Intelligent Automation & Soft Computing
DOI:10.32604/iasc.2022.022569
Article

Tech Science Press

Analyzing the Big Data Security Through a Unified Decision-Making Approach

Abdulaziz Attaallah[1], Hassan Alsuhabi[2], Sarita Shukla[3], Rajeev Kumar[3,*], Bineet Kumar Gupta[3] and Raees Ahmad Khan[4]

[1]Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[2]Department of Mathematics, Al-Qunfudah University College, Umm Al-Qura University, Mecca, 24381, Saudi Arabia
[3]Department of Computer Applications, Shri Ramswaroop Memorial University, Barabanki, 225003, India
[4]Department of Information Technology, Babasaheb Bhimrao Ambedkar University, Lucknow, 226025, India
*Corresponding Author: Rajeev Kumar. Email: rs0414@gmail.com
Received: 11 August 2021; Accepted: 14 September 2021

Abstract: The use of cloud services, web-based software systems, the Internet of Things (IoT), Machine Learning (ML), Artificial Intelligence (AI), and other wireless sensor devices in the health sector has resulted in significant advancements and benefits. Early disease detection, increased accessibility, and high diagnostic reach have all been made possible by digital healthcare. Despite this remarkable achievement, healthcare data protection has become a serious issue for all parties involved. According to data breach statistics, the healthcare data industry is one of the major threats to cyber criminals. In reality, healthcare data breaches have increased at an alarming rate in recent years. Practitioners are developing a variety of tools, strategies, and approaches to solve healthcare data security concerns. The author has highlighted the crucial measurements and parameters in relation to enormous organizational circumstances for securing a vast amount of data in this paper. Security measures are those that prevent developers and organizations from achieving their objectives. The goal of this work is to identify and prioritize the security approaches that are used to locate and solve problems using different versions of two approaches that have been used to analyze big data security in the past. The Fuzzy Analytic Hierarchy Process (Fuzzy AHP) approach is being used by authors to examine the priorities and overall data security. In addition, the most important features in terms of weight have been quantitatively analyzed. Experts will discover the findings and conclusions useful in improving big data security.

Keywords: Big data; ioT; security assessment; fuzzy AHP

1 Introduction

The IoT is a current development in the present world in which all objectives are connected to the internet for increasing the quality of our daily lives without human interaction to share information [1,2]. The development of internet-enabled objectives has accelerated dramatically. As per the analysis report of

CISCO-2020, more than 50 billion objectives will be needed to be interconnected. IoT is a noble and sensible approach that makes physical things more accessible with minimum human effort. IoT has been used in a variety of applications, including smart healthcare, smart farming, and smart cities.

The amount of data acquired by these IoT objectives in numerous formats including unstructured and structured is growing by the day. Maintaining the security of huge data is a critical issue. The 4Vs of big data (Velocity, Value, Volume, and Veracity) reduce the level of security that is necessary. Several research works have been acknowledged for comprehending and categorizing the approaches for assessing security in order to improve big data security services [3–5].

The authors are attempting to re-establish big data security by evaluating it using several security approaches [6–10]. Although there is a lot of study in this field, there isn't much in the way of assessing the properties of big data security to world problems in the literature. The factors of large data security are crucial in ensuring security. With CIAAE (Confidentiality, Integrity, Authenticity, Availability and Efficiency), security factors may have an impact on big data security services. These factors each have a unique role to play in ensuring security [11–14]. Further, neglecting security properties is not an option for assessment.

Even so, assessing big data security is a management issue because each firm has its own approach [15–21]. The evaluation requires decision-makers to recognize preferences while maintaining large data security. Although, the Fuzzy-AHP approach is being used by the authors in this paper for the assessment [22–27]. It is critical to developing a hierarchy that identifies the properties that are impacted by big data security for this assessment. Big data security has been analyzed using the hierarchy and the Fuzzy-AHP. Finally, the findings and conclusions of this paper can be employed to improve big data security.

The following is the outline for this paper: The second segment goes over previous big data security work. The third segment covers the fundamentals of IoT, big data, healthcare big data analytics, and big data security. The procedure was provided in the fourth section, and the findings were discussed. Section five contains the conclusion of the proposed task.

2 Literature Review

Kambatla [4] presented a functional structure that recognizes the capture, processing, mining, and managing of IoT big data. Data storage module, a data processing module, data management module, data acquisition, and integration module, and application optimization module are some of the technical modules discussed. In addition, IoT applications, experiments, prospects, and certain open issues were highlighted, along with some important examples.

Hashem [5] updated the existing situation and planned guidelines for using remote health monitoring in combination. Before systems can be designed for seamless integration into clinical practice, various obstacles in sensing, analytics, and visualization must be addressed, according to this study.

Ali [6] presented research work on storage models, data privacy, data kinds, data security, and big data application analysis approaches. The recommended work also discussed the issues and expansion of big data in terms of forecasting current and future trends. Finally, in the era of big data, possible issues of big data security and machine learning approaches are addressed.

Hashem [7] projected a novel framework to process and analyze the massive amount of big data in a cloud context based on Hadoop from the healthcare perspective. This research work defined the significance of big data that is used to take accurate results for health practitioners by choosing the right care. Several approaches based on cryptography were employed to secure the framework. Map Reduce approach was employed for healthcare-based big data to increase performance. Further, Hadoop was employed to analyze the massive healthcare-based big data more appropriately.

Elijah [8] discussed a summary of big data analytics and IoT in the agricultural field. Precision agriculture, agriculture machinery, observing, tracing, tracking, and production were all briefly discussed as areas where IoT in agriculture is being developed. Demonstrated the significance of big data analytics and how it can assist with insurance, farm management, storage, decision-making, precision farming, and prediction from an agriculture perspective. Finally, the future issues and challenges that are open have been acknowledged and deliberated. Security and cost were identified as factors that this study found to be challenging in the farming industry.

Nalini [9] presented big data analytics and IoT. The challenges of big data and IoT were deliberated such as data security issues, big data storage, big data analytics, impact on day-to-day living. The inefficiencies occur in data collection including money, time, efforts, and loss of status when security was being compromised.

Siddiqa [10] revisited the advanced features of big data including background, history, and related technologies. Mainly, this work motivated big data analytic classes, architecture, stages, and explained big data issues or challenges specifically security, privacy, accurate decisions, heterogeneous sources, minimum energy consumption, and within time. More, the various applications of big data were described with some instances: government, healthcare, agriculture, telecommunication, transportation insurance, banking, manufacturing, retail, and customer products.

