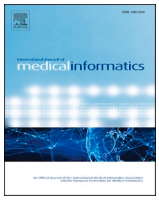
International Journal of Medical Informatics 205 (2026) 106107

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf



Security practices and insider threats in Spanish healthcare centers: a survey-based risk assessment

Isabel Herrera Montano [a,*], Susel Góngora Alonso [a], Soledad Sañudo García [b], José Javier García Aranda [c], Joel J.P.C. Rodrígues [d,e], Isabel de la Torre Díez [a]

[a] Department of Signal Theory and Communications and Telematics Engineering University of Valladolid, Paseo de Belén, 15, 47011 Valladolid, Spain
[b] Admission and Clinical Documentation Service, Hospital Universitario Río Hortega, Valladolid, Spain
[c] Department of Innovation, Nokia, Maria Tubau Street, 9, 28050 Madrid, Spain
[d] Federal University of Piauí, Teresina-PI, Brazil
[e] Instituto de Telecomunicações, Covilhã, Portugal

ARTICLE INFO

Keywords:
Cybersecurity
Healthcare
Insider threats
Information security
Survey

ABSTRACT

Introduction: Insider threats pose a critical risk in healthcare environments, where Hospital Information Systems (HIS) manage sensitive patients data. Authorized users may intentionally or accidentally compromise data confidentiality, integrity, and availability. This study assessed information security practices from the perspective of healthcare professionals in Spanish medical centers.

Methods: A descriptive, analytical, cross-sectional study was conducted using a survey administered to 41 healthcare professionals with access to confidential data. The survey covered access control, encryption at rest and in transit, communication channels, and data usage control. Descriptive statistics, Chi-square tests, and Cramér's V were applied to identify significant associations. K-means clustering and Silhouette coefficient were used to define user profiles. Principal Component Analysis (PCA) was used to visualize behavior patterns. A Random Forest model identified the most relevant predictive variables.

Results: Critical security gaps were detected, 31.7 % reported no control over data usage. Only 29.3 % encrypted data at rest and 36.6 % during transmission. Over 40 % used personal email or messaging apps to share sensitive data, and 97.6 % relied solely on passwords for authentication. These practices are inadequate to mitigate insider threats.

Conclusion: There is an urgent need to strengthen insider data protection. Security strategies should be tailored to user risk profiles. Measures must include strong authentication, full encryption, and stricter control of data transmission to reduce exposure to insider threats (intentionally or unintentionally) in healthcare settings. Additionally, there is a need to promote continuous cybersecurity training.

1. Introduction

In recent years, telemedicine and e-health have played a prominent role in the provision of health services. They have made remote medical care possible, especially in areas with limited mobility and during exceptional situations such as the COVID-19 pandemic period. In this period, telemedicine increased and gained special relevance, enabling continuity of care without the need for physical exposure of the patient [1–4]. However, the increased use of interconnected devices and networks for the transfer of medical data poses significant challenges in terms of data security and privacy [5]. Especially in the context of insider threats, the healthcare sector is a critical concern [1]. In this context, employees can intentionally or accidentally compromise the integrity and confidentiality of sensitive patient data [6,7]. Some human errors, such as the use of weak passwords or lack of knowledge of security protocols, are frequent causes of incidents in public health institutions [7–9].

In Spain, the magnitude of the problem has been reflected in some

* Corresponding author at: Department of Signal Theory and Communications and Telematics Engineering, University of Valladolid. Paseo de Belén, 15. 47011 Valladolid, Spain.
E-mail addresses: isabel.herrera.montano@uva.es (I. Herrera Montano), susel.gongora@uva.es (S. Góngora Alonso), ssanudo@saludcastillayleon.es (S. Sañudo García), jose_javier.garcia_aranda@nokia.com (J.J. García Aranda), joeljr@ieee.org (J.J.P.C. Rodrígues), isator@uva.es (I. de la Torre Díez).

https://doi.org/10.1016/j.ijmedinf.2025.106107
Received 3 April 2025; Received in revised form 19 August 2025; Accepted 31 August 2025
Available online 2 September 2025

recent cases, for example: In *Castilla y León* region, a patient denounced unauthorized access to his medical records, demanding that the health administration identify the person responsible [10]. In Zamora, the head nurse accessed five times without authorization the medical records of a worker and was prosecuted for the crime of disclosure of secrets [11]. Internationally, the cases are also alarming. According to the February 2025 Insider Threat Incidents Report [12], a business manager of a medical diagnostics company participated in a $70.6 million Medicare fraud scheme. This fraud was based on illegal agreements with professionals for the generation of unnecessary medical tests. Such incidents demonstrate that the insider threat not only compromises privacy but can also be associated with large-scale economic crime.

The literature on security in e-health and telemedicine has addressed various dimensions of secure information management and insider threat mitigation. In the study [5], they assessed the perceived risks in telemedicine and noted that both patients and professionals face difficulties in ensuring security in the use of virtual platforms during and after the pandemic. In 2015, Fernández-Alemán et al. conducted a study in a Spanish hospital. The study highlighted the insufficiency of secure passwords and the lack of training in security protocols among healthcare professionals [13]. Evans and other authors [8] used the "Information Security Core Human Error Causes" technique to identify human errors that lead to security breaches. The latter highlights causes such as workload or task repetitiveness.

In the last decade, the use of approaches based on artificial intelligence and machine learning to improve information security in the healthcare environment has increased. In this regard, the study [7] provides a review of techniques to prevent information leaks caused by insider threats, including cryptographic approaches and predictive models. In the study [14], the Random Forest (RF) algorithm was used to analyze cultural dimensions of that influence patient safety. This approach helped to identify critical variables such as managerial support and perception of the work environment. Research combines K-means clustering with statistical techniques (Principal Component Analysis (PCA) and Gaussian density) to detect insider threats. Offermann et al. in [15] identified three emotional profiles towards telemedicine (sceptics, undecided and supporters) using K-means clustering, showing that these attitudes influence the acceptance of remote consultations in nursing homes. However, exploring the existing literature revealed a significant information gap due to insufficient evidence on the security methods currently used to mitigate insider threats in the Spanish healthcare context.

Due to this knowledge gap, the main objective of this research is to identify and analyze the security methods applied to mitigate insider threats in the Spanish healthcare sector. In order to meet this objective, a survey was carried out among professionals in the Spanish healthcare sector. The following Research Questions (RQ) were posed in this study:

RQ1: What are the main insider threat security methods presented in the study sample?
RQ2: What are the characteristics and security levels of different groups with similar behaviors in the data obtained from the sample?
RQ3: What are the most prevalent insider threat security gaps in the study sample?

2. Background and State of the Art

The digitization of the healthcare sector has brought about a profound transformation in management, transmission, and storage of clinical information. Technologies such as Electronic Health Records (EHR), telemedicine, Internet of Medical Things, and Hospital Information Systems (HIS) have enabled substantial improvements in operational efficiency, patient-centered care, and data-driven clinical decision-making [16–19]. However, this evolution has been accompanied by growing challenges in terms of security and privacy, placing the healthcare sector among the most vulnerable to cyberattacks [20]. Within this scenario, insider threats, have become particularly relevant due to their impact on the confidentiality of clinical information and patient trust [21,22].

Surveys administered through institutional digital platforms have proven effective for collecting anonymous, large-scale data on healthcare professionals' attitudes toward digital health, cybersecurity awareness, and perceived barriers in clinical environments. In academic hospital settings, such instruments have enabled the stratification of responses by professional role and region, and the combination of quantitative metrics with qualitative content analysis to extract context-specific insights into user behavior and system-level challenges [23].

2.1. Insider threats and human factors

Insider threats in healthcare institutions can stem from both deliberate actions and unintentional mistakes, with human factors playing a key role in the latter. Research has consistently shown that common staff practices, such as using weak passwords, being unaware of internal policies, or lacking clear incident response procedures, often lead to security breaches [13,24,25].

