This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2024.3467179

A Global Cybersecurity Standardization Framework for Healthcare Informatics

Kishu Gupta, *Member, IEEE,* Vinaytosh Mishra, and Aaisha Makkar, *Member, IEEE*

***Abstract*— Healthcare has witnessed an increased digitalization in the post-COVID world. Technologies such as the medical internet of things and wearable devices are generating a plethora of data available on the cloud anytime from anywhere. This data can be analyzed using advanced artificial intelligence techniques for diagnosis, prognosis, or even treatment of disease. This advancement comes with a major risk to protecting and securing protected health information (PHI). The prevailing regulations for preserving PHI are neither comprehensive nor easy to implement. The study first identifies twenty activities crucial for privacy and security, then categorizes them into five homogeneous categories namely: C_1 (Policy and Compliance Management), C_2 (Employee Training and Awareness), C_3 (Data Protection and Privacy Control), C_4 (Monitoring and Response), and C_5 (Technology and Infrastructure Security) and prioritizes these categories to provide a framework for the implementation of privacy and security in a wise manner. The framework utilized the Delphi Method to identify activities, criteria for categorization, and prioritization. Categorization is based on the Density-Based Spatial Clustering of Applications with Noise (DBSCAN), and prioritization is performed using a Technique for Order of Preference by Similarity to the Ideal Solution (TOPSIS). The outcomes conclude that C_3 activities should be given first preference in implementation and followed by C_1 and C_2 activities. Finally, C_4 and C_5 should be implemented. The prioritized view of identified clustered healthcare activities related to security and privacy, are useful for healthcare policymakers and healthcare informatics professionals.**

***Index Terms*— Clustering, Healthcare Security, Medical Standards, Privacy, Prioritization, Security.**

I. Introduction

HEALTHCARE has witnessed rapid digitalization in recent years [1]. Industry 4.0 and its main enabling information and communication technologies are completely changing services and production. This is especially true for the health domain, where the Internet of Things, cloud, and big data technologies are revolutionizing eHealth and its whole ecosystem, moving it towards [2], [3]. Technologies such as Artificial Intelligence (AI), Internet of Things (IoT), and Cloud Computing are transforming the way healthcare has been delivered in the recent past [4], [5]. AI enables systems to augment medical staff in each aspect of care, including diagnosis, prognosis, and treatment. These technologies impact the efficiency of nursing and the managerial activities of hospitals [2], [6], [7]. Healthcare is one of the most prominent fields utilizing IoT. This technology enables medical practitioners and hospital staff to perform their duties efficiently and intelligently. With the latest advanced technologies, most of the challenges of using IoT have been resolved, and this technology can be a great revolution and has many benefits in the future of digital healthcare [2], [8], [9]. Similarly, cloud computing has transformed the traditional way of healthcare delivery [10]. It offers numerous healthcare advantages, significantly impacting how healthcare data is managed, accessed, and leveraged for patient care [2]. Cloud-based Electronic Health Records (EHRs) are key enablers of digital health. It facilitates the digital retrieval of patient records and the extraction of clinical information. Consequently, various additional uses of this technology have become available, including quality management, healthcare administration, and translational research [11].

Chenthara et al. discuss the risks to privacy and security in cloud-based healthcare records, suggesting mitigation strategies [12]. They suggest pseudonymizing Electronic Health Record (EHR) data to safeguard Personal Health Information (PHI) privacy and security. Recent research, including multiple studies, points to blockchain technology as a promising solution for privacy and security issues in EHRs [13], [14]. Tang et al. describe an efficient blockchain-based scheme [15], and Guo et al. offer a secure, attribute-based signature scheme for blockchain in EHRs involving multiple authorities [16]. Al Mamun et al. thoroughly examine blockchain's application in EHRs and outline areas for future research [17].

Ensuring privacy and security necessitates a comprehensive overhaul of infrastructure, processes, and practices within healthcare organizations, alongside allocating necessary resources for effective implementation. Moreover, prioritizing these efforts helps implement these measures in resource-constrained settings. Based on the review of the extant literature, we observe there is a lack of a comprehensive healthcare-specific privacy and security framework. While HIPAA is widely accepted, it falls short in addressing challenges posed by emerging technologies like tele-health, AI, and cross-border data sharing. A high need emerges to provide a technical solution that is sufficient; considering behavioral and awareness factors are equally crucial in managing PHI's privacy and

This work is supported by National Sun Yat-sen University, Kaohsiung, Taiwan; Thumbay Institute for AI in Healthcare, Gulf Medical University, Ajman, UAE; University of Derby, UK. (Corresponding author: Vinaytosh Mishra.)

Kishu Gupta is with the Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, 80424, Taiwan (e-mail: kishuguptares@gmail.com).

Vinaytosh Mishra is with the Thumbay Institute for AI in Healthcare, Gulf Medical University, Ajman, United Arab Emirates (e-mail: dr.vinaytosh@gmu.ac.ae).

Aaisha Makkar is with the College of Science & Engineering, University of Derby, UK (e-mail: aaisha.makkar@ieee.org).

Authorized licensed use limited to: UNIV OF DERBY. Downloaded on December 24,2024 at 13:48:54 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2024.3467179

security in healthcare.

Paper Contributions and Outline: This study aims to develop a tailored framework to enhance data protection in modern healthcare systems by proposing a **G**lobal **C**yber-**S**ecurity **S**tandardization Framework for **H**ealthcare **I**nformatics (GCS-HI). The key contributions of the outlined model are as follows:

1) The study proposes a novel, three-fold structured GCS-HI framework to enhance data privacy and security in the healthcare system. It strategically identifies, categorizes, and prioritizes the activities crucial towards privacy and security using brain-storming for identification, clustering for categorization, and the multi-criteria decision-making (MCDM) approach for prioritization.
2) The framework offers a systematic method for healthcare organizations to adopt comprehensive security measures in a post-COVID digitalized environment.
3) The experimental work and extensive comparison with state-of-the-art approaches highlight the importance and efficiency of the proposed GCS-HI framework.

