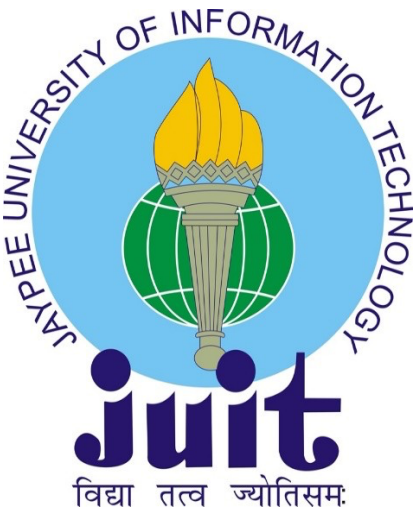
SECURITY AND PRIVACY PRESERVATION IN ENVIRONMENTAL MONITORING AND HEALTHCARE APPLICATIONS

Thesis submitted in fulfilment for the requirement of the Degree of

DOCTOR OF PHILOSOPHY

By

NIDHI SHARMA



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING AND INFORMATION TECHNOLOGY

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT, SOLAN-173234, HIMACHAL PRADESH, INDIA

AUGUST 2024

SUPERVISOR'S CERTIFICATE

This is to certify that the work reported in the Ph.D. thesis entitled "**Security and Privacy Preservation in Environmental Monitoring and Healthcare Applications"** by **Nidhi Sharma** at **Jaypee University of Information Technology, Waknaghat, Solan (HP), India,** is a bonafide record of her original work carried out under my supervision. This work has not been submitted elsewhere for any other degree or diploma.

Dr. Ravindara Bhatt

Associate Professor

Department of Computer Science & Engineering and Information Technology,

Jaypee University of Information Technology, Waknaghat, Solan (HP), India.

Date: 28/08/2024

DECLARATION BY THE SCHOLAR

I hereby declare that the work reported in my Ph.D. thesis, "Security and Privacy Preservation in Environmental Monitoring and Healthcare Applications," submitted at Jaypee University of Information Technology, Waknaghat, Solan (HP), India, is authentic and carried out under the supervision of **Dr. Ravindara Bhatt**. I have not submitted this work elsewhere for any other degree or diploma. I am fully responsible for the contents of my Ph.D.thesis.

Nidhi Sharma
Enrollment Number: 166206
Department of Computer Science & Engineering and
Information Technology
Jaypee University of Information Technology, Waknaghat, Solan
(HP), India.

Date: 28/08/2024

ACKNOWLEDGEMENT

I would like to express my deep respect and gratitude to my guide, Dr. Ravindara Bhatt, Associate Professor in the Department of Computer Science and Engineering & Information Technology, for his support throughout this research work. I want to thank him for introducing me to the field of Wireless Sensor Networks and Privacy and allowing me to work under him. Without his invaluable advice and assistance, it would not have been possible for me to complete this thesis. I am greatly indebted to him for his constant encouragement and invaluable advice in every aspect of my academic life. I consider it my good fortune to have had an opportunity to work with such a wonderful person.

I would like to express my gratitude to our Honorable Vice Chancellor, Prof. (Dr.) Rajendra Kumar Sharma and Dean of Research Prof. (Dr.) Ashok Kumar Gupta will promote the research and facilitate the institution's resources. I would also like to thank Prof. (Dr.) Vivek Kumar Sehgal, Head Department of CSE & IT, for constant guidance on research facilities and resources to carry out my research work. I would also like to thank my doctoral committee members, Dr. Amit Kumar, Dr. Pankaj Dhiman and Prof. (Dr.) Rakesh Kumar Bajaj, thank you for your valuable feedback, critical reviews during presentations, and time-to-time help. I would also like to thank Mr Amit Kumar Shrivastava for his extended and unparalleled support throughout the Ph.D.

I am also grateful for the support from JUIT, Waknaghat (Himachal Pradesh). I thank all staff at the Department of Computer Science and Engineering & Information Technology, JUIT, Waknaghat (Himachal Pradesh), who have been extremely helpful on numerous occasions. I also thank my friends for their consistent help and valuable discussions. Last, I would like to thank my family for spiritually supporting me.

Nidhi Sharma

ABSTRACT

WSNs are extensively used for monitoring healthcare, environment, and other objectives. Autonomous sensor nodes in a WSN-based medical system employ wireless communication to communicate. These nodes gather physical data from the area of interest, such as motion, temperature, and pressure, among other available input parameters. Medical data is considered sensitive and secret since it contains personal information about the patient. As a result, data security and privacy are critical challenges for medical applications. Privacy protection is a significant concern when using IoT-enabled event-driven wireless sensor networks for monitoring applications. In our thesis, we have presented a methodology for preserving privacy in IoT-enabled WSNs-based medical applications. The proposed study establishes a foundation for a privacy-protection strategy. Device authentication for physicians and patients remains a key concern in medical IoT networks. We designed an authenticating session key mechanism for smart IoT healthcare networks to improve security. Our proposed scheme uses multi-factor authentication that protects the doctors' and patients' data and provides an authentication mechanism.

