Applying CHAID algorithm to investigate critical attributes of secured interoperable health data exchange

Shalini Bhartiya*
Institute of Information Technology and Management,
GGSIP University,
New Delhi 110027, India
Email: shalinibhartiya69@gmail.com
*Corresponding author

Deepti Mehrotra
Amity School of Engineering and Technology,
Amity University,
Noida, Uttar Pradesh, India
Email: mehdeepti@gmail.com

Abstract: With enhanced interoperability in healthcare environment the exchange of electronic health records (EHRs), both intra and inter organisations, has increased manifold. Sharing of the EHR creates room for illegal disclosures and confidentiality breaches. Interoperable healthcare is a complex system with many independent components. To design a secured framework for such a system, one need to understand the most important security attributes and predict various dependencies among them. The security attributes selected for statistical analysis are taken from the real-time study of patient-doctor relationship existing in any hospital or clinic. Hospitals with functional EHR-systems are the prerequisite of this study. The dependencies in the obtained data are generated through classification technique, chi-squared automatic interaction detection (CHAID). The decision tree obtained is analysed and verified using regression. The paper enabled the identification of the salient feature controlling which would maximally reduce security threats while sharing EHRs in interoperable healthcare environment.

Keywords: CHAID; chi-squared automatic interaction detection; classification; confidentiality; EHRs; electronic health records; interoperability; privacy; security issues; sharing.


Biographical notes: Shalini Bhartiya is a postgraduate from Kurukshetra University and pursuing PhD in Computer Science and Engineering from Amity University. Currently, she is working as an Assistant Professor at IITM, affiliated to Guru Gobind Singh Indraprastha University, New Delhi, India. She has more than 10 years of experience in teaching and industry. She had published about 10 papers in national and international conference proceedings.
1 Introduction

Electronic health record (EHR) is a repository in which all clinical and diagnostic information (Conklin and McLeod, 2009) of the patient is stored on the electronic device and made available for processing and managing patient’s record whenever required. The electronic medium enables hassle-free approach to fetch the required information in a fraction of seconds. This has remarkably improved the services in the healthcare sector. The patient can have an access to the best medical assistance and consultation anywhere across globe even remotely. But developing interoperable systems for smooth and effective sharing of healthcare information is a major concern (Virtanen et al., 2011) especially maintaining the security aspect of data.

With simplicity of access comes the challenges of maintaining security, i.e., privacy and confidentiality of EHR (Bhartiya and Mehrotra, 2013a), particularly when data is accessed across different hospitals or clinics. These records contain different sensitivity with respect to disclosure to intended users. A multi-level disclosure and confidentiality is demanded, which depends upon several factors, like type of ailment (psychological), societal and professional environment (epilepsy, HIV) and individual attitude. Apart from interoperability issues (Bhartiya and Mehrotra, 2013b) controlling and managing authorisation, accountability, integrity and confidentiality of data during transit require much attention. There is a need to have a common standard data exchange policy that may help to resolve these issues. Walsh et al. (2008), Young et al. (2010) and Dust and Lin (2003) have discussed federal and state regulations like Health Insurance Portability and Accountability Act (HIPAA) identifying unresolved challenges to patient privacy in e-health.

Extensive literature survey (Table 3) has been performed to have a deep understanding of various parameters that affect the security of EHRs in interoperable healthcare exchange. To further instigate the study, a questionnaire was designed covering the identified parameters. It was distributed to various healthcare stakeholders that are involved in sharing of data. To identify heterogeneity involved in the health data, the selected parameters are classified in a manner that predicts the model based on obtained dependencies among them. As the given dataset is categorical in nature, hence, chi-squared automatic interaction detection (CHAID) was considered as the most suitable multivariate analysis (MVA) technique for generating a decision tree based on observed dependencies. Further, regression relationship was obtained on the generated hierarchies and their relevance discussed with given standards and policies detailing the guidelines.
Applying CHAID algorithm to investigate critical attributes

for sharing of EHRs. An automated tool, SPSS, is utilised in the laboratories of Institute of Information Technology and Management, New Delhi and Amity University, NOIDA for obtaining the statistics and interpretation.

Remaining paper is divided into five sections. Section 2 introduces the usage of EHRs and concentrates on the security aspects of sharing of EHRs with respect to different stakeholders. It gives a roadmap of security breaches occurred in the past and also identifies the quantity of access available to each user of EHR. Section 3 briefs about the methodology adopted in studying the data picked up through questionnaire. The statistical techniques used for providing crisp and clear visual sense on the deciding factor for the enhancement of security controls in an interoperable sharing is discussed. In Section 4, analysis of the observations and results is done. The reflections and prognostications are formed on the basis of the solutions generated. Section 5 interprets the generated classifications trees with respect to identified security concerns and EHR sharing, using regression analysis method. Section 6 concludes stressing on the most viable control deemed necessary for smooth and secured sharing of records and discusses its implications with current standards and policies of EHR.

2 Literature review

Healthcare providers differ in their demands, preferences and authority in determining access to the EHRs. These stakeholders do not have similar authority to access EHR and are normally independent and not restricted or controlled by any centralised organisation. HIMSS Survey 2011–2014 indicates the amount of access made available to each stakeholder of the EHR.

The numbers in Table 1 signifies physicians accessing the major component of the EHR with patient being the lowest accessor. The trend from 2011 to 2014 depicts doctors having consistent access to the data with an increasing trend of patients’ access over time. This demonstrates the enhanced awareness of the patients demanding right to information and the authority of their health records. The figures reveal, with every passing year, nurses are gaining more access to EHRs. This may lead to unauthorised data exposure to nurses and other paramedical staff which is also a matter of security concern.

Table 1: Remote access available to various stakeholders of EHR (HIMSS Survey 2011–2014)

<table>
<thead>
<tr>
<th>Stakeholders accessing EHR from a distant location</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physicians</td>
<td>95</td>
<td>97</td>
<td>97</td>
<td>96</td>
</tr>
<tr>
<td>Physician extender</td>
<td>76</td>
<td>85</td>
<td>81</td>
<td>75</td>
</tr>
<tr>
<td>Nurses/nurse practitioners</td>
<td>56</td>
<td>60</td>
<td>57</td>
<td>70</td>
</tr>
<tr>
<td>Professionals (i.e., Occupational therapists)</td>
<td>–</td>
<td>–</td>
<td>60</td>
<td>63</td>
</tr>
<tr>
<td>Other clinical non-clinical staff (i.e., transcriptionists)</td>
<td>55</td>
<td>59</td>
<td>52</td>
<td>61</td>
</tr>
<tr>
<td>Patients</td>
<td>19</td>
<td>23</td>
<td>28</td>
<td>36</td>
</tr>
</tbody>
</table>

An extensive literature review has been made for identifying the security issues pertaining to sharing of EHRs in interoperable healthcare domain. Annual HIMSS Leadership Survey, 2010, 2011, 2012, 2013, 2014 identifies major areas of security breach in healthcare data (Table 2). Internal and external breaches, information leakage,
unauthorised access to data and accessing information on mobiles involve data passing through various stakeholders.

