Intelligent Automation & Soft Computing
DOI:10.32604/iasc.2022.023460
Article

Tech Science Press

Analyzing the Implications of Healthcare Data Breaches through Computational Technique

Ahmed H. Almulihi[1], Fawaz Alassery[2], Asif Irshad Khan[3], Sarita Shukla[4], Bineet Kumar Gupta[4] and Rajeev Kumar[4,*]

[1]Department of Computer Science, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia
[2]Department of Computer Engineering, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia
[3]Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[4]Department of Computer Applications, Shri Ramswaroop Memorial University, Barabanki, 225003, Uttar Pradesh, India
*Corresponding Author: Rajeev Kumar. Email: rs0414@gmail.com
Received: 09 September 2021; Accepted: 18 October 2021

Abstract: The contributions of the Internet of Medical Things (IoMT), cloud services, information systems, and smart devices are useful for the healthcare industry. With the help of digital healthcare, our lives have been made much more secure and effortless and provide more convenient and accessible treatment. In current, the modern healthcare sector has become more significant and convenient for the purpose of both external and internal threats. Big data breaches affect clients, stakeholders, organisations, and businesses, and they are a source of concern and complication for security professionals. This research examines the many types and categories of big data breaches that companies face. In addition, the study's main purpose is to investigate and draw conclusions from healthcare big data breaches, with the goal of improving healthcare big data confidentiality. From the beginning of 2020–21, both years, practically our entire world moved online due to the COVID-19 pandemic. The coronavirus pandemic dramatically increased the extent of use of technology. Hacking/IT incidents, followed by unauthorised internal disclosures, are the most typical attacks behind healthcare data breaches, according to the report. This has become a main enticement for healthcare data theft and misuse. The number of healthcare big data breaches, the number of records exposed, and the resulting significant commercial losses are increasing. In the report, they analyze and investigate the number of healthcare big data breaches, the number of records exposed during the COVID-19 pandemic. Also, the study assesses the correctness of the datasets in order to achieve a dynamic digital healthcare data breaches environment using fuzzy based computational technique.

Keywords: Big data security; healthcare big data breaches; cost effectiveness; big data analysis

1 Introduction

In the healthcare business organization, Electronic Health Record (EHR) systems have changed paper-based systems, providing additional economical and better services to their clients. EHRs improve patient care, increase practice efficiency, improve disease diagnosis, foster patient cooperation, and provide constant access to patient health information [1]. Moreover, mobile phones and other web-based smart devices have altered communication methods. These devices enable customers to access online services provided by many companies in a simple and comfortable manner, like in the healthcare sector. Healthcare big data has become increasingly digitised in recent years [2]. So the Internet of Medical Things (IoMT) plays a significant character. Healthcare organizations collect and store sensitive data on servers to make it more accessible at all times. However, privacy is increasingly being affected by the use of smartphones and other digital devices [3].

Moreover, when these big databases are exposed by unauthorized individuals due to software vulnerabilities and security breaches, then the result is sensitive data is exposed as a result of data breaches. This sensitive big data is protected by loss and theft from sensitive data disclosure or attackers of in perspective of healthcare [4]. In evaluation to other big data industries, the healthcare industry is one of the most affected [5].

According to several practitioners, the overall number of people affected by healthcare data breaches was 249.09 million from 2005 to 2019. In the previous five years alone, 157.40 million people have been affected [6]. In 2018, a total of 2216 big data breaches were stated from 65 nations. The health industry was hit with 536 of the 2216 breaches. This means that, as compared to other industries, the health industry has experienced the most data breaches [7]. In the year 2019, there were 13 big data breaches stated from 86 nations [8] and there were 41.2 million healthcare records exposed, stolen, or illegally released in 505 healthcare data breaches [8]. According to an analysis by IBM, the average cost of a big data breach was $3.92 million in 2019, with a $6.45 million cost in the healthcare industry [9]. The onslaught of COVID-19 data breaches has left a trail of victims, both businesses and individuals, globally.

In the year 2020, there were 17 big data breaches reported from 17 countries [8] and there were 29.3 million healthcare records exposed in 642 healthcare data breaches [8]. According to an IBM analysis, the average cost of a big data breach in 2020 was $3.86 million, with a $9.23 million cost in the healthcare industry [9]. Accordingly, in the year 2021, there were 17 big data breaches reported from 17 countries [8] and there were 3.350 million healthcare records exposed in 674 healthcare big data breaches [8]. According to an IBM analysis, the average cost of a big data breach in 2020 was $4.24 million, with a $499 million cost in the healthcare industry [9]. In comparison to other countries, the United States had the highest expense. A big data breach would normally be worth $8.19 million. The average cost of a healthcare data breach in the United States (average breach size of 38,000 records) is $10 million [10].

Big data assets are at risk in the form of individuals and organizations, as demonstrated by the data presented above. The most concerning, the target of cybercriminals are healthcare businesses, and thus they are the most vulnerable. As a result, both individuals and businesses are concerned about big data privacy and confidentiality. Healthcare big data needs expanded privacy and security, as well as the ability to withstand data breaches. Our main goal in this research was to discuss healthcare big data breaches that had been reported or published by a variety of reputable and reliable sources and use the findings to enhance healthcare big data security. The healthcare big data breaches, led by elements, will be addressed in future research work targeted at improving healthcare big data security.

The remainder of this research is separated into the segments below. The applied methodology is defined in the second part. The third section contains information regarding the sources of big data. The fourth section contains the investigation of big data breaches, providing perceptions into big data breaches that

are relevant to the healthcare industry. The next portion contains data analysis and related results through Fuzzy Analytic Hierarchy Process (fuzzy-AHP) based methodology. The ninth portion contains a discussion and a summary of the work's findings, while the last section details the conclusion.