3 Big Data and Internet of Things

Sensors, gadgets, temperature sensors, healthcare applications, digital devices, and software programs generate massive amounts of unstructured, or semi-structured, and organized data that are all producing more data. Further, big data is the outcome of this tremendous data creation [6]. When it comes to keeping, treating, and analyzing rapidly rising amounts of large data, traditional database systems are inefficient [7]. The term “big data” has been used in the past, but it is comparatively different from the perspective of Information Technology (IT) [9]. The next frontier for improvement, competition, and production is an example of big data-related theories; were, [13] defined big data as the number of datasets that are a better database system tool than the typical tools for obtaining, keeping, treating, and analyzing such data [8]. According to the IT world [11], extracting value from a massive volume of data in many formats by allowing high-speed data discovery, analysis and capture are the main objectives of big data technology. Also, big data is divided into three categories including data sources, data analytics, and analytics findings presentations. The 3Vs (Volume, Variety, and Velocity) paradigm produced by Steed [11] that is used to define big data. The 3Vs model illustrates an e-commerce trend during managing the data that encounter volume or size management difficulties.

3.1 Types of Big Data

The big data are categorized into three ways which are given below:

3.1.1 Structured Big Data

Structured data is defined as data that can be monitored, accessed, and managed in a stable way [12]. It denotes to extremely structured material that can be stored and accessed from a database with ease using simple search engine approaches. For example, the personnel table in a firm database will be designed so that employee information, such as job titles, salaries, and so on, is organized.

3.1.2 Unstructured Big Data

Unstructured data refers to information that does not have a clear shape or organization [13]. Processing and analyzing unstructured data becomes extremely difficult and time-consuming according to the results. A

heterogeneous data source including a variety of simple text files, photos, videos, and email, for example, is one example of unstructured data.

3.1.3 Semi-Structured Big Data

Structured and unstructured large data can both be found in semi-structured data [14]. We can acquire semi-structured data in a structured format, but it is not defined by a table definition in a relational database management system. A data set contained in an Extensible Markup Language (XML) file is an example of semi-structured data.

3.2 Internet of Things

IoT provides a stage for sensors, things, and equipment that allows for unfettered communication in a smart environment and facilitates data sharing between sites [15]. The most recent version of wireless-based numerous technologies employs IoT as the next breakthrough tool, and internet technology will provide the opportunity. With evolving smart systems include smart offices, smart agriculture, smart retail, smart transportation, smart healthcare, smart water supply, and smart energy, IoT has gained different acceptance during constructing smart cities [16,17].

In recent years, the IoT has developed as a different style that may be utilized to collect data in mobile devices, transportation services, home applications, and general utilities [17–19]. Emergency alarms, wristwatches, garage doors, and vending machines as well as home appliances such as refrigerators, air conditioners, water heaters, and microwave ovens, are all connected to an IoT network and managed remotely [18–20]. Huge numbers of communication devices are incorporated in sensors in the IoT paradigm. Data is captured via sensors, and it will be sent via embedded communication devices. Several communication options, such as ZigBee, Bluetooth, Global System for Mobile Communications (GSM), and Wireless Fidelity (WiFi), are used to keep devices and objects connected [21].

The communication devices receive commands and exchange data from remote-controlled equipment, allowing for direct incorporation with the outside world via computers to enhance the quality of life. More than fifty billion gadgets, including smart sensors, laptops, phones, and game consoles are expected to be connected to the internet using emerging technologies [22–24]. IoT's most recent version of numerous wireless technologies is the next innovative technology that helps utilize all possibilities.

4 Big Data Analytics

Big data analytics is the procedure of mining, finding, and analyzing large databases in order to restore decision-making performance [1,3–5]. The capability to analyze large amounts of data can support a corporation deal with significant data that can have a negative impact on the company. As an outcome, the primary purpose of big data analytics is to assist corporations in better comprehending data and making well-informed decisions. Data miners and scientists can use big data analytics to investigate enormous amounts of data that aren't accessible with outdated tools [2,4].

Big data analytics necessitates a variety of technologies and approaches that may be used to convert massive amounts of semi-structured, unstructured, and structured data into a format that can be analyzed. Algorithms are frequently used in these analytical tools to find diverse trends, patterns, and interactions in data at different time intervals. The systematic tools depict the outcomes in the form of graphs, spatial charts, and tables for operational decision-making when the data analysis is completed. As the outcomes of the scalability and complexity of data, as well as the underlying algorithms that allow such approaches, big data analysis is a significant issue in many applications.

As a result, the focus of the issue is on the performance of current algorithms employed in big data analysis as the number of computational inputs continues to expand rapidly. While some technologies can

handle enormous data collections in a reasonable amount of time, big data analytics approaches take a long time to provide users with feedback and assistance. Contrary to popular belief, the majority of the remaining tools rely on complicated approaches such as trial-and-error to deal with these massive datasets and data heterogeneity.

4.1 Categories of Healthcare Big Data Analytics

Different forms of analytics are available based on the needs of IoT applications [5–7]. The real-time, memory-level, off-line, large level, and Business Intelligence (BI) level analytics categories are described in this subsection. Fig. 1 also includes a comparison based on the types of analytics and their levels.

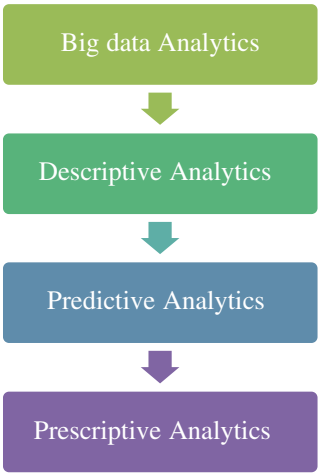

Figure 1: Healthcare big data analytic types

- Using business intelligence and data mining, descriptive analytics helps to demonstrate a picture of the past. Experience, as we all know, teaches us a lot. Using this data aids in the delivery of an approach for drawing a plan to achieve the goal.
- When compared to traditional business approaches, predictive analytics using large datasets helps to improve the client experience while also increasing the results. It enables the evaluation of large volumes of transactional and unstructured data at the same time, with the results assisting in the prediction of the future. Predicting the future based on usable datasets has been a difficult task since the dawn of computing till now. This form of business intelligence application aids in the computation of data streams with a broader reach, such as shopping experiences, social media content, survey results, and everyday user actions.
- The term "prescriptive analytics" refers to a tool that makes recommendations. By learning patterns, prior approaches, and data inputs, the system delivers suggestions based on the outcomes it has generated (unstructured and structured). This approach also permits for analyzing and recommending based on the conclusions of any other study on a similar action. By entering data from other research, it aids in connecting the dots and delivering answers based on both ascribed efforts. Since we cannot completely rely on machines, which are human inventions, they are only seen as possibilities. Though, depending on the findings, a thorough investigation into alternative solutions to a problem can be conducted.