Privacy is a major problem in IoT-enabled incident wireless sensor networks for monitoring applications. The thesis investigates a practical framework for protecting source private information in incident wireless sensing networks. The thesis suggests a grid-based deployment security research that provides three event detection strategies: Source Location Privacy for Event Detection (SLP ED), Chessboard Alteration Pattern (SLP ED CBA), and Grid-based Source Location Privacy (GB SLP). The suggested approach for preserving source location privacy in incident wireless sensor networks is discussed in the thesis. In environmental monitoring, a source node detects two categories of events. These occurrences might be crucial or insignificant. When an event is detected, the collected data is transferred towards the sink node. For nominal occurrences, the algorithm chooses a low-energy consumption strategy. For key events, the algorithm selects high-energy consumption pathways. The proposed work creates a technique for protecting privacy when monitoring applications in WSN.

Medical cyber-physical systems involve the healthcare critical integration of a network of medical devices. Machine Learning (ML) based applications can provide valuable information to all stakeholders in the healthcare system. The data can facilitate patient care and diagnose diseases at an initial stage. The early detection of the disease leads to better medication. For

example, it would be helpful if a doctor knows a patient's risk for a particular disease based on lab test results and family history. The machine learning algorithms for healthcare applications are privacy-sensitive and require large quantities of training data. The challenges for data privacy in healthcare applications are associated with machine learning algorithms. The thesis presents various Machine Learning Classification Techniques for a healthcare dataset. We compare six machine learning classification algorithms and observe that Support Vector Machine (SVM) performs better than other available techniques. Further, we give privacy preservation techniques for the healthcare dataset. The original dataset is preserved by applying privacy-preservation techniques to the data. We observe that employing a single privacy-preserving technique could not provide optimal results.

The thesis's proposed work provides a privacy preservation framework for healthcare applications and privacy preservation techniques for monitoring applications. In addition, the thesis proposes privacy preservation in machine-learning healthcare applications, which may be exploited in other application areas.

Keywords: Wireless Sensor Networks (WSNs), Internet of Things (IoT), Machine Learning, Security, Privacy, Environmental Monitoring, Healthcare.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

WSN	Wireless Sensor Networks
IoT	Internet of Things
BSN	Body Sensor Networks
SLP_ED	Source Location Privacy for Event Detection
SLP_ED_CBA Alteration	Source Location Privacy for Event Detection with a Chessboard Pattern
GB_SLP	Grid-Based Source Location Privacy
ED	Event Detection
CBA	Chessboard Alteration Pattern
EeSP	Energy-efficient Source Location Privacy Protection
SVM	Support Vector Machine
WBAN	Wireless Body Area Network
KG	Knowledge Graph
RDF	Resource Description Framework
DSS	Decision Support System
MAKA	Mutual Authentication and Key Agreement
BP	Blood Pressure
WBSN	Wireless Body Sensor Network
AI	Artificial Intelligence
FRW	Forward Random Walk
ML	Machine Learning

PPML	Privacy Preservation Machine Learning
PII	Protecting Personally Identifiable Information
GDPR	Global Data Protection Regulation
LPU	Local Processing Unit
DES	Data Encryption Standard
ROR	Real-or-random
BAN	Burrows-Abadi-Needham
AVISPA	Automated Validation of Internet Security Protocols and Applications
Kser	Server's Master Key
DY	Dolev-Yao
SN	Sensor Node
HER	Electronic Health Record
MPC	Multi-Party-Computation
PCA	Principal Component Analysis
AUC	Accuracy
ADMM	Alternating Direction Method of Multipliers
OCR	Optical Character Recognition
PPPCA	Privacy Preserving Principal Component Analysis
ICT	Information and Communication Technology
DNN	Deep Neural Networks
HealthIIoT	Healthcare Industrial Internet of Things
PTFS	Privacy-aware and Traceable Fine-grained System

EMSA	Euclidean L3P-based Multi-Objective Successive Approximation
PAASH	Privacy-preserving authentication and fine-grained access control system, for smart health
HitL-aided	Human-in-the-loop-aided
ECG	Electrocardiogram
CKA	Chosen Keyword Attacks
KGA	Keyword Guessing Attacks
SPE	Searchable Public Key Encryption
WMSN	Wireless Medical Sensor Networks
DOS	Denial of Service
DDOS	Distributed Denial of Service
WEP	Wireless Equivalent Privacy
SPREAD	Secure Protocol for Reliable Data Delivery
H()	Hash Function
M	Message
H(M)	Hashing of Message
R(M)	Received Message
RH(M)	Received the Message of Hashing
HR(M)	Hashing of Received Message
S_i	Identification of Server
U_j	Identification of User Node
$H(u_j)$	Hash Value of User Node Identification