2 Applied Methodology

The main purpose was to analyse healthcare data breaches. The goal of this review was to study more about the causes and effects of big data breaches on individuals and businesses. In this study, the authors used the sources to analyze information about healthcare data breaches and also other sectors in more detail. These sources are:

- The Privacy Rights Clearinghouse (PRC),
- HIPAA journals, Human Services (hss.gov) USA, and the Office for Civil Rights (OCR) Department of Health,
- The Ponemon Institute reports on big data breach costs, and
- Verizon Big Data Breach Investing to do this (Verizon-DBIR).

In the following segment, we'll go over various sources of information in detail. The format steps of big data analysis applied in this research are as follows:

Ist step-The above-mentioned sources are used to compile big data, which is then displayed in a tabular format.

IInd step-Then arithmetic operations like percentage, sum, and average are applied to this big data, and various categories of patterns are extracted.

IIIrd step-These patterns will aid in understanding the sources and repercussions of healthcare data breaches, as well as the increase and fall of big data breaches, the behaviour of various forms of attacks, and other significant topics included in the study's analysis section.

3 Big Data Sources

The following sources were used to obtain relevant data for the current study project shown in Tab. 1.

All of the big data sources are respectable and trustworthy, as well as globally recognised information sources that provide data breach reports. For this research project, we've also used the sources listed above to look at healthcare data breaches, their causes, and significances. These materials aided us in gaining a better sympathetic of the patterns of big data breaches.

4 Big Data Breaches Analysis

A big data breach is defined as the unauthorised disclosure or use of information and is also well-defined by the US Department of Health and Human Services as "the unauthorised use or disclosure of confidential health information that compromises the privacy or security of that information under the privacy rules and poses a sufficient risk of financial, reputational, or other type of harm to the affected person" [11].

Data breaches can cause harm to individuals and organisations in a variety of procedures. Apart from the major financial losses that organizations face as a result of big data theft, such incidents also damage the organizations' brand niche and brand value. There are two most common forms of data breaches, i.e, internal and external big data breaches. All cases of big data breach that are exposed through an organization's internal sources are referred to as internal big data breach. Privilege abuse, unauthenticated disclosure/access, inappropriate disposal of redundant but sensitive big data, theft or loss, or accidental delivery of secret big data to the wrong address are all scenarios that could occur. A third-party company

or source is responsible for external big data breaches. Among other hacking/IT problems, it could be a ransomware attack, malware attack, a spyware, phishing, or card fraud.

Table 1: Sources of big data

Sources	Descriptions
PRC big database	PRC is a non-profit organisation in the United States, formed in 1992 by Beth Givens. The organization's chief motivation is to protect users' information and advocacy services. Its main goal is to increase user awareness about the properties of technology on personal privacy and to establish guidelines for data management. It offers a comprehensive database of big data breaches. Various organisations have reported a total of 9016 big data breaches to the big database. According to the PRC big database, about 10 billion user records have been compromised since 2005.
HIPAA journal	The HIPAA Act of 1996 resulted in the publication of a HIPAA journal. It's a US-based publication that covers healthcare data breaches in depth, as well as HIPAA compliance and practical data breach prevention strategies. Since September 2009, it has been giving thorough information about data breaches in perspective of healthcare.
OCR reports	Every year, bi-yearly, or tri-yearly, the US Department of Health and Human Services' Office for Civil Rights publishes a "Report to Congress on Breaches of Unsecured Protected Health Information." From 2009 to 2017, these big data breach reports [11–13] provide detailed information about healthcare big data breaches.
Ponemon institute reports	The Ponemon Institute is a well-known research organization that emphases on subjects such as big data security, privacy, and information security, as well as legislation. It was founded in Michigan in 2002 by Dr. L. Ponemon. IBM supported the institute's reports, which include a compilation of verifiable records of big data breach expenses [13].
Verizon-DBIR	Verizon Enterprises' big data breach investigation reports are based on yearly assessments of big data breaches. Verizon published the first report of its kind in 2008. The papers detail big data breaches in both private and public organisations around the world.

From 2020 to 2021*, the Privacy Rights Clearinghouse (PRC), a non-profit organisation in the United States, recorded 9208 big data breaches across several sectors. Intentional Disclosure and Unknown Approach (UNKN), such as sending Bi (DISC). The following types of enterprises have been impacted by these big data breaches:

MED; Educational Organizations are denoted by EDU; Businesses-Financial, Insurance Institutes and Organizations are denoted by BSF; Businesses-Other are denoted by BSO; Businesses-Retail, including Online Retail, are denoted by BSR; Government and Defense Institutes are denoted by GOV; and Non-Governmental Organizations are represented by NGO [6].

Each sector's big data breach instances were likewise included in the PRC big database. Because no records were compromised in these attacks, the authors did not include these numbers in their reference to sector-wise picture of big data breaches. Subsequently extensive research on the PRC big database, the accumulated data is shown in Tab. 2.

Table 2: Sector-wise picture of big data breaches

Sectors	Big data breaches in last 15 years (2005–2019)		Big data breaches in last 5 years (2015–2019)		Big data breaches in last 2 years (2020–2021)	
	Breaches	Percentages	Breaches	Percentages	Breaches	Percentages
EDU	671	10.55	64	3.08	572	12.82
BSF	410	6.45	194	9.36	1143	25.63
BSO	426	6.70	113	5.45	132	02.96
MED	3912	61.55	1587	76.59	993	22.26
GOV	561	8.82	45	2.17	108	02.42
NGO	75	1.18	7	0.33	77	01.72
BSR	300	4.72	62	2.99	331	07.42
TOTAL	6355	99.97	2071	99.97	4459	100.41

In three situations, Tab. 2 depicts the sector-by-sector breakdown of big data breach instances. The first scenario is based on a collection of data breaches over the previous 15 years. The second situation, which is the emphasis of our investigation, charts the number of breaches in the healthcare industry over the last five years, and the third & last scenario is based on a collection of data over recent 2 years.

According to an assessment of the entire 17-year period, the healthcare (MED) sector had the largest number of big data breaches in three scenarios: (2005 to 2019), (2015 to 2019), and (2020–2021). There were 3908 breaches in the healthcare industry alone between 2020 and 2021, out of 6355 overall breaches. This accounts for 87.65 percent of the total. This is a huge cause for concern, and immediate action is required.