The roadmap of the paper is as follows: Section II outlines the related literature for the study. Section III describes the proposed model framework. Followed by Section V that discusses the results. Finally, Section VI concludes with a summary, research implications, and future direction for research.

II. Related Work

The terms ”privacy,” ”confidentiality,” and ”security” in the context of healthcare are interrelated but distinct concepts. The privacy means that a patient's personal health information is only accessible to the patient and those authorized by the patient [18], [19]. Meanwhile, confidentiality obligates healthcare providers to protect patient information and disclose it only with the patient's consent or under legally permissible circumstances [19], [20]. Security refers to the measures, protocols, and procedures that protect personal or sensitive information from unauthorized access, disclosure, alteration, and destruction [18]. The literature study is divided into two major aspects; first, the threats in healthcare related to privacy and security, and second, the widely accepted healthcare standards standards in practice.

A. Privacy and Security Threats

Due to the sensitive nature of data, privacy and security threats are a major concern in healthcare. Some of the major threats reported in the extant literature are:

1) Data Breaches and Cyberattacks: Cybercriminals attack health information systems to steal PHI, which is in high demand in the market [21]. One of the common methods is an attack of ransomware, where attackers encrypt data and demand payment for its release [22].

2) Insider Threats: Healthcare staff can misuse their access to PHI, which can result in privacy breaches. They can do it either maliciously or accidentally. This can include viewing patient records without a legitimate reason or inadvertently disclosing information [23]. Prabhu & Thompson propose a unified classification model of insider threats to information security [24].

3) Phishing and Social Engineering Attacks: Phishing attacks intend to trick healthcare employees into revealing sensitive information, such as login credentials. It is a major privacy invasion in which an attacker poses as a legitimate entity to gain access [25]. The approach, such as social engineering, is popular among hackers with malicious intentions [26].

4) Inadequate Security Measures and Policies: Healthcare providers can fail to implement sufficient security measures such as data encryption or access controls [27]. Lack of planned security audits and training can expose an organization to privacy and confidentiality risks [28].

5) Mobile Device Security: Ubiquitous personal devices and hospital EHRs on these systems expose a health system to privacy and security risks [29]. The increasing use of mobile devices in healthcare, such as tablets and smartphones, can create security vulnerabilities, especially if these devices are lost, stolen, or used over unsecured networks [30].

6) Unsecured IoT Devices: Medical IoT devices are designed foremost for usability, but with this simplicity of design, most fail to support encryption [31]. This means that whenever a medical IoT device is used to connect with a hospital network or healthcare database, there is a risk of interception or infiltration [32].

7) Lack of Patient Awareness: Actions like sharing personal health information on unsecured platforms or falling for scams result in the breach of privacy and security of patients [33]. A need for patient education program is required to mitigate the risk of security breaches originating because of the patient's irresponsible behavior [20].

B. Privacy and Security Standards and Law

The most widely used laws for ensuring privacy and security include the Health Insurance Portability and Accountability Act (HIPAA) from the US and the General Data Protection Regulation (GDPR) from the European Union [34]. Other major laws include the Personal Information Protection and Electronic Documents Act (PIPEDA) of Canada, the Health Records and Information Privacy Act (HRIPA) of Australia, the Data Protection Act (DPA) of the UK, and Digital Personal Data Protection Act (DPDPA) of India. Most of these laws are implementations of GDPR with provisions of HIPAA in healthcare contexts. These laws have evolved and have been mostly reactive towards the privacy and security threats [35]. Moreover, the continually evolving state-of-the-art techniques in Machine Learning (ML), Data Analytics (DA), and hacking are making it even more difficult to protect a patient's privacy absolutely [36].

III. Proposed Framework

The GCS-HI framework considers the condition of various worldwide healthcare standards ($\hbar$) such as: $\{$HIPAA, GDPR, PIPEDA, HRIPA, DPDPA, GDPR$\}$ comprising considerable activities: $\{\Lambda_1, \Lambda_2, \ldots, \Lambda_n\} \in \Lambda$ crucial towards privacy and security of healthcare data as illustrated in Fig. 1. A lot of healthcare privacy and security-related activities (Λ) are common in all $\hbar$ while each ($\hbar$) has a few unique activities as well. The proposed framework identifies the most pivotal Λs among

Authorized licensed use limited to: UNIV OF DERBY. Downloaded on December 24,2024 at 13:48:54 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2024.3467179

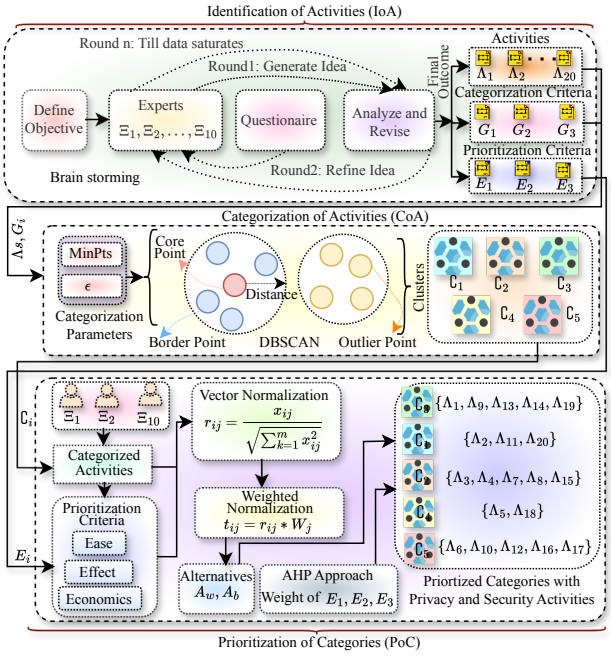

Fig. 1: GCS-HI schematic overview.

these and the categorization criteria $\{G_1, G_2, G_3\}$ for the Λs by performing brainstorming to utilize the expertise of a focus group of ten experts: $\{\Xi_1, \Xi_2, \ldots, \Xi_{10}\} \in \Xi$ having different educational backgrounds with diverse professions. Further, the categorization parameters such as ϵ and MinPts are employed to categorize Λs into different categories: $\{\complement_1, \complement_2, \ldots, \complement_5\} \in \complement$. These Ξs prioritizes the $\complement$s by utilizing the three identified prioritization criteria: $\{E_1, E_2, E_3\}$ means the {Ease, Effect, Economics} on a weighted scale of [1,10] for each E. The detailed description regarding the identification of activities (IoA), categorization of activities (CoA), and prioritization of categories (PoC) are described in the following subsections: III-A, III-B and, III-C, respectively.