Table 2  Percentage of security breaches observed through HIMSS survey

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal breach of security</td>
<td>Authorised data disclosure, access control</td>
<td>51 36 32 27 30</td>
</tr>
<tr>
<td>Inadequate support for security process</td>
<td>Access control, awareness, organisation policies, performance</td>
<td>20 8 32 4 7</td>
</tr>
<tr>
<td>Compliance with HIPAA security regulations</td>
<td>Data interoperability, awareness, organisation policies</td>
<td>18 30 34 28 38</td>
</tr>
<tr>
<td>Data leakage</td>
<td>Data interoperability, access control</td>
<td>17 16 12 16 15</td>
</tr>
<tr>
<td>Limits of existing technology</td>
<td>Data interoperability, organisation policies, performance</td>
<td>16 10 10 8 8</td>
</tr>
<tr>
<td>Connecting I.T. at hospital/remote locations</td>
<td>Authorised data disclosure, access control, internet access</td>
<td>16 6 5 4 5</td>
</tr>
<tr>
<td>Unauthorised use of data</td>
<td>Data interoperability, access control</td>
<td>13 8 7 5 8</td>
</tr>
<tr>
<td>External breach of security</td>
<td>Data interoperability, authorised data disclosure, access control, organisation policies, internet access</td>
<td>12 11 13 10 17</td>
</tr>
<tr>
<td>Securing information on mobile devices</td>
<td>Data interoperability, internet access</td>
<td>Not Not 6 36 24</td>
</tr>
</tbody>
</table>

Each of the stated breach is due to lack or inefficient implementation of one or many security controls. As observed, external sharing of data imposes higher threat to security as compared with internal sharing. Literature was explored to identify the most common and important parameters (Table 2) for obtaining the optimal solutions in mitigating various security breaches while sharing EHRs in different healthcare setups. The parameters are studied solely from healthcare perspective, so as to obtain the relevant reasons contributing to security gaps in interoperable sharing of EHRs. The parameter that may cause the security breaches are identified and mapped with different types of breaches (Table 2). Examining the major cause of the above mentioned security breaches, access controls and policies take a lead (Hu et al., 2006) in determining consistent and workable solutions for the sharable healthcare environment. In Table 3, the effect of these security parameters on interoperable environment was explored. Designing of a secured framework that could control and monitor these parameters is the need of the hour.
Table 3
Cause-effect matrix representing the security issues in interoperable health data sharing

<table>
<thead>
<tr>
<th>Cause</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Redundant and inconsistent data storage</td>
<td>Friesen et al. (2008)</td>
</tr>
<tr>
<td>Authorised data disclosure</td>
<td>Friesen et al. (2008), Fleeter (2013) and Coma et al. (2008)</td>
</tr>
<tr>
<td>Access control</td>
<td>Friesen et al. (2008), Appari and Johnson (2010) and Hu and Scarfone (2012)</td>
</tr>
<tr>
<td>Awareness</td>
<td>Ghazvini and Shukur (2013)</td>
</tr>
<tr>
<td>Organisation policies</td>
<td>Hu and Scarfone (2012)</td>
</tr>
<tr>
<td>Internet access</td>
<td>Hu and Scarfone (2012)</td>
</tr>
<tr>
<td>Performance</td>
<td>Hu and Scarfone (2012)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frequent shifting of duties and department</th>
<th>Friesen et al. (2008)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obtaining patient’s consent</td>
<td>Appari and Johnson (2010) and Fleeter (2013)</td>
</tr>
<tr>
<td>Mode of data access</td>
<td>Friesen et al. (2008)</td>
</tr>
<tr>
<td>Access to paramedical staff</td>
<td>Renner (2001), Knott (2012) and Mundl et al. (2001)</td>
</tr>
<tr>
<td>Application response time</td>
<td>Bourke and Wessely (2008) and Walsh (2008)</td>
</tr>
<tr>
<td>Security policies and standards</td>
<td>Susilo and Win (2006)</td>
</tr>
</tbody>
</table>

| Fabbri et al. (2011) and Hupferich et al. (2012) |
| Flierer (2013) and Lee and Gostin (2009) |
| Wiljer et al. (2008) |
| Lo and Parham (2010) |
| Beard et al. (2012) |
| Hu and Scarfone (2012) |

| Coulkin and McLeod (2009) |
| Lee and Gostin (2009) |
| Hu and Scarfone (2012) and Knott (2012) |

| Gajamayake et al. (2011) |
| Varshney (2007) |
| Hu and Scarfone (2012) |

| Anton et al. (2007) |
| Hu and Scarfone (2012) |

| Borovicka (2008) |
3 Methodology

A multi-site cross-sectional questionnaire (Appendix A) was designed and piloted for use in administering the opinions of health domain stakeholders. The questionnaire composed of questions related to major security parameters – confidentiality, integrity and availability. The questionnaire was categorised on certain parameters (Table 4), surmised crucial with respect to the secured interoperable sharing of health data.