Fig. 1 depicts a graphical depiction of Tab. 2. The graph displays that, in the second scenario (2015–2019), the slope of each sector has dropped, with the exception of the MED sector, which has slightly grown. According to the graph, the healthcare engineering is the most popular goal of attackers due to the amazing financial significance of EHRs.

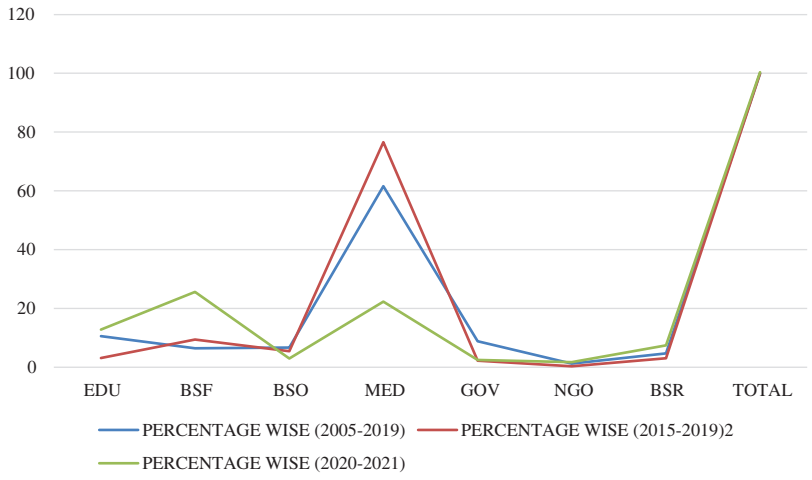

Figure 1: Picture of incidents of big data breach

4.1 HIPAA Big Data Breach Report Analysis

The authors categorized the big data on healthcare data breaches available in the HIPAA journal based on analysis results between 2010 and 2021*. To get results from these analyzed reports, the data was analyzed from a number of yearly, monthly, and other reports published by HIPAA magazine. It is not likely to deliver a link to every monthly & yearly HIPAA journal report that authors used to collect the data. We only mentioned valid sources of big data journals' key references in the results. After collecting records of big data from the same sources, we found calculable differences in various reports, such as the number of big data breaches recorded in 2019 being 505 in one HIPAA report and 512 in the other. A new study cites the increase in health data breaches during the COVID-19 pandemic. According to the studies of healthcare data breaches in the first six months of 2021, the authors found that the healthcare industry saw a 51% increase in breaches/leakages compared to 2019. In such circumstances, we select to use the big data from the most current report, which is shown in Tab. 3 [8,14–18].

Table 3: Healthcare big data breaches according to HIPAA

Year	No. of big data breaches	Exposed records (Millions)
2010	199	5.530
2011	200	13.150
2012	219	2.800
2013	280	6.950
2014	314	17.450
2015	270	113.270
2016	330	16.400
2017	360	5.100
2018	371	33.200
2019	512	41.200
2020	642	29.300
2021*	674	3.350
Total	4371	287.7

4.2 OCR Big Data Breach Report Analysis

From 2010 to 2019, the united big data of various reports published by the Office for Civil Rights with the label "Report to Congress on Breaches of Unsecured Protected Health Information" was given in Tab. 4. ORC has not yet made available the OCR big data breach reports for the years 2020 and 2021 [19–23].

Since ORC big data for the years 2020–2021 is not available, the united big data of HIPAA and ORC reports was compared from 2010 to 2019. According to a comparison of HIPAA and OCR healthcare data breach reports, we found a small difference in the number of data breaches which are reported each year.

According to HIPPA and ORC's big data reports, there were 3055 data breaches and 255.05 total reports of these healthcare breaches from 2010 to 2019, with the uppermost number of big data breaches stated in 2019, and the maximum number of records exposed in 2015, as shown in Tab. 4.

Tabs. 3 and 4 illustrate that in 2015, the healthcare business experienced a tremendous increase in big data breach cases, with more than 40% of health information being compromised. After 2015, 2019, was the

year with the most disclosed health records. The number of cases amounted to 16.14 percent of the overall of 255.18 million exposed health data between 2010 and 2019. In 2017, there were much fewer healthcare data breaches, with either 5.1 or 5.7 million records breached, according to the combined big data. The big data breach trend began to pick up speed in 2014, according to the overall report, and has been continuously expanding since then.

Table 4: Reported breaches of healthcare big data

Year	No. of big data breaches	Individuals affected (millions)
2010	207	5.400
2011	236	11.410
2012	222	3.270
2013	294	8.170
2014	277	21.340
2015	289	110.700
2016	334	14.570
2017	385	5.740
2018	302	12.200
2019	408	38.735
2020	————	————
2021*	————	————
Total	2954	231.535

5 Healthcare Big Data Disclosure Types

HIPAA big data breach reports have been thoroughly examined, and it has been discovered that the most common types of secured healthcare information leaks are illegal access (internal), hacking incidents, theft or loss, and inappropriate disposal of superfluous data. The approach to references [24–29] that we outlined in section 4.2 is likewise followed in this case. The following are brief descriptions of the various disclosure kinds listed below:

5.1 Incidents of Hacking

All cyber-attacks that are employed to acquire unauthorised access to confidential data are classified as hacking incidents. The most common hacking techniques used to disclose protected health information are ransomware and malware [30–33].

5.2 Unauthorized Access (Internal)

This term refers to all forms of attacks that result in the disclosure of confidential health data using any internal source within a company. It could be unauthenticated access/disclosure, misuse of privileges, and so on.

5.3 Theft or Loss

This refers to any occurrence involving the theft or loss of a hard drive, laptop, or other portable device carrying secured data of healthcare sector, which results in the disclosure of secured health information. This could be the result of catastrophic device failure or loss.

Big data that isn't needed but is sensitive and confidential should be carefully disposed of so that it can't be recovered. Sensitive health information could be revealed if big data is handled incorrectly. Improper disposal attacks are defined as incidents that occur as a result of the inappropriate disposal of avoidable but confidential and sensitive data in healthcare perspective. The number of healthcare big data breaches that happened as a result of the above-mentioned disclosure types is shown in Tab. 5.