K s_iu_j	Pairwise Session Key
T_{id}	Time-stamp id
MAC_u	Message Authentication Code of the msg
RFID	Radio Frequency Identification
D_i	Doctor Identity
P_i	Patient Identity
PWD_i	Password of the doctor
$BIO_{i.}$	Biometric identity of the Doctor
randN	Random Number
Key_{secret}	Secret Key of the Doctor's end
$DID_{Gateway}$	Doctor gateway identity
$MAS_{gatewayid}$	Gateway identity of the medical authentication server
$Token_{i.}$	Short authentication identity
||	Concatenation operation
SMG	Smart Medical Gateway
RP	Registration Phase
LP	Login and Authentication Phase
SIoT	Social Internet of Things
FIoT	Future of the Internet of Things
TCP	Transmission Control Protocol
TP	True Positive

TN	True Negative
FP	False Positive
FN	False Negative
KNN	K-Nearest Neighbors Classifier
Rn	No Readmission
Rl	Readmission in less than 30 days
Rg	Readmission in greater than 30 days
SVR	Support Vector Regression
RDF	Resource Description Framework
DSS	Decision Support System
HER	Electronic Health Record
NER	Named Entity Recognition
PPDM	Privacy-Preservation in Data Mining

CHAPTER 1

INTRODUCTION

1.1 WSN-BASED HEALTHCARE SYSTEM

WSNs are commonly used for healthcare monitoring, environmental monitoring, and other purposes. Autonomous sensor nodes use wireless communication in a WSN-based healthcare system to communicate with one another. These nodes collect physical data from the region of interest, including motion, temperature, pressure, etc. The medical system helps keep track of each patient's condition and monitors their disease. The healthcare monitoring system, as a result, offers home assistance and support for patients who are elderly or have special needs [1]. Information about the patient should never be made public since it might be used inappropriately, or privacy concerns might prevent people from fully utilizing technology. A new discipline called wireless body area networks has emerged to deal with the expanding use of sensor technology [2], as shown in Figure 1.1. The information gathered from a person's body can be processed over a communication network and delivered to a medical facility using a wireless body area network so that personnel in the medical division can remotely monitor the patients.