A. Identification of Activities (IoA)

To identify the key activities (Λs), the proposed framework utilizes the brainstorming-based-delphi approach. The existing literature recommends a considerable size of eight to sixteen experts (Ξ). It selects the participants Ξ using purposive sampling and thus considers a focus group of ten Ξs to perform the brainstorming activities. The delphi study involved three iterative rounds: identifying activities, establishing categorization criteria, and prioritizing them. In cases of indecision, a simple majority was used to reach a decision. This process ensured a thorough, expert-driven consensus, producing a well-structured and prioritized set of activities and criteria. Table I elaborate the details of considered focus group of Ξs

The brainstorming process as illustrated in Fig. 1 comprises major steps like defining the objective, selecting of a focus group of Ξs, multiple rounds of analysis till data saturates, and lastly utilizing these results for identification of Λ_i : $\{\Lambda_1, \Lambda_2, \ldots, \Lambda_n\}$, $G_i : \{G_1, G_2, G_3\}$, and $E_i : \{E_1, E_2, E_3\}$.

TABLE I: Characteristics of focus group of experts (Ξ)

Expert (Ξ)	Profession	Education	Experience (years)
Ξ_1	SE	B.Tech	5
Ξ_2	SE	B.Tech	6
Ξ_3	SE	M.Tech	5
Ξ_4	SE	MCA	6
Ξ_5	WA	MCA	8
Ξ_6	DA	B.Tech	6
Ξ_7	HM	B.Tech	6
Ξ_8	HM	MBA	8
Ξ_9	HM	MBA	9
Ξ_{10}	HM	MBA	8

SE: Software Engineer; WA: Web Analyst; DA: Data Analyst; HM: Healthcare Manager; B.Tech: Bachelor of Technology; M.Tech: Master of Technology; MCA: Master of Computer Applications; MBA: Master of Business Administration

Before initiating the brainstorming session, a focus group of participants (Ξs) are briefed on diverse global medical standards ($\hbar$) such as HIPAA and GDPR. During the refinement phase, any recurring themes were identified and eliminated. A voting method is conducted in case a clear decision can not be made. The iteration required for the identification of Λ_i, G_i, and E_i were 5, 3, and 2, respectively.

B. Categorization of Activities (CoA)

To categorize the identified activities $\{\Lambda_1, \Lambda_2, \ldots, \Lambda_n\}$ into effectively manageable groups, the proposed GCS-HI framework deployed a density-based spatial clustering of applications with noise (DBSCAN) clustering algorithm because of its capability in handling the noise, cluster shape flexibility, proficiency in automatically detecting clusters, resilience against outliers, capability to manage clusters of irregular shapes, effectiveness in identifying noise, adaptability to different data types and dimensions, efficiency, scalability, and ease of interpretation. The clustering process involved two key parameters: ϵ and MinPts, where ϵ is the distance threshold that defines the neighborhood around a data point. At the same time, MinPts is the minimum number of points required to form a dense region. A typical depiction of the clustering approach is given in Fig. 2.

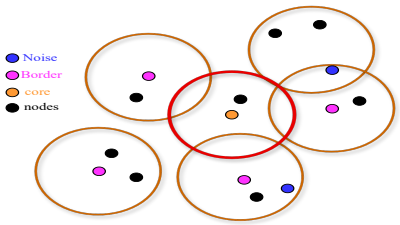

Fig. 2: Illustration of density based categorization approach.

It assumes, D is the dataset containing n points: $D = \{p_1, p_2, ..., p_n\}$. Consider all points P_i in the dataset D are visited. All the border points in ϵ-Neighborhood of point P_i ($P_i : N_\epsilon(P_i)$) are computed by using Eq. (1) and all P_i are added to the new cluster $\complement_k$. Further, the $\complement_k$ is expanded for un-visited points P_j in $N_\epsilon(P_i)$ and to mark P_j as visited. Now, to find all points in ϵ-Neighborhood of point P_j by using Eq.

Authorized licensed use limited to: UNIV OF DERBY. Downloaded on December 24,2024 at 13:48:54 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2024.3467179

(2) and these points are added to $N_\epsilon(P_i)$ for further processing and ifP_j is not yet a member of any cluster add P_j to $\complement_k$.

$$|N_\epsilon(P_i)| < MinPts \tag{1}$$

$$N_\epsilon(P_j) if |N_\epsilon(P_j)| \geq MinPt \tag{2}$$

For any possible clustering $\{\complement_1, \complement_2, \ldots, \complement_5\}$ out of all clusters $\complement$, it minimizes the number of clusters under the condition that every pair of points in a cluster is densely reachable, which corresponds to two main properties, connectivity, and maximality of the cluster. The main objective is to optimize the loss function as computed using Eq. (3).

$$\begin{aligned} &\min |\complement| \\ &\complement \subset \complement \\ &d_{dc}(p,q) \leq \epsilon, \forall\, p,q \in \complement_i \forall \complement_i \in \complement \end{aligned} \tag{3}$$

wherein, $d_{db}(p,q)$, gives the smallest ϵ such that two points p and q are density-connected.