Table 5: Category of healthcare big data breaches with different disclosure

Year	Hacking/ IT incidents	Unauthorized access/ disclosure incidents	Theft/Loss incidents	Improper disposal incidents
2010	8	10	148	10
2011	17	29	137	7
2012	18	28	148	8
2013	29	64	150	13
2014	39	87	149	12
2015	56	103	106	6
2016	115	130	79	7
2017	148	128	71	11
2018	168	140	54	10
2019	312	141	53	6
2020	429	143	54	16
2021*	—	———	———	——
Total	1339	1003	1149	106

The number of disclosure types is shown in the Tab. 5, which spans the years 2010 to 2020. Healthcare big data breaches with different disclosure types are not yet available in the healthcare data breaches report for the year 2021*. The following facts are highlighted:

- Between 2010 and 2020, the above-mentioned disclosure categories were used to carry out a total of 3597 breaching events.
- Separate hacking/IT incidents accounted for 37.22 percent of the breaches.
- Internal unauthorised disclosures were responsible for 27.88 percent of the breaches.
- Theft/loss cases accounted for 31.94% of the cases.
- 2.94 percent of the incidents were caused by the inappropriate disposal of sensitive big data that was no longer needed.
- Over a ten-year period, the statistics reveal that theft/loss is the most common, followed by IT incidents/hacking and unauthorised internal disclosure, with only a few instances of inappropriate disposal.
- When we looked back over the last four years, we noticed a sharp growth in IT incidents/hacking. Out of the 1339 IT incidents/hacking recorded during a ten-year period (2010–2020), 1057 occurrences

were reported in the last four years alone (2017–2020), accounting for 78.93% of the total, with 32.03 percent occurring in 2019.

According to our findings, other IT-related vulnerabilities/hacking have become a major worry for the healthcare big data industry in recent years. Internal disclosure and unauthorized access have risen in recent years as well, but not at the same rate as hacking incidents. Out of a total of 843 illegal internal disclosure instances, 542 were stated in the last four years. This represents 55.03 percent of the total, with 14.25 percent of the events taking place in 2019. When this percentage (16.84%) is compared to the same period last year (2019), it is apparent that hacking has climbed by 32.23 percent. The number of instances of improper internal disclosure has more than doubled. We also uncovered how the frequency and intensity of hacking attacks has increased, posing a severe threat to the healthcare business.

Theft/loss and improper disposal, on the other hand, have been reduced dramatically during the last four years. In the last four years, just 232 theft/loss incidents were reported, accounting for 23.86 percent of the total. Furthermore, in the last four years, just 43 instances of improper disposal have been documented, accounting for 40.56 percent of the total. According to their calculations, theft/loss and improper disposal have a minor negative impact on the healthcare industry. Fig. 2 is a graphical picture of the various categories of disclosures.

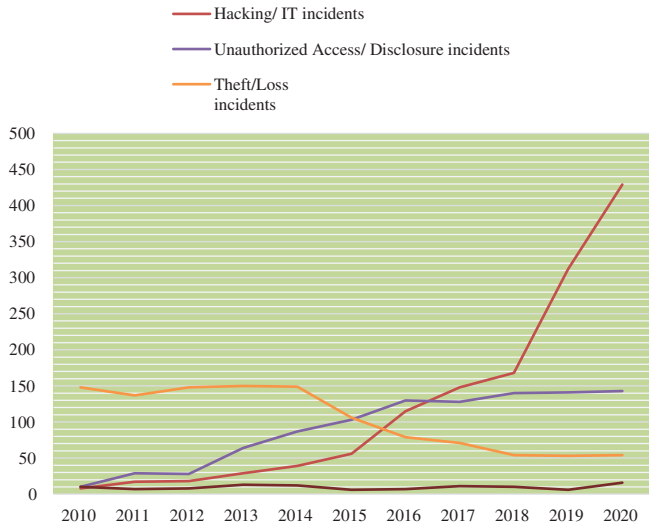

Figure 2: Graphical picture of different big data disclosure type

While theft/loss and improper disposal incidents have reduced in frequency, IT incidents/hacking and unauthorised access occurrences have improved in frequency, as shown in the graph. In recent years, there has been a significant growth in the number of hacking and IT mishaps. The next subsection will go through where information has been leaked and where sensitive health information has been disclosed/ breached.

6 Breached Locations

Papers or electromechanical storage devices are used to store protected health information. This section describes the sites where protected health information is breached using various methods. Tab. 6 shows year-by-year data on the locations of big data breach occurrences. Further, Tab. 6 contains big data derived from HIPAA and OCR reports. We used the similar technique in this case as we did in section 4.2. Because we referenced 24 distinct reports in 2018 and 2019, we have only supplied the most important references [34–37]. It would be impossible to include all of those sources in this study.

Table 6: PHI breached location

Year	Email	Laptop	Desktop computer	Other PED	Paper/ Films	Network server	EMR	Other	Total
2021*	19	0	2	1	3	42	6	2	75
2020	231	16	20	17	90	268	31	44	717
2019	214	24	34	15	61	132	39	52	571
2018	115	25	33	20	62	66	26	35	382
2017	92	20	38	18	64	82	34	39	385
2016	51	25	26	16	73	83	30	30	334
2015	32	34	29	19	70	52	27	26	289
2014	42	43	25	20	57	49	13	30	279
2013	26	70	40	23	60	31	16	32	298
2012	8	60	27	20	50	30	5	22	222
2011	3	48	32	30	65	21	5	32	236
2010	2	50	26	37	21	12	0	60	208
Total	835	415	332	236	676	868	368	404	3996

The sites where secured health information has been hacked are listed in Tab. 6 as Electronic Medical Records (EMR), Laptops, Desktop computers, Other Portable electronic devices, Network Servers, Email, Paper documents, and others. According to the investigation, the Paper/Film site is the most vulnerable to a breach of the eight. Out of a total of 3253 events, it has experienced 676 breaches. From 2010 to 2021*, this amounted to 16.91% of the total number of episodes. Paper/Films have taken the lead due to the inappropriate disposal of superfluous yet sensitive healthcare data. Films/Paper is followed by email, which accounts for 28.250.89 percent of the total, and network servers, which account for 21.17 percent.

