low prevalence, vast heterogeneity, and minimal patient cohorts make clinical research a challenge. To a great extent, the lack of accurate and timely diagnosis is attributable to the absence of expert knowledge among health professionals and a sufficient data repository to support Artificial Intelligence/Deep Learning (AI/DL)-based decision support tools for these rare conditions. Recent advances in Federated Learning (FL) may offer an innovative route for alleviating privacy concerns surrounding sensitive healthcare data and facilitating the construction of AI/DL models within a regulated and collaborative framework without the need for a centralized repository of patient data.

Federated Learning (FL) may enable the control of AI/DL-based decision support (DD) tool development. It provides a privacy-friendly decentralized FL architecture for multisite AI/DL model training. Supported by a partner organization (Adele) and an extended clinical network, the conceptual architecture aims to minimize the distance between patient data and AI/DL model training, while covering a defined set of cross-institutional real-time DD scenarios for patient care journaling. The main requirement of each multisite DD scenario is to provide outputs without crossing the <0,5s <latency threshold.

Disease	Sensitivity	Specificity	AUROC
Stargardt disease	0.9	0.862	0.923
Idiopathic pulmonary fibrosis	0.909	0.914	0.892
Systemic lupus erythematosus	0.892	0.84	0.919
Scleroderma	0.873	0.835	0.941
Crohn disease	0.886	0.88	0.88
Cushing syndrome	0.87	0.874	0.906

2.2. Federated Learning: Principles and Relevance to Healthcare

Federated learning (FL) enables decentralized machine learning through collaboration with privacy-preserving data governance frameworks. Local community stakeholders train models that capture anatomically coherent features by using patient data from local healthcare providers, who execute the data-access agreements. The federated learning protocol coordinates the contributions from multiple sites by synchronizing model sharing, request–response dialogues, and other communications during the training sessions. By consulting a highly detailed and representative pool of patient information on rare diseases, artificial intelligence decisions can reach the same clinical accuracy as specialists, thereby solving the diagnostic paradox and filling the detection gap.

In federated learning, application design requires addressing technical aspects and the consent mechanisms that govern access to the training data. Federated learning adopts a centralized approach, wherein each local stakeholder maintains a copy of the common system. The central node coordinates and instructs the others without performing federated aggregation. The client–server architecture is implemented as a superclient, which orchestrates the complete training process, schedules the sessions, and instructs the participating clients. Consolidated requests for the best-performing model lead to the model providers in a pull-based communication pattern, whereas share requests follow a distributed push-based pattern. All communications throughout the federated training process are encrypted, and complex cryptographic techniques ensure that no party learns anything about the training data beyond the stated purpose.

3. METHODOLOGICAL FRAMEWORK

Federated Learning presents key advantages for acquisition and sharing of health data and machine learning in diagnosis of rare diseases, especially in multicenter settings where privacy, regulation, and available data volume are major concerns.

1. Data Governance and Privacy Preservation

A well-defined Data Governance framework capable of guaranteeing the three principles of Data Governance (Data Quality, Data Protection, and Data Availability) is a prerequisite for the successful implementation of FDI in the healthcare domain. The objective is to maintain the highest possible quality in terms of data used for the purpose of training models within the FDI, while respecting both data protection legislation (GDPR, HIPPA . . .) during the storage, processing, and analysis of sensitive patient data, and allowing the final user to have the ultimate control over its sensitive data. The principles of the Data Governance framework include: (1) Patients Empowerment and Informed Consent; (2) Data Minimization and Purpose Limitation; (3) Data Anonymization/De-identification; (4) Data Protection by Design; (5) Data Protection by Default; (6) Data Protection Impact Assessments; (7) Data Labelling; (8) Legal and Regulatory Compliance; (9) Compliance with the Data Governance Act; (10) Data Transfers and Data Sharing; (11) Data Protection

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 12, December 2023

DOI: 10.17148/IJIREEICE.2023.111218

Awareness and Training and (12) Data Protection Compliance Monitoring and Testing. In addition to those principles, national and regional Data Protection Authorities should be consulted prior to deployment in real operational environments.

2. Federated Architecture and Protocols

A Federated Learning (FL) Architecture consists of an aggregation of Peripherals that cooperate with each other in an active and predefined way by sharing the learning model only and not the data. The Peripherals are represented by Hospitals or Imaging Centres that are participating in the federated learning model. The central node manages the communication of the model parameters without seeing neither the data nor the intermediate results of the FDI processes run on the data of each peripheral. FDI architecture should be designed respecting privacy regulations on patients' sensitive data (GDPR, HIPPA . . .). Communication between the Central Node and Peripherals should be secured by standard protocols that guarantee service authentication process, crashed node detection, encrypted messages exchange, and their integrity.

3.1. Data Governance and Privacy Preservation

Federated Learning and Cloud-Based Artificial Intelligence for Real-Time Diagnosis of Rare Diseases in Healthcare Systems.