In this sense, the IS-CHEC framework, adapted from the HEART model, has been used to categorize the human causes of safety incidents in the public domain. Its application in the UK healthcare sector revealed that the most common errors are related to time pressure, repetitive task execution and the inability to reverse unwanted actions [26]. Alanazi (2023) identified that, although most clinical staff are aware of the importance of protecting patient data, there are barriers such as lack of time for training, work overload, and lack of leadership in institutional security [27].

Additionally, limited cybersecurity training and low levels of digital literacy among some healthcare personnel increase exposure to risk. This issue becomes more critical in systems such as HIS, widely used in hospital settings, where an access error or omission can compromise multiple sensitive records [28]. In this context, a data breach analysis methodology applied in the healthcare sector identified insider threat patterns through the examination of publicly available incident descriptions. Using a five-step approach that combined tools such as VOSviewer for objective keyword extraction and NVivo for contextual analysis. This method revealed frequent co-occurrences of terms like "employee" and "PHI" (Protected Health Information), emphasizing how human errors, such as the mishandling of individually identifiable clinical data, are recurrent contributors to healthcare security incidents [29].

2.2. Technological approaches to mitigation

Recent literature has proposed multiple technological approaches to mitigate the risk of insider threats. These include anomaly detection systems based on machine learning, advanced cryptographic mechanisms [30,31], the use of blockchain for access traceability, and predictive models based on artificial intelligence [24,32,33]. However, a systematic review shows that many of these solutions are reactive and require historical incident data to be trained, which limits their prevention capabilities in environments with poor traceability [21,34–36].

Among the most widely studied methods for detecting anomalies are those based on decision trees, recurrent neural networks, autoencoders, and clustering algorithms such as K-means. These approaches make it possible to identify deviations from normal user behavior patterns within systems, which can alert to unusual access or malicious behavior at an early stage [24,34,37,38]. Recent findings underscore the need for integrated mitigation strategies that combine anomaly detection, behavioral analysis, and access control within real-time monitoring systems. These multi-tiered approaches offer more robust and proactive protection against both malicious and negligent insider threats while also preserving employee privacy [21,23,39–41].

2.3. Sociotechnical approaches and Emerging needs

Given that insider threats arise from the interaction between people, processes, and technology, it is necessary to adopt socio-technical approaches that integrate these three elements in a balanced way. Recent literature highlights the importance of strengthening institutional governance, promoting continuous training in cybersecurity, and building an organizational culture oriented towards information protection [17,18].

This approach suggests that it is not enough to install technological tools, but that workflows must be redesigned, individual responsibilities clarified, and effective channels for reporting and responding to incidents established. In particular, the lack of alignment between formal policies and actual practices, as evidenced during the WannaCry attack on the NHS, shows that written protocols are insufficient if they are not integrated into the daily operations of clinical staff [34].

A significant gap has been detected in the literature applied to the Spanish context. Although there are relevant studies in local hospitals, such as those by Fernández-Alemán et al. most research focuses on general diagnoses and does not delve into specific mitigation methods or the segmentation of personnel according to their risk profile. This gap hinders the application of personalized strategies and limits the impact of security policies in the real world [42].

2.4. Analytical techniques for insider threat detection in healthcare

In recent research on the security of HIS, both descriptive and inferential statistical techniques have been used to assess users' perceptions of privacy, confidentiality, information security, and patient safety. In particular, Alipour et al. (2023) [28] applied basic descriptive analyses, such as means, standard deviations, and frequencies, combined with analytical tests such as Pearson and Spearman correlation, as well as the chi-square test, to examine associations between dimensions perceived by clinical staff and sociodemographic variables [43].

In addition, more complex tools such as PCA and Exploratory Factor Analysis have made it possible to reduce the dimensionality of the data collected in questionnaires and extract latent factors related to the perception of technological usability in older adults [43]. These techniques have been complemented by multivariate analyses, such as ANOVA and structural equation modeling, to validate theoretical constructs and model relationships between variables including trust, privacy, and willingness to adopt digital health systems [19]. Additionally, studies focusing on the cybersecurity of EHR have employed anomaly detection methods, such as the Local Outlier Factor, alongside advanced statistical and machine learning techniques to detect atypical access to clinical systems and enhance privacy surveillance mechanisms [33,37].

3. Methods

3.1. Data collection

This descriptive and analytical study was based on an online questionnaire designed according to ISO 27002 [44] and HIPAA [45] standards. The instrument included five questions on security practices related to authentication, encryption (at rest and in transit), information sharing, and usage control, as well as two demographic questions on participant role and healthcare center.

The survey (see supplementary file and Ref. [45]) was distributed between April 2021 and November 2024 through more than 100 publicly available institutional email addresses from Spanish healthcare centers. It was also promoted via social networks and the official website of the Telemedicine and e-Health Research Group. In total, 41 valid responses were obtained from professionals actively handling confidential health data across 18 public and private institutions in seven regions of Spain. Respondents included medical, nursing, administrative, and IT staff.

Inclusion criteria required regular access to sensitive health information and employment in a Spanish healthcare institution. Incomplete surveys or those from ineligible participants were excluded. All responses were anonymous, and confidentiality was ensured in compliance with the Declaration of Helsinki [46]. The center name was collected solely for statistical analysis and never linked to individual responses.

3.2. Sample description

An initial sample of 48 surveys was collected, of which 7 were excluded due to missing values or because the respondents did not belong to a Spanish healthcare center. The final sample included 41 participants with valid data for analysis. Due to the confidential nature of the survey, only demographic information on participants' occupations and affiliated healthcare centers was available. Occupational characteristics of the sample are presented in Fig. 1.

Fig. 1 shows that 63 % of participants belong to medical staff, followed by nursing staff, representing 22 % of respondents. Together, these two groups account for 85 % of the total sample, consistent with the main objective of the survey, which focuses on professionals directly involved in healthcare delivery.

The professionals who completed the survey came from 18 healthcare centers across 7 Spanish different regions. Notably, 56 % of the participating centers are in the Castilla y León region, indicating a higher representation from this autonomous community in the study sample.

3.3. Data analysis

For data analysis, the security score was first analyzed in terms of access, usage control, sharing tools, and encryption at rest and in motion. The methodology and justification used for the analysis are shown in Table 1.

Secondly, the Security_Score variable was created, as described in Eq. (1). This variable represents the level of security obtained by each participant according to their responses in the survey.

$$Security\_Score = SA + CI + ST + SRI + STI \quad (1)$$

Where, SA is Secure_Access, CI is Control_Information, ST is Secure_Tools, SRI is Secure_Repos_Info and STI is Secure_Transfer_Info.

Finally, categorical variables were created from each item in the survey questions, as shown in Table 2.

3.4. Statistical analysis

Descriptive and nonparametric statistics were used to assess variable distributions. The chi-square (Eq. (2)) test and Cramer's V Coefficient

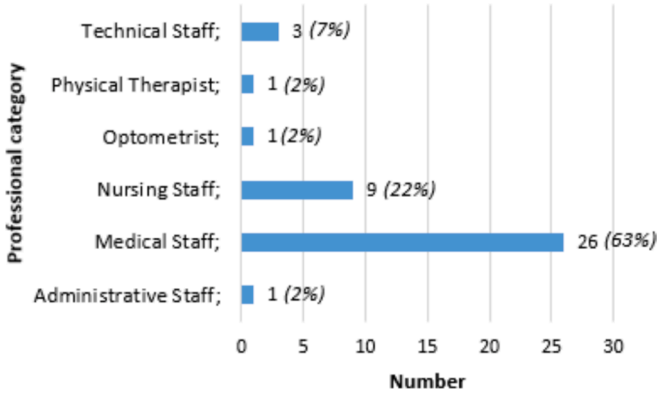


Fig. 1. Professional category of survey participants.

Table 1
Methodology and justification used for security score analysis.