<table>
<thead>
<tr>
<th>Security parameters</th>
<th>Related questions</th>
<th>Question -count</th>
<th>Weightage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data interoperability</td>
<td>Access in interoperable systems leads to superfluous and inconsistent information storage, higher chances of breaches when shared between independent organisations</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Authorised data disclosure</td>
<td>Frequency of sharing of password, frequent shift of duties and departments, allowing additional access rights, obtaining patient’s consent, frequency of sharing between departments, method of accessing EHR between different hospitals</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>Access control</td>
<td>The Doctor has full access rights to all records, access available to experts and specialists, separate logins for doctors-on-call, patient control, the person authorising access, data accessed by paramedical staff</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>Awareness</td>
<td>Patient’s awareness</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Organisational policies</td>
<td>Frequency of change of password, Patient identity measures, frequency of breaches experienced, level of physical security deployed, the person deciding the quantity of data to be shared, Amount of data to be shared, monitoring and controlling shares, data unavailability experienced, response time during heavy rush, EHR systems compliant with international security standards, monitoring policies deployed</td>
<td>11</td>
<td>37</td>
</tr>
<tr>
<td>Internet access</td>
<td>Media used to access the data, level of system security deployed, security tools available, data unavailability experienced, response time during heavy rush, EHR systems compliant with international security standards, monitoring policies deployed</td>
<td>7</td>
<td>23</td>
</tr>
<tr>
<td>Performance</td>
<td>Detracted system’s performance due to strict security policies, data unavailability experienced, response time during heavy rush</td>
<td>3</td>
<td>10</td>
</tr>
</tbody>
</table>

A stratified random sampling method was used for taking the samples from a healthcare population. The samples as that of HIMSS Survey were referred to decide the population for this study. The population composed of physicians, laboratory technicians, paramedical staff, patients, pharmacists, interns and pathologists as also emphasised in (New York State Department of Health, 2007). Google drive and emailing were used as the medium of accumulating the responses apart from manual conduct of interviews. A pilot study (30 respondents) was conducted for over a period of three months to bring
Applying CHAID algorithm to investigate critical attributes

forth clarity and ease in understanding each question. Modified questionnaire was sent to over 1200 healthcare stakeholders. A total of 550 responses were received.

The focal point is to identify the most suitable solution enabling secure sharing of EHR in interoperable healthcare environment. Holding on the target in proximity, variables have been identified related to practical situations that touch on the confidentiality and privacy of EHRs during sharing. The dependent variables frequently observed in real-time healthcare environment are taken under the consideration of the regulations (HIPAA Quick Reference Guide for Employees, 2009; Young et al., 2010) set by standard bodies for accessing and sharing EHRs among independent healthcare units. For example, obtaining patient consent is required before giving any health maintenance services, but its feasibility is questionable.

The data collected comprises multi-attribute values. Since MVA is considered to give reliable results on such data, various techniques are studied and a suitable technique (CHAID) is adopted to analyse the data. An important step in CHAID prediction model is the selection of relevant parameters for classification. The aim is to perform sensitivity analysis to interpret the importance of a variable in the decision tree. The analysis identifies the dependencies related to the following security parameters and further analyses whether predicted values in reality have any substantial relationship with each other. Regression is applied to identify relationships between obtained dependencies. The results obtained are then discussed with respect to security issues in consultation with various standards and policies like HIPAA and HITECH.

3.1 Security parameters

The questionnaire was designed to cover the security parameters and the relative distribution of these parameters in the questionnaire is shown in Table 4.

3.1.1 Data interoperability

The study (Walker et al., 2005) explores the health data exchange and interoperability challenges when shared by various stakeholders. Data interoperability is linked with data integrity. Assuming that with all interoperable challenges and technological constraints, data can be shared. The information goes through several clients with different access rights even with similar designation and uses in the authorised hierarchy. Ace of the observations reveals that doctors visit multiple hospitals attending patients as per the availability, convenience and requirement of facilities. Occurrence of breaches and thefts drastically multiply with undefined roles in interoperable exchange of information.

3.1.1.1 Doctor treats patients at various sites

The doctor is usually on panel of more than one hospital. The patient can approach any of the hospitals to which the doctor is associated with. As a result, this gives rise to redundant or inconsistent storage of EHRs (Appari and Johnson, 2010), being stacked in different hospital information systems (HIS), which becomes a major challenge maintaining confidentiality and integrity of EHRs. Deficiency of central controlling authority, multiple ways of access and conflicting security policies opens a gamut of risks requiring deployment of integrity checks and access control mechanisms commonly accepted and embraced by independent hospitals or clinics.
3.1.2 Authorised data disclosure

To minimise the disclosure risk of patient’s EHR, concrete authorisation and disclosure policies must exist in the hospitals. People accessing EHRs in sharable environment must be bound with identification, proof along with authorisation and admission limitations. Who should bear the agency to permit/disallow access, how much information should be portioned out, how the access can be controlled are fundamental decisions ought to be looked after while disclosing the sensitive and confidential health information. Certain disclosure control parameters have been identified for applying statistical analysis and infer the importance and significance of each with regard to secured sharing of EHRs.

3.1.2.1 Patient’s consent

HIPAA (Walsh et al., 2008) mandates patient’s consent before enabling anyone the access of his or her EHR except in emergency cases. The heterogeneity and technicality of the data and varied sharing requirements makes it difficult to clearly obtain the relationship of this mandate with actual patterns. Extrapolating the concept of getting the patient’s consent every time before accessing data by the concerned health professionals, say, doctors and paramedical staff, can suffer grave repercussions (Fleeter, 2013) on allowing timely availability of their services to the patients. The patient may or may not be in the state of giving consent or approval every time his records need be accessed by respective doctor or hospital thereby delaying the quality and time of care. It requires seeking a robust solution enabling relevant and authorised access of EHRs to care providers.

3.1.2.2 Frequency of sharing

There exists the need of accessing patient’s records at multiple sites or departments usually managed independently. Sharing of EHR between hospitals, usually in referral cases and otherwise is a usual phenomenon. The frequency of sharing EHR by health providers is an important characteristic to measure its confidentiality and integrity. The higher the sharing, the higher are the chances of confidentiality and integrity loss. Availability of EHRs must encompass authorisation check and consolidation of relevant information permitted to be transported.

3.1.2.3 Frequent shifting of duties

Shifting of duties from one medical department to another is a common phenomenon, especially among paramedical staff and interns. This internal scheduling of a hospital lays heavy on the confidentiality and privacy of patient’s EHRs. Sharing of data goes along without any consent or control of the patient. With the shift of duties, access rights also must be changed. As discussed (Coma et al., 2008), it needs a restricted platform which can dynamically ascertain granting and simultaneous revoking irrelevant or unwanted access rights of such users of EHR.

3.1.3 Patients’ awareness

On one hand, technical advancements in health informatics enable easy access of health records to doctors, while at the other, make them vulnerable to malicious intrusion.
As per the guidelines by NIST and other statutory bodies, the patient should have complete dominance and mastery over his or her health records. It requires to be complemented with measuring the consciousness of a patient in not only ensuring confidentiality and secrecy of their EHRs but also overseeing and holding it over the time. Not every human can be tech savvy. Medical terminologies are highly complex and beyond the agreement of a layperson. Patient-centric command and monitoring might result in unmanaged, incomplete EHRs thereby adversely affecting treatments and quality of maintenance.