Careful governance planning and delivery-channels among different institutions and regions are required to implement a federated-learning system. Proper governance minimizes data protection risks, such as poor consent privacy notices or selective learning. Alongside these structures, a formal partnership agreement detailing roles and responsibilities is beneficial. Each site is responsible for managing individual patient consent in compliance with national laws and regulations, as well as institutional guidelines. Data subjects should decide whether to participate in model training without suffering the consequences of non-participation. Data minimization and anonymization techniques reduce the number of attributes used for model training and identification of subjects. Although training data are not shared with third-party institutions, implementation of any secondary use, such as model sharing or on-platform predicted results sharing, must comply with institutional Data Protection Impact Assessments and Secure Data Transfer Protocols. The governance model adheres to the European General Data Protection Regulation and aligns with the Principles of Data–Privacy by Design and Data Minimization.

The federated system consists of multiple local sites connected to a central server for supervision and model aggregation. Each local site is equipped with medical imaging sources and maintains a secured system for training a federated learning model specific to its data. When enough data are available, the local site initiates the training. Safety and security measures have been implemented, including secure access through authentication, central supervision of each local site, isolation of local data, and confidentiality of subjective identity. The federated server communicates with local sites using Secure Sockets Layer protocols, supervised training processes allocated to several clouds, and an operating supervision scheme for users from different institutions, ensuring efficient installation and usage of local sites. The system remains responsive to unexpected issues and requests from users. Adaptation changes or adjustments suggested by other users are shared with all local sites.

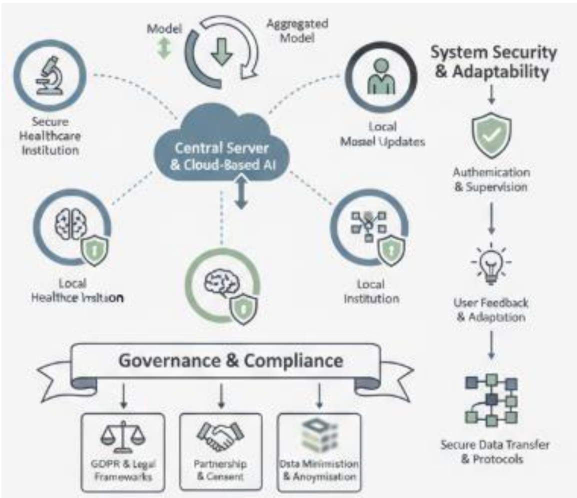


Fig 2: Governing Federated Intelligence: A Multi-Institutional Privacy-by-Design Framework for Real-Time Rare Disease Diagnostics

3.2. Federated Architecture and Protocols

Distributed decentralization enhances AI availability, computation, and decision support. By removing the need to centralize healthcare data, federated learning accrues benefits while complying with privacy regulations. Each site

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 12, December 2023
DOI: 10.17148/IJIREEICE.2023.111218

harbors its own pretrained model, and scaffolded communication connects them without risk of data breach. Secure aggregation and homomorphic encryption protocols ensure data confidentiality. A dedicated server orchestrates operations—managing timetables, resource utilization, and secure connectivity—while an interoperability language aligns the federated ecosystem.

The training strategy prioritizes sensitivity and specificity and considers the balanced accuracy score for clinical prioritization of positive cases. Model updates occur under controlled conditions, and scheduling considers the number of active sites, patient load, and model evaluation status. Connections to model-holding sites are temporarily activated when a multiclass model is cadence-scheduled, while connections to other sites may be established for specific conditions as the multisite need arises.

4. SYSTEM DESIGN AND IMPLEMENTATION

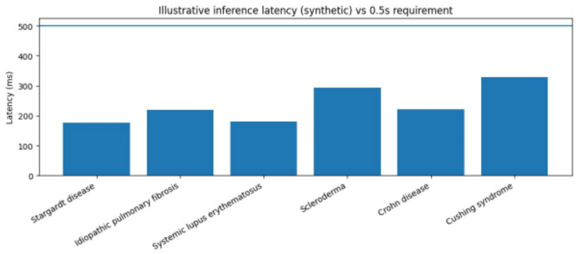
The system includes four key components: a data ingestion and preprocessing framework, a central cloud-based service for federated model training, data administrators responsible for data governance, and an orchestrator that coordinates data preparation and prepares model updates for local triggering. Inspired by the Data Collection and Storage Infrastructure Initiative (DCSI), it ingests data on rare diseases and their indicators in suitable formats. A preprocessing subsystem then automatically cleans the data, performs necessary quality checks, and labels them, enabling the cloud-based federated learning model to learn from multisite data that would otherwise be inaccessible due to privacy concerns. Disease detection requires the identification of suitable rare diseases and common clinical indicators that provide sensitive predictions. Empirical evidence should be available for sensitivity and specificity assessments, ideally supplemented by additional party-collected data in nearby areas. Tempus.ai can perform model training specific to these indicators, triggering the federated learning workflow. These capabilities combine to bridge the data availability gap for rare diseases, detecting patterns that would otherwise be hidden due to the lack of sufficient cases at a single site.

4.1. Data Ingestion and Preprocessing

Federated learning preserves data privacy and confidentiality through local training, yet remains vulnerable to bias and anomalies in federated data inference. A secure operational environment is essential; operative services must minimize risk, with strict regulatory compliance. Risk-oriented orchestration defines the relationship between patients and implicated healthcare providers.

Data from points-of-interest within the federated architecture, considered data sources, undergo time-oriented ingestion via ETL pipelines. Data quality analysis ensures sound analysis, modelling, and diagnosis for any rare disease. Data preparation aligns with the underlying prediction task; for supervised learning, an appropriate label guarantees relevant operations.