With only 368 incidents, EMR are the least vulnerable. This accounts for barely 9.20 percent of the 3996 events reported during the same time period. Other Portable Electronic Devices (PED) come in second, accounting for 5.90 percent of the total. Desktop computers make up 8.25% of all computers. According to big data, attacks on network server locations and emails have increased significantly between 2017 and 2020. In the last four years, according to studies, the most prevalent causes of data breaches include obsolete security software, big database servers without passwords, and emails with weak or no passwords. Our research also found that the number of breaches of protected health information through paper, desktop computers, laptops, and films has decreased little over the last four years.

Our research confirms that cyber criminals are currently targeting sensitive healthcare data by employing various strategies, such as ransomware, malware, and phishing attacks [38–39] to prey on EHRs. Hackers are increasingly targeting email and network servers in order to gain access to sensitive health data. Fig. 3 depicts a comparison of different places based on the number of breached occurrences each site experiences each year. The results of this investigation will be better understood by the readers if they are presented in a graphical format. It will also assist readers in mapping the variance of healthcare data breach instances that occurred in certain areas over a ten-year period.

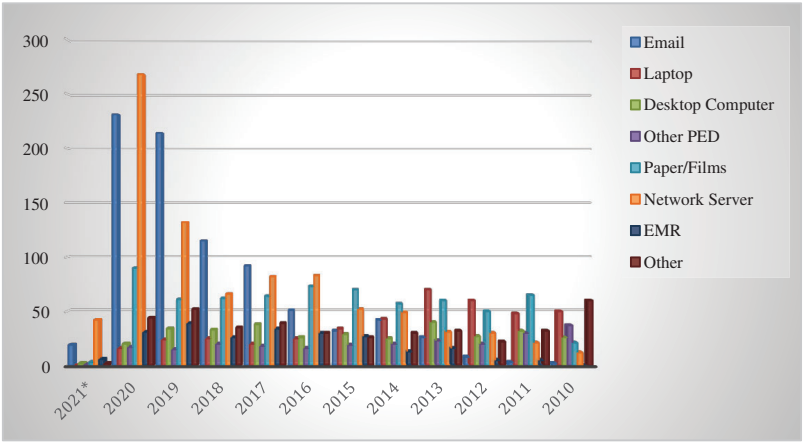

Figure 3: Graphical picture of comparative analysis

7 Financial Effect of Big Data Breaches

It's impossible to figure out just how much a data breach will cost. Many institutes have set characteristics and used various methodologies to assess the average cost of big data breaches. The Ponemon Institute collects both indirect and direct charges sustained by an organisation to calculate the average cost of a big data breach. This segment deliberates the financial effects of big data breaches, with a focus on healthcare data breaches. The Ponemon Institute's big data breach cost reports, which were supported by IBM, were used to calculate the financial effects of big data breaches on individuals, organisations, and countries. From 2010 to 2019, Tab. 7 outlines the cost of data breaches. Big data gathered from several Ponemon-IBM supported big data breach cost reports [40–43] is presented in the Tab. 7.

Table 7: Cost of big data breaches

Year	Average cost of breach (Millions)	Average cost per record	Cost per record in healthcare
2010	$7.24	$214	$294
2011	$5.50	$194	$240
2012	$3.20	$136	$233
2013	$3.29	$140	$296
2014	$3.50	$145	$359
2015	$3.79	$154	$363
2016	$4.00	$158	$355
2017	$3.62	$141	$380
2018	$3.86	$148	$408
2019	$3.92	$150	$429
2020	$3.86	$146	$7.13
2021*	$4.24	$161	$499

According to a big data breach cost study, the cost of a healthcare breached record has risen rapidly in comparison to the average cost of a breached record. The average cost of a record in 2010 was $214. That cost decreased by 10% in 2011. In comparison to the previous year, it decreased by 42.64 percent in 2012. After that, it continued to rise or fall year after year, with a 1.55 percent increase in 2019 over the previous year. Between 2010 and 2019, the average cost of healthcare enhanced by 45.91 percent, from $294 to $429. The cost of each breached healthcare record was $294 in 2010, and it continued to fall until 2012.

Between 2014 and 2015, it increased by 1.11 percent, 7.04 percent between 2016 and 2017, and 5.14 percent between 2018 and 2019. According to the Verizon DBIR research from 2018, financial gain fuelled 76 percent of big data breaches in 2018 [7]. According to the Verizon DBIR 2019 analysis, financial gain motivates 83 percent of healthcare big data breaches [17]. Fig. 4 depicts a cost assessment between average healthcare breached record expenses and breached record costs over the course of a year. In the following step of this research work, we'll employ time series investigation to figure out the pattern of healthcare data breaches and their cost.