The distance function is computed using Eq. (4) denotes the distance between two points, p and q in m-dimensional space. The neighborhood function is computed using Eq. (5) denotes ϵ-neighborhood of point p. Algorithm 1 presents a summarized form of the approach used for categorization.

$$dist(p,q) = \sqrt[2]{\sum_{i=1}^{m}(p_i - q_i)^2} \tag{4}$$

$$N_\epsilon(p) = \{q \in D \mid dist(p,q) \leq \epsilon\} \tag{5}$$

C. Prioritization of Categories (PoC)

Followed by the classification of the most admissible categories ($\complement$)s, the prioritization of these categories is performed. To perform the prioritization a multi-criteria decision-making approach called the Technique for Order of Preference by Similarity to the Ideal Solution (TOPSIS) is utilized. TOPSIS is used for multi-criteria decision-making because it provides a clear, systematic method for ranking alternatives based on their relative closeness to an ideal solution. TOPSIS is a robust and reliable tool in scenarios requiring the careful balancing of various factors. The given decision data is normalized by deploying the vector normalization approach using Eq. (6). The normalization matrix represents the normalization of evaluation matrix comprising n alternatives activities (Λ) and m criteria ($\mathfrak{C}$) such that $(x_{ij})_{m*n}$.

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{k=1}^{m} x_{ij}^2}} \tag{6}$$

here, $i = 1, 2, ..., m$ and $j = 1, 2, ..., n$. The weighted normalized decision matrix (t_{ij}) is computed using Eq. (7) and the weight of criteria for prioritization is computed using the analytical hierarchy process (AHP) Approach.

$$t_{ij} = r_{ij} * W_j \tag{7}$$

Further, the worst (A_w) and best alternatives (A_b) are computed using Eq. (8). Moreover, Eqs. (9) and (10), computes the distance between target i and worst or bad condition, respectively.

Algorithm 1 CoA: Summary for Categorization.

```
Label all points as un-visited
for each point P_i in the dataset D do
  if P_i is visited then
    Skip to next point
  end if
  Mark P_i as visited
  find all points in ε-Neighborhood of point P_i. Such that P_i : N_ε(P_i)
  if |N_ε(P_i)| < MinPts then
    Mark P_i as noise (later it may be classified as border point)
  else
    create a new cluster, C_k and add P_i to C_k
  end if
  Expand Cluster C_K
  for each point P_j in N_ε(P_i) do
    if P_j is unvisited then
      Mark P_j as visited
    end if
    find all points in ε-Neighborhood of point P_j. Such that P_j : N_ε(P_j)
    if |N_ε(P_j)| ≥ MinPts then
      add these points to N_ε(P_i) for further processing (expand the neighborhood)
    end if
  end for
  if P_j is not yet a member of any cluster then
    add P_j to C_k
  end if
end for
```

$$\begin{aligned} A_w = \Big\{ &\langle max(t_{ij} \Big| i = 1,2,...,m)\ j \epsilon J_- \rangle, \\ &\langle min(t_{ij} \Big| i = 1,2,...,m)\ j \epsilon J_+ \rangle \Big\} \\ &\equiv \{t_{wj} | j = 1,2,...,n\} \\ A_b = \Big\{ &\langle min(t_{ij} \Big| i = 1,2,...,m)\ j \epsilon J_- \rangle, \\ &\langle max(t_{ij} \Big| i = 1,2,...,m)\ j \epsilon J_+ \rangle \Big\} \\ &\equiv \{t_{bj} | j = 1,2,...,n\} \end{aligned} \tag{8}$$

wherein, $J_+ = \{j = 1, 2, ..., n | j\}$ and $J_- = \{j = 1, 2, ..., n | j\}$ are associated with a criteria having positive impact and negative impact, respectively.

$$d_{iw} = \sqrt{\sum_{j=1}^{n}(t_{ij} - t_{wj})^2} \tag{9}$$

$$d_{ib} = \sqrt{\sum_{j=1}^{n}(t_{ij} - t_{bj})^2} \tag{10}$$

here, $i = 1, 2, ..., m$.

Authorized licensed use limited to: UNIV OF DERBY. Downloaded on December 24,2024 at 13:48:54 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2024.3467179

The effectiveness of prioritization depends on the proper identification of criteria and accurate weighting. The steps used for the prioritization process are demonstrated in Algorithm 2. The prioritization using TOPSIS has several advantages over other multi-criteria decision-making such as simplicity and understand ability, balance between ideal and negative ideal solutions, clear ranking of alternatives, and compatibility with other methods.

Algorithm 2 PoC: Summary for Prioritization.

Start
Create an evaluation matrix consisting of n activities and n criteria $(x_{ij})_{m*n}$
The matrix $(x_{ij})_{m*n}$ is normalized to obtain $R = (r_{ij})_{m*n}$ as shown in Eq. (6)
Calculate the weighted normalized decision matrix (t_{ij}) using Eq. (7)
Determine the worst alternative A_w and best alternative A_b using Eq. (8)
Calculate the distance between target alternative i and worst condition A_w using Eq. (9)
Similarly, calculate the distance between target alternative i and best condition A_b using Eq. (10)
End

IV. Operational Design and Complexity

The overall summary of the proposed GCS-HI framework is illustrated in Algorithm 3. The study used a naive implementation for the categorization of activities without any indexing or optimized search for the neighborhood.

Algorithm 3 GCS-HI: Operational Summary.

Start
Define the objective
Select Experts : $\{\Xi_1, \Xi_2, \ldots, \Xi_{10}\}$
Perform IoA by deploying brain-storming approach to identify the pivotal activities: $\Lambda_1, \Lambda_2, \ldots, \Lambda_n$
Perform CoA by using Eqs. (1)-(5) as described in Algorithm 1
Perform PoC by using Eqs. (6)-(10) as explained in Algorithm 2
Output: Pivotal activities in the order of their importance
End

Complexity: The time complexity is $O(n^2)$, where n is the number of points. This is because, for each of the n points, the algorithm needs to compute the distance to every other point to determine if they fall within the specified ϵ-neighborhood. The algorithmic complexity of prioritization depends on the number of alternatives (n) and the number of criteria (m) involved in the decision-making process. The method consists of several computational steps, including normalization, weighting, determining the positive and negative ideal solutions, calculating distances, and finally, computing the similarity to the ideal solution. All steps have complexity $O(mn)$, while the similarity calculation step has complexity $O(m)$. The overall complexity of prioritization is thus $O(mn)$, as each step must be performed for every element in the decision matrix, and the steps are generally sequential. Therefore, the total time complexity of the GCS-HI framework is $\mathcal{O}(mn^3)$.