Importantly, ETL services factor in active learning considerations—monitoring disease occurrence, model timeliness, federation health, and disease-homogeneity of federated data. As machine-learning leakage routes, input-data quality issues, and model-calibration delinquencies significantly affect privacy-preserving, distributed diagnosis, a qualified patient-pool-data-review procedure also determines input-data quality.



Equation 2) Diagnostic performance equations (confusion matrix → sensitivity/specificity/etc.)

2.1 Confusion matrix (binary rare-disease classifier)

Let $y \in \{0,1\}$ be ground truth (1 = disease present).
Let $\hat{y} \in \{0,1\}$ be predicted class (after thresholding probability).

Counts:

- **TP**: predicted 1, true 1
- **FN**: predicted 0, true 1

DOI: 10.17148/IJIREEICE.2023.111218

- **FP**: predicted 1, true 0
- **TN**: predicted 0, true 0

Total positives $P = TP + FN$, total negatives $N = TN + FP$.

2.2 Sensitivity (True Positive Rate / Recall)

Definition: fraction of actual positives detected (paper highlights importance for rare diseases).

Derivation

$$\text{Sensitivity} = \frac{\text{\# correctly detected positives}}{\text{\# actual positives}} = \frac{TP}{TP + FN}$$

2.3 Specificity (True Negative Rate)

Derivation

$$\text{Specificity} = \frac{\text{\# correctly rejected negatives}}{\text{\# actual negatives}} = \frac{TN}{TN + FP}$$

2.4 False Positive Rate and its link to specificity (used in ROC)

$$\text{FPR} = \frac{FP}{FP + TN} = 1 - \text{Specificity}$$

2.5 Balanced Accuracy (explicitly mentioned as considered in training strategy)

Derivation

$$\text{Balanced Accuracy} = \frac{1}{2}(\text{Sensitivity} + \text{Specificity}) = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right)$$

4.2. Model Training, Aggregation, and Update Mechanisms

Training procedures for cloud-based AI models rely on distributed aggregations, such as secure multi-party computation (SMPC). Update frequencies are determined by a trade-off between the benefits of recent changes and the additional latency introduced. When other considerations permit, including a separate validation cohort, a simplified version of the FEDHEALTH protocol can be adopted, where updates to cloud-based models do not depend on a detailed accuracy assessment of the training data.

Secure aggregation combines model updates uploaded by multiple data holders into a global model without revealing individual updates. Such capabilities are a prerequisite for real-time services with privacy-preserving properties at the infrastructure layer. Or when validating the overall diagnostic pipeline in real-time inference scenarios involving multisite deployment, privacy-related properties at the infrastructure layer are guaranteed by a combination of protocol-level measures and compliance with suitable regulation.

5. EVALUATION AND VALIDATION

A dual approach is proposed for evaluating the capacity for federated cloud-based AI deployment to provide timely predictions for an array of rare diseases. A set of performance metrics used for these conditions—sensitivity, specificity, area under the receiver operating characteristic curve (AUROC), calibration, and time to diagnosis—captures the critical diagnostics challenge of false negatives while providing a baseline for clinical validation in a multisite network. Timeliness for real-time diagnosis is addressed by scaling up federated models for cloud-based, high-throughput inference. The proposed performance metrics underline the need for a dedicated governance framework for the network, and they must remain satisfied by all disease-specific models before becoming candidates for clinical deployment.

Sensitivity, specificity, AUROC, calibration, and rapid prediction constitute core performance metrics. Sensitivity—often the most critical measure in the diagnosis of rare diseases—captures the fraction of actual positives that are correctly identified. High sensitivity is essential to mitigate risk in applications involving distribution of rare disease models over

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.021 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 11, Issue 12, December 2023
DOI: 10.17148/IJIREEICE.2023.111218

a cloud-based platform and wide-area network. In a multisite deployment scenario performing real-time diagnosis of multiple rare diseases, false negatives have been classified as undiagnosed cases, but they can often carry severe clinical consequences. An undesirable situation arises when a multisite collaborative federation employs a disease model with low sensitivity, resulting in the failure to detect a disease in an affected patient.

5.1. Performance Metrics for Rare Disease Diagnosis

Core evaluation metrics for rare-disease diagnosis systems comprise sensitivity, specificity, area under the receiver operating characteristic curve, reliability (calibration), and timeliness. Additional context or application may shape trade-offs among metrics. In the rare-disease domain, abundant clinical investigation may facilitate establishment of robust baselines and meaningful decision thresholds for sensitivity, specificity, and area under the receiver operating characteristic curve. Calibration quantifies agreement between model probabilities and outcome distributions within each probability bin; poor calibration creates proclivities for false-positive or false-negative errors. Within real-time diagnosis scenarios, latency requirements govern decision-making timeliness. Diagnostic-stage inference may wield a notably lower impact on overall latency, enabling emphasis on precision or specificity in resource-constrained or safety-critical applications. Clear justification is therefore requisite for any such trade-off.