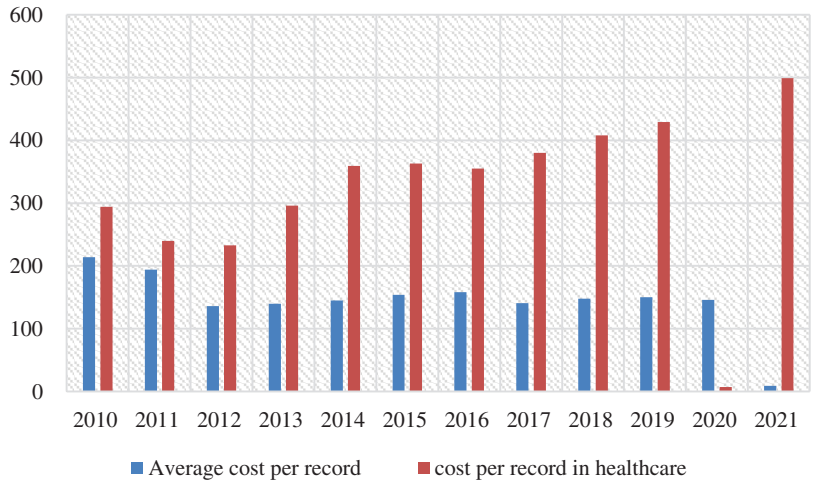

Figure 4: Graphical comparison of average record cost and healthcare record cost

8 Data Analysis and Results

AHP under the fuzzy environment has been used in this work for greater accuracy and efficiency. To determine the overall idealness and performance nature of the different sources of big data breaches in healthcare sectors, five above-mentioned sources of healthcare data breaches as attributes have been considered for this experiment. These attributes are symbolized as PRC Big Database (D1), HIPAA Journal (D2), OCR Reports (D3), Ponemon Institute Reports (D4), and Verizon-DBIR (D5) in the following tables. From 2010 to present year, the data analysis of these sources of big data breaches related to healthcare sectors and others are explained. With the help of [25–29], assessment of the supervised and the unsupervised approaches under fuzzy-based AHP environment has been examined as follows and Tab. 8 demonstrates fuzzy pair-wise comparison matrix based on collected expert's judgment.

Defuzzification is used after the comparison matrix is built to generate a measurable value based on the derived TFN values. The defuzzification approach used in this study was developed from [35,44–45], which is often referred to as the alpha cut method, as defined in [30–33]. The set of all items in a fuzzy set is called the alpha cut. Any number between 0 and 1 is used as the alpha threshold value. As a result, the alpha threshold value of 0.5 was chosen. Which have a membership value that is more than or equal to an alpha threshold value, denoted by α. A fuzzy set may be described as a collection of crisp sets using the

alpha cut technique. Crisp sets simply state whether or not an element belongs to the set. The alpha cut method is depicted in [34–37]. Tab. 9 shows the defuzzified final fuzzy pair wise comparison matrix.

Table 8: Fuzzy pair-wise comparison matrix based on collected expert's judgment

	D1	D2	D3	D4	D5
D1	1.00000, 1.00000, 1.000000	1.000000, 1.515007, 1.900331	0.480096, 0.630072, 1.000000	0.410052, 0.500743, 1.000000	0.220015, 0.200871, 0.415002
D2	–	1.00000, 1.00000, 1.000000	0.570043, 0.600657, 0.802200	0.300039, 0.393006, 0.500661	0.260079, 0.350021, 0.500176
D3	–	–	1.00000, 1.00000, 1.000000	1.000000, 1.310095, 1.550018	0.300009, 0.430052, 0.802007
D4	–	–	–	1.00000, 1.00000, 1.000000	0.530086, 0.910043, 1.583006
D5	–	–	–	–	1.00000, 1.00000, 1.000000

Table 9: Defuzzified final fuzzy pair wise comparison matrix

Datasets	D1	D2	D3	D4	D5
D1	1.000000	1.490012	0.690010	0.640010	0.300027
D2	0.671136	1.000000	0.670070	0.410043	0.372004
D3	1.449250	1.492380	1.000000	1.290077	0.490035
D4	1.562480	2.438770	0.775148	1.000000	0.960036
D5	3.333030	2.688140	2.040670	1.041630	1.000000
C.R. = 0.0286547000					

Tab. 9 shows the normalised values and defuzzified local weights of characteristics. The weights and priority of the characteristics have been established with the assistance of the local attribute-weights. Tab. 10 shows the results.

Table 10: Weight and priority of attributes

S. No.	Datasets	Weights	Percentages	Ranks
1	D1	0.132225	13.22%	4
2	D2	0.107036	10.70%	5
3	D3	0.197540	19.75%	3
4	D4	0.229325	22.93%	2
5	D5	0.333874	33.38%	1

The most important task for every researcher [36] is to validate the evaluated outcomes. The authors to this study compared the results to those of another comparable approach called the traditional Fuzzy-AHP in order to accomplish validation and offer a clear perspective of the acquired outcomes. The authors calculated the same data using the traditional Fuzzy-AHP approach with the same data. Tab. 11 and Fig. 5 show the outcomes of both methods. The findings of both approaches are strongly linked (person correlation coefficient is) [44–45], as shown in Tab. 11. Further, Tab. 11 shows that the fuzzy-based technique outperforms the classical methodology.

Table 11: Difference between fuzzy AHP and AHP

S. No.	Datasets	Fuzzy AHP		AHP	
		Weights	Percentages (%)	Weights	Percentages (%)
1	D1	0.132225	13.22	0.134575	13.45
2	D2	0.107036	10.70	0.115454	11.54
3	D3	0.197540	19.75	0.187740	18.77
4	D4	0.229325	22.93	0.230655	23.06
5	D5	0.333874	33.38	0.326584	32.65

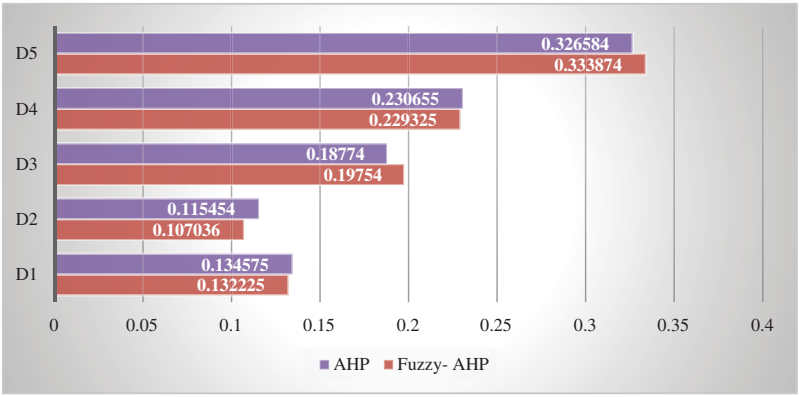

Figure 5: Comparison between the results through fuzzy AHP and classical AHP

Fig. 5 shows the graphical representation of the comparison between the results obtained from the Fuzzy AHP and Classical AHP including global weights and priority of the attributes in healthcare big data breaches. It can be clearly seen that Fuzzy-AHP is more efficient than Classical AHP.