V. Performance Evaluation

A. Experimental Setup

The experiment was carried out on machines with 11th Gen Intel® Core ™ i7-1195G7 CPU, which employs a clock speed of 2.92 GHz. This computational system utilizes a 64-bit Windows 11 Home 22H2 version Operating system. The installed RAM of the system is 32.0 GB (31.8 GB usable). The software used for the DBSCAN was IBM SPSS 26.

B. Description of Data

The focus group was asked to rate identified activities based on the criteria identified. The activities are rated on criteria using an ordinal scale of [1 to 10], here 1 is the least, and 10 is the highest. The response average was taken and approximated to get the response matrix. The response matrix data was used for the categorization of activities using DBSCAN. Finally, categories were ranked using criteria of prioritization. The raw data used in the study was uploaded to an online repository, and DOI was generated [37].

C. Results

Based on brainstorming sessions, experts identified twenty activities required to preserve data privacy and security in healthcare. The focus group is asked to come up with twenty activities. The objective is achieved after five iterations. Secondly, the group is asked to identify criteria, for grouping these activities in meaningful ways. After three iterations, the focus group agreed on three criteria: Functional Focus (G1), Stakeholder Engagement (G2), and Strategic Objective (G3). The number of criteria was limited to three because of the inability of the DBSCAN method to handle multidimensional data. These twenty activities with their description are listed in Table II. The value of two parameters, ϵ and MinPts, is adjusted to zero noise points. Finally, a ϵ = 0.5 and MinPts = 2 are taken for the clustering assignment. The result of clustering is given in Table III. The clusters are further given descriptive names for further analysis. The profile plot for clusters using the mean value of criteria is calculated to see whether the category/cluster is distinct. Fig. 3 shows that categories are different from each other.

Further, the results of prioritizing categories using the TOPSIS method are explained. The three criteria taken for prioritization problems are (1) ease of implementation (Ease), (2) effectiveness (effect), and (3) economics (cost). All these criteria are beneficial as a high rating of economics means less cost of implementation. The weight of ease, effectiveness, and economy were calculated using AHP and found to be 0.11,0.63, and 0.26, respectively. The consistency ratio for pairwise comparison was 4%, less than the recommended value of 10%. Finally, the ranking categories identified are

Authorized licensed use limited to: UNIV OF DERBY. Downloaded on December 24,2024 at 13:48:54 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2024.3467179

TABLE II: Pivotal activities to ensure privacy and security

Code	Activity (Λ)	Description of the Activity
Λ_1	Regular Risk Assessment	Conduct frequent risk assessments to identify potential vulnerabilities in the healthcare system.
Λ_2	Employee Training	Provide continuous training for employees on data privacy and security protocols.
Λ_3	Strong Access Controls	Implement strict access controls to ensure only authorized personnel can access sensitive data.
Λ_4	Data Encryption	Encrypt patient data in transit and at rest to protect against unauthorized access.
Λ_5	Audit Trails	Maintain detailed audit trails to monitor access and changes to patient data.
Λ_6	Anti-Malware Software	Install and regularly update anti-malware software to protect against cyber threats.
Λ_7	Secure Data Storage	Use secure storage solutions, such as encrypted databases, for patient data.
Λ_8	Data Minimization	Only collect and retain the minimum amount of patient data necessary for healthcare purposes.
Λ_9	Incident Response Plan	Develop and regularly update an incident response plan for potential data breaches.
Λ_{10}	Regular Software Updates	Keep all software and systems updated to protect against vulnerabilities.
Λ_{11}	Multi-Factor Authentication	Implement multi-factor authentication for accessing patient data systems.
Λ_{12}	Secure Communication Channels	Use secure communication channels, such as encrypted email, for transmitting patient data.
Λ_{13}	Patient Consent Management	Regularly obtain and manage patient consent for data use and sharing.
Λ_{14}	Third-Party Vendor Assessment	Conduct thorough assessments of third-party vendors who have access to patient data.
Λ_{15}	Data Anonymization Techniques	Apply data anonymization techniques where appropriate for research and analysis.
Λ_{16}	Physical Security Measures	Enhance physical security measures to protect data storage and access areas.
Λ_{17}	Mobile Device Management	Implement policies for secure use of mobile devices in accessing patient data.
Λ_{18}	Cybersecurity Insurance	Consider obtaining cybersecurity insurance to mitigate financial risks associated with data breaches.
Λ_{19}	Regular Compliance Audits	Conduct audits to ensure ongoing compliance with healthcare data protection regulations.
Λ_{20}	Patient Education	Educate patients about their data rights and how to protect their health information.

TABLE III: Summary of categorization results

Cluster ($\mathbb{C}$)	Member Activities	Name of Category
$\mathbb{C}_1$	$\{\Lambda_1, \Lambda_9, \Lambda_{13}, \Lambda_{14}, \Lambda_{19}\}$	Policy and Compliance Management
$\mathbb{C}_2$	$\{\Lambda_2, \Lambda_{11}, \Lambda_{20}\}$	Employee Training and Awareness
$\mathbb{C}_3$	$\{\Lambda_3, \Lambda_4, \Lambda_7, \Lambda_8, \Lambda_{15}\}$	Data Protection and Privacy Control
$\mathbb{C}_4$	$\{\Lambda_5, \Lambda_{18}\}$	Monitoring and Response
$\mathbb{C}_5$	$\{\Lambda_6, \Lambda_{10}, \Lambda_{12}, \Lambda_{16}, \Lambda_{17}\}$	Technology and Infrastructure Security

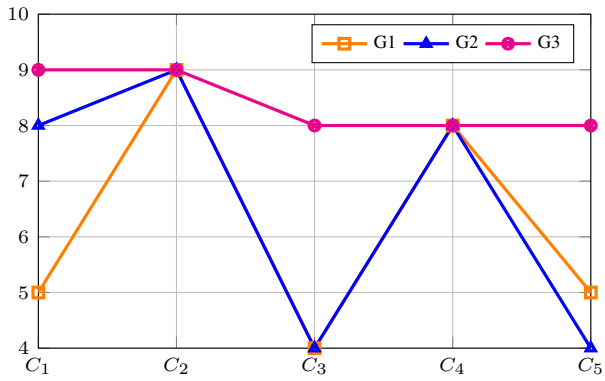

Fig. 3: Profile plot of the identified clusters.