3.1.4 Access control

Access control policies specify how access is managed and who, when and why and under what conditions can access what information. Access control models apply a layer of abstraction between the resource and the user. It involves a mapping from the model to mechanism depending on the user’s requirement and permits. An access control request is generated in the system and granted on the basis of access policies and rules defined for each user in the system. Accessing EHRs in interoperable, but independent systems demand well-defined data ownership policies (Ministry of Health and Family Welfare, 2013) regardless of any region or nation.

3.1.4.1 Doctor control

Considering doctors as prime users of the EHR, strengthen their ability to hold complete control on the availability of information. Hospitals or clinics maintain, retains and controls all the medical records of a patient under treatment, that too more specifically applicable in IPD patients. The patient usually gets a consolidated summary of treatment on discharge. Patient consent is obtained briefing him/her or the attendants for every service or treatment provided, but the control of entire records remains with health providers. A mass of debate is leading on to ensure the custody and ownership of EHRs.

3.1.4.2 Patient control

EHR contains the observations and investigations of patient’s diseases. Accessing EHR requires patient’s authorisation (Hupperich et al., 2012). Much, the viability of patient, controlling access to healthcare professionals delay healthcare services and raises discontentment among them. With regard to security, patient’s control may enhance self-efficacy (Ilioudi et al., 2010) along with confidentiality and integrity, adversely affecting the availability and completeness of records accessed for sharing.

3.1.5 Organisational policies

From a small clinic to full-fledged hospital, there exist organisational regulations and policies (Ministry of Health and Family Welfare, 2013) for standardised and organised functioning of the system (Conklin and McLeod, 2009). Policies and rules contradict while integrating different schemas, attributes, constraints and languages. The conflicting access control policies weaken the linkage probability between disparate organisations. The repercussion of conflicts exposes sensitive health information to confidentiality and privacy breaches. A market study conducted and analysed by Gaynor et al. (2012) models the occurrence and severity of data breach incidents.
3.1.6 Internet access

Accessibility to EHR is affected by the source or the medium through which the data travels. Securing data on the net is always questionable. Monitoring, standardising and keeping pace with the advancement in technology is challenging for IT people. It increases the complexity, sharing information in a secured manner in such an environment. With an advent of mobiles and tablets, the technology has been well exploited enabling quick and comfortable access to required information in a timely fashion. Varshney (2007) describes various challenges and emerging problems in ubiquitous pervasive healthcare.

3.1.7 Performance

Healthcare activities are time-critical. Disparities in access control policies and standards deployed in EHR systems affect the reaction time, thereby delaying data availability to the respective users. Performance assessment of EHRs access device during heavy rushes of patients and also when the device is to be shared among health providers at the same time needs to be addressed.

3.2 Classification technique: CHAID

Health data’s heterogeneity and huge volume makes it difficult to track and secure while sharing electronically between different hospitals. Each hospital has its own set of rules and policies related to various security measures for sharing the EHRs. Thus, designing a model where all these security needs are met is a complex process. Few of these objectives are even contradictory to each other. Thus, there is a need to identify the relatively important parameter and their relation with other parameters. The statistical analysis allows better decision-making by unfolding associated risks and uncertainties which might remained unseen otherwise. The classification or decision trees make powerful use of conditional information in managing heterogeneous data.

Generating a model using predictive grouping of selective variables is referred as supervised learning. It is the statistical technique correctly interpreting or predicting data and building classification models from an input dataset. It is possible to identify variable relationships and interactions more directly from the data. The given set of data can be categorised and generalised using decision trees best describing mathematical and computational techniques in documentation with the psychoanalysis. This technique is used to estimate error rates which are defined as the probability of present classification function misclassifying the future observation. The bias in error rates can be avoided through various classification techniques. It includes various techniques (Wilkinson, 2004) of classifying variable dependencies, namely, predictive discriminate analysis, classification and regression trees (C&RT), CHAID, multilayer perceptrons (MLPs), radial basis function (RBF) and others. A discriminant analysis technique (Rencher, 2002), partitioning, divides the sample into two parts, training and validation, calculates linear or quadratic classification function with training sample evaluated using validation sample. The technique is best suitable on large sample and results in variable estimate of error rate. Another technique (Rencher, 2002), hold-out or cross-validation, computes the classification rules based on all $N$ observations. It is an improved method over partitioning but increases the computation load.
Applying CHAID algorithm to investigate critical attributes

Other decision-tree techniques such as nearest-neighbour technique (Bhatia and Vandana, 2010) are usually used in text retrieval seeking similar documents from huge database. Our concern is to predict the probable dependencies among various independent tasks while accessing health data in interoperable environment. Moreover, nearest-neighbour technique often result in computational complexities and memory requirements as it is based on computing distance from an observation to all other predictor variables. Clustering is a method of grouping records into segments to be used for further decision-making. It is best utilised with identical values for the particular predictors. Clustering (Rokach, 2010) comprising hierarchical and partitioning methods requires the knowledge of the number of clusters in advance and their use is limited to numeric attributes.

C&RT (Berson et al., 1999) uses entropy metrics to determine different predictions by comparing different predictors. It is capable of using predictors with continuous values and can validate the model while building it. CHAID (Kass, 1980) identifies the sampling invariability inherent in the data. It is an efficient method to heuristically search through large tables, that too, in computationally less expensive time. Some other similar techniques, namely, AID (automatic interaction detector) algorithm applied by Grimm et al. (2013), generate binary regression tree but require one outcome variable to be quantitative.

The data obtained through the questionnaire need to be grouped in a manner to generate classification rules that best describe dependencies between independent and dependent variables enabling reduction of redundant variables and leading to improved error rates.

Since the predictor variables are categorical, the classification method CHAID (Grimm et al., 2013) is applied to engender a decision tree by performing multi-level splits on the chosen predictor. It divides the data without losing any of the data. It generates easily understandable model and has proven to be easy to integrate with existing IT processes. The task is to determine the explanatory variable expressing significant effect on confidentiality and privacy of patient’s EHR. The data thus require segmented into groups to see all possible distributions of the explanatory variable.