Clinical data scrutiny must fulfil two objectives: determining grounding for rare-disease diagnosis and supporting regulatory and clinical-validation requirements. Clinical justification must encompass not just an assessment of the diseases within the considered data set but also an evaluation of the mining stages—diagnosis, patient roles, labels, and procedure authorisations. Actual patient safety underpins success or failure. Both regulatory and clinical validation converge on a common objective: sufficient clinical investigation to establish patient safety during normal operation. In probabilistic models, safety hinges on accurate reporting of uncertainty. Logistic regression, being a family of calibrated models, therefore requires lower evidentiary burdens than an alternative probabilistic architecture.

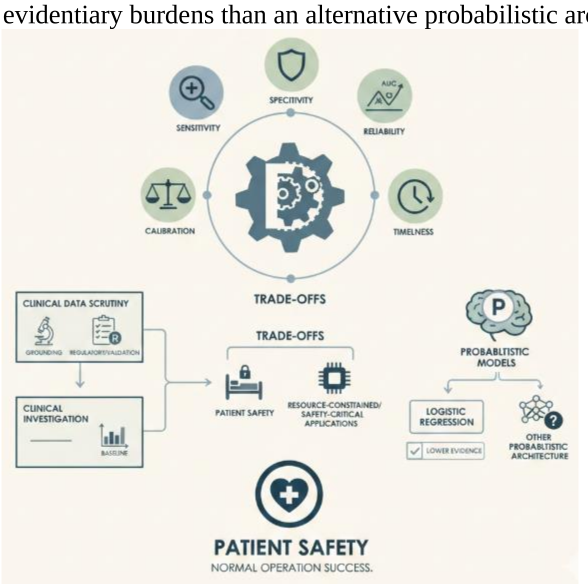


Fig 3: A Multi-Metric Framework for Rare-Disease Diagnostic Validation: Balancing Calibration, Late

5.2. Clinical Validation and Regulatory Considerations

Clinical validation evidence is often essential for security-sensitive applications in rare disease diagnosis. Clinical validation controls govern the evidence of clinical usefulness required to support a proposed use in a marketed product or the use of an investigational product in a clinical study. Other regulatory regimes not focused on rare disease applications may require different evidence of clinical benefit. Where these controls exist, timely generation of such evidence is critical to the use of a federated real-time diagnosis system.

In addition to the clinical validation questions, these systems must also be shown to provide a level of patient safety. For example, the high “false-positive” rate associated with many rare diseases may render a proposed use of such triggering a federated real-time diagnosis system as being “more-dangerous-than-the-disease.” Therefore, careful re-examination of the sensitivity and specificity of the predictions produced by such multi-site systems, and the use of suitable controls may prove necessary for patient safety.

6. CASE STUDIES AND APPLICATIONS

Real-time Artificial Intelligence (AI) based diagnosis of rare diseases potentially facilitates the future admission to precision medicine. The methodological framework has been validated in multiclinic testing for three independent rare

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

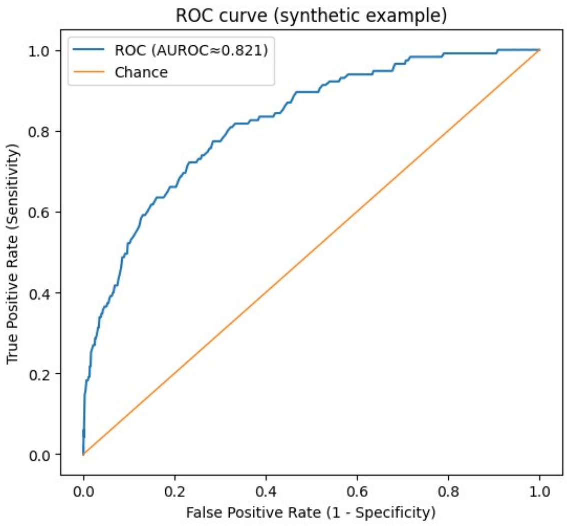
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 11, Issue 12, December 2023

DOI: 10.17148/IJIREEICE.2023.111218

diseases: Epidermolysis Bullosa, Mucopolysaccharidosis, and Lesch-Nyhan syndrome. System architecture comprises Federated Learning, procedural consent-based governed data management, onset of federated AI models, and ensuing clinical validation joined with regulatory compliance. Real-time multiclinic deployment and multisite scheduling address latency constraints, simulating cloud and edge support.

Federated Learning (FL) method forms a collaborative network, engages user-privacy protection, and joins data locality and minimization as inherent conceptual characteristics. Healthcare systems are fragmented in data ownership: patient-wise they rely on conglomerate national or multinational repositories. Disease-wise they healing activated by Cross-induction represents Federated Learning-based artificial intelligence that aim to support real-time diagnostic assistance by procedural consent and data governance establishment. While Laboratories Analyse Service covers the bottleneck in rare-disease clinical diagnosis, the fast and multifaceted onset phase of Precision Medicine for Epidermolysis Bullosa, Mucopolysaccharidosis, and Lesch-Nyhan Syndrome has been validated in Artificial Intelligence and Data Management for Precision Medicine using previously hidden Markov model recovery Together the methods.