Score_Variable	Methodology	Justification
Secure_Access (SA)	A score of 0 was assigned when access to confidential information was achieved using a username and password or neither method. A score of 1 was assigned when access was achieved using other methods or a combination of methods.	A single security mechanism for accessing confidential information is considered insufficient against insider threats because once an insider has the access or decryption keys, the information is no longer secure [7,9,30,47].
Control_Information (CI)	A 0 was assigned when the answer to R2 was none. A 1 was assigned when the answer was other than none.	Security against insider threats is considered insufficient if the use of confidential information is not controlled by any security mechanism mentioned in RQ2 or another mechanism described by the user [7,9,30,47].
Secure_Tools (ST)	A score of 0 was assigned when the response to R3 was other than institutional email or none. Otherwise, a score of 1 was assigned.	Any sharing of information through tools other than institutional email is considered insufficient security against the threat of insiders. The response "None" indicates that no confidential information is shared [7,9,30,47].
Secure_Repos_Info (SRI)	It is assigned 0 if the answer to R4 is "Plain Text" and 1 if the answer is "Encrypted".	It is considered that if the confidential information stored is in plain text, it does not have sufficient security against insider threats [7,9,30,47].
Secure_Transfer_Info (STI)	It is assigned 0 if the answer to R5 is "Plain Text" and 1 if the answer is "Encrypted".	It is considered that if confidential information is transferred in plain text, it does not have sufficient security against insider threats [7,9,30,47].

Table 2
Results of the Chi-square test between security methods and the Security Score variable.

Variable	Description (Categories = YES/NO)
Access with User and Password	Authentication based on username and password credentials to access confidential information.
Mechanism	Mechanism embedded in the information (watermark on confidential documents or medical images that allows tracking where it has been used).
Port blocking	Blocking information output (via Internet tools, emails, online repositories).
Output blocking	Blocking USB output ports on devices containing confidential information to protect the information.
None_U_C	Controlling the Use of Information.
Social Networks	Using social media as a means of sharing or managing confidential information.
Private groups or chats	Using private groups or chats on messaging platforms to share data.
Personal mail	Using personal email for exchanging and storing information.
Institutional mail	Using institutional or corporate email for managing sensitive information.
Removable devices	Using removable devices, such as USB drives or external hard drives, for storing or transferring data.
Data at Rest	Encrypting information stored on devices or servers.
Data in Transit	Encrypting information in transit or in motion.

(Eq. (4)) were applied to examine associations between practices and the overall security score. A significance level of $p < 0.05$ was adopted.

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{2}$$

Where O_{ij} represents the frequency observed in cell i, j and E_{ij} is the expected frequency, calculated as shown in the Eq. (3):

$$E_{ij} = \frac{((Rowtotal_i)(Columntotal_j))}{Granttotal} \tag{3}$$

$$V = \sqrt{\frac{\frac{X^2}{n}}{min(k-1, r-1)}} \tag{4}$$

Where X^2 is the chi-square statistic, n is the total number of observations, y $(k-1, r-1)$ represents the lowest value between the number of rows minus one and the number of columns minus one in the contingency table.

3.5. Cluster analysis

K-Means clustering segmented participants based on their security behaviors. Data preprocessing included One-Hot Encoding and normalization. The optimal number of clusters was determined using the Silhouette Coefficient (See Eq. (5)). PCA was used to visualize the clusters in two dimensions.

$$S(i) = \frac{b(i) - a(i)}{(a(i), b(i))} \tag{5}$$

Where, $a(i)$ is the average distance of a point within its own cluster and $b(i)$ is the average distance from the same point to the nearest cluster to which it does not belong.

3.6. Identification of key security gaps

A Random Forest model [21] ranked the importance of variables in predicting information security. This classification model was chosen for its robustness to noisy data, its ability to handle non-linear interactions and its low risk of overfitting. Previous research in hospital and cybersecurity contexts supports its use for the identification of critical factors and security breaches [7,14]. The survey design, data collection and analysis procedures are described below. The model used out-of-bag error estimation and the Gini index (Eq. (6)) reduction to identify the most influential security weaknesses.

$$G(D) = 1 - \sum_k f_k^2 \tag{6}$$

Where, f is a set of normalized frequencies $(f_1 + f_2, + \cdots + f_k)$ depending on the classes k. At each node split, the variable and threshold that generated the lowest combined impurity among the resulting subsets were chosen [21].

To complement the procedural details presented in the previous sections, Fig. 2 provides a visual summary of the methodological workflow applied in this study. It outlines the sequential structure of the research process, starting with the design of the survey instrument and its dissemination through institutional and digital channels, followed by the collection and filtering of valid responses. The diagram further illustrates the transformation of raw data into five security-related variables and the calculation of a composite Security Score. Finally, it shows the main stages of statistical and analysis, including bivariate testing such as Chi-square and Cramér's V, clustering using the K-Means algorithm, dimensionality reduction via PCA, and variable importance estimation through Random Forest modelling.

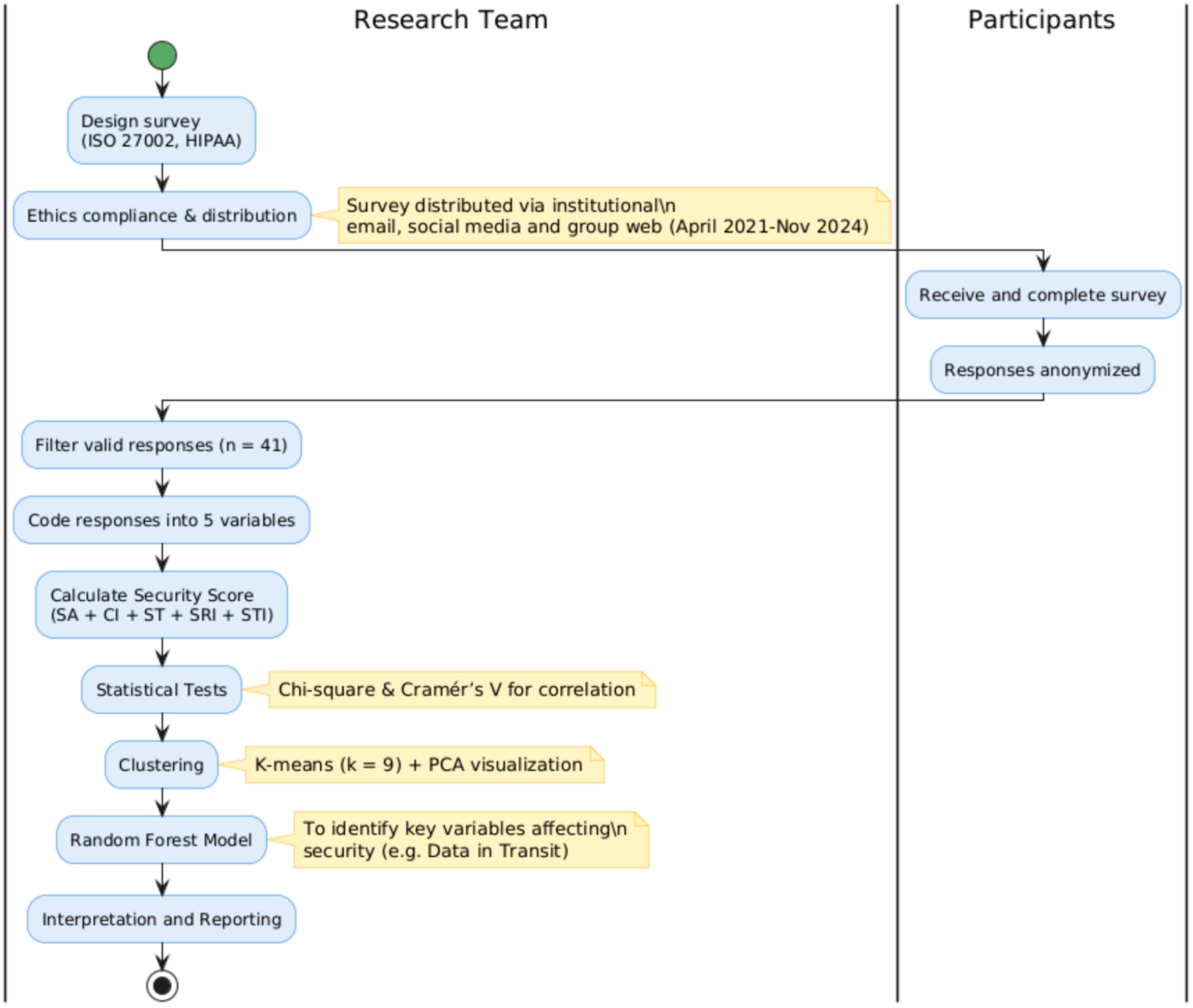


Fig. 2. Methodological workflow.