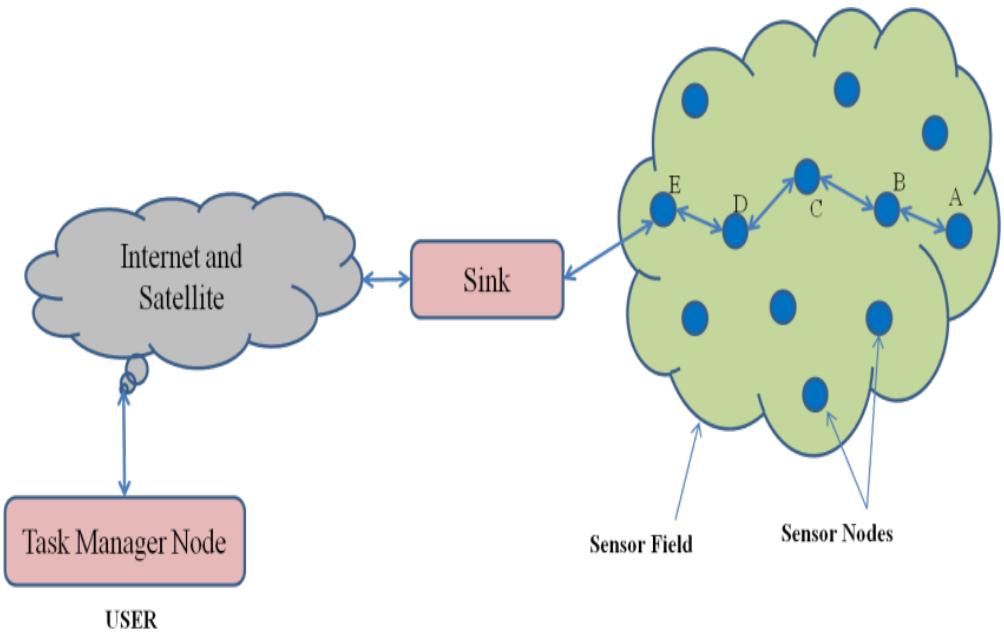


Figure 1.1: WSN environment

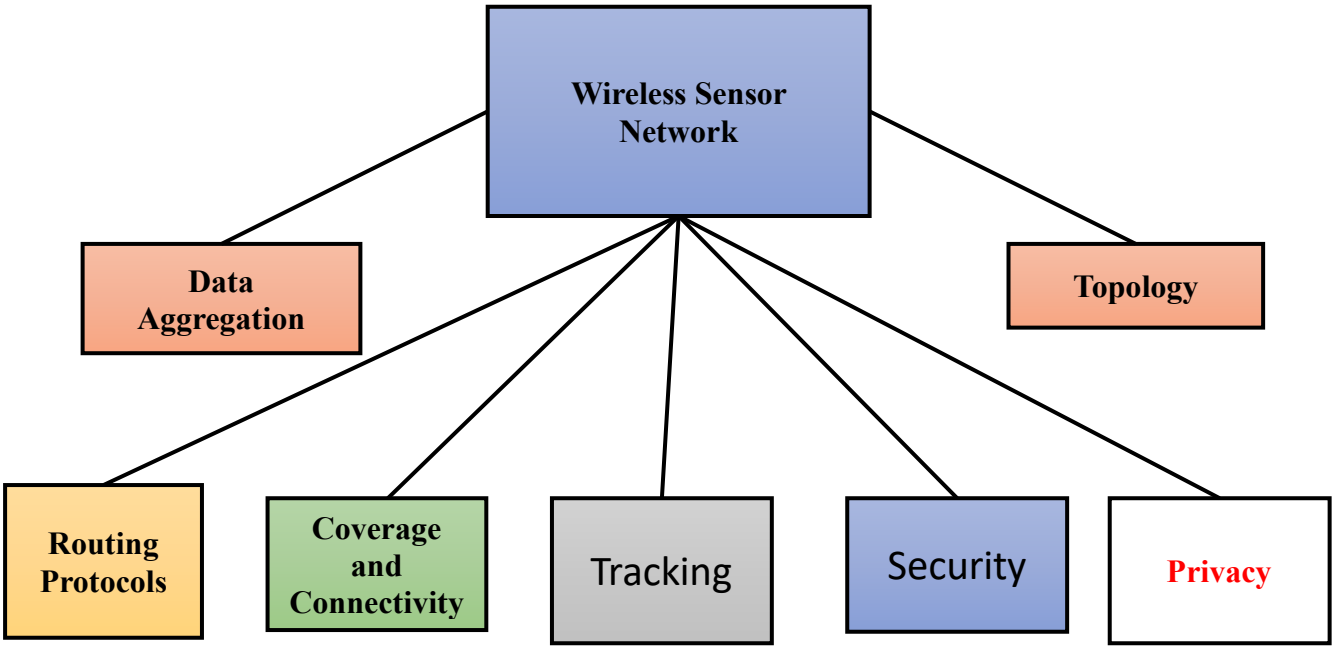


Figure 1.2: Design issues of WSN

Military purposes, monitoring of geographical areas, environmental monitoring, earth sensing, industrial monitoring, monitoring of health care, etc., are some of the uses for wireless sensor networks. Today, several applications, including those for monitoring blood pressure and heart rate, are widely used across the globe by different patients, doctors and caregivers [2]. The input gathered from these sensors is subsequently processed in the processing unit, where the system analyses the parameters [3]. The base station receives the processed data across the internet from the system's many stakeholders. Specialists from the health department remotely monitor the patient using information acquired from the base station. At the time of data sharing or data transmission, an attacker may use a strong receiver to intercept the data gathered by the medical sensors. There might be consequences if the collected data is published on additional social networking websites. To preserve these, the authors look into a privacy-preservation method that protects patient data from insider attacks [4, 5]. Wireless communication allows easy eavesdropping; hackers can quickly introduce and send harmful messages into networks. Figure 1.2 presents the design issue of the sensor network. Privacy is the capacity to keep oneself or information about oneself private, allowing one to express oneself only in certain circumstances [6].