4.2 Big Data Analytics Life Cycle

For data science projects and big data problems, the data analytic lifecycle is taken into account. To handle the many requirements for executing big data analysis, a step-by-step procedure is required to arrange the activities and tasks associated with data acquisition, processing, analysis, and repurposing (Tab. 1).

Table 1: Lifecycle of data analytics

Data Analytics Lifecycle	Descriptions
Discovery	• Data science investigates and learns about the problem. Improve context and comprehension. • Learn about the data sources that will be required and accessible for the assignment. • The team develops early hypotheses that can be verified with data later.
Model Planning	• Data science creates datasets for producing, training, and testing at this level. • Based on the work done during the model planning phase, the team builds and executes models. • Explores data to discover links between variables, then chooses crucial variables and the best appropriate models. • MATLAB and STATISTICA are two popular tools for this step.
Data Preparations	• Before modeling and analysis, there are steps to discover, pre-process, and condition data. • It necessitates the creation of an analytic sandbox, as well as the execution, loading, and transformation of data into the sandbox by the team. • Tasks for data preparation are anticipated to be completed several times and in no particular order. • Hadoop, Alpine Miner, Open Refine, and other tools are normally used for this phase.
Communication Results	• Next the execution of the model, the team must relate the results to the success and failure criteria defined. • The team evaluates how to communicate funding to numerous team members and stakeholders in the most effective way possible, taking into consideration suggestions and assumptions. • The team should categorize major financing sources, calculate the business value, and create a narrative to summarise and communicate funding sources to participants.
Model Building	• Create datasets for use in testing and production. • Teams also examine whether their current tools are adequate for running models or if they require a more robust environment for model execution. • Rand PL/R, Octave, and WEKA are examples of free and open-source software. • MATLAB and STATISTICA are commercial tools.
Operationalize	• The team communicates the project's benefits more clearly and establishes an experimental project to organize work in a controlled manner before expanding the project's complete enterprise of consumers. • This tactic allows the team to learn about the model's performance and restrictions in a production setting on a modest scale before deploying it fully. • Final reports, briefings, and codes are delivered by the team. • Octave, SQL, and MADLIB are examples of free and open-source software.

4.3 Connection Between Big Data Analytics and IoT

In the IoT, big data analytics is rapidly evolving for active decision-making. The study of "connected items" is one of the most essential components of IoT. Further, IoT and big data analytics necessitate the processing of a large amount of data and its storage in a variety of storage systems. Because the majority of unstructured data is collected in an open manner from web-enabled devices, big data solutions must combine light intelligence analysis with larger queries in order to obtain fast insights, make fast decisions, and communicate with humans and machines.

The integrated architecture and built operating system enable sophisticated applications by allowing groups of sensing and activating devices to share information on sites. In general, IoT expands the quantity and variety of data. In addition, it opens the door to big data analytics applications and development. The goal of big data technologies in IoT is to speed up the development of IoT and commercial models. Fig. 2 depicts the association between IoT and big data. To handle IoT data, three processes are involved. The first step is to manage IoT databases, which are used by networked sensors to communicate with one another.

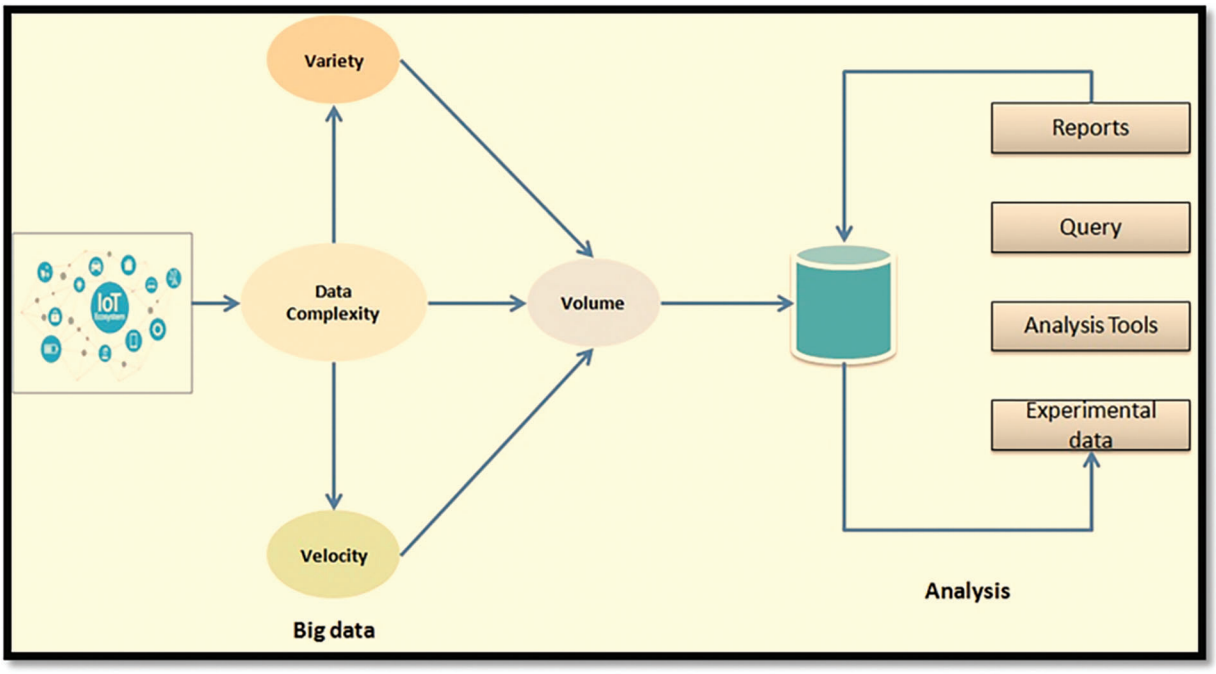

Figure 2: IoT and big data analytics relationship

For example, communication equipment like smart traffic lights, smart home devices, and Closed-Circuit Television (CCTV) cameras generate massive amounts of data in a variety of forms. The data will be kept in a low-cost cloud storage system. Data is created in the second phase, which is referred to as "big data" because of its volume, velocity, and variety. Shared distributed error handler databases will be used to store this vast amount of data. In the final phase, analytics technologies such as Map-Reduce are used to analyze big IoT datasets. The four levels of analysis start with experimental data and go to analytic queries, tools, and reports at the next level.