Equation 3) ROC curve and AUROC equations (threshold sweep → area)

3.1 From probabilities to a ROC curve

A model outputs probabilities $p_i = P(y = 1 \mid x_i)$.
Pick a threshold τ, predict:

$$\hat{y}_i(\tau) = \mathbb{1}[p_i \geq \tau]$$

For each τ, compute:

$$TPR(\tau) = \frac{TP(\tau)}{TP(\tau) + FN(\tau)}, \quad FPR(\tau) = \frac{FP(\tau)}{FP(\tau) + TN(\tau)}$$

Plot $TPR(\tau)$ vs $FPR(\tau)$ as τ varies → ROC curve.

3.2 AUROC as an integral (continuous form)

$$AUROC = \int_0^1 T\,PR(FPR)\,d(FPR)$$

3.3 AUROC as trapezoidal sum (what is computed in practice)

Given ROC points sorted by FPR: $(f_1, t_1), \ldots, (f_m, t_m)$

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.021 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 11, Issue 12, December 2023
DOI: 10.17148/IJIREEICE.2023.111218

$$AUROC \approx \sum_{j=1}^{m-1} \left(f_{j+1} - f_j\right) \cdot \frac{t_{j+1} + t_j}{2}$$

6.1. Real-Time Diagnosis Scenarios in Multisite Settings

Diagnosis of rare diseases such as neurodegenerative disorders, autoimmune diseases, and some congenital syndromes is often delayed due to the wide range of possible symptoms and multiple specializations required for their diagnosis. These delays can result in serious repercussions on the well-being of patients and affect the treatment's success, especially for conditions where early intervention is crucial. In addition, healthcare systems face the challenge of reducing latency to preserve patient safety, become more efficient, and limit costs. By streamlining diagnosis, successfully diagnosticians can gain credibility with the patients, and the risk of misdiagnosis-induced errors can be mitigated.

The deployed architecture can therefore be orchestrated and configured as a distributed real-time federated learning system for rare diseases in multisite settings. Several hospitals can join the local real-time diagnosis effort by implementing either the complete federated environment or standalone platforms offering only the FedAI service. Users wishing to request the diagnosis of an unconfirmed rare disease carry out the same steps as previously described and receive the real-time result without any extra cost. When such an user submits the request, she is redirected to the local Fed-AI server, where the diagnosis is processed using either the locally stored model or the local model combined with the available cross-institutional knowledge.

6.2. Scalability to Diverse Rare Diseases

Research and development on rare diseases remain scattered and limited to few medical and healthcare institutions. The scope of the proposed methods for real-time rare disease diagnosis through federated learning is therefore examined—specifically, whether the proposed solution can be utilized for epidemic or endemic rare diseases allocated to different geographical regions. Rare diseases that have a cross-effect, as well as medical or clinical symptoms that are associated with non-rare diseases (e.g., COVID-19), are also considered. Two potentials for a more general and scalable utilization are envisaged.

First, different sites can rapidly create a factory of rare disease diagnostic models that are orchestrated, motorcycles of rapid inference model (even retranslation by a local engine), and hotspots of latency and real-time inference deployment—leading to slightly more (or less) than real-time response while still complying with clinical patient-centered SLAs). Second, the federated solution can be associated with a medical image analysis over middleware concept, where proxies or validation services act like simple containers of light federated services within a local cloud, validating the adopted models in real life and proposing local medical or clinical experts' reports without heavy medical or clinical chains. Overall, more general and scalable utilization is feasible to a certain degree, with the wall or the barrier being the disease.

7. CONCLUSION

Federated Learning and Cloud-Based Artificial Intelligence for Real-Time Diagnosis of Rare Diseases in Healthcare Systems

A summary of findings, limitations, policy implications, and directions for future research is offered. Despite their low incidence, rare diseases affect over 350 million people worldwide and account for 10–20% of hospitalizations in high-income countries. Early diagnosis can improve patient quality of life, yet many cases remain undiagnosed. Federated learning offers an avenue to close the data gap by enabling artificial-intelligence-based diagnostic support without sharing sensitive patient information. However, most proposed solutions lack practical validation and fail to consider the comprehensive infrastructure required for real-time diagnosis.

Real-time disease diagnosis via federated learning using multisite patient cohort orchestration is illustrated. The approach takes advantage of a cloud-based artificial-intelligence framework that conforms to the federated-learning paradigm. Privacy-preserving governance structures manage the flow of data and related sensitive patient information, while model training, aggregation, and update mechanisms ensure that epidemic-timing controls are respected. Latency to diagnosis is minimized by supporting a regional-access delegation service.

Key elements of the solution, including multisite deployment scenarios, diagnosis-latency orchestration, real-time inference workflows, and cross-disease transfer considerations, are discussed. The proposed approach highlights the potential of cloud-based artificial-intelligence solutions to harness multimodal diagnostic data sets during epidemics and other time-critical scenarios. The method supports not only rare diseases but also other conditions requiring diagnosis across geographically dispersed sites. Future research should refine policy frameworks and optimize algorithmic capabilities for disease diagnosis across the spectrum of disorder incidence.