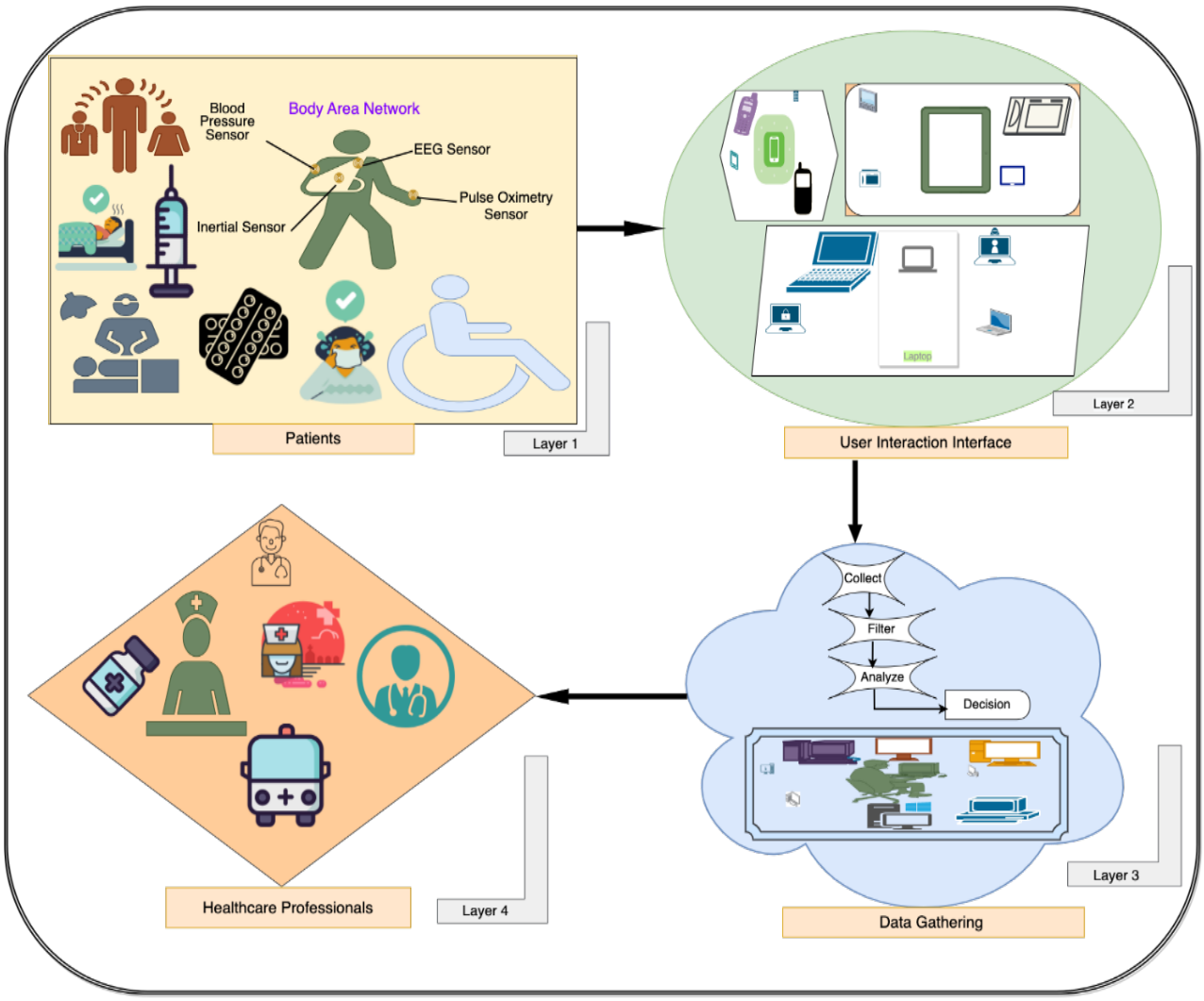


Figure 1.3: Basic framework BSN

Figure 1.3 illustrates the basic framework of the Body Sensor Network (BSN) and its environment. Heartbeats, blood pressure, temperate, and brain impulses are just a few examples of the features of a person or patient that are regularly seen in a sensing area. The base station with the connected network receives the data that was collected. The data is subsequently processed in the processing unit, where the system analyses the parameters [3]. The system offers many stakeholders the ability to access the processed information with the help of the base station. Based on the data from the base station, the health department's personnel remotely monitor the patient [7]. This, in turn, requires high-end computing and storage facilities with low latency and better Quality of service for healthcare applications [4]. Medical data is confidential because it contains patient personal information. Therefore, data protection is an important issue for medical applications. The healthcare component (WBAN) of the Internet of Things is focused on improving people's quality of life [5]. Several authentication

protocols have been developed for the Medical Internet of Things to ensure user privacy. These protocols do not protect against verification theft or table leak attacks nor provide secure mutual authentication, anonymity, or untraceability [8]. Wireless sensor networks (WSNs) are an important technological backbone of the Internet of Things, providing data sources for Internet of Things applications [9]. The WSN architecture has three participants: users, gateway, and sensor nodes. Authorised users can access the data, and the combination and analysis of this data help managers make the right decisions [10].

1.1.1 NEED OF PRIVACY

- Patient privacy protection in the healthcare industry [36].
- Respect for privacy is likewise required by ethical, moral, and scientific principles [21].