4.4 Big Data Analytics in Healthcare

By identifying connections from a large volume of healthcare data, big data analytics has the prospective to enhance the quality of care and lower patient medical costs by providing a broader viewpoint of clinical competence based on medical evidence and numerous tests. Advanced analytical approaches and tools used in healthcare systems deliver services that meet an increasing need by allowing healthcare organizations to handle large amounts of data in real-time, analyze in real-time, and remove data from all patients' medical records. The primary goal of Big Data Analytics (BDA) is to enhance the functioning of a healthcare facility so that people can live healthier lives. This includes numerous analytical tools for understanding factors linkages and discovering knowledge, including machine learning, pattern identification, visualization, and data mining.

Big data analytics is founded on the notion of data mining, which entails using a variety of analytical approaches to evaluate and examine massive amounts of data in order to extract relevant and usable information. The publications [8–11] may provide example proof concerning big data analytics and healthcare to the authors of this work. The goal of big data analytics in smart healthcare the volume of data in the healthcare business is always growing from numerous sources, but processing this data in hard or soft copy formats is quite difficult [12]. Data digitalization can help solve this problem, but one of the most difficult tasks is analyzing all of the data. When dealing with vast amounts of data, data analysis is critical because it will be utilized to make judgments.

Big data analysis creates a new method for healthcare systems that requires estimating the reasonable time for producing reasonable discretions, organizing future perspectives, and maximizing time value. At the same time, it assists health institutions in providing cognitive knowledge on their planning, management, and measurements. Finally, the estimated results are a tool to help managers improve their executive abilities. The usage of IoT data analytics in the healthcare business is depicted in Fig. 3.

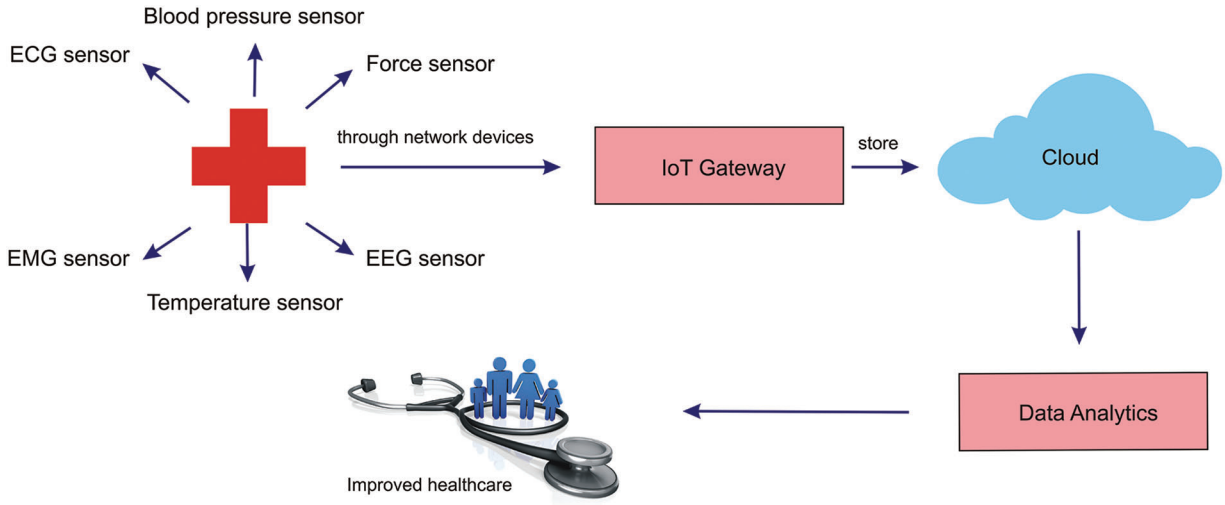

Figure 3: IoT data analytics concerning healthcare

5 Big Data Security: A Point of Concern

Big data is primarily classified into four elements based on the data: volume, variety, velocity, and value, which is referred to as the 4 V's model represented by Kumar et al. [13] (Fig. 4). This approach is primarily highlighted by an e-commerce trend in which data management issues include managing enormous amounts of data, a variety of data, and the speed with which data is generated.

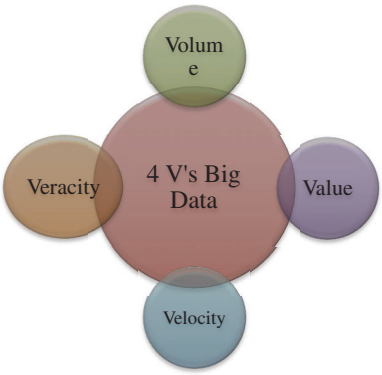

Figure 4: Big data's 4 V's in healthcare

5.1 Volume

Some factors contribute to the data volume and it could be transactional data that has been collected over time or data flowing through social media [14,15]. The data volume refers to the total amount of mass data in an organization. The volume of data created in an organization grows at a random rate, which can be measured in petabytes and zeta bytes depending on the organization's making actions and kind.

5.2 Velocity

The data in the full data transferred currently in an organization or motion is referred to as velocity [16]. The rate at which an organization generates, processes, and analyses data are usually increasing. It has an impact on how data is formed and sent from one place to the next. It's frequently time-sensitive.

5.3 Value

We must concentrate on value as a major issue. The amount of data we save or process isn't the only factor that determines value [17]. It is the amount of valuable, dependable, and trustworthy data that must be saved, processed, and analyzed in order to uncover insights.

5.4 Veracity

Veracity, which is made up of data about which the organization is unsure [18,19]. It examines the degree to which various types of data are regarded as reliable. Factors such as weather and consumer feedback, as well as reaching a choice, typically delay an organization's executing of processes to guarantee the quality and reliability of data.

Security has several significant factors of big data security that have an indirect impact. CIAAE is a set of factors. Auxiliary, CIAAE is one of the security pillars [3,5,11–14]. Confidentiality denotes to the permission of authorized access to sensitive and secure data in the context of security. Integrity is a demand-driven factor that is recognized by ethical assurance and resolutions. In the context of a computer system, availability denotes to a customer's ability to access information or assets for a specific amount of time. This work contributes to a fuzzy AHP-based assessment of big data security. Fig. 5 depicts a tree structure of big data security features [15–17].