3.2.1 CHAID algorithm for categorical predictors

For categorical predictor variables, splits are made between categorical values. Since these values are unordered, all possible splits between the categories need to be considered. CHAID incorporates a sequential merge and split procedure based on chi-square test statistic. CHAID proposed by Kass, also known as Kass’s algorithm, works on cross-tabulation giving due consideration to the computation time required for each predictor. It is a step-wise splitting process (Wilkinson, 1995) that performs cross-tabulation, merging, splitting and stopping based on critical value and defines test statistics that could be chi-square, t-test, F-test, etc. Another major characteristics of CHAID is the use of Bonferroni-adjusted P-values during multiple testing. It is an approximation set on n-categories to obtain the final number of optimally merged k-categories.
Steps taken:
At each node, CHAID determines for each potential predictor the optimal \( n \)-ary split it would produce, and selects the predictor on the basis of these optimal splits. CHAID uses \( P \)-values with a Bonferroni correction as splitting criteria.

- **Cross-tabulation:** The \( n \)-categories of predictor variable are cross-tabulated with \( k \)-categories of the dependent variable. Find the pair of categories of the predictor (only considering allowable pairs as determined by the type of the predictor) whose \( 2 \times k \) sub-table (\( k \) being the number of categories of the dependent variable) is least significantly different.

- **Merging:** Based on significant chi-square test, if two categories of a predictor variable in the sub-table are significantly different on the chi-square test, the categories are merged. Merging is repeated until no non-significant chi-square is found in a sub-table.

- **Splitting criteria:** Specifically, it uses the phi-coefficient (\( P \)-value) of the chi-square. The phi-coefficient (Wilkinson, 1995) is \( \frac{c^2}{n} \) for a \( 2 \times k \) table by the split of \( k \) categories of the dependent variable using Bonferroni multiplier (\( m \)). The way of calculating \( m \) depends on the type of predictor (Ritschard, 2010), ordinal, nominal or floating. Essentially, the predictor having highest chi-square value is split into \( l \leq n \) subsets, where \( l \) is the number of categories obtained by merging process. In the general case where the dependent variable has \( k \) categories, the \( P \)-values are obtained from a chi-square distribution with \( k - 1 \) degrees of freedom.

- **Stopping:** Various criteria and test values can specify when to stop further merging and splitting: For example, if all the cases in the nodes have identical values of the dependent variable, it is stopped for further split. Also, if the size of the node is less than defined minimum, the node does not split further.

### 3.3 Reliability assessment

The CHAID algorithm uses chi-squared tests for the decision-tree building. It saves calculation time, but does not guarantee to recover the best splits predicting accurate models. Further reliability to the study is ensured through regression analysis on the classified tree thus obtained. The significant linear relationship between independent and dependent variables is made using the \( t \)-test and \( F \)-test method. The slope significantly different from zero proves a substantial relationship between autonomous and dependent variables. The accompanying tables show the regression equation based on the obtained classification hierarchy. Regression explores associations between the classified hierarchy and overall view of the respondents. Regression analysis finds the slope (\( \text{Coef} \)), standard error (\( \text{SE Coef} \)), test statistics (\( F \)) and \( P \)-value (\( P \)) of the classified variables. The results obtained are further discussed with the risk estimate on observed classification. These estimates enable the identification of the security parameter(s) being affected and the standards and rules required to be addressed in finding the deciding factor for designing of a viable secured framework for sharing EHRs in interoperable environment.
4 Observations and results

Statistical analysis reported in this paper covers few of the above-mentioned security concerns. The thought process behind the observations presented below is to find the most suitable method of ensuring confidentiality and privacy of EHRs while sharing with varied healthcare professionals. Applying CHAID on the data reveals the following dependencies. The significance of the predictors is tested using chi-square. Table 5 shows each of the observed dependency classification with well-known chi-square statistics and significance value ($P$-value at 95% confidence) for each factor. Each of chi-square obtained impacts the concerned factor contributing to the enhancement of security controls while sharing of EHRs.

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Security control</th>
<th>Dependency classification</th>
<th>Chi-Square</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obtaining patient’s consent</td>
<td>Accountability and authorisation</td>
<td>The method used to access patient’s EHRs</td>
<td>78.240</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Increase in data breach</td>
<td>8.108</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Controlling authority</td>
<td>13.485</td>
<td>0.001</td>
</tr>
<tr>
<td>Patient’s awareness</td>
<td>Accountability and confidentiality</td>
<td>Frequency of sharing EHR by doctors</td>
<td>38.024</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Securing identity of patient</td>
<td>29.798</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Decide on the data content to be transferred</td>
<td>32.880</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Patient’s control</td>
<td>Authorisation, accountability and confidentiality</td>
<td>The method used to access patient’s EHRs</td>
<td>19.688</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Increase in data breach</td>
<td>41.765</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Full access rights to doctors for all EHRs</td>
<td>24.868</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Response time of application</td>
<td>15.282</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Access to experts and specialists</td>
<td>11.391</td>
<td>0.002</td>
</tr>
<tr>
<td>Doctor’s control</td>
<td>Authorisation, accountability and confidentiality</td>
<td>The patient has full control</td>
<td>46.228</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Patient’s awareness</td>
<td>24.428</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Higher sharing of EHR by doctors</td>
<td>Integrity and availability</td>
<td>Controlling authority</td>
<td>25.071</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The method used to access patient’s EHRs</td>
<td>20.619</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Accessing EHR at multiple sites</td>
<td>Integrity</td>
<td>The method used to access patient’s EHRs</td>
<td>25.457</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Policies adopted for monitoring</td>
<td>18.037</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Impact of strict security policies</td>
<td>18.875</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Table 5  Sub-nodes in each classification tree with their significant values (continued)

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Security control</th>
<th>Dependency classification</th>
<th>Chi-Square</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequent shift of</td>
<td>Confidentiality</td>
<td>EHR under control of patient</td>
<td>23.062</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>duties and</td>
<td></td>
<td>Security tools in place</td>
<td>14.073</td>
<td>0.001</td>
</tr>
<tr>
<td>departments</td>
<td></td>
<td>Maturity level of security in place</td>
<td>10.533</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Policies adopted for monitoring</td>
<td>17.575</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

5 Interpretation of generated classification hierarchies

Each node says N0, N1 and so on in the figures comprise multiple states like Strongly Agree, Agree, Somewhat Agree, Disagree, Strongly Disagree and others. The most prominent state as classified using CHAID and its impact on other variables is shown in the figures. For example, the state – Agree (43%) signifies the necessity to obtain the patient’s consent (Figure 2) before allowing any disclosure of his health records. Other states – Strongly Agree (25%), Somewhat Agree (17%), Disagree (12%) and Strongly Disagree (2%) were comparatively having low impact while drawing the decision tree. The full arrow represents sub-nodes and the dashed arrow represents leaf nodes. For example, in Figure 1, N3 and N4 are sub-nodes, whereas N1, N5, N6, N7, N8 and N9 are leaf nodes. A similar approach has been applied on further nodes and dependencies.