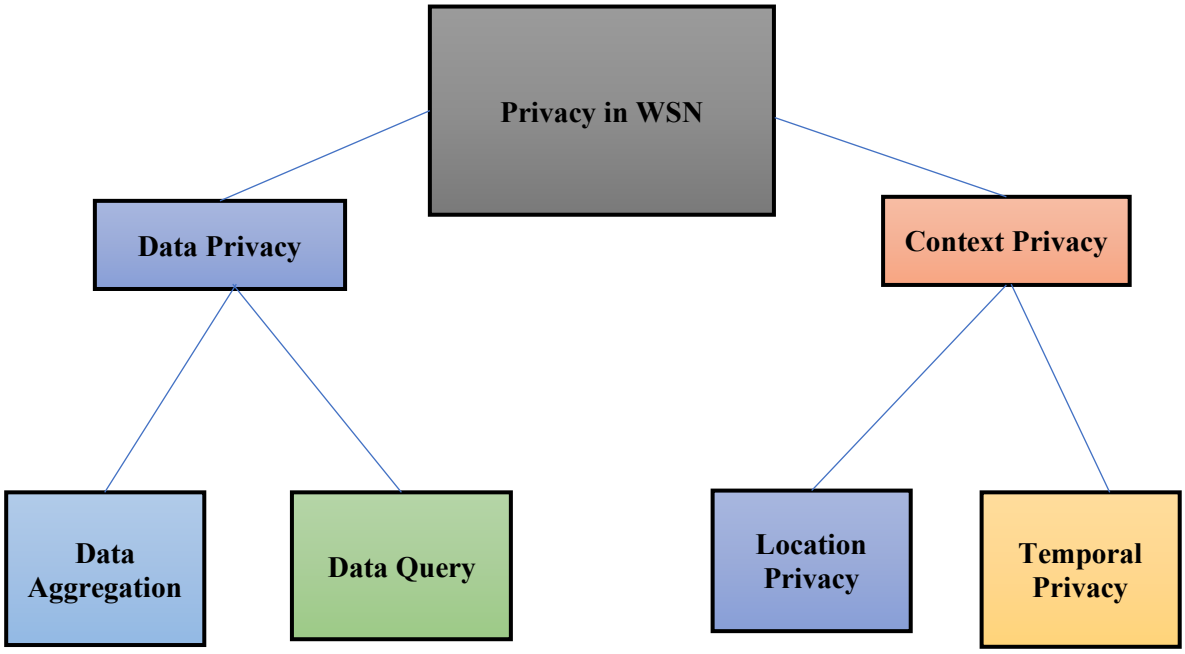


Figure 1.4: Privacy in WSN

Figure 1.5: Security and Privacy Entities

- A lot of organizations collaborate and exchange data. Data exchange between companies necessitates the protection of individual privacy [9].
- Location data should be protected as an individual right [6].
- Let's assume a hospital wishes to disclose certain patient information that is distinct to each individual.
- Information continues to be valuable in real life.
- An individual's identity cannot be established.
- Whereas security is concerned with safeguarding data, privacy is concerned with preserving user identity [1].

Machine learning has become quite popular in extracting valuable data for commercial and scientific research applications in the healthcare industry. Since it contains the patient's personal information, healthcare data is sensitive. Privacy is, therefore, a major concern for

healthcare apps, as shown in Figures 1.4 and 1.5. Machine learning enables medical professionals to identify disease at its earliest stage and provide the appropriate therapy. The three primary subfields of machine learning are supervised, unsupervised, and reinforcement learning [11]. "Predictive learning" is a term used to describe supervised learning. Using data from comparable objects that correspond to the class of the unknown item, a machine may predict the class of the item. Descriptive learning is another name for unsupervised learning. By combining related things, a system may identify patterns in unlabeled objects [110]. In reinforcement learning, a computer learns to behave autonomously to accomplish objectives. Concerns around privacy in healthcare and other applications have grown over the last few years. On the other hand, the current e-healthcare systems lack user trust and privacy [12]. The privacy of patient data would be seriously threatened by sharing personal information related to the person. Additionally, the violations of patients' privacy cause ethical, legal, economic, psychological and social issues [115].

1.2 PRIVACY PRESERVATION IN MONITORING APPLICATION

Privacy protection is a major challenge with IoT-enabled event-driven wireless sensor networks for monitoring applications. Content privacy and context privacy are the two main categories of privacy [12], as shown in Figure 1.4. Content privacy safeguards the data transferred between sensor nodes. Privacy covers message creation times, source and destination locations, and other context-sensitive data [12][13]. The monitoring involves IoT-enabled event-driven wireless sensor networks for monitoring applications. These networks are tasked with safeguarding data transferred between sensor nodes, ensuring content privacy and protecting sensitive context-related information such as message creation times and source/destination locations. Monitoring is crucial for enhancing security and detecting unauthorised access attempts or breaches in real time, thereby improving the security of sensitive data and systems. The early detection of anomalies and continuous monitoring enables the early detection of irregularities in data transmission, which can reduce potential security threats and system malfunctions. For instance, in habitat monitoring applications, the nodes collect details of the endangered species and report them to the central controller, i.e., the base station. Preserving the privacy of these assets from attackers is imperative [16]. The privacy of the object(s), event(s), or asset(s) being monitored may be jeopardized due to the wireless nature of information communications to the Base Station (BS), also known as a sink [136]. An example of such an event is when an adversary (also called an attacker or hunter)

equipped with sophisticated technology can detect the message flows, trace back to the source originating the messages in the reverse path, and locate the events or objects being monitored. Habitat or asset monitoring is not just essential; it is urgent. It plays a pivotal role in preventing species' extinction and understanding their movement patterns, which can provide valuable insights into their behaviour and survival strategies.