4. Results

4.1. Security measures against insider threats

The exploratory analysis revealed a strong reliance on conventional protection methods among healthcare professionals. As shown in Fig. 3, 97.6 % of users access confidential information using only a username and password, while 70.7 % rely on institutional Additionally, 41.5 % of participants use personal email, and 24.4 % utilize private messaging groups or chats to share sensitive data. In contrast, stricter security strategies, such as port blocking, are used by only 7.3 % of respondents. These findings suggest that most users depend on basic authentication and traditional communication platforms, potentially increasing the risk of unauthorized access and data leaks.

To explore the relationship between these security practices and overall information protection, a chi-square test was performed. Statistically significant associations were identified between several variables and the Security Score, particularly data in transit ($p = 0.000022$), data at rest ($p = 0.026598$), personal email ($p = 0.013962$), and private messaging groups ($p = 0.023512$). These results indicate that how users manage the storage and transmission of information significantly impacts their level of security, underlining the importance of stronger protocols in these areas.

Table 3 presents summarizing the correlations. A strong correlation ($r = 0.70$) was found between the "Data in Transit" variable and the Security Score, indicating that implementing protection mechanisms such as encryption during data transmission is a key factor in achieving a higher level of information security. A moderate correlation ($r = 0.36$) was also found between "Data at Rest" and the Security Score, suggesting that users who encrypt stored data tend to have stronger security practices overall.

Other notable correlations include a relationship between the use of personal email and Security Score ($r = 0.40$), implying that personal accounts may be used with varying degrees of security depending on user behavior. Similarly, a correlation of $r = 0.37$ between the use of private messaging platforms and the Security Score suggests that the choice of encrypted or privacy-focused tools may influence security outcomes. Additionally, the use of structured mechanisms, such as watermarking or content tracking, showed a moderate association with Security Score ($r = 0.30$), highlighting their contribution to better information control. Conversely, variables such as the use of social networks showed near-zero correlation with other security measures,

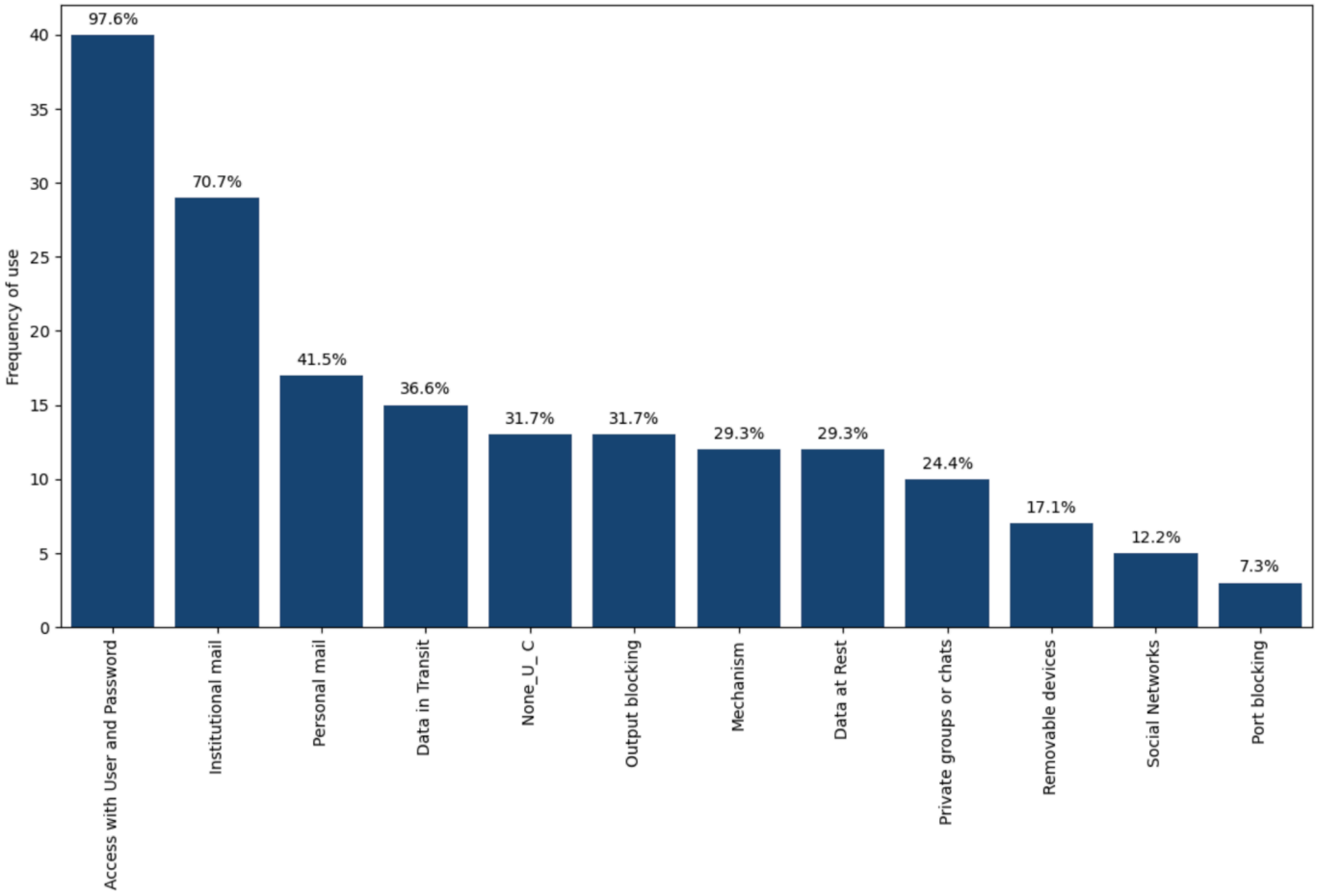


Fig. 3. Frequency distribution of security methods.

Table 3
Correlation values between security variables.

Variable	Cramér's V (r)
Data in Transit	0.70
Personal mail	0.40
Private groups or chat	0.37
Data at Rest	0.36
None_U_C	0.34
Mechanism	0.30

indicating that data exposure through these channels may not reflect broader protection practices.

In conclusion, these findings demonstrate that protecting data in transit and at rest plays a critical role in strengthening overall information security. The strong correlation between data transmission protection and the Security Score reinforces the need for robust encryption protocols and secure communication channels. Moreover, the varied impact of tools such as personal email and messaging apps emphasizes the importance of user education and context-specific strategies. These insights support the implementation of comprehensive security frameworks that go beyond authentication, incorporating data protection mechanisms and continuous training on secure information practices.

4.2. Segmentation using K-means clusters

The K-Means clustering algorithm was used to segment participants according to their information security behaviors. The optimal number of clusters was determined using the Silhouette coefficient, which indicated that nine clusters provided the best structure (Fig. 4).

This segmentation revealed clear distinctions between user groups based on their adoption of security practices. PCA was applied to reduce the dimensionality of the data and visualize the clusters in two dimensions (Fig. 5). Each point represents a respondent, colored according to their security score, allowing a visual assessment of similarity and protection levels among the groups.