Figure 1  Patients controlling the access to Hisher health records
Applying CHAID algorithm to investigate critical attributes

5.1 Patients controlling the access to his or her health records

HIPAA Privacy Rule states patient holds the ownership rights and authority of his or her health records to allow anyone sharing or accessing the records. Respondents were asked to provide their opinion on patient having full control of his EHRs. The aggregated result varied between Disagree and Agree selecting Somewhat Agree as unified choice. Other questions in the classification hierarchy include increased chances of data breach while sharing between independent healthcare units as compared within the same hospital and should doctors hold full access rights on any EHR? Also, how much access should be allowed to experts and specialists called on case-to-case basis? Figure 1 shows the decision tree obtained using CHAID classification technique. The classification hierarchy identifies the dependencies affected when patient is given complete control and authority on his EHRs. A total of 33% respondents disagree on higher possibilities of data breaches. A total of 32% patients agree to provide full access to doctors. Patients who have the opinion of reduced or relevant access to doctors want visiting doctors to have only partial access (42%) of their records. Doctors holding complete access rights with patient controlling the access result in system’s low response time probably due to multiple authentication and role assessment ensuring authorised availability of the records.

On asking whether doctors should have full access rights on any patient’s EHR, 34% disagree whereas 29% viewed its need in some cases. People who agreed to allow the doctors enjoy full access on the health records found the patient’s awareness towards maintaining confidentiality of his records by the concerned authorities, as rare. The decision rules obtained conclude, if the patient control complemented with patient awareness to security of EHRs increases, doctors rights would automatically reduce. The predicted acceptance of this model is merely 31% indicating the researchers to take care of some more aspects concerning availability and security of EHRs.

Regression analysis in Table 6 reveals the most important parameter of the classification hierarchy with strongly influencing the access rights available to the doctors. The analysis reveals inverse effect on the access rights and privileges to doctors allowing them to access any patient’s records anytime. Following this model would reduce the data breaches and increase the probability of data availability to visiting doctors, i.e., enabling patient to obtain second opinions without going for repetition of laboratory tests.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef.</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>3.306</td>
<td>0.295</td>
<td>11.193</td>
<td>0.000</td>
</tr>
<tr>
<td>Chances of data breach</td>
<td>–0.131</td>
<td>0.048</td>
<td>–2.734</td>
<td>0.007</td>
</tr>
<tr>
<td>Full access rights to doctors</td>
<td>–0.195</td>
<td>0.043</td>
<td>–4.565</td>
<td>0.000</td>
</tr>
<tr>
<td>Application response time</td>
<td>0.066</td>
<td>0.054</td>
<td>1.222</td>
<td>0.222</td>
</tr>
<tr>
<td>Access to visiting doctors</td>
<td>0.132</td>
<td>0.053</td>
<td>2.482</td>
<td>0.013</td>
</tr>
</tbody>
</table>

Regression equation: Patient’s control = 3.3 – 0.13 * Data breaches – 0.19 * Full access rights + 0.07 * Response time + 0.13 * Visiting doctor access rights.
5.2 Mandatorily obtaining patient’s consent

Figure 2 illustrates the classification obtained keeping patient’s consent as dependent variable. It identifies that 43% stakeholders agree to the necessity of acquiring patient’s consent before sharing with legitimate users. Patient’s awareness and control on his data were found having no related impact on the observation. Rather, the mode of access, i.e., how the records would be accessed claim to be an important observation with maximum using unencrypted emails (57%) as a medium of sharing. This mode of access is susceptible to chances of data breach (70%) as opined by the respondents (Ahmad et al., 2009).

Figure 2 Mandatorily obtaining patient’s consent

Analysing the data obtained in Figure 2, since the $P$-value for each predictor is less than the significance level (0.05), there exist a linear relationship between the observed variables. A change in one independent variable keeping others constant would bring a relative change in dependent variable, i.e., obtaining patient’s consent for sharing his health records by legitimate users. Higher $T$ score and smaller $P$-value (Table 7) signifies the obtained relationship as most unlikely to be due to chance. From each of the predictor, the mode of access reflects high probability factor ($T$-value = 7.880) deciding the need of obtaining patient’s consent while sharing his EHR by various health professionals. As per the results obtained, if the mode of access is web, patients’ consent is moderate irrespective of the controlling authority that could be nurse, doctor,
Applying CHAID algorithm to investigate critical attributes

IT-team or HR-team. According to the analysis of variance, there appears a significant difference between obtaining patient’s consent and deciding on mode of access and the controlling authority taking into account the chance of occurrence of breaches \( F(3, 548) = 34.788 \) \( P < 0.001 \).

### Table 7  Significance of classification tree for dependent variable-patient’s consent using \( t \)-test

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef.</th>
<th>SE Coef.</th>
<th>( T )</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.915</td>
<td>0.207</td>
<td>9.258</td>
<td>0.000</td>
</tr>
<tr>
<td>Mode of Access</td>
<td>0.200</td>
<td>0.025</td>
<td>7.880</td>
<td>0.000</td>
</tr>
<tr>
<td>Chances of Data Breach</td>
<td>0.145</td>
<td>0.040</td>
<td>3.627</td>
<td>0.000</td>
</tr>
<tr>
<td>Controlling Authority</td>
<td>0.159</td>
<td>0.053</td>
<td>2.985</td>
<td>0.003</td>
</tr>
</tbody>
</table>

*Regression equation:* Patient consent = 1.9 + 0.2 * Mode of access + 0.14 * Data breaches + 0.16 * Controlling authority.

### 5.3 Frequent sharing of EHRs by doctors

Figure 3 illustrates the classification based on the frequency of sharing EHRs by doctors. A total of 33% respondents observed doctors often share patient’s EHR with colleagues. Other best observation in the same node N0 is Quite Frequently (29%). Frequent sharing seeks access controller determining the authority delegated to provide access to only relevant and required part of EHRs. The respondents found doctor-in-charge (38%) or the primary doctor treating the patient as the first choice holding the authority of such decision-making. Other prominent actors considered were patient itself and H.O.D of concerned department both rated equally (32%).