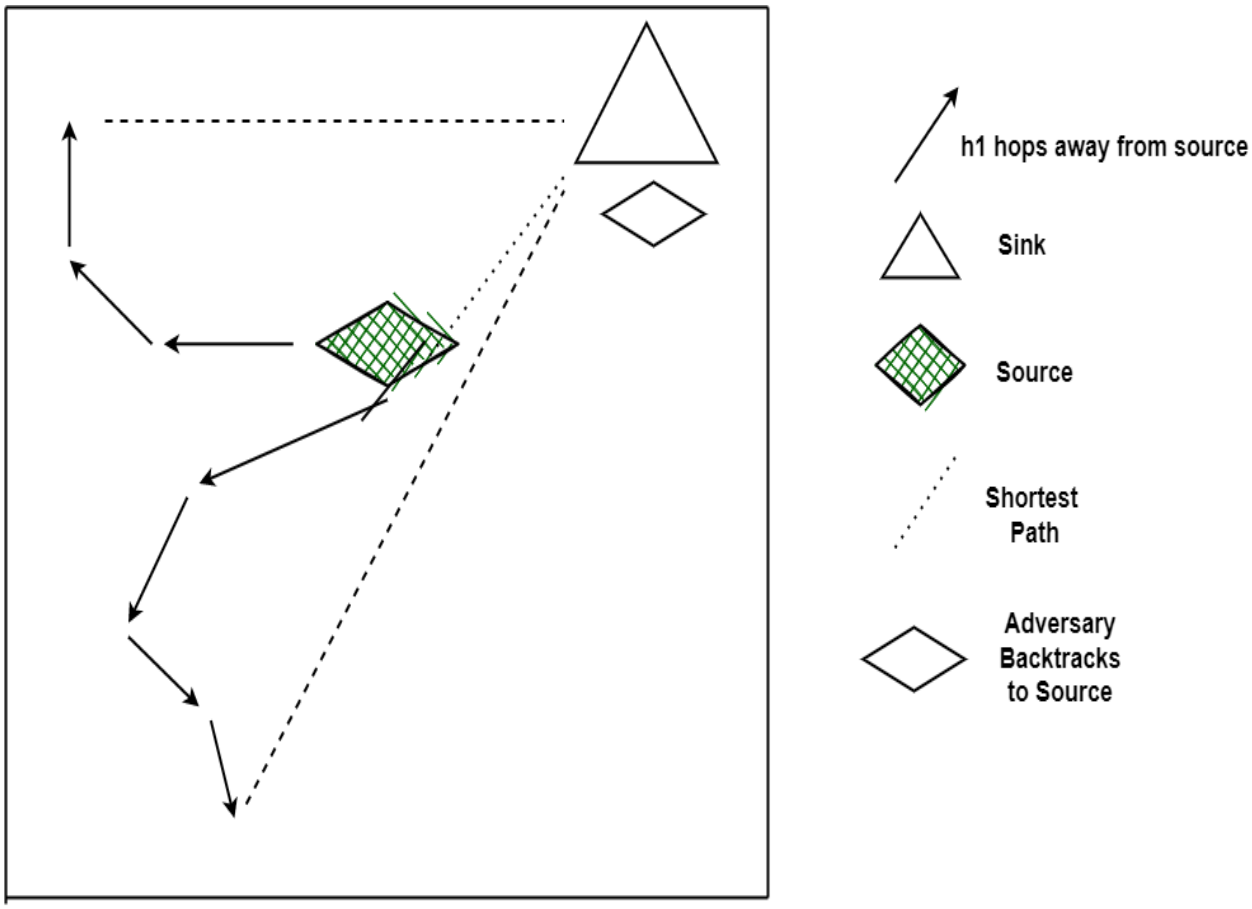


Figure 1.6: Privacy Preservation using Routing

Context-based privacy refers to sink or source node privacy in environmental monitoring applications. Safeguarding temporal and locational private information is the main goal of context-based privacy [121][147]. Without interpreting the message's information, the adversary might violate private information. The authors suggest phantom routing, which offers more secrecy from the source location than straightforward routing techniques [14][15], as shown in Figure 1.6. Figure 1.2 uses different paths for transferring gathered data to the base station. Figure 1.3 shows the deployment of WSN nodes in a two-dimensional plane utilising a checkerboard pattern of switching between the active and sleep regions, as shown in Figure 1.7.