TABLE IV: Final decision matrix for prioritization

Category ($\mathbb{C}$)	Name of Category	Ease (E1)	Effect (E2)	Economics (E3)
$\mathbb{C}_1$	Policy and Compliance Management	4	7	5
$\mathbb{C}_2$	Employee Training and Awareness	4	8	6
$\mathbb{C}_3$	Data Protection and Privacy Control	5	9	6
$\mathbb{C}_4$	Monitoring and Response	6	7	6
$\mathbb{C}_5$	Technology and Infrastructure Security	7	7	5

listed in Table III. The initial decision matrix based on focus group discussion is given in Table IV.

Now the decision matrix is normalized to obtain the matrix R, which is further multiplied with the criteria weight vector $W = [0.11, 0.63, 0.26]$ to get matrix T is given by:

$$T = \begin{bmatrix} 0.04 & 0.29 & 0.12 \\ 0.04 & 0.29 & 0.12 \\ 0.05 & 0.32 & 0.12 \\ 0.06 & 0.25 & 0.12 \\ 0.06 & 0.25 & 0.10 \end{bmatrix}$$

Based on the matrix T, the best alternative computed are represented as A_b = [0.06, 0.32, 0.12] while the worst alternatives computed are given by A_w = [0.04,0.25,0.10]. Now, the distance of alternatives (categories) from the best and worst alternatives is given by vectors d_b = [0.05, 0.05, 0.020, 0.07, 0.07] and d_w = [0.04, 0.04, 0.08, 0.03, 0.03], respectively. Now, the vector representing similarity from the worst alternative S_w = [0.48, 0.48, 0.80, 0.27, 0.27]. This helped to conclude that $\mathbb{C}_3$ (data protection and privacy control) activities should be given first preference in implementation. $\mathbb{C}_1$ and $\mathbb{C}_2$ activities (policy and compliance management and employee training and awareness) should be implemented next. Finally, $\mathbb{C}_4$ and $\mathbb{C}_5$ (monitoring and response and technology and infrastructure security) should be implemented. The results of the prioritization show that all the identified categories have significant importance, and none of these categories can be left without implementation.

D. Comparison

To further analyze the efficiency of the GCS-HI framework, it is deployed with comparable schemes including DTEM [38],

Authorized licensed use limited to: UNIV OF DERBY. Downloaded on December 24,2024 at 13:48:54 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2024.3467179

Soni et al. [39], Zahrani et al. [40], Ansari et al. [41], and Mishra et al. [10] as described in Table V. The GCS-HI is a comprehensive framework that integrates the identification, categorization, and prioritization of activities to safeguard PHI. It identifies the key activities to ensure privacy and security in healthcare and provides a clear road map for implementing these measures in a resource-constrained environment.

VI. Conclusion

This study developed a comprehensive, easy-to-implement framework for the phase-wise implementation of measures for ensuring privacy and security. It identified and categorized key activities essential for maintaining data privacy and security, ultimately prioritizing them based on ease of implementation, effectiveness, and economic feasibility. The application of DBSCAN effectively grouped activities into distinct clusters, while TOPSIS provided a clear ranking, emphasizing the importance of data protection and privacy control. The result conclude that C_3 activities should be given first preference in implementation. C_1 and C_2 activities should be implemented next. Finally, C_4 and C_5 should be implemented. Implementing privacy and security in healthcare faces challenges like balancing data access with protection, managing complex regulations (e.g., HIPAA, GDPR), and addressing technological vulnerabilities. Limited resources and varying organizational capabilities further complicate efforts, making it difficult to ensure consistent, comprehensive safeguards across diverse healthcare settings.

Future research should focus on adapting and evolving this framework in response to emerging technologies and threats, ensuring it remains relevant and effective in a rapidly changing digital landscape. Additionally, further validation of the framework in real-world healthcare settings would be invaluable, contributing to its practical utility and effectiveness in safeguarding patient data. Also, the usability and interoperability of the healthcare information systems can be explored.

References

[1] V. Mishra, "Telemedicine in chronic care –a case of diabetes management," vol. 9, pp. 7–12, 03 2020.

[2] G. Aceto, V. Persico, and A. Pescapé, "Industry 4.0 and health: Internet of things, big data, and cloud computing for healthcare 4.0," *Journal of Industrial Information Integration*, vol. 18, p. 100129, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2452414X19300135

[3] K. Gupta, D. Saxena, R. Gupta, and A. K. Singh, "Maids: malicious agent identification-based data security model for cloud environments," *Cluster Computing*, vol. 27, no. 5, pp. 6167–6184, 2024. [Online]. Available: https://doi.org/10.1007/s10586-023-04263-9

[4] V. Mishra and M. G. Sharma, "Digital transformation evaluation of telehealth using convergence, maturity, and adoption," *Health Policy and Technology*, vol. 11, no. 4, p. 100684, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2211883722000910

[5] K. Gupta, D. Saxena, R. Gupta, J. Kumar, and A. K. Singh, "Fedmup: Federated learning driven malicious user prediction model for secure data distribution in cloud environments," *Applied Soft Computing*, vol. 157, p. 111519, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S156849462400293X

[6] D. Lee and S. N. Yoon, "Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges," *International Journal of Environmental Research and Public Health*, vol. 18, no. 1, 2021. [Online]. Available: https://www.mdpi.com/1660-4601/18/1/271

[7] K. Gupta and A. Kush, "A forecasting-based dlp approach for data security," in *Data Analytics and Management*, A. Khanna, D. Gupta, Z. Pólkowski, S. Bhattacharyya, and O. Castillo, Eds. Singapore: Springer Singapore, 2021, pp. 1–8.