**Figure 3** Frequent sharing of EHRs by doctors
Requiring frequent access to health records by different specialisations and departments foster the need to determine other factors before making the data accessible. The regression equation in Figure 3 displays mode of access, with acceptable mode as web portal, having significant effect on managing legitimate access to all the professionals involved. The analysis (Table 8) shows no significant relationship between the authority controlling the access of EHRs and doctors frequently sharing these records thereby dropping it for further prediction. But this is an important observation that requires deep analysis of the behavioural and professional attitude of the doctor which is out of the scope of this paper.

Higher the sharing lower is monitoring. This is true in healthcare domain that dynamically generates the need and amount of data access needs from case to case. The complexities in applying monitoring techniques increase manifolds. The frequency with which the doctors share the records demands robust and reliable monitoring controls protecting illegal disclosures and confidentiality breaches. The results shown in Table 9 identify significant need of monitoring authorities. It may be IT-team or HR-team or IT and hospital-admin staff jointly, playing pivotal role in ensuring secured transmission of records between disparate systems.

Table 8  Significance of classification tree for dependent variable-doctors’ frequent sharing using T-test

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef.</th>
<th>SE Coef.</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>3.385</td>
<td>0.207</td>
<td>9.258</td>
<td>0.000</td>
</tr>
<tr>
<td>Controlling authority</td>
<td>−0.067</td>
<td>0.059</td>
<td>−1.121</td>
<td>0.263</td>
</tr>
<tr>
<td>Mode of access</td>
<td>0.088</td>
<td>0.031</td>
<td>2.883</td>
<td>0.004</td>
</tr>
</tbody>
</table>

*Regression equation: Doctors’ Frequent Sharing = 3.9 – 0.07 * Controlling authority + 0.09 * Mode of access.*

Table 9  Significance of classification tree for dependent variable-doctors’ frequent sharing using T-test

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef.</th>
<th>SE Coef.</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>3.499</td>
<td>0.236</td>
<td>14.816</td>
<td>0.000</td>
</tr>
<tr>
<td>Monitoring Authority</td>
<td>−0.219</td>
<td>0.052</td>
<td>−4.181</td>
<td>0.000</td>
</tr>
<tr>
<td>Used Monitoring Policy</td>
<td>0.082</td>
<td>0.037</td>
<td>2.242</td>
<td>0.026</td>
</tr>
</tbody>
</table>

*Regression equation: Doctors’ frequent sharing = 3.45 – 0.22 * Monitoring authority + 0.08 * Monitoring policy compliance.*

6 Conclusion

Healthcare dynamics and data sharing requirements by varied users in different and frequent-shifting roles threaten the confidentiality and privacy of patient’s health records. Stakeholders, primarily patient and doctor or organisation where the doctor is employed, are the owners of EHRs. The paper majorly covers the security issues keeping these two users in proximity. Frequency of sharing, types of doctors with varied needs and capacities, monitoring access bounded with different organisation policies are analysed.
Applying CHAID algorithm to investigate critical attributes

and tested using statistical techniques. The computational convenience given by minimum sum of squared errors (SSE) criteria (Table 10) provides the estimate that is unbiased and efficient. The variance of each predictor listed in the table clearly explains the statistical significance of the estimated model with respect to obtained dependencies. Risk estimated on the predicted model indicates the need of further investigating the affected parameter under the guidelines stated by various standards and policies. The point of focus is the access control parameter complemented with authorisations controls and monitoring during frequent sharing of EHRs by doctors.

Table 10  Summarised findings of each classified dependencies

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Minimum sum of squared errors criteria</th>
<th>Analysis of variance (F value)</th>
<th>Probability (F)</th>
<th>Risk estimate (if model is accepted) (%)</th>
<th>Security parameter(s) affected</th>
<th>Standards and policies needs to be addressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient’s control</td>
<td>56.849</td>
<td>12.24</td>
<td>0.000</td>
<td>60</td>
<td>Access control</td>
<td>Patient’s Authorisation. (HIPAA Quick Reference Guide for Employees, 2009), 164.312(a)(1) (Scholl et al., 2008)</td>
</tr>
<tr>
<td>Patient’s consent</td>
<td>85.86</td>
<td>34.78</td>
<td>0.000</td>
<td>45</td>
<td>Authorised data disclosure</td>
<td>NYSDH (2007), NIGB (2009), DH Informatics (2010), Disclosure Tracking- Purpose of Disclosure Definition Table (PDF) (HIPAA Quick Reference Guide for Employees, 2009), NIST SP 800-66 (Scholl et al., 2008)</td>
</tr>
<tr>
<td>Doctor’s frequent sharing-controlling authority</td>
<td>12.313</td>
<td>4.89</td>
<td>0.008</td>
<td>63</td>
<td>Access control, authorised data disclosure</td>
<td>Information Access Management – 164.308(a)(4), NIST SP 800-66 (Scholl et al., 2008)</td>
</tr>
<tr>
<td>Doctor’s frequent sharing-monitoring authority</td>
<td>28.127</td>
<td>5.64</td>
<td>0.008</td>
<td>57</td>
<td>Organisational policies</td>
<td>Assigned security responsibility 164.308(a)(8) – Evaluation 164.308(a)(1)(ii)(D) – Information System Activity Review (Scholl et al., 2008)</td>
</tr>
</tbody>
</table>

Analysis and interpretation of the obtained datasets reveal the importance of access controls from various perspectives. Controlling the accessibility of data is judged by considering two separate cases, doctor and patient holding complete access rights and data sharing authorities. Another effort is made to classify and analyse patient’s consent and control ensuring security of health records while sharing. Observations point towards determining and controlling the access rights requirements at each interactive node.
A clearly stated authorisation needs be defined between the users demanding sharing of data for legitimate purposes.

Hence, limiting and controlling access of EHRs to each independently participating user would ensure protection against confidentiality and integrity loss while dealing. The paper highlights the need to design and develop authorisation-controlled dynamic access control mechanism for enhancing the security of EHRs. Many other solutions persist like standardisation of organisation policies (Scholl et al., 2008); encrypting (Susilo and Win, 2006) data both during transit and while at rest, but the challenges of data interoperability and compatibility are yet to overcome. Moreover, there is a wide gap between the technological complexities and their usages in healthcare industry. Neither healthcare professionals nor the patients can be bounded to fulfil technological constraints and requirements for secured sharing of the data.

To conclude, the paper strictly recommends the need of designing viable access control techniques enabling decision-making through clear authorisations for sharing the data. Significant decisions when two legitimate users wish to share the EHR for relevant and defined purpose would be taken in more logical and timely manner. It would certainly cater to complex healthcare environment composing of dynamically shifting of roles, differently allocated access rights between two independent EHR systems, incompatible or conflicting organisation policies, undefined or differently defined authority and multiple mode of access any time and anywhere.