9 Discussion

Smart phones, cloud services, information systems, internet access, IOMT, and other web-connected smart gadgets have enabled the healthcare industry to transition from paper-based systems to electronic health record systems. In information and communication technologies, healthcare big data has grown more digitised, distributive, and mobile. Despite the numerous benefits of EHRs, the digital health data of patients is currently under threat. As revealed in our investigation, the big data breach patterns for infiltrating sensitive big data have also undergone a significant transformation. The healthcare industry is certainly a target for cyber thieves, as demonstrated in this research. Furthermore, in order to gain

expertise and use it in our future studies, we analysed many big data breach reports created by various corporations and institutes. The study's final findings are, as summarized by us:

- From 2005 to 2021, more than 12 billion records from several industries were exposed. MED, EDU, NGO, BSF, BSO, BSR, and GOV are the types of these sectors.
- In the healthcare industry alone, there have been 4371 documented big data breaches. From 2020 to 2021*, about 85 percent of health data was compromised, the highest rate in any industry.
- According to HIPAA and OCR statistics, hacking/IT incidents are the leading cause of healthcare data breaches. According to HIPAA data, 4371 healthcare data breaches harmed 287.7 million people between 2010 and 2021* and an average of 3,343,448 healthcare records were breached.
- IT incidents/hacking, internal disclosure/unauthorised access, loss/theft, and improper disposal are the types of assaults employed to disclose protected health data. Theft/loss and inappropriate disposal, on the other hand, have seen a decline in the graph during the last three to four years.
- From 2020 to 2021, the number of hacking/IT incidents climbed by 37.22 percent. However, from 2020 to 2021, illegal internal disclosure, theft/loss, and inappropriate disposal fell by 29.88 percent, 31.94%, and 22.22 percent, respectively.
- By concentrating on the characteristics of healthcare data breaches, security measures will be improved, evaluated, and prioritised.
- When compared to AHP, MCDM methods like as Fuzzy-AHP give more efficient findings, and hence develop as a suitable technique for estimating big data breaches in healthcare.

The healthcare business is the most expensive, with a cost of $4.24 million for a big data breach, but the average total cost of a big data breach in 2021 was only $4.24 million [9]. A breached record costs on average $161. In the healthcare industry, however, the cost of each breached record in 2021* was $499 [24].

10 Conclusions

The authors believe that e-health data is extremely vulnerable because it is the most frequently targeted by attackers, based on their investigation of healthcare data breaches. According to a long-term investigation into big data breaches, healthcare records were exposed due to both internal and external threats such as hacking, theft/loss, unauthenticated internal disclosure, and inappropriate disposal of redundant yet sensitive data. According to the findings, both the quantity of big data breaches and their cost will enhance in the future. As a result, researchers, security experts, and the healthcare industry must prioritise and work on all preventive measures available. In addition, when creating a research study that attempts to provide perceptions of healthcare data breaches, there are a lot of other aspects to consider. Only the most important ones have been listed by the writers of this study.

- Identify and address the reasons for the majority of cyber-attack victims in the healthcare sector/organizations in recent years.
- The classification of IT incidents/hacking that result in healthcare data breaches, as well as the preventive steps that should be implemented to avoid them.

As research is a dynamic process, we cannot claim that our identified attribute collection is perfect, but it is a good starting point. Furthermore, the suggested assessment method, fuzzy-based AHP, is an effective but not optimum MCDM method. As a result, if possible, researchers can use additional approaches to improve their results. We shall concentrate our efforts in the future on the development of a theoretical framework.

Acknowledgement: We deeply acknowledge Taif University for supporting this study through Taif University Researchers Supporting Project Number (TURSP-2020/344), Taif University, Taif, Saudi Arabia.

Funding Statement: Funding for this study was received from the Taif University, Taif, Saudi Arabia under the Grant No. TURSP-2020/344.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] F. Kamoun and M. Nicho, "Human and organizational factors of healthcare big data breaches: The Swiss cheese model of big data breach causation and prevention," *International Journal of Healthcare Information Systems and Informatics*, vol. 9, no. 1, pp. 42–60, 2014.

[2] A. H. Seh, M. Zarour, M. Alenezi, A. K. Sarkar, A. Agrawal *et al.,* "Healthcare big data breaches: Insights and implications," *Healthcare*, vol. 8, no. 5, pp. 1–18, 2020.

[3] Team Health IT, 2020. [Online]. Available: https://www.healthit.gov/topic/health-it-basics/benefits-ehrs.

[4] T. T. Smith, "Examining big data privacy breaches in healthcare," *Walden University Journal*, vol. 4, no. 6, pp. 144–160, 2016.

[5] M. Chernyshev, S. Zeadally and Z. Baig, "Healthcare big data breaches: Implications for digital forensic readiness," *Journal of Medical Systems*, vol. 43, no. 1, pp. 1–28, 2019.

[6] V. Liu, M. Musen and T. Chou, "Big data breaches of protected health information in the United States," *JAMA*, vol. 313, no. 14, pp. 1471–1473, 2015.

[7] Internal Expert Team, 2020. [Online]. Available: https://privacyrights.org/Big.

[8] Expert Team, 2018. [Online]. Available: https://enterprise.verizon.com/resources/reports/DBIR_2018_Report.pdf.

[9] Expert Team, 2020. [Online]. Available: https://www.hipaajournal.com/december-2019-healthcare-BigData-breach-report/.

[10] Expert Team, 2020. [Online]. Available: https://www.ibm.com/security/Big.

[11] Expert Team, 2019. [Online]. Available: https://www.hipaajournal.com/2019-cost-of-a-BigData-breach-study-healthcare-BigData-breach-costs/.

[12] S. B. Wikina, "What caused the breach? An examination of use of information technology and health big data breaches," *Perspectives in Health Information Management*, vol. 14, no. 6, pp. 1–21, 2014.