[8] Z. N. Aghdam, A. M. Rahmani, and M. Hosseinzadeh, "The role of the internet of things in healthcare: Future trends and challenges," *Computer Methods and Programs in Biomedicine*, vol. 199, p. 105903, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260720317363

[9] K. Gupta and A. Kush, "A learning oriented dlp system based on classification model," *INFOCOMP Journal of Computer Science*, vol. 19, no. 2, pp. 98–108, 2020. [Online]. Available: https://infocomp.dcc.ufla.br/index.php/infocomp/article/view/1008

[10] V. Mishra, K. Gupta, D. Saxena, and A. K. Singh, "A global medical data security and privacy preserving standards identification framework for electronic healthcare consumers," *IEEE Transactions on Consumer Electronics*, pp. 1–1, 2024.

[11] C. P. Friedman and M. L. Rigby, "Conceptualising and creating a global learning health system," *International journal of medical informatics*, vol. 82 4, pp. e63–71, 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:2390589

[12] S. Chenthara, K. Ahmed, H. Wang, and F. Whittaker, "Security and privacy-preserving challenges of e-health solutions in cloud computing," *IEEE Access*, vol. 7, pp. 74 361–74 382, 2019.

[13] A. Mayer, C. André da Costa, and R. Righi, "Electronic health records in a blockchain: A systematic review," *Health Informatics Journal*, vol. 26, p. 146045821986635, 09 2019.

[14] A. Shahnaz, U. Qamar, and A. Khalid, "Using blockchain for electronic health records," *IEEE Access*, vol. 7, pp. 147 782–147 795, 2019.

[15] F. Tang, S. Ma, Y. Xiang, and C. Lin, "An efficient authentication scheme for blockchain-based electronic health records," *IEEE Access*, vol. PP, pp. 1–1, 03 2019.

[16] R. Guo, H. Shi, Q. Zhao, and D. Zheng, "Secure attribute-based signature scheme with multiple authorities for blockchain in electronic health records systems," *IEEE Access*, vol. 6, pp. 11 676–11 686, 02 2018.

[17] A. Mamun, S. Azam, and C. Gritti, "Blockchain-based electronic health records management: A comprehensive review and future research direction," *IEEE Access*, vol. PP, pp. 1–1, 01 2022.

[18] J. J. Hathaliya and S. Tanwar, "An exhaustive survey on security and privacy issues in healthcare 4.0," *Computer Communications*, vol. 153, pp. 311–335, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140366419311880

[19] H. Ghayvat, S. Pandya, P. Bhattacharya, M. Zuhair, M. Rashid, S. Hakak, and K. Dev, "Cp-bdhca: Blockchain-based confidentiality-privacy preserving big data scheme for healthcare clouds and applications," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 5, pp. 1937–1948, 2022.

[20] W. Bani-Issa, A. Ibrahim, A. Akour, A. Marzouqi, Abbas, Hisham, and griffith, "Confidentiality, security and patient safety concerns about electronic health records," *International Nursing Review*, vol. 67, pp. 218–230, 04 2020.

[21] L. Wasserman and Y. Wasserman, "Hospital cybersecurity risks and gaps: Review (for the non-cyber professional)," *Frontiers in Digital Health*, vol. 4, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:251497628

[22] S. Kiser and B. Maniam, "Ransomware: Healthcare industry at risk," *Journal of Business and Accounting*, vol. 14, pp. 64–81, 2021.

[23] N. Saxena, E. Hayes, E. Bertino, P. O. Ojo, K.-K. R. Choo, and P. Burnap, "Impact and key challenges of insider threats on organizations and critical businesses," *Electronics*, vol. 9, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:221640971

[24] S. Prabhu and N. Thompson, "A primer on insider threats in cybersecurity," *Information Security Journal: A Global Perspective*, vol. 31, pp. 602–611, 09 2022.

[25] P. K. Yeng, M. A. Fauzi, B. Yang, and P. Nimbe, "Investigation into phishing risk behaviour among healthcare staff," *Information*, vol. 13, no. 8, 2022. [Online]. Available: https://www.mdpi.com/2078-2489/13/8/392

[26] W. Syafitri, Z. Shukur, U. A. Mokhtar, R. Sulaiman, and M. A. Ibrahim, "Social engineering attacks prevention: A systematic literature review," *IEEE Access*, vol. 10, pp. 39 325–39 343, 2022.

[27] P. Prince and S. Lovesum, "Privacy enforced access control model for secured data handling in cloud-based pervasive health care system," *SN Computer Science*, vol. 1, 07 2020.

[28] P. Sarosh, S. A. Parah, G. M. Bhat, and K. Muhammad, "A security management framework for big data in smart healthcare,"

Authorized licensed use limited to: UNIV OF DERBY. Downloaded on December 24,2024 at 13:48:54 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2024.3467179

TABLE V: GCS-HI framework v/s state-of-the-art approaches

Study	Framework					Feature		Computational Complexity
	Factors Selection	Prioritization Method	Relation Establishment	Suggested Framework	Methodology	Ξ	f	
DTEM [38]	SLR	⁕	MDS	TTFA	Delphi	8	9	$\mathcal{O}(nr\|\Xi\|)$
Soni et al. [39]	SLR	⁕	⁕	Not Holistic	Experimental Design	5	6	$\mathcal{O}(n^2.m\log m)$
Zahrani et al. [40]	SLR	MCDM	TFN	Descriptive	Fuzzy, ANP, TOPSIS	6	13	$\mathcal{O}(n.m^2)$
Ansari et al. [41]	SF	MCDM	ISO 27005	Prescriptive	Fuzzy, TOPSIS	25	7	$\mathcal{O}(n^2.m)$
GDSPS [10]	RS	MCDM	×	Descriptive	K-Means, OPA	7	20	$\mathcal{O}(nkm\log m)$
GCS-HI	Delphi	Structural	×	Prescriptive	DBSCAN, AHP-TOPSIS	10	20	$\mathcal{O}(mn^3)$