References
Applying CHAID algorithm to investigate critical attributes


Healthcare Information and Management Systems Society (HIMSS) Security Survey (2011) supported by Medical Group Management Association, Available at: www.himss.org

Healthcare Senior IT Executive (2012) 23rd Annual HIMSS Leadership Survey, Available at: www.himss.org

Healthcare Senior IT Executive (2013) 24th Annual HIMSS Leadership Survey, Available at: www.himss.org

Healthcare Senior IT Executive (2014) 25th Annual HIMSS Leadership Survey, Available at: www.himss.org


Applying CHAID algorithm to investigate critical attributes


Appendix A: Questionnaire

1. How frequently do doctors or others change their passwords?
   - Always
   - Quite Often
   - Sometimes
   - Rarely
   - Never

2. How frequently does doctors or others in healthcare services, share their passwords with colleagues?
   - Always
   - Quite Often
   - Sometimes
   - Rarely
   - Never

3. Do you think that doctors and paramedical staff would gain additional access rights over the time due to frequent shift of duties from one department to another?
   - Strongly Agree
   - Agree
   - Somewhat Agree
   - Disagree
   - Strongly Disagree

4. Do you think that every doctor of the hospital should have full access rights on any patient’s data?
   - Always
   - Quite Often
   - Sometimes
   - Rarely
   - Never
5. In your opinion, how much access is provided to experts and specialists who are called per-case basis?

<table>
<thead>
<tr>
<th>Complete Access</th>
<th>Partial Access</th>
<th>Only Required Details</th>
<th>Read-Only copy of required details</th>
</tr>
</thead>
</table>

6. Do you think that doctors who visit on call should be given separate logins for ascertaining secured access to the electronic health records of the patients?

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Somewhat Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
</table>

7. The doctor on-panel of more than one hospital can treat the patient at any location resulting in redundant and inconsistent storage of patient’s health information. Give your opinion on it.

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Somewhat Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
</table>

8. Do you think that patient should be given full control over his health data?

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Somewhat Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
</table>

9. What security measures are used in your hospital to identify the patients on every visit?

<table>
<thead>
<tr>
<th>Smart Cards</th>
<th>Unique Identifier</th>
<th>Biometric Devices</th>
<th>Any Other:……………….</th>
</tr>
</thead>
</table>

10. Give your opinion about obtaining patient’s consent for accessing or sharing his or her health information by various legitimate users.

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Somewhat Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
</table>

11. How many times have you experienced any kind of security breach to health data of the patients in your hospital/clinic?

<table>
<thead>
<tr>
<th>Quite Frequently</th>
<th>Often</th>
<th>Sometimes</th>
<th>Less Often</th>
<th>Never</th>
</tr>
</thead>
</table>

12. How much will you rate the level of physical security deployed in your hospital for preventing theft and illegal entry into the premises? (Point Scale: 1–7 with 1 as low and 7 as high)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |

13. How frequently do the professionals need to share health records between departments of the hospital?

<table>
<thead>
<tr>
<th>Quite Frequently</th>
<th>Often</th>
<th>Sometimes</th>
<th>Less Often</th>
<th>Never</th>
</tr>
</thead>
</table>

14. Do you think that chances to data breach are higher in data sharing between different hospitals as compared with sharing of data within the same hospital?

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Somewhat Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

15. What is the media used to access the health data of the patient? (Multiple Selection Possible)

- CDs
- Secured Email
- USB
- Third Party Services
- Thumb Drives
- Third Party Services
- Unencrypted Email Web Portals
- Any Other

16. In case of patient switching clinics/hospitals from time to time, how his or her electronic medical records should be made accessible to the doctors?

- Direct Access to the Data-store
- Partial Access to Data-store
- Required Data send as Message from the data-store
- Any Other

17. How much data should be transferred between hospitals/clinics for providing healthcare services?

- Complete Medical Records
- Only Current Records
- Relevant Records
- As Per Demand
- Any Other

18. According to you, who should decide about the amount of information to be provided to other hospital or doctor in case of referral cases?

- Doctor-in-charge of taking care of the patient
- Patient himself
- Doctor-in-charge and HOD of the Department
- Any Other

19. Do you think that patient is conscious about how well the hospital maintains the confidentiality and privacy of his or her health records?

- Absolutely
- Quite Often
- Sometimes
- Rarely
- Never

20. Who is authorising access of data to paramedical staff?

- Head Nurse
- Doctor
- IT Team
- HR Dept
- Any Other

21. What kind of data is accessed by paramedical staff? (Multiple Selection Possible)

- Complete Data
- Prescriptions
- Diagnostic Data
- Personal Details
- Any Other

22. How much mature is the security environment of your hospital? Rate your opinion on a point scale of 1–7 with 1 as least mature and 7 as highly mature.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
23. What are the security tools that are in place to prevent data loss and breach in your organisation?

- Firewall
- Encryption
- Single Sign-on
- Electronic Signatures
- Biometric Technologies
- Wireless Security
- Protocols
- Any Other: ………

24. According to you should the privacy of data in electronic storage be monitored and controlled by IT and Hospital staff jointly?

- Strongly Agree
- Agree
- Somewhat Agree
- Disagree
- Strongly Disagree

25. Do you think that strict security policies bring down the performance of the system in terms of speed and time?

- Strongly Agree
- Agree
- Somewhat Agree
- Disagree
- Strongly Disagree

26. How often do you experience unavailability of data due to some problem in the hospital management system installed at your workplace?

- Always
- Quite Often
- Sometimes
- Rarely
- Never

27. Rate the response time of the application during heavy rush of the patients specially in OPDs

- Very High
- High
- Moderate
- Low
- Very Low

28. Does the hospital management system incorporate international standards like ISO/IEC 27001 for ensuring confidentiality and integrity of health information?

- Always
- Quite Often
- Sometimes
- Rarely
- Never

29. Rate your satisfaction level in terms of performance of the system for displaying the required data from data-centre to the hospital/clinic over the network.

- Highly Satisfied
- Moderately Satisfied
- Less Satisfied
- Not Satisfied
- Not Applicable

30. What is most widely used for policy compliance monitoring in your hospital?

- Audit Logs of Firewall
- Information from servers
- Information from Applications
- Intrusion Detection Systems
- Network Devices
- Any Other: ………