DOI: 10.17148/IJIREEICE.2023.111218

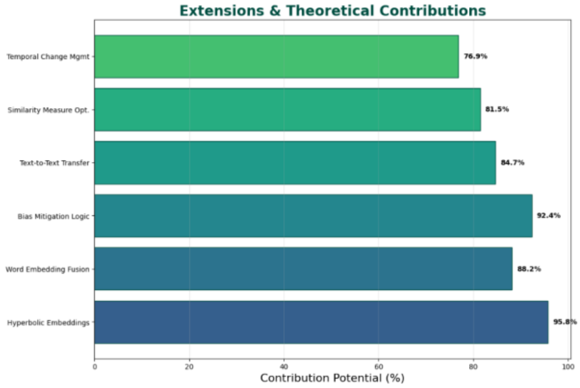


Fig 4: Infrastructure & Governance Readiness

7.1. Summary of Findings and Future Directions

Real-time and accurate diagnosis of rare diseases remain critical challenges in healthcare, underlined by the low prevalence yet high aggregate impact of these conditions. Large volumes of meticulously collected clinical data exist but not at sufficient scale, quality or density for individual institutions to offer reliable diagnostic support. A novel, cloud-based diagnostic-support system is proposed using federated learning principles to realize collaborative learning across datasets and societies while sensitive data remain within local control. Collaboration among judges and data providers is firmly governed to ensure authoritative permission, minimization, anonymization and compliance with regulations such as GDPR, HIPAA, LaVIR. A federated architecture is designed to enable the proposed approach and its features documented; components include a local/shared Cloud agent, data-preparation/inference blocks and an intranet for communication incompatible with Internet use. Diagnostic support is illustrated for a multisite deployment in neuro-muscular diseases, orchestrating an ensemble of dedicated classifiers in sub-seconds under real-time conditions. Expressive potential across rare conditions and transfer from one disease to another with scarce data have been verified; the solution can be generalized to other rare diseases and domains of application.

Fostering shared intelligence across distributed healthcare data sources using cloud-based AI offers extraordinary potential for real-time closed-loop support yet exposes substantial socio-ethical concerns. The proposed diagnostic-support solution bridges patient safety with advanced security/governance provisions for real-world operational validation, fostering acceptance and effective wider deployment. These principles promote both patient safety and clinician welfare, and can be extended to other safety-critical contexts (e.g. aviation) through real-time closure of critical decision loops. Completing the second pillar, a cloud-based federated-learning methodology enables identity-agnostic sharing of knowledge across disparate data holdings scattered in diverse jurisdiction and ownership/control domains, preserving privacy and complying with regulations such as GDPR. Speed and quality of rare-disease diagnosis is promoted through shared-learning boost, even among sites with insufficient data for quality-assured diagnosis. The federated scheme is further specialized in support of a cloud-based diagnostic solution explicitly designed to operate at sub-second latencies, allowing incorporation into real-time workflows.

REFERENCES

[1]. World Health Organization. Early detection, assessment and response to acute public health events
[2]. Centers for Disease Control and Prevention. Surveillance strategies for emerging infectious diseases
[3]. Amistapuram, K. (2022). Fraud Detection and Risk Modeling in Insurance: Early Adoption of Machine Learning in Claims Processing. Available at SSRN 5741982
[4]. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection using online data
[5]. Salathé M, Bengtsson L, Bodnar TJ, et al. Digital epidemiology
[6]. Kiran Reddy Burugulla, J. (2023). Transforming Payment Systems Through AI And ML: A Cloud-Native Approach. Educational Administration: Theory and Practice. https://doi.org/10.53555/kuey.v29i4.10144
[7]. Yang W, Lipsitch M, Shaman J. Inference of seasonal and pandemic influenza transmission dynamics
[8]. Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data
[9]. Rongali, S. K. (2023). Explainable Artificial Intelligence (XAI) Framework for Transparent Clinical Decision Support Systems. International Journal of Medical Toxicology and Legal Medicine, 26(3), 22-31
[10]. Paul MJ, Dredze M. Social media for health surveillance
[11]. Garapati, R. S. (2023). Optimizing Energy Consumption in Smart Build-ings Through Web-Integrated AI and Cloud-Driven Control Systems
[12]. Shaman J, Karspeck A. Forecasting seasonal outbreaks of influenza