[13] J. D. Collins, V. A. Sainato and D. N. Khey, "Organizational big data breaches 2005–2010: Applying SCP to the healthcare and education sectors," *International Journal of Cyber Criminology*, vol. 5, no. 1, pp. 1–21, 2011.

[14] Expert Team, 2020. [Online]. Available: https://www.hhs.gov/hipaa/for-professionals/breach-notification/reports-congress/index.html.

[15] Expert Team, 2020. [Online]. Available: https://www.hipaajournal.com/largest-healthcare-BigData-breaches-of-2016-8631/.

[16] Expert Team, 2019. [Online]. Available: https://www.hipaajournal.com/january-2019-healthcare-BigData-breach-report/.

[17] Expert Team, 2020. [Online]. Available: https://www.hipaajournal.com/healthcare-Big.

[18] Expert Team, 2019. [Online]. Available: https://enterprise.verizon.com/resources/executivebriefs/2019-dbir-executive-brief.pdf.

[19] Expert Team, 2018. [Online]. Available: https://www.hipaajournal.com/january-2018-healthcare-BigData-breach-report/.

[20] Expert Team, 2018. [Online]. Available: https://www.hipaajournal.com/december-2018-healthcare-BigData-breach-report/.

[21] Benchmark Research sponsored by Symantec, 2015. [Online]. Available: https://www.ponemon.org/local/upload/file/2015%20Global%20CODB%20FINAL%203%20copy.pdf.

[22] Expert Team, 2020. [Online]. Available: https://www.ibm.com/downloads/cas/7VMK5DV6.

[23] Expert Team, 2020. [Online]. Available: https://www.ibm.com/downloads/cas/ZYKLN2E3.

[24] Expert Team, 2020. [Online]. Available: https://www.ibm.com/downloads/cas/AEJYBPWA.

[25] Expert Team, 2020. [Online]. Available: https://www.ibm.com/downloads/cas/ZBZLY7KL.

[26] Benchmark Research sponsored by Symantec, 2013. [Online]. Available: https://www.ponemon.org/blog/2013-cost-of-BigData-breach-global-analysis.

[27] Benchmark Research sponsored by Symantec, 2011. [Online]. Available: https://www.ponemon.org/local/upload/file/2011_US_CODB_FINAL_5.pdf.

[28] Expert Team, 2020. [Online]. Available: https://citadel-information.com/wp-content/uploads/2010/12/2010-ponemon-report-global-cost-of-BigData-breach.pdf.

[29] A. Almulihi, 2020. [Online]. Available: https://pdfs.semanticscholar.org/6b6b/e8acb6cab384296 af67d636dda0891d5302a.pdf.

[30] Eva OSTERTAGOV, 2020. [Online]. Available: https://www.researchgate.net/publication/256086712_Forecasting_Using_Simple_Exponential_Smoothing_Method.

[31] Expert Team, 2021. [Online]. Available: https://enterprise.verizon.com/resources/reports/2021/2021-data-breach-investigations-report.pdf/.

[32] Expert Team, 2020. [Online]. Available: https://enterprise.verizon.com/resources/reports/2021/2021-data-breach-investigations-report.pdf/.

[33] Expert Team, 2020. [Online]. Available: https://www.hipaajournal.com/2020-healthcare-data-breach-report-us/.

[34] Expert Team, 2021. [Online]. Available: https://www.ibm.com/downloads/cas/OJDVQGRY.

[35] B. Beckles, V. Welch and J. Basney, "Mechanisms for increasing the usability of grid security," *International Journal of Human Computer Studies*, vol. 63, no. 12, pp. 74–101, 2005.

[36] K. Sahu, F. A. Alzahrani, R. K. Srivastava and R. Kumar, "Hesitant fuzzy sets based symmetrical model of decision-making for estimating the durability of web application," *Symmetry*, vol. 12, no. 6, pp. 1770–1792, 2020.

[37] A. Attaallah, M. Ahmad, M. Tarique, A. K. Pandey, R. Kumar *et al.,* "Device security assessment of internet of healthcare things," *Intelligent Automation & Soft Computing*, vol. 27, no. 2, pp. 593–603, 2021.

[38] S. A. Khan, M. Alenezi, A. Agrawal, R. Kumar and R. A. Khan, "Evaluating performance of software durability through an integrated fuzzy-based symmetrical method of ANP and TOPSIS," *Symmetry*, vol. 12, no. 4, pp. 1–15, 2020.

[39] R. Kumar, M. Zarour, M. Alenezi, A. Agrawal and R. A. Khan, "Measuring security durability of software through fuzzy-based decision-making process," *International Journal of Computational Intelligence Systems*, vol. 12, no. 2, pp. 627–642, 2019.

[40] R. Kumar, S. A. Khan and R. A. Khan, "Analytical network process for software security: A design perspective," *CSI Transactions on ICT*, vol. 4, no. 2, pp. 255–258, 2016.

[41] K. Sahu, F. A. Alzahrani, R. K. Srivastava and R. Kumar, "Evaluating the impact of prediction techniques: Software reliability perspective," *Computers, Materials & Continua*, vol. 67, no. 2, pp. 1471–1488, 2021.

[42] K. Sahu and R. K. Srivastava, "Soft computing approach for prediction of software reliability," *ICIC Express Letters*, vol. 12, no. 12, pp. 1213–1222, 2018.

[43] K. Sahu and R. K. Srivastava, "Revisiting software reliability," *Advances in Intelligent Systems and Computing*, vol. 802, pp. 221–235, 2019.

[44] W. Alosaimi, A. Alharbi, H. Alyami, M. Ahmad, A. K. Pandey *et al.,* "Impact of tools and techniques for securing consultancy services," *Computer Systems Science and Engineering*, vol. 37, no. 3, pp. 347–360, 2021.

[45] K. Sahu and R. K. Srivastava, "Needs and importance of reliability prediction: An industrial perspective," *Information Sciences Letters*, vol. 9, no. 1, pp. 33–37, 2020.