Clusters 1 and 8 stood out for having 100 % implementation of security measures during data transmission. This suggests that participants in these groups prioritize secure communication, likely using encryption or protected networks. In contrast, Clusters 4 and 6 showed a complete absence of information usage control mechanisms, with all users reporting no implemented protections in that area, positioning them as the most vulnerable to insider threats.

Other clusters demonstrated more varied patterns. Cluster 5 combined institutional and personal email use with a high reliance (75 %) on removable devices, increasing the risk of data leakage if not properly managed. Cluster 0 was characterized by exclusive use of private messaging groups without additional safeguards, while Cluster 3 relied mainly on username and password authentication (80 %) with no further security enhancements.

These findings directly address RQ2, confirming that users can be categorized into groups with high, medium, and low levels of information security, depending on the tools and strategies they apply. While some clusters exhibit proactive behaviors by implementing multiple security layers, others operate with minimal or no protective measures. Intermediate clusters reflect inconsistent practices, which may depend on context, access to resources, or institutional culture.

User segmentation based on security behavior offers a structured

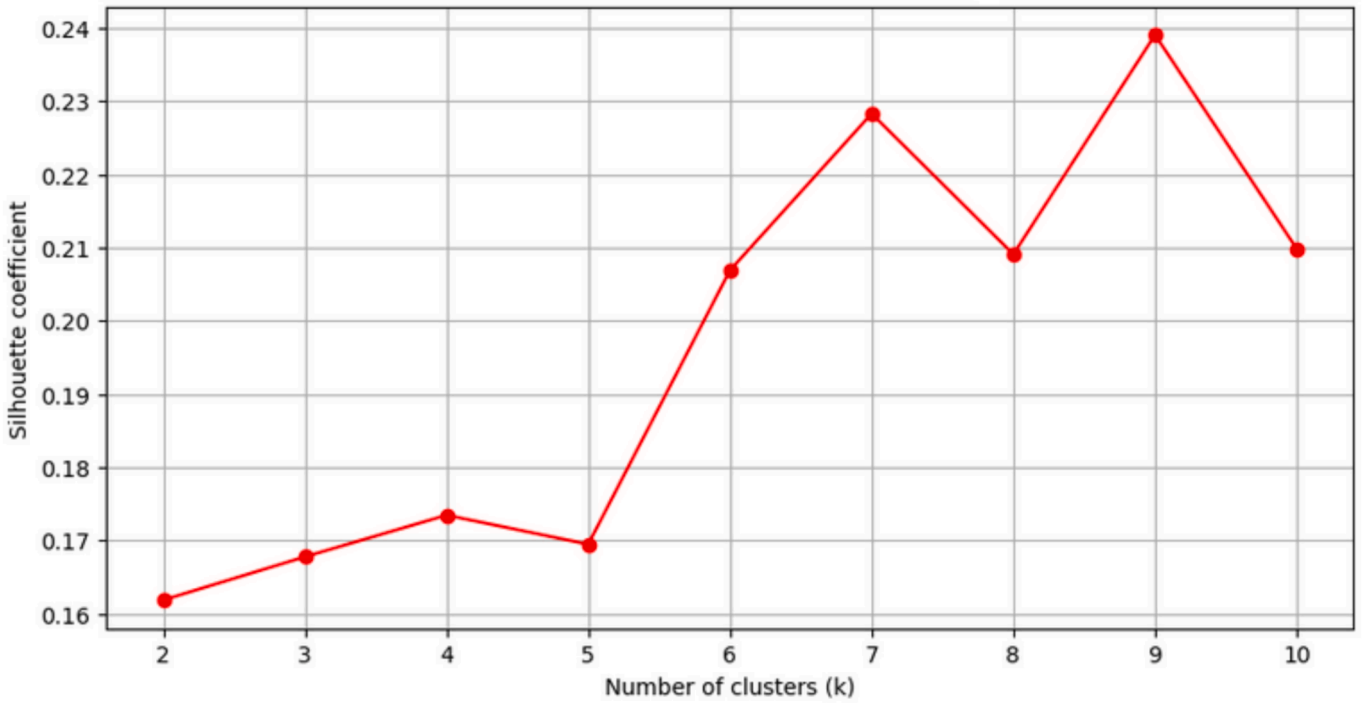


Fig. 4. Silhouette coefficient for different k.

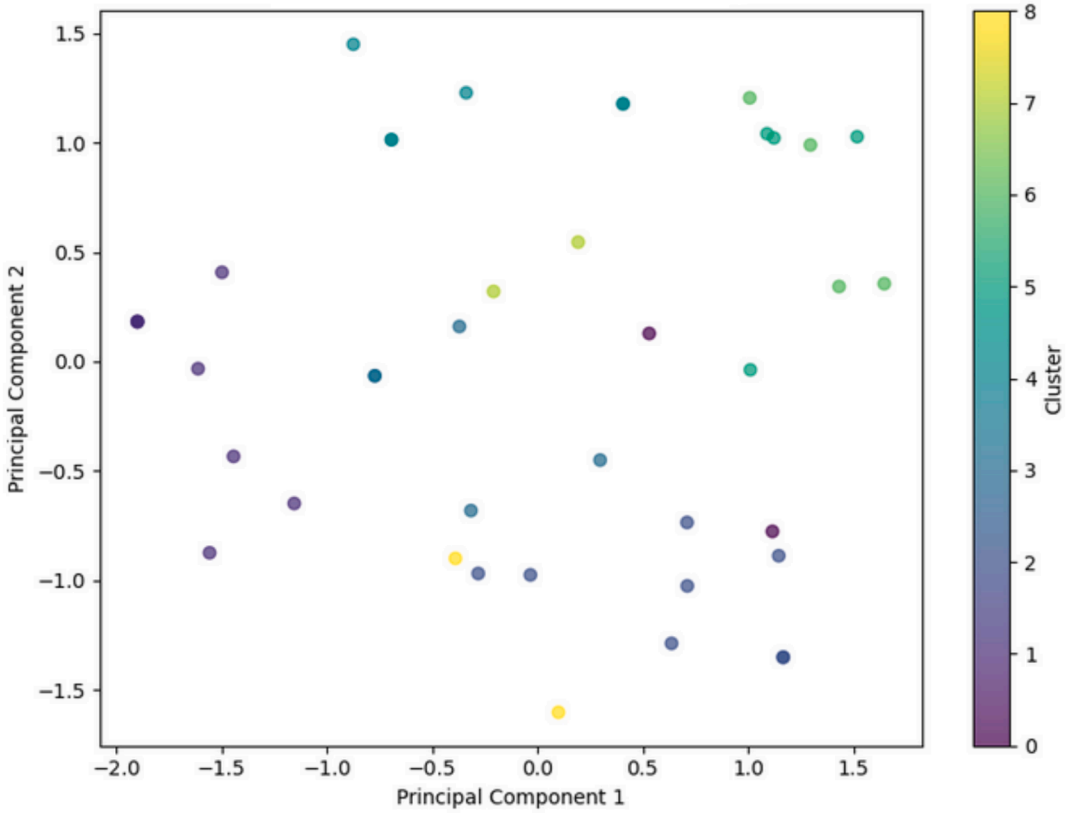


Fig. 5. K-Means Clustering Visualization k = 9.

understanding of risk exposure. It also enables the development of tailored strategies that respond to the specific vulnerabilities of each group. The results highlight the importance of targeted cybersecurity awareness initiatives, especially for high-risk clusters, promoting practices such as secure credential use, encryption of data in transit, and control over the use of information. Ultimately, the analysis confirms the existence of diverse security profiles and underscores the need for adaptive interventions to address the varying needs identified across the

sample.

4.3. Security gaps to insider threats

To determine the most prevalent security breaches in the sample studied, a Random Forest model was used. This model allows us to identify the most influential variables in determining the level of information security (See Fig. 6). The analysis identifies "Data in Transit" as the most critical vulnerability, with a relative importance of 25.14 %. This highlights the need for robust transmission security, such as encryption and protected networks. However, this does not imply that employees actively scan communications for sensitive content. There is no evidence of such behavior in the sample studied. The insider threat is primarily linked to unauthorized access or unintentional sharing of confidential data. Prevention should therefore focus on controlling access, enforcing usage policies, and ensuring secure communication channels.