×: Absent; √: Present; ⁕: Not applicable; Ξ: Experts; f: Factors; SLR: Systematic Literature Review; SF:Survey Form; RS: Review of Standards; $MCDM$: Multi Criteria Decision Making; $PRISMA$: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; OPA: Ordinal Priority Approach; $DBSCAN$: Density-Based Spatial Clustering of Applications with Noise; AHP: Analytical Hierarchical Approach; $TOPSIS$: Technique for Order of Preference by Similarity to Ideal Solution; MDS: Multi-Dimensional Scaling; $TTFA$: Task Technology Fit Analysis; TFN: Triangular Fuzzy Number; ANP: Analytic Network Method; n: Number of Experts; m: Number of Criteria/Factors; k: Number of Clusters; r: Number of Rounds

Big Data Research, vol. 25, p. 100225, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214579621000423

[29] B. Liao, Y. Ali, S. Nazir, L. He, and H. U. Khan, "Security analysis of iot devices by using mobile computing: A systematic literature review," *IEEE Access*, vol. 8, pp. 120 331–120 350, 2020.

[30] E. Di Minin, C. Fink, A. Hausmann, J. Kremer, and R. Kulkarni, "How to address data privacy concerns when using social media data in conservation science," *Conservation Biology*, vol. 35, pp. 437–446, 03 2021.

[31] A. Mavrogiorgou, A. Kiourtis, K. Perakis, S. Pitsios, and D. Kyriazis, "Iot in healthcare: Achieving interoperability of high-quality data acquired by iot medical devices," *Sensors*, vol. 19, p. 1978, 04 2019.

[32] M. Elhoseny, N. N. Thilakarathne, M. I. Alghamdi, R. K. Mahendran, A. A. Gardezi, H. Weerasinghe, and A. Welhenge, "Security and privacy issues in medical internet of things: Overview, countermeasures, challenges and future directions," *Sustainability*, vol. 13, no. 21, 2021. [Online]. Available: https://www.mdpi.com/2071-1050/13/21/11645

[33] K. Pushpalatha, D. Saha, N. Gudi, and R. Sinha, "Awareness about privacy and security of patient health information," *Indian Journal of Public Health Research and Development*, vol. 9, p. 190, 12 2018.

[34] K. Koeninger, R. Bradshaw, H. PA, and J. Conley, "International health data: How hipaa interacts with the eu gdpr," 2020.

[35] W. Moore and S. Frye, "Review of hipaa, part 2: Limitations, rights, violations, and role for the imaging technologist," *Journal of Nuclear Medicine Technology*, vol. 48, no. 1, pp. 17–23, 2020. [Online]. Available: https://tech.snmjournals.org/content/48/1/17

[36] S. M. Shah and R. A. Khan, "Secondary use of electronic health record: Opportunities and challenges," *IEEE Access*, vol. 8, pp. 136 947–136 965, 2020.

[37] V. Mishra, "Data for framework for implementation of privacy and security in healthcare."

[38] V. Mishra and M. G. Sharma, "Digital transformation evaluation of telehealth using convergence, maturity, and adoption," *Health Policy and Technology*, vol. 11, no. 4, p. 100684, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2211883722000910

[39] P. Soni, J. Pradhan, A. K. Pal, and S. H. Islam, "Cybersecurity attack-resilience authentication mechanism for intelligent healthcare system," *IEEE Transactions Industrial Informatics*, vol. 19, no. 1, pp. 830–840, 2023. [Online]. Available: https://doi.org/10.1109/TII.2022.3179429

[40] F. A. Al-Zahrani, "Evaluating the usable-security of healthcare software through unified technique of fuzzy logic, anp and topsis," *IEEE Access*, vol. 8, pp. 109 905–109 916, 2020.

[41] M. T. J. Ansari, F. A. Al-Zahrani, D. Pandey, and A. Agrawal, "A fuzzy topsis based analysis toward selection of effective security requirements engineering approach for trustworthy healthcare software development," *BMC Medical Informatics and Decision Making*, vol. 20, p. 236, 2020.

Kishu Gupta (Member, IEEE) is working as a Post Doctoral Research Fellow at the Cloud Computing Research Center, Department of Computer Science & Engineering, National Sun Yat-sen University (NSYSU), Kaohsiung, Taiwan. She earned her Ph.D. from India in 2023 with prestigious *INSPIRE Fellowship* sponsored by the Department of Science & Technology (DST), under Ministry of Science and Technology (MOST), Govt. of India. Also, she is a recipient of the *Gold Medal* for securing Ist rank in overall university during M.Sc. Her major research interest includes Data Security and Privacy, Cloud Computing, Federated Learning, Machine Learning, Quantum Computing, etc. Some of her research findings are published with top-notch venues including IEEE TASE, IEEE TCE, Applied Soft Computing, Cluster Computing, etc.

Vinaytosh Mishra is an Associate Professor and Director of Thumbay Institute for AI in Healthcare, Gulf Medical University, Ajman, UAE. He earned PostDoc in AI in Healthcare from the University of Arizona, USA and PostDoc in Ethical AI in Healthcare from the University of Ben Gurion, Israel. He has a PhD in Healthcare Management and a Bachelor of Technology in Electronics Engineering from the Indian Institute of Technology (BHU), India. He has over 19 years of experience in industries such as Information Technology, Manufacturing, Finance, Healthcare, and Education alongwith one Australian and One German Patent in AI in Healthcare. He has published more than 70 research papers in different journals and conferences of high repute. His current research interests include Statistics, Quality Assurance Engineering, Supply chain management, and Healthcare Digital Health.

Aaisha Makkar (Member, IEEE) is a Lecturer in computer science at the University of Derby, UK. She is an experienced researcher with more than 8 years of cutting-edge research and teaching experience in prestigious higher education institutions, including University of Derby (UK), Seoul National University of Science and Technology (South Korea), and Thapar Institute of Engineering and Technology (India). She has authored and co-authored more than 40 research papers in high-ranked international journals (SCI indexed) and conferences. She has a track record of collaborations with industries, delivering innovative Artificial Intelligent based solutions to various emergent problems.

Authorized licensed use limited to: UNIV OF DERBY. Downloaded on December 24,2024 at 13:48:54 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.