Further, Fig. 5 shows that CIAAE and EDS (Efficiency, Durability, and Storage Capacity) affect the security of big data [13–17]. Big data security may be enhanced by targeting CIAAE with EDS together [15,17,19]. Hence, these factors should be involved in the assessment of big data security. We must protect data from leakage since confidentiality is the foundation of big data security and privacy. The value of the data will be lost if it is leaked [16]. If hackers target the data and change it or unearth secret

information, the value of the big data could be lost. Because big data security and privacy demand a lot of network bandwidth, efficiency is very important [17,18]. Authenticity is required to ensure that data sources, processors, and authorized data requesters are trustworthy [19,20]. Authenticity can help you avoid bad analysis results and get the most out of your big data.

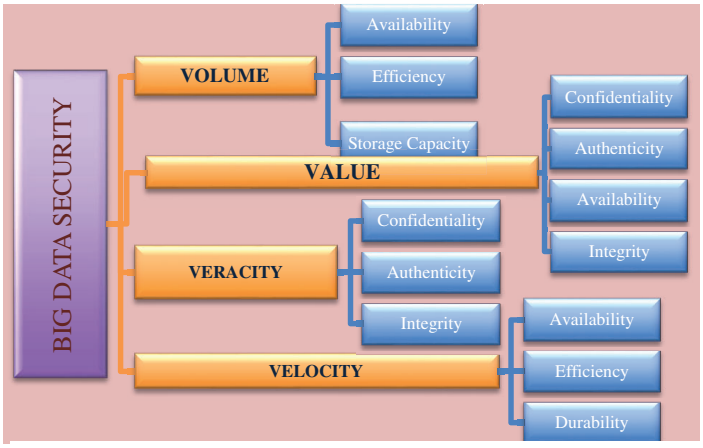

Figure 5: A tree structure of big data security factors

Big data should be available whenever we require it. Otherwise, its worth may dwindle. Integrity is also necessary for obtaining useful and reliable data. We cannot analyze the correct conclusion with wrong or partial data, especially when the missing data is the most sensitive and useful. The longevity of big data security for a given time period is characterized as durability [23]. Security durability has a considerable impact on big data security because the time limit of security has a substantial impact on total big data security. Storage capacity is a sub-aspect of big data security that works with various storage systems to determine how much data must be available in storage nodes at the same time.

6 Fuzzy Based Decision Making Approach

The major goals are to establish goals based on the weight and categorization of the security attribute of big data through a multi-criteria decision-making process that demonstrates the usage of analytical hierarchy, in order to create a secure and dependable network [3]. Because no efforts to objectively categorize and grade the big data security factors have been made, the Fuzzy AHP can be utilized to prioritize them [3,22]. Because it is useful in decision-making procedures to mitigate the problem of uncertainty and uncertainty, Fuzzy AHP produces more accurate findings than classical AHP [22]. This will improve the security and early detection of vulnerabilities, which will help consumers and organizations by increasing secure network capacity and durability. The AHP was used to explore the prioritization of large data security factors using a multi-criteria decision-making system [22,23].

Multiple-Criteria Decision-Making (MCDM) is an operation research sub-discipline that aids in the decision-making process by allowing for multiple assessments of conflicting criteria [24]. Further, when compared to other MCDM approaches, AHP is a superior approach for measuring the objective and subjective values of the qualities [24,25]. Regardless, the underlying vagueness and ambiguity found by drawing the consciousness of precise no's of a decision-maker are impossible to measure with AHP [25]. The author discovered that practitioners consolidated fuzzy theory with AHP since the real world is especially shaky for investigating uncertain real situations [25,26]. Furthermore, the AHP uses the matrix for pair-wise comparisons in "MCDM" situations, as shown in Eq. (1).

$$A = [a_{ij}] = \begin{array}{c} \\ C_1 \\ C_2 \\ \vdots \\ C_{n\,1} \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & \dots & C_n \end{array} \\ \begin{bmatrix} 1 & a_{12} & \dots & a_{1n} \\ 1/a_{12} & 1 & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1/a_{1n} & 1/a_2 n & \cdots & 1 \end{bmatrix} \end{array} [n*n] \tag{1}$$

where a_{ij}= 1 and a_{ij}j = $\frac{1}{a_{ij}}$, i, j = 1, 2… n

Here a_{ij} = 1 and a_{ij} = 1/aij, i, j = 1, 2… n. A n-by-n matrix, A can be articulated as per appeared in Eq. (1). Let C1, C2… Cn indicates the arrangement of the feature while a_{ij} stand for an evaluated decision on a couple of features Ci, Cj. The relative consequences of the two factors are appraised utilizing a scale [22,23]. The Fuzzy AHP approach entails four important steps, which are detailed below. The issue is then organized into a hierarchical graded framework for Fuzzy AHP to address. The study's purpose is to develop a hybrid model of MCDM large data security approaches. The huge data security factor structure is depicted in the diagram. This hierarchy can be created by employing expert opinions and replies in questionnaires, as well as brainstorming and other approaches. The Triangular Fuzzy Numbers (TFN) are then used to define the hierarchy.

To begin, the problem is divided into hierarchical scaled arrangements in order to be tackled with Fuzzy AHP. It should be represented very clearly, and a systematic balanced structure, as shown in Fig. 5, is created (Tree diagram of big data security factors). The goal of this research is to gather information for a hybrid model of MCDM approaches for large data security. The structure of large data security factors is depicted in the diagram. This hierarchy is created by employing expert opinions and responses in a questionnaire/opinion poll, or by brainstorming and other approaches.

There are a lot of unknown details in fuzzy set theory. A fuzzy collection is made up of membership range artifacts. A membership function defines such a factor, assigning a membership grade to each item that falls between zero and one (1). A fuzzy triangular form is depicted in Fig. 6.