The variable "Data at Rest" shows an importance of 13.6 %, highlighting the importance of having security mechanisms in place for information storage. The lack of controls on data at rest can facilitate unauthorized access or the leakage of sensitive information, reinforcing the need to adopt strategies such as storage encryption and access control.

A significant result is the high importance of the variable "None_U_C" (12.68 %), which groups users who do not have known security mechanisms to control the use of confidential information. This result indicates that a significant portion of the sample operates without documented protection strategies, which represents one of the main insider security breaches. The lack of structured measures can expose data to uncontrolled access, increasing the risk of leaks or insider attacks.

On the other hand, the use of personal email (10.34 %) and private groups or chats (9.11 %) are also relevant security factors. These results suggest that the use of these channels without adequate protection measures can be a risk vector within the organization, facilitating the leakage of sensitive information or the spread of threats. Other variables, such as the presence of security mechanisms embedded in the information (8.27 %), the use of institutional email (6.92 %), and the use of removable devices (6.19 %), also influence security, although to a lesser extent. In this sense, the use of removable devices constitutes a potential risk if adequate restrictions are not applied to prevent the loss or extraction of information.

These results respond to RQ3, showing that the main security breaches in the sample are associated with the lack of protection of information in transit and storage, the absence of control over information use, and the use of insecure communication channels. Considering these factors is essential to reducing the risk of insider threats. Consequently, the implementation of security protocols for data transmission, restrictions on the use of personal email and private chats for sharing information, and the establishment of security standards for data storage are recommended. These findings provide key information for designing risk mitigation strategies and improving cybersecurity in the environments analyzed.

4.4. Limitations

The main limitations of this study are as follows: The sample is small ($n = 41$), which reduces the generalizability and limits the statistical power of the analyses. This constrains the detection of more subtle associations and increases the risk errors. Furthermore, there is a significant territorial concentration, with the majority of participants from the region of *Castilla y León*. This may bias the geographic representation of the healthcare system. The available demographic information is limited. Only occupations and healthcare centers are recorded. Key variables such as age, education, professional experience, or level of responsibility are lacking. This prevents the analysis of individual factors associated with the adoption of security measures. The data are based on self-reported responses. This introduces potential interpretation or social desirability biases. Some participants do not report applying any control over their use of information, making it difficult to determine whether this is due to ignorance, omission, or a true absence of measures. The study also does not incorporate qualitative methods. Perceptions, barriers, or motivations surrounding information security are not explored, which limits our understanding of user behavior.

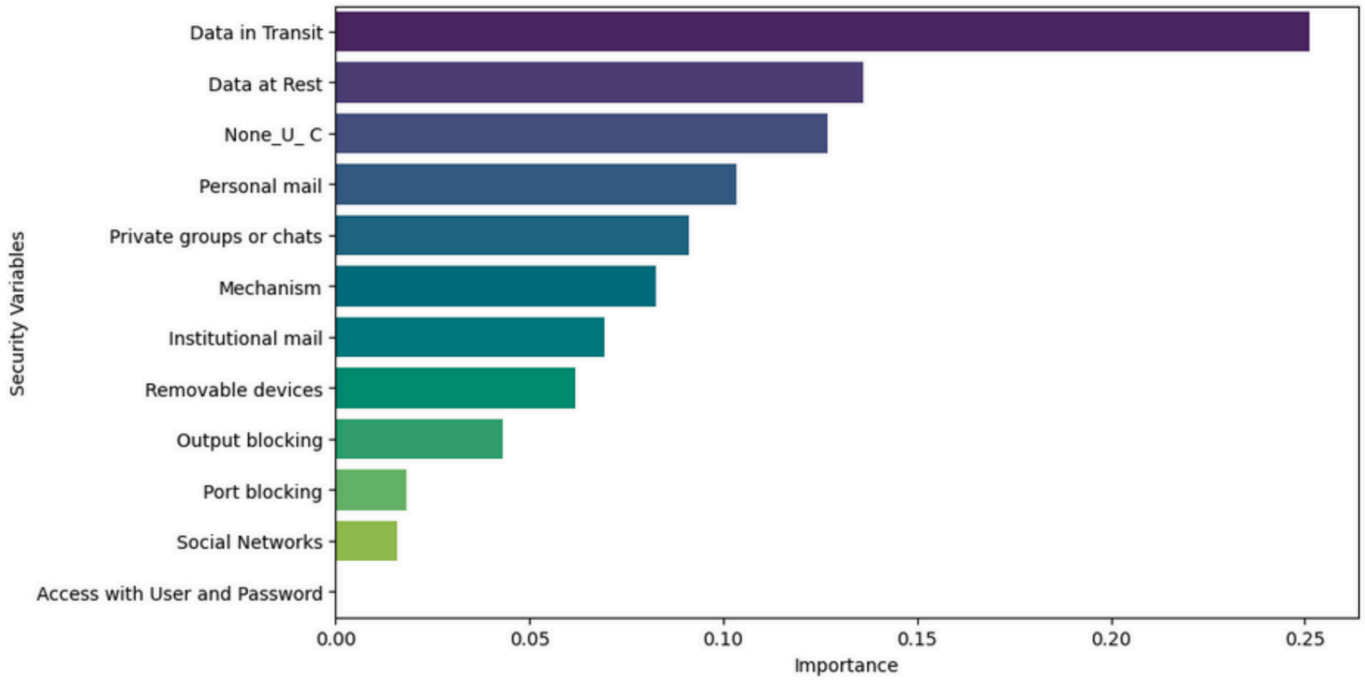


Fig. 6. Security gaps are prevalent in the study.

5. Conclusions

This study assessed the level of information security against insider threats in Spanish healthcare centers, addressing three key RQs. Findings show a predominance of basic protection mechanisms such as passwords and institutional emails (RQ1), while encryption and usage controls are still limited.

User segmentation (RQ2) revealed significant variation in security practices, identifying high-risk groups lacking structured measures. These profiles are not merely descriptive but provide actionable insight for practice. High-risk groups could be targeted with mandatory encryption and usage-control tools. While intermediate clusters may benefit from awareness programs or reinforced authentication, and low-risk groups can act as peer-leaders or early adopters of advanced technologies.

The most critical vulnerabilities (RQ3) include the absence of usage control (31.7 %), low encryption adoption (29.3 % at rest and 36.6 % in transit), and the use of personal communication channels (41.5 % email; 24.4 % messaging apps). Although passwords were not flagged by the Random Forest model, 97.6 % of users rely solely on them, an insufficient safeguard against insider threats.

These results highlight the need for adaptive, context-sensitive strategies rather than one-size-fits-all measures. Practical implementations include configurable authentication challenges to generate dynamic encryption keys. Additionally, the implementation of secure, secure file systems based on virtual file systems offers transparent protection compatible with clinical workflows.

Future studies should expand the sample to include diverse regions and professional roles, and incorporate contextual variables such as training, responsibility level, and cybersecurity awareness. Combining quantitative and qualitative methods (e.g., interviews, focus groups) would deepen understanding of user behavior.

Another promising direction is the integration of data protection principles into medical training. Virtual patients and chatbots, already used to simulate clinical encounters, could also present scenarios involve confidentiality, consent, and ethical data handling. For example, a simulated patient might ask whether their records remain stored, requiring the trainee to justify secure practices. Grounded in ongoing studies of EHR access routines, such simulations would embed DLP principles into early medical education, fostering both cybersecurity awareness and professional values.

Finally, implementing differentiated strategies, such as specific training, stronger authentication, data usage controls, and comprehensive encryption—based on cluster segmentation can enhance the effectiveness of cybersecurity policies and reduce exposure to insider threats in healthcare environments.