DOI: 10.17148/IJIREEICE.2023.111218

[13]. Rongali, S. K. (2021). Cloud-Native API-Led Integration Using MuleSoft and .NET for Scalable Healthcare Interoperability. Available at SSRN 5814563
[14]. Ribeiro M, Silva R, Prata D. Machine learning techniques for disease outbreak prediction
[15]. Guntupalli, R. (2023). Optimizing Cloud Infrastructure Performance Using AI: Intelligent Resource Allocation and Predictive Maintenance. Available at SSRN 5329154
[16]. Gao H, Barbier G, Goolsby R. Harnessing the crowdsourcing power of social media for disaster relief
[17]. Varri, D. B. S. (2022). AI-Driven Risk Assessment And Compliance Automation In Multi-Cloud Environments. Available at SSRN 5774924
[18]. Wang L, Wang G, Alexander CA. Big data and predictive analytics in healthcare
[19]. Challa, K. Dynamic Neural Network Architectures for Real-Time Fraud Detection in Digital Payment Systems Using Machine Learning and Generative AI
[20]. Nsoesie EO, Kluberg SA, Brownstein JS. Online reports of disease outbreaks
[21]. Chen M, Mao S, Liu Y. Big data: A survey
[22]. Kummari, D. N., & Burugulla, J. K. R. (2023). Decision Support Systems for Government Auditing: The Role of AI in Ensuring Transparency and Compliance. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 493-532
[23]. Bishop CM. Pattern recognition and machine learning
[24]. Gottimukkala, V. R. R. (2022). Licensing Innovation in the Financial Messaging Ecosystem: Business Models and Global Compliance Impact. International Journal of Scientific Research and Modern Technology, 1(12), 177-186
[25]. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning
[26]. LeCun Y, Bengio Y, Hinton G. Deep learning
[27]. Uday Surendra Yandamuri. (2023). An Intelligent Analytics Framework Combining Big Data and Machine Learning for Business Forecasting. International Journal Of Finance, 36(6), 682-706. https://doi.org/10.5281/zenodo.18095256
[28]. Vapnik V. The nature of statistical learning theory
[29]. Meda, R. (2023). Data Engineering Architectures for Scalable AI in Paint Manufacturing Operations. European Data Science Journal (EDSJ) p-ISSN 3050-9572 en e-ISSN 3050-9580, 1(1)
[30]. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A survey of deep learning techniques for electronic health records
[31]. Uday Surendra Yandamuri. (2022). Cloud-Based Data Integration Architectures for Scalable Enterprise Analytics. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), 472–483. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/8005
[32]. Zhao S, Musa SS, Lin Q, et al. Estimating the unreported number of infections
[33]. Garapati, R. S. (2022). AI-Augmented Virtual Health Assistant: A Web-Based Solution for Personalized Medication Management and Patient Engagement. Available at SSRN 5639650
[34]. Zhang Q, Wang D, Zhang M. Forecasting infectious disease outbreaks using machine learning
[35]. Inala, R. Revolutionizing Customer Master Data in Insurance Technology Platforms: An AI and MDM Architecture Perspective
[36]. Wagner MM, Moore AW, Aryel RM. Handbook of biosurveillance
[37]. Varri, D. B. S. (2023). Advanced Threat Intelligence Modeling for Proactive Cyber Defense Systems. Available at SSRN 5774926
[38]. Pollett S, Johansson MA, Reich NG. Big data and predictive analytics in epidemic preparedness
[39]. Segireddy, A. R. (2021). Containerization and Microservices in Payment Systems: A Study of Kubernetes and Docker in Financial Applications. Universal Journal of Business and Management, 1(1), 1–17. Retrieved from https://www.scipublications.com/journal/index.php/ujbm/article/view/1352
[40]. Santillana M, Zhang DW, Althouse BM, Ayers JW. What can digital disease detection learn from Google Flu Trends
[41]. Cowling BJ, Leung GM. Epidemiological research priorities for public health control
[42]. Chan EH, Sahai V, Conrad C, Brownstein JS. Using web search query data to monitor disease outbreaks
[43]. Olson DR, Konty KJ, Paladini M, et al. Reassessing Google Flu Trends data
[44]. Lu FS, Hattab MW, Clemente CL, et al. Improved state-level influenza forecasting
[45]. Amistapuram, K. (2021). Digital Transformation in Insurance: Migrating Enterprise Policy Systems to .NET Core. Universal Journal of Computer Sciences and Communications, 1(1), 1–17. Retrieved from https://www.scipublications.com/journal/index.php/ujcsc/article/view/1348
[46]. Yang S, Santillana M, Kou SC. Accurate estimation of influenza epidemics using search data
[47]. Reyné B, Castillo-Salgado C. Trends in syndromic surveillance