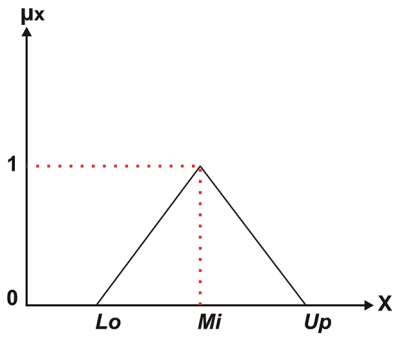

Figure 6: Comparison of the results

TFN shown mainly as (Lo, Mi, Up). The Eqs. (2–4) are used to convert the numeric values into "TFN" [24,25] and signified as "(Lo_{ij}, Mi_{ij}, Up_{ij}) where Loij is lower value, Mi_{ij} is middle value and Up_{ij} are the highest level events". Further, TFN [η_{ij}] is setup as the accompanying:

$$\eta_{ij} = [Lo_{ij}, \quad Mi_{ij}, \quad Up_{ij}] \quad Here \quad Lo_{ij} \leq Mi_{ij} \leq Up_{ij} \tag{2}$$

$$Lo_{ij} = min(J_{ijk}) \tag{3}$$

$$Mi_{ij} = (J_{ij1}, \quad J_{ij2} \ldots\ldots J_{ijk})^{1/k} \tag{4}$$

$$Up_{ij} = max(J_{ijk}) \tag{5}$$

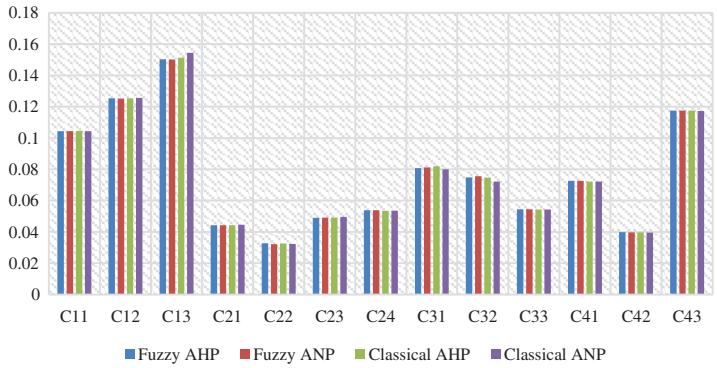

Figure 7: Triangular fuzzy numbers

Jijk displays the relative value of two parameters provided by the expert z in the example above. Where i and j are calculated parameters by experts. The geometric mean for a set of relationships determines this value. Stakeholder consensus, which reflects the lowest and greatest ratings for the relative value of the two parameters, will be accurately summarized and described. After collecting the TFN value for each pair of reference points, a fuzzy matrix of (n x n) dimensions is created. This research combines analysts and engineers with big data security experience. AHP analysis is completely reliant on volunteers who have been carefully chosen. After qualitative evaluation, the "TFN" membership task and pair-wise comparisons are determined to construct the confusing choice matrix in the third step.

Additional, when the matrix has been detected, defuzzification is performed to produce a quantifiable value based on assessed TFN values. Reverse fuzzification is defuzzification. The defuzzification approach used in this work was adapted from [26,27], which was created in Eqs. (5–7), and is known as alpha cutting.

$$\rho\alpha\beta(\tilde{A}) = [\beta.\tilde{A}\alpha(Lo_{ij}) + (1-\beta).\tilde{A}\alpha(Up_{ij})] \tag{6}$$

where $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$

Such that,

$$\tilde{A}\alpha(Lo_{ij}) = (Mi_{ij} - Lo_{ij}).\alpha + Lo_{ij} \tag{7}$$

$$\tilde{A}\alpha(Up_{ij}) = Up_{ij} - (Up_{ij} - Mi_{ij}).\alpha \tag{8}$$

For professional states, these computations employ and. These two numbers range from 0 to 1. The alpha cut of a fuzzy set is an all-element package. An integer between 0 and 1 is used as the alpha-threshold. It has an alpha threshold value that is more than or equal to the membership value set. The lesser and higher defused values are indicated by (Lo ij) and (Up ij). Eq. (8) depicts the matrix that was created after the participants' decisions were evaluated.

$$\rho\Gamma,\ \eta'(\tilde{A}) = \dot{\imath}'A\ \pm,^{2}\ [\tilde{a}ij] = \begin{array}{c} \\ C_1 \\ C_2 \\ \vdots \\ C_n \end{array} \begin{array}{c} \begin{array}{cccccc} C_1 & & & C_2 & \ldots\ldots\ldots\ldots & C_n \end{array} \\ \begin{bmatrix} 1 & & A'\pm,^{2} & (\tilde{a}\ 11)\ldots\ldots & A'\pm,^{2} & (\tilde{a}\ 1i) \\ 1/A'\pm,^{2} & (\tilde{a}\ 21) & 1\ldots.. & & A'\pm,^{2} & (\tilde{a}\ 2i) \\ . & & . & & . & \\ . & & . & & . & \\ 1/A'\pm,^{2} & (\tilde{a}\ j1) & 1/A'\pm,^{2} & (\tilde{a}\ j2)\ldots\ldots & 1 & \end{bmatrix} \end{array} \tag{9}$$

For a fuzzy number, aij indicates the relative relevance of two C1 factors. The ability to view a hazy package as a series of flat sets is sometimes aided by alpha reduction. Crisp sets, (a), effectively represent whether or not an entity belongs to a group. In Eq. (9), a single pair matrix is used. According to the outcomes, evaluating the peer-to-peer reference matrix is an additional step in this approach. Own vector computing is used to select the aggregated weight of specified norms. Expect that aij's own vector is denoted by, and that the unusual factor of aij is denoted.

$$[\rho\alpha,\ \beta(\tilde{A}) - \lambda I].\rho = 0 \tag{10}$$

Eq. (10) is based on a linear vector transformation, which I discuss in more detail below. Weights with specified standards could be constructed for every other potential criteria by adding Eqs. (1–9) to the equation. Now, calculate the Ratio of Consistency (CR). If the CR value is less than 0.1, the AHP measurement is calculated accurately a second time.

7 Data Analysis and Results

Big data security qualities are multi-dimensional and, in most cases, qualitative. Assessing the big data security aspects statistically becomes a difficult task. As a result, the relative weights and rankings of big data security criteria validate a major responsibility for network design that is highly secure. Tab. 2 shows a built aggregated fuzzy pair-wise comparison matrix that displays the priorities derived from many pair-wise comparisons at level 1. The fuzzified aggregated pair-wise comparison matrix at level 2 for sub-factors (C1, C2, C3, C4) is created by utilizing the geometric typical approach to analyze expert opinions (Eqs. (1–4)). Also, Tabs. 3–6 show the produced matrix fuzzy aggregated pair-wise comparison matrix at level 1.