CRediT authorship contribution statement

Isabel Herrera Montano: Writing – original draft, Software, Methodology, Investigation, Data curation, Conceptualization. **Susel Góngora Alonso:** Writing – original draft, Visualization, Software, Data curation. **Soledad Sañudo García:** Investigation. **José Javier García Aranda:** Supervision. **Joel J.P.C. Rodrígues:** Validation, Supervision. **Isabel de la Torre Díez:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research has been carried out in a collaborative stay between the Telemedicine and e-Health group of the University of Valladolid and the Instituto da Telecomunicações da Delegação da Covilhã, Portugal. This work is partially funded by Brazilian National Council for Scientific and Technological Development - CNPq, via Grant No. 306607/2023-9. We thank Nokia Spain for the close collaboration to achieve successful results, and Banco Santander for their funding.

Additionally, we wants to thank the Spanish Ministry for Science, Innovation and Universities (MICINN), *Agencia Estatal de Investigacion (AEI)*, as well as to the *Fondo Europeo de Desarrollo Regional funds (FEDER, EU)*, under grant number PID2021-122210OB-I00, by M0CIN/AEI/10.13039/501100011033 and "ERDF A way of making Europe", European Union, for partially funding this work in the framework of the project "Advanced Artificial Intelligence Techniques to Detect and Combat Unknown and Adversary Cybersecurity Attacks", within *Plan Estatal de Investigacion Cientifica, Tecnica y de Innovacion 2021-2023, Ministerio de Ciencia e Innovacion,* Spain.

Funding sources

Isabel Herrera Montano has been funded through the UVa 2022 predoctoral contracts call, co-financed by Santander Bank.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijmedinf.2025.106107.

Data availability statement

The datasets generated and analyzed during the current study are not publicly available due to confidentiality agreements with participating healthcare professionals and institutions.

References

[1] G. Marques, N. Drissi, I. de la T. Díez, B.S. de Abajo, S. Ouhbi, Impact of COVID-19 on the psychological health of university students in Spain and their attitudes toward Mobile mental health solutions, Int. J. Med. Inform. 147 (2021) 104369, https://doi.org/10.1016/j.ijmedinf.2020.104369.

[2] M.S. Jalali, A. Landman, W.J. Gordon, Telemedicine, privacy, and information security in the age of COVID-19, J. Am. Med. Informatics Assoc. 28 (2021) 671–672, https://doi.org/10.1093/jamia/ocaa310.

[3] Deloitte, Securing the promise of virtual health care Addressing cyber risk in a new era of medicine, 2018.

[4] D. AlOsail, N. Amino, N. Mohammad, Security issues and solutions in E-health and telemedicine, in: Lect. Notes Data Eng. Commun. Technol., 2021: pp. 305–318. Doi: 10.1007/978-981-16-0965-7_26.

[5] K. Andreadis, K.A. Muellers, J.J. Lin, R. Mkuu, C.R. Horowitz, R. Kaushal, J. S. Ancker, Navigating privacy and security in telemedicine for primary care, Am. J. Manag. Care 30 (2024) SP459–SP463, https://doi.org/10.37765/ajmc.2024.89553.

[6] E. Smith, J.H.P. Eloff, Security in health-care information systems – current trends, Int. J. Med. Inform. 54 (1999) 39–54, https://doi.org/10.1016/S1386-5056(98)00168-3.

[7] I. Herrera Montano, J.J. García Aranda, J. Ramos Diaz, S. Molina Cardín, I. de la Torre Díez, J.J.P.C. Rodrigues, Survey of techniques on data leakage protection and methods to address the insider threat, Cluster Comput. 25 (2022) 4289–4302, https://doi.org/10.1007/s10586-022-03668-2.

[8] M. Evans, Y. He, L. Maglaras, I. Yevseyeva, H. Janicke, Evaluating information security core human error causes (IS-CHEC) technique in public sector and comparison with the private sector, Int. J. Med. Inform. 127 (2019) 109–119, https://doi.org/10.1016/j.ijmedinf.2019.04.019.

[9] I. Herrera Montano, J. Ramos Diaz, J.J. García Aranda, S. Molina-Cardín, J. J. Guerrero López, I. de la Torre Díez, Securecipher: an instantaneous synchronization stream encryption system for insider threat data leakage protection, Expert Syst. Appl. 254 (2024) 9, https://doi.org/10.1016/j.eswa.2024.124470.

[10] A. Santiago, Un paciente reclama saber quién accedió a su historia clínica de forma irregular, El Norte Castilla (2023). https://www.elnortedecastilla.es/castillayleon/paciente-reclama-saber-accedio-historia-clinica-forma-20231113195119-nt.html (accessed March 27, 2025).

[11] Radio Zamora, Piden dos años de cárcel por acceder sin permiso a la historia clínica de una enfermera., Cadena SER (2025). https://cadenaser.com/castillayleon/2025/01/28/piden-dos-anos-de-carcel-para-el-responsable-del-servicio-de-enfermeria-de-sayago-por-revelacion-de-secretos-radio-zamora/ (accessed March 27, 2025).