DOI: 10.17148/IJIREEICE.2023.111218

[48]. Gottimukkala, V. R. R. (2021). Digital Signal Processing Challenges in Financial Messaging Systems: Case Studies in High-Volume SWIFT Flows
[49]. Kang M, Zhong H, He J, Rutherford S, Yang F. Using Google Trends for influenza surveillance
[50]. Avinash Reddy Aitha. (2022). Deep Neural Networks for Property Risk Prediction Leveraging Aerial and Satellite Imaging. International Journal of Communication Networks and Information Security (IJCNIS), 14(3), 1308–1318. Retrieved from https://www.ijcnis.org/index.php/ijcnis/article/view/8609
[51]. Olson DR, Simonsen L, Edelson PJ, Morse SS. Epidemiological implications of internet-based surveillance
[52]. Kalisetty, S. (2023). Harnessing Big Data and Deep Learning for Real-Time Demand Forecasting in Retail: A Scalable AI-Driven Approach. American Online Journal of Science and Engineering (AOJSE)(ISSN: 3067-1140), 1(1).
[53]. Morse SS. Factors in the emergence of infectious diseases
[54]. Goutham Kumar Sheelam, Hara Krishna Reddy Koppolu. (2022). Data Engineering And Analytics For 5G-Driven Customer Experience In Telecom, Media, And Healthcare. Migration Letters, 19(S2), 1920–1944. Retrieved from https://migrationletters.com/index.php/ml/article/view/11938
[55]. Wolfe ND, Dunavan CP, Diamond J. Origins of major human infectious diseases
[56]. Unifying Data Engineering and Machine Learning Pipelines: An Enterprise Roadmap to Automated Model Deployment. (2023). American Online Journal of Science and Engineering (AOJSE) (ISSN: 3067-1140) , 1(1). https://aojse.com/index.php/aojse/article/view/19
[57]. Deo RC. Machine learning in medicine
[58]. Garapati, R. S. (2022). Web-Centric Cloud Framework for Real-Time Monitoring and Risk Prediction in Clinical Trials Using Machine Learning. Current Research in Public Health, 2, 1346
[59]. Ristevski B, Chen M. Big data analytics in medicine and healthcare
[60]. Segireddy, A. R. (2020). Cloud Migration Strategies for High-Volume Financial Messaging Systems
[61]. Eysenbach G. Infodemiology and infoveillance
[62]. Meda, R. (2023). Developing AI-Powered Virtual Color Consultation Tools for Retail and Professional Customers. Journal for ReAttach Therapy and Developmental Diversities. https://doi. org/10.53555/jrtdd. v6i10s (2), 3577
[63]. Chen J, Zhang Y, Zhang M. Neural network models for epidemic prediction
[64]. Meda, R. (2023). Intelligent Infrastructure for Real-Time Inventory and Logistics in Retail Supply Chains. Educational Administration: Theory and Practice
[65]. Huang Z, Ling CX. Using AI to detect emerging infectious diseases
[66]. AI Powered Fraud Detection Systems: Enhancing Risk Assessment in the Insurance Sector. (2023). American Journal of Analytics and Artificial Intelligence (ajaai) With ISSN 3067-283X, 1(1). https://ajaai.com/index.php/ajaai/article/view/14
[67]. Keller M, Blench M, Tolentino H, et al. Use of unstructured event-based reports for global infectious disease surveillance
[68]. Ramesh Inala. (2023). Big Data Architectures for Modernizing Customer Master Systems in Group Insurance and Retirement Planning. Educational Administration: Theory and Practice, 29(4), 5493–5505. https://doi.org/10.53555/kuey.v29i4.10424
[69]. Mandl KD, Overhage JM, Wagner MM, et al. Implementing syndromic surveillance
[70]. Kummari, D. N. (2023). Energy Consumption Optimization in Smart Factories Using AI-Based Analytics: Evidence from Automotive Plants. Journal for Reattach Therapy and Development Diversities. https://doi. org/10.53555/jrtdd. v6i10s (2), 3572
[71]. Zou Q, Qu K, Luo Y, et al. Predicting disease outbreaks with machine learning
[72]. Varri, D. B. S. (2022). A Framework for Cloud-Integrated Database Hardening in Hybrid AWS-Azure Environments: Security Posture Automation Through Wiz-Driven Insights. International Journal of Scientific Research and Modern Technology, 1(12), 216-226
[73]. Xu Q, Gel YR, Ramirez Ramirez LL, et al. Forecasting influenza in the US using machine learning
[74]. Held L, Meyer S, Bracher J. Probabilistic forecasting in infectious disease epidemiology
[75]. Yandamuri, U. S. (2022). Big Data Pipelines for Cross-Domain Decision Support: A Cloud-Centric Approach. International Journal of Scientific Research and Modern Technology, 1(12), 227–237. https://doi.org/10.38124/ijsrmt.v1i12.1111
[76]. Grassly NC, Fraser C. Mathematical models of infectious disease transmission
[77]. Guntupalli, R. (2023). AI-Driven Threat Detection and Mitigation in Cloud Infrastructure: Enhancing Security through Machine Learning and Anomaly Detection. Available at SSRN 5329158
[78]. Daley DJ, Gani J. Epidemic modelling: an introduction
[79]. Rongali, S. K. (2022). AI-Driven Automation in Healthcare Claims and EHR Processing Using MuleSoft and Machine Learning Pipelines. Available at SSRN 5763022

DOI: 10.17148/IJIREEICE.2023.111218

[80]. Riley S. Large-scale spatial-transmission models of infectious disease
[81]. Avinash Reddy Segireddy. (2022). Terraform and Ansible in Building Resilient Cloud-Native Payment Architectures. Keerthi Amistapuram. (2023). Privacy-Preserving Machine Learning Models for Sensitive Customer Data in Insurance Systems. Educational Administration: Theory and Practice, 29(4), 5950–5958. https://doi.org/10.53555/kuey.v29i4.10965
[82]. Chen Y, Li J, Xiao L. Big data driven disease surveillance
[83]. Uday Surendra Yandamuri. (2022). Cloud-Based Data Integration Architectures for Scalable Enterprise Analytics. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), 472–483. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/8005
[84]. Wang H, Ding Y, Tang J, et al. Identifying outbreaks using big data analytics
[85]. Inala, R. AI-Powered Investment Decision Support Systems: Building Smart Data Products with Embedded Governance Controls
[86]. Ribeiro MT, Singh S, Guestrin C. Model-agnostic interpretability of machine learning
[87]. Gottimukkala, V. R. R. (2023). Privacy-Preserving Machine Learning Models for Transaction Monitoring in Global Banking Networks. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 633-652
[88]. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence
[89]. Aitha, A. R. (2023). CloudBased Microservices Architecture for Seamless Insurance Policy Administration. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 607-632
[90]. Beam AL, Kohane IS. Big data and machine learning in health care
[91]. Kushvanth Chowdary Nagabhyru. (2023). Accelerating Digital Transformation with AI Driven Data Engineering: Industry Case Studies from Cloud and IoT Domains. Educational Administration: Theory and Practice, 29(4), 5898–5910. https://doi.org/10.53555/kuey.v29i4.10932