Table 2: Pair-wise judgment matrix in fuzzy form at level 1

	C1	C2	C3	C4
C1	1.000000, 1.000000, 1.000000	1.75500, 2.34500, 3.03600	1.48500, 1.950075, 2.520063	1.120098, 1.550051, 1.989005
C2	-	1.000000, 1.000000, 1.000000	0.500700, 0.780060, 1.160000	0.560000, 0.720000, 0.960099
C3	-	-	1.000000, 1.000000, 1.000000	0.628006, 0.817005, 1.000756
C4	-	-	-	1.000000, 1.000000, 1.000000

Table 3: Pair-wise judgment matrix in fuzzy form for F1 at level 2

	C11	C12	C13
C11	1.000000, 1.000000, 1.000000	0.230075, 0.287009, 0.367005	0.342001, 0.447700, 0.824007
C12	-	1.000000, 1.000000, 1.000000	0.661004, 1.172005, 1.690036
C13	-	-	1.000000, 1.000000, 1.000000

Table 4: Pair-wise judgment matrix in fuzzy form for C2 at level 2

	C21	C22	C23	C24
C21	1.000000, 1.000000, 1.000000	0.690041, 0.895003, 1.110024	0.234005, 0.280078, 0.364001	0.710012, 0.950041, 1.350012
C22	-	1.000000, 1.000000, 1.000000	0.490031, 0.642003, 1.241004	0.271003, 0.350015, 0.520016
C23	-	-	1.000000, 1.000000, 1.000000	1.085004, 1.329007, 1.558002
C24	-	-	-	1.000000, 1.000000, 1.000000

Table 5: Pair-wise judgment matrix in fuzzy form for C3 at level 2

	C31	C32	C33
C31	1.000000, 1.000000, 1.000000	0.665003, 1.172003, 1.697004	1.157006, 1.447002, 1.704003
C32	-	1.000000, 1.000000, 1.000000	1.007007, 1.524007, 1.934003
C33	-	-	1.000000, 1.000000, 1.000000

Table 6: Pair-wise judgment matrix in fuzzy form for C4 at level 2

	C41	C42	C43
C41	1.000000, 1.000000, 1.000000	1.197008, 1.580083, 2.150064	0.491001, 0.642002, 1.009009
C42	-	1.000000, 1.000000, 1.000000	0.224001, 0.290056, 0.427009
C43	-	-	1.000000, 1.000000, 1.000000

With the help of Eqs. (5–8), this paper uses the cut approach of defuzzification. Furthermore, CR values are less than 0.1 (Eqs. (1–10)). Tabs. 7–11 display the defuzzified aggregated pair-wise comparison matrix and local weights at level 1 and level 2 factors, according to the hierarchical structure (C1, C2, C3, C4). The dependent or overall weights and hierarchy ranking are shown in Tab. 12. Further, Tab. 12 shows the weighted evaluation standards for the most important concerns to big data security, according to the hierarchy. The difference between Fuzzy-AHP and AHP findings is seen in Tab. 13 and Fig. 7.

Table 7: Defuzzified matrix and local weight at level 1

	C1	C2	C3	C4	Weights
C1	1.0000000	2.3723000	1.9810090	1.5560040	0.3800000
C2	0.4215000	1.0000000	0.8240030	0.7440070	0.1800000
C3	0.5046000	1.2100320	1.0000000	0.8300090	0.2100000
C4	0.6425000	1.3420080	1.2035000	1.0000000	0.2300000
CR = 0.0015004					

Table 8: Defuzzified matrix and local weight at level 1 for C1

	C11	C12	C13	Weights
C11	1.0000000	1.1700300	0.4900400	0.2748000
C12	0.8520050	1.0000000	1.1700200	0.3297000
C13	2.0240030	0.8530020	1.0000000	0.3955000
CR = 0.00245				

Table 9: Defuzzified matrix and local weight at level 1 for C2

	C21	C22	C23	C24	Weights
C21	1.0000000	0.8920000	1.1730000	0.9900400	0.2462000
C22	1.1210010	1.0000000	0.6900100	0.3700200	0.1821000
C23	0.8520050	1.4470020	1.0000000	1.2900800	0.2723000
C24	1.0060010	2.6880020	0.7700040	1.0000000	0.2994000
CR = 0.0025400					

Table 10: Defuzzified matrix and local weight at level 1 for C3

	C31	C32	C33	Weights
C31	1.0000000	1.1720000	1.3630000	0.3842000
C32	0.8533000	1.0000000	1.4910000	0.3564000
C33	0.7337000	0.6707000	1.0000000	0.2594000
CR= 0.0025000				

Table 11: Defuzzified matrix and local weight at level 1 for C4

	C41	C42	C43	Weights
C41	1.0000000	1.6330000	0.6910000	0.3158000
C42	0.6124000	1.0000000	0.3030000	0.1731000
C43	1.4472000	3.3003000	1.0000000	0.5111000
CR= 0.0052000				

Table 12: Global weights through the hierarchy

Factors of Level 1	Independent Weights	Factors of Level 2	Independent Weights	Global Weights	Final Ranking
C1	0.3800000	C11	0.2748000	0.1044240	4
		C12	0.3297000	0.1252860	2
		C13	0.3955000	0.1502900	1
C2	0.1800000	C21	0.2462000	0.0443160	11
		C22	0.1821000	0.0327780	13
		C23	0.2723000	0.0490140	10
		C24	0.2994000	0.0538920	9
C3	0.2100000	C31	0.3842000	0.0806820	5
		C32	0.3564000	0.0748440	6
		C33	0.2594000	0.0544740	8
C4	0.2300000	C41	0.3158000	0.0726340	7
		C42	0.1731000	0.0398130	12
		C43	0.5111000	0.1175530	3

Table 13: Comparisons between the outcomes

Factors	Fuzzy AHP	Fuzzy ANP	Classical AHP	Classical ANP
C11	0.1044240	0.1044120	0.1045440	0.1044250
C12	0.1252860	0.1251250	0.1253650	0.1255470
C13	0.1502900	0.1501240	0.1512540	0.1544510
C21	0.0443160	0.0442540	0.0442560	0.0445620
C22	0.0327780	0.0322140	0.0325540	0.0323670
C23	0.0490140	0.0491140	0.0491140	0.0495670
C24	0.0538920	0.0538540	0.0533560	0.0535540
C31	0.0806820	0.0811450	0.0819650	0.0800050
C32	0.0748440	0.0755470	0.0747540	0.0722310
C33	0.0544740	0.0545240	0.0542360	0.0542360
C41	0.0726340	0.0725640	0.0722250	0.0722360
C42	0.0398130	0.0397430	0.0397540	0.0395470
C43	0.1175530	0.1175240	0.1175590	0.1172540

As shown in Tab. 13, the modification between the findings of big data security assessment using Fuzzy-AHP and traditional AHP approaches is insignificant [22–24]. While employing Fuzzy-AHP rather than conventional AHP, it is possible to obtain results that are more efficient and enhanced. It's because utilizing Fuzzy with AHP produces more exact inputs and, crisper results.