[12] I.T.D. Group., Insider Threat Incidents Report For March 2023 Produced By National Insider Threat Special Interest Group Insider Threat Defense Group, 2023. https://www.insiderthreatdefense.us/.
[13] J.L. Fernández-Alemán, A. Sánchez-Henarejos, A. Toval, A.B. Sánchez-García, I. Hernández-Hernández, L. Fernandez-Luque, Analysis of health professional security behaviors in a real clinical setting: an empirical study, Int. J. Med. Inform. 84 (2015) 454–467, https://doi.org/10.1016/j.ijmedinf.2015.01.010.
[14] M.C.E. Simsekler, A. Qazi, M.A. Alalami, S. Ellahham, A. Ozonoff, Evaluation of patient safety culture using a random forest algorithm, Reliab. Eng. Syst. Saf. 204 (2020) 107186, https://doi.org/10.1016/j.ress.2020.107186.
[15] J. Offermann, A. Rohowsky, M. Ziefle, Emotions of scepticism, trust, and security within the acceptance of telemedical applications, Int. J. Med. Inform. 177 (2023) 105116, https://doi.org/10.1016/j.ijmedinf.2023.105116.
[16] N.N. Basil, S. Ambe, C. Ekhator, E. Fonkem, Health records database and inherent security concerns: a review of the literature, Cureus (2022), https://doi.org/10.7759/cureus.30168.
[17] S.T. Argaw, J.R. Troncoso-Pastoriza, D. Lacey, M.-V. Florin, F. Calcavecchia, D. Anderson, W. Burleson, J.-M. Vogel, C. O'Leary, B. Eshaya-Chauvin, A. Flahault, Cybersecurity of hospitals: discussing the challenges and working towards mitigating the risks, BMC Med. Inform. Decis. Mak. 20 (2020) 146, https://doi.org/10.1186/s12911-020-01161-7.
[18] P. Ewoh, T. Vartiainen, Vulnerability to cyberattacks and sociotechnical solutions for health care systems: systematic review, J. Med. Internet Res. 26 (2024) e46904, https://doi.org/10.2196/46904.
[19] O. Enaizan, B. Eneizan, M. Almaaitah, A.T. Al-Radaideh, A.M. Saleh, Effects of privacy and security on the acceptance and usage of EMR: the mediating role of trust on the basis of multiple perspectives, Informatics Med. Unlocked 21 (2020) 100450, https://doi.org/10.1016/j.imu.2020.100450.
[20] L. Nemec Zlatolas, T. Welzer, L. Lhotska, Data breaches in healthcare: security mechanisms for attack mitigation, Cluster Comput. 27 (2024) 8639–8654, https://doi.org/10.1007/s10586-024-04507-2.
[21] U. Inayat, M. Farzan, S. Mahmood, M.F. Zia, S. Hussain, F. Pallonetto, Insider threat mitigation: systematic literature review, Ain Shams Eng. J. 15 (2024) 103068, https://doi.org/10.1016/j.asej.2024.103068.
[22] C.V.P. Herrera, J.S.M. Valcarcel, M. Díaz, J.L.H. Salazar, L. Andrade-Arenas, Cybersecurity in health sector: a systematic review of the literature, Indones. J. Electr. Eng. Comput. Sci. 31 (2023) 1099, https://doi.org/10.11591/ijeecs.v31.i2.pp1099-1108.
[23] A. Wernhart, S. Gahbauer, D. Haluza, eHealth and telemedicine: Practices and beliefs among healthcare professionals and medical students at a medical university, PLoS One 14 (2019) e0213067, https://doi.org/10.1371/journal.pone.0213067.
[24] S. Yuan, X. Wu, Deep learning for insider threat detection: Review, challenges and opportunities, Comput. Secur. 104 (2021) 102221, https://doi.org/10.1016/j.cose.2021.102221.
[25] L.H. Yeo, J. Banfield, Human factors in electronic health records cybersecurity breach: an exploratory analysis, Perspect. Heal. Inf. Manag. 19 (2022) 1i. http://www.ncbi.nlm.nih.gov/pubmed/35692854.
[26] M. Hedda, B.A. Malin, C. Yan, D. Fabbri, Evaluating the effectiveness of auditing rules for electronic health record systems, AMIA Annu. Symp. Proceedings. AMIA Symp. 2017 (2017) 866–875.
[27] A.T. Alanazi, Clinicians' perspectives on healthcare cybersecurity and cyber threats, Cureus (2023), https://doi.org/10.7759/cureus.47026.
[28] J. Alipour, Y. Mehdipour, A. Karimi, M. Khorashadizadeh, M. Akbarpour, Security, confidentiality, privacy and patient safety in the hospital information systems from the users' perspective: a cross-sectional study, Int. J. Med. Inform. 175 (2023) 105066, https://doi.org/10.1016/j.ijmedinf.2023.105066.
[29] I. Lee, Analyzing web descriptions of cybersecurity breaches in the healthcare provider sector: a content analytics research method, Comput. Secur. 129 (2023) 103185, https://doi.org/10.1016/j.cose.2023.103185.
[30] I.H. Montano, J.R. Diaz, S. Molina-Cardín, J.J.G. López, J.J.G. Aranda, I. de la T. Díez, SecureMD5: a new stream cipher for secure file systems and encryption key generation with artificial intelligence, Comput. Stand. Interfaces 25 (2025) 104047, https://doi.org/10.1016/j.csi.2025.104047.
[31] I. Herrera Montano, J. Ramos Diaz, J. Javier García Aranda, S. Molina-Cardín, J. José Guerrero López, I. de la Torre Díez, Securecipher: an instantaneous synchronization stream encryption system for insider threat data leakage protection, Expert Syst. Appl. (2024) 124470, https://doi.org/10.1016/j.eswa.2024.124470.
[32] H. Hindy, R. Atkinson, C. Tachtatzis, J.-N. Colin, E. Bayne, X. Bellekens, Towards an effective zero-day attack detection using outlier-based deep learning techniques, (2006).
[33] W. Hurst, B. Tekinerdogan, T. Alskaif, A. Boddy, N. Shone, Securing electronic health records against insider-threats: a supervised machine learning approach, Smart Heal. 26 (2022) 100354, https://doi.org/10.1016/j.smhl.2022.100354.
[34] S. Walker-Roberts, M. Hammoudeh, A. Dehghantanha, A systematic review of the availability and efficacy of countermeasures to internal threats in healthcare critical infrastructure, IEEE Access 6 (2018) 25167–25177, https://doi.org/10.1109/ACCESS.2018.2817560.
[35] A. Kim, J. Oh, J. Ryu, K. Lee, A review of insider threat detection approaches with IoT perspective, IEEE Access 8 (2020) 78847–78867, https://doi.org/10.1109/ACCESS.2020.2990195.
[36] C.S. Kruse, B. Frederick, T. Jacobson, D.K. Monticone, Cybersecurity in healthcare: a systematic review of modern threats and trends, Technol. Heal. Care 25 (2017) 1–10, https://doi.org/10.3233/THC-161263.
[37] W. Hurst, A. Boddy, M. Merabti, N. Shone, Patient privacy violation detection in healthcare critical infrastructures: an investigation using density-based benchmarking, Futur. Internet 12 (2020) 100, https://doi.org/10.3390/fi12060100.
[38] A. Amod Agarkar, M. Karyakarte, R.V. Kulkarni, V. Bag, S.P. Abhang, B. Sule, Assessing the impact of user behavior and insider threats on critical infrastructure, Inf. Secur. J. A Glob. Perspect. (2025) 1–12, https://doi.org/10.1080/19393555.2025.2502553.
[39] J. Yuan, B. Malin, F. Modave, Y. Guo, W.R. Hogan, E. Shenkman, J. Bian, Towards a privacy preserving cohort discovery framework for clinical research networks, J. Biomed. Inform. 66 (2017) 42–51, https://doi.org/10.1016/j.jbi.2016.12.008.
[40] P. Shojaei, E. Vlahu-Gjorgievska, Y.-W. Chow, Security and privacy of technologies in health information systems: a systematic literature review, Computers 13 (2024) 41, https://doi.org/10.3390/computers13020041.
[41] S.H. Houser, C.A. Flite, S.L. Foster, Privacy and security risk factors related to telehealth services – a systematic review, Perspect. Heal. Inf. Manag. 20 (2023) 1f. http://www.ncbi.nlm.nih.gov/pubmed/37215337%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC9860467.
[42] J.L. Fernández-Alemán, A. Sánchez-Henarejos, A. Toval, A.B. Sánchez-García, I. Hernández-Hernández, L. Fernandez-Luque, J. Offermann, A. Rohowsky, M. Ziefle, M. Evans, Y. He, L. Maglaras, I. Yevseyeva, H. Janicke, E. Bønes, P. Hasvold, E. Henriksen, T. Strandenæs, E. Smith, J.H.P. Eloff, J. Alipour, Y. Mehdipour, A. Karimi, M. Khorashadizadeh, M. Akbarpour, J.I. Fernando, L. L. Dawson, Analysis of health professional security behaviors in a real clinical setting: an empirical study, Int. J. Med. Inform. 84 (2023) 105116, https://doi.org/10.1016/j.ijmedinf.2015.01.010.
[43] S. Góngora Alonso, J.M. Toribio Guzmán, B. Sainz de Abajo, J.L. Muñoz Sánchez, M.F. Martín, I. de la Torre Díez, Usability evaluation of the eHealth long lasting memories program in Spanish elderly people, Health Informatics J. 26 (2020) 1728–1741, https://doi.org/10.1177/1460458219889501.
[44] G. Spini, M. van Heesch, T. Veugen, S. Chatterjea, Private hospital workflow optimization via secure k-means clustering, J. Med. Syst. 44 (2020) 8, https://doi.org/10.1007/s10916-019-1473-4.
[45] M. Shoffner, P. Owen, J. Mostafa, B. Lamm, X. Wang, C.P. Schmitt, S.C. Ahalt, The secure medical research workspace: an it infrastructure to enable secure research on clinical data, Clin. Transl. Sci. 6 (2013) 222–225, https://doi.org/10.1111/cts.12060.
[46] Asociación Médica Mundial (AMM), Declaración de Helsinki de la AMM, Asoc. Médica Mund. (2017) 1–7. https://www.wma.net/es/policies-post/declaracion-de-helsinki-de-la-amm-principios-eticos-para-las-investigaciones-medicas-en-seres-humanos/.
[47] I.H. Montano, I. de La Torre Díez, J.J.G. Aranda, J.R. Diaz, S.M. Cardín, J.J.G. López, Secure File Systems for the Development of a Data Leak Protection (DLP) Tool Against Internal Threats, in: 2022 17th Iber. Conf. Inf. Syst. Technol., IEEE, 2022: pp. 1–7. Doi: 10.23919/CISTI54924.2022.9820170.