In big data security, Fig. 7 shows a graphical representation of the comparison of the outcomes derived from the Fuzzy-AHP and Classical AHP approaches, including global weights and the final ranking of the factors. Fuzzy-AHP is clearly superior to Classical AHP in terms of efficiency [25–27]. Big data security and IoT have gotten a lot of attention recently. In this day, big data security is a burgeoning field of study. The goal of the proposed research is to improve security aspects and compare fuzzy based approaches.

8 Conclusions

In this day, big data security is a burgeoning field of study. Contributing factors in the area of big data security play an important role in determining security. Next, choosing the most important factor from a list of options is vital for effective big data security. Based on large data security, this paper expressed viewpoints on the straight associated qualities and sub-factors. The assessment found that the directed dependency on Fuzzy Analytical Network Process (ANP), the ecological instance of validation, and trustworthiness are the most planned factors of all. The comprehensive values have been calculated as the results of the effective application of Fuzzy AHP. The priority-based attribute listing aids in determining the utmost significant factors among a large number of factors in big data security. Furthermore, periodization will aid in devising approaches for determining the most effective use of big data security in explicit analysis and mitigation mechanisms. We are excited to conduct numerous experiments employing various MCDM approaches on the same research problem and in various IoT applications including smart healthcare. The suggested effort is a quantitative assessment that aims to improve the strength of big data security features. The following are the primary implications of the proposed work:

- The security approaches will be improved, evaluated, identified, and prioritised by focusing on big data security factors.
- MCDM approaches, such as Fuzzy-AHP, have been shown to produce more efficient findings than Classical AHP, making them a viable approach for assessing large data security.

Funding Statement: The authors received no specific funding for this study.

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

References

[1] D. Miorandi, "Internet of things: Vision, applications and research challenges," *Ad Hoc Networks,* vol. 10, no. 7, pp. 1497–1516, 2012.

[2] D. Whitmore, "The internet of things- a survey of topics and trends," *Information Systems Frontiers,* vol. 17, no. 2, pp. 261–274, 2015.

[3] M. T. J. Ansari, D. Pandey and M. Alenezi, "STORE: Security threat oriented requirements engineering methodology," *Journal of King Saud University-Computer and Information Sciences,* vol. 6, no. 3, pp. 1–18, 2018.

[4] K. Kambatla, "Trends in big data analytics," *Journal of Parallel and Distributed Computing,* vol. 74, no. 7, pp. 2561–2573, 2014.

[5] I. A. T. Hashem, "The rise of big data‖ on cloud computing: Review and open research issues," *Information Systems,* vol. 47, no. 5, pp. 98–115, 2015.

[6] M. Ali, "Big data-driven smart policing: Big data-based patrol car dispatching," *Journal of Geotechnical and Transportation Engineering,* vol. 1, no. 2, pp. 1–18, 2016.

[7] I. Hashem, "The rise of _big data on cloud computing: Review and open research issues," *Information Systems,* vol. 47, no. 8, pp. 98–115, 2015.

[8] O. Elijah, "An overview of internet of things (iot) and data analytics in agriculture: Benefits and challenges," *IEEE Internet of Things Journal,* vol. 5, no. 5, pp. 3758–3773, 2018.

[9] N. Nalini, "A study on data analytics: Internet of things & healthcare," *International Journal of Computer Science and Mobile Computing,* vol. 6, no. 3, pp. 20–27, 2017.

[10] A. Siddiqa, "A survey of big data management: Taxonomy and state-of the-art," *Journal of Network and Computer Applications,* vol. 71, no. 7, pp. 151–166, 2016.

[11] C. A. Steed, "Big data visual analytics for exploratory earth system simulation analysis," *Computers and Geosciences,* vol. 61, no. 1, pp. 71–82, 2013.

[12] S. Arulananda, J. Jothi, A. Samath and K. Anandharaj, "A study on iot and data analytics in healthcare systems," *International Journal of Scientific & Technology Research,* vol. 8, no. 10, pp. 2277–8616, 2019.

[13] S. Kumar and M. Singh, "Big data analytics for healthcare industry: Impact, applications, and tools," *Big Data Mining and Analytics,* vol. 2, no. 1, pp. 1–20, 2019.

[14] S. Revanth, "Improving healthcare using big data analytics," *International Journal of Scientific & Technology Research,* vol. 6, no. 3, pp. 2277–2289, 2017.

[15] V. Palanisamy and R. Irunavukarasu, "Implications of big data analytics in developing healthcare frameworks-a review," *Journal of King Saud University-Computer and Information Sciences,* vol. 31, no. 4, pp. 415–425, 2019.

[16] 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.

[17] A. Celesti, O. Amft and M. Villari, "Guest editorial special section on cloud computing, edge computing, internet of things, and big data analytics applications for healthcare industry 4.0," *IEEE Transactions on Industrial Informatics,* vol. 15, no. 1, pp. 454–456, 2019.

[18] 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.

[19] J. Qadir, M. M. Rahman and M. H. Rehmani, "IEEE access special section editorial: Health informatics for the developing world," *IEEE Access,* vol. 5, no. 8, pp. 27818–27823, 2017.

[20] M. Abomhara and G. M. Køien, "Cyber security and the internet of things: Vulnerabilities, threats, intruders and attacks," *Journal of Cyber Security,* vol. 4, no. 4, pp. 65–88, 2015.

[21] K. Sahu and R. K. Srivastava, "Predicting software bugs of newly and large datasets through a unified neuro-fuzzy approach: Reliability perspective," *Advances in Mathematics: Scientific Journal,* vol. 10, no. 1, pp. 543–555, 2021.

[22] F. A. Alzhrani, "Evaluating the usable-security of healthcare software through unified technique of fuzzy logic, ANP and TOPSIS," *IEEE Access,* vol. 8, no. 8, pp. 109905–109916, 2020.

[23] 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.

[24] V. Torra, "Hesitant fuzzy sets," *International Journal of Intelligent Systems,* vol. 25, no. 6, pp. 529–539, 2010.

[25] R. Kumar, A. Baz, H. Alhakami, W. Alhakami, A. Agrawal *et al.*, "A hybrid fuzzy rule-based multi-criteria framework for sustainable-security assessment of web application, "*Ain Shams Engineering Journal,* vol. 6, no. 3, pp. 1–21, 2021.

[26] 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.

[27] M. T. J. Ansari, A. Baz, H. Alhakami, W. Alhakami, R. Kumar *et al.*, "P-STORE: Extension of store methodology to elicit privacy requirements," *Arabian Journal for Science and Engineering,* vol. 64, no. 3, pp. 1–24, 2021.