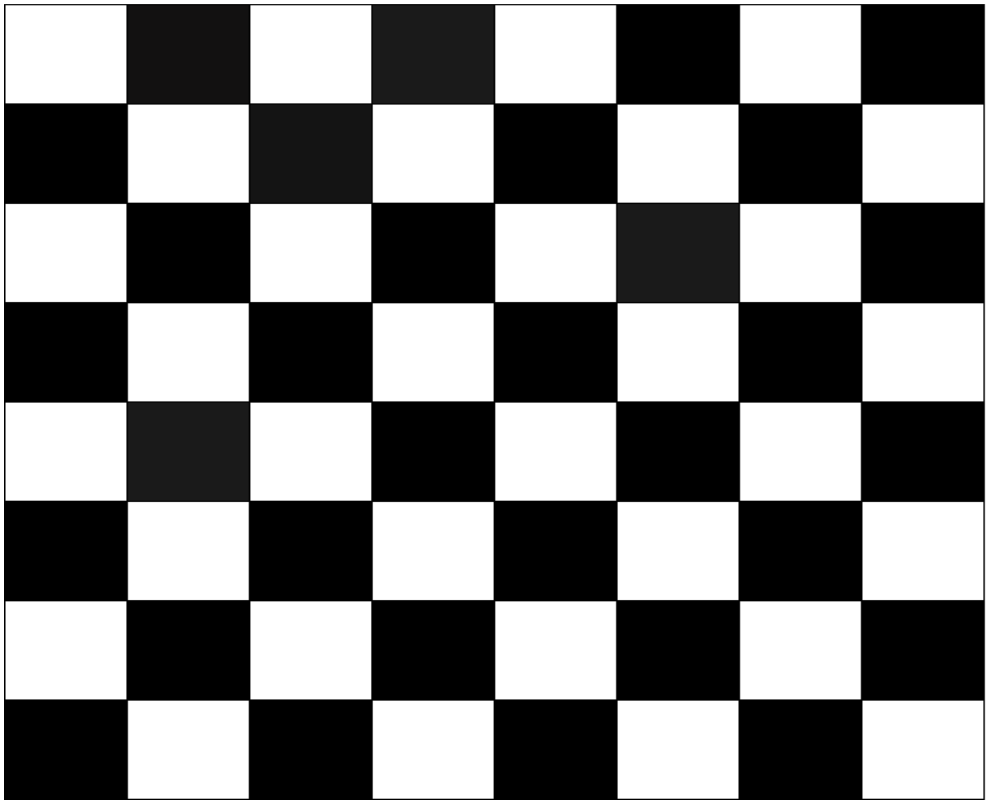


Figure 1.7: Chessboard deployment plane (400×400)

1.3 PRIVACY PRESERVATION IN MACHINE LEARNING

Machine learning has recently been widely used to extract useful information for healthcare applications for scientific research and business purposes. Healthcare data is sensitive as it contains the patient's personal information. Thus, privacy is a significant issue for healthcare applications [16]. Machine learning helps doctors notice disease at the initial stage of the disease, leading to necessary medication. Supervised, unsupervised, and reinforcement learning are the three main subfields of machine learning. Predictive learning is a term that refers to supervised learning [161]. Using data from similar items that correspond to the class of the unknown item, a machine can predict the class of the item. Descriptive learning is another name for unsupervised learning. By combining related objects, a system can identify patterns in unidentified objects. In reinforcement learning, a computer learns to act independently to accomplish objectives.

People have recently become increasingly concerned about privacy issues in healthcare and other applications. Current e-healthcare systems, on the other hand, lack privacy and user trust [178]. Sharing a patient's information would bring a severe threat to data privacy. Further, breach of privacy leads to moral, legal, and social problems. Summarization, data separation,

and data obfuscation are popular privacy-preserving techniques. Data anonymization and encryption are imperative approaches to privacy preservation [17]. Suppression and generalization are two k-anonymity strategies. The provided data may become less helpful to receivers of excessive anonymization [16]. Secure multiparty computation is a mechanism to calculate a function without disclosing its private inputs [19], [20]. Health systems share the data horizontally or vertically partitioned [21].

1.4 PRIVACY PRESERVATION IN KNOWLEDGE GRAPH

A knowledge graph (KG) is similar to a traditional graph regarding nodes and edges. However, KG includes semantics in addition to entities from the real world and relationships among nodes [24]. Some representational frameworks used in knowledge graphs are property graphs and Resource Description Frameworks (RDF). Neo4j uses a labelled property graph to represent a knowledge graph. An entity with zero or more characteristics is called a node in the knowledge graph [25]. A relationship between two nodes could have zero, one, or more characteristics. The RDF framework stores the triples as subject-predicate-object triples. For instance, the triple (abc, patient's name) is used. For healthcare applications, RDF supports both relational and hierarchical data models. To aid in making difficult healthcare decisions, a healthcare knowledge graph is a connected graph with an entity and relationships that have been semantically improved [159]. Doctors and nurses can also benefit from the knowledge graph-based Decision Support System (DSS) [26]. Healthcare is experiencing a significant workforce shortage compared to the population-to-worker ratio, especially in India. Medical or healthcare knowledge graphs are advantageous to the healthcare system.

The medical knowledge graph's design involves techniques for protecting privacy. The authors discuss creating big data-based information systems for diabetes management [27]. The system gathers data from many sources, preprocesses it, and then saves it in a database. A knowledge graph can be added to the system to benefit all parties involved in the medical industry.

1.5 PROBLEM STATEMENT

1.5.1 PRIVACY PRESERVATION IN BSN

Understanding an e-healthcare system's architecture is fundamental for comprehending the intricacies of its privacy features and functionalities.

E-healthcare systems, being IoT-enabled, need secure authentication and robust security measures to function effectively and protect patient data. This urgent need must be addressed, underlining the issue's importance.

It encompasses a range of components and processes, each playing a crucial role in ensuring the security and confidentiality of sensitive medical data. Critical security issues such as access control, authentication, non-repudiation, and accountability demand meticulous attention to mitigate potential threats and address privacy concerns effectively [7]. Achieving end-to-end data protection necessitates the implementation of robust measures across the system.

The Internet of Things (IoT) has transformed the healthcare landscape, enabling valuable data collection for medical research, patient care, and commercial purposes. However, this integration also brings significant challenges, particularly in safeguarding patients' confidential information. To address this, we propose the establishment of a robust mutual authentication and key agreement (MAKA) system tailored to the Internet of Medical Things (IoMT) [9]. This system, which works by [specific functionality of the MAKA system], is crucial for ensuring the integrity of user health information and enabling the seamless delivery of healthcare services while upholding stringent privacy standards.

The privacy preservation comprehending the architecture of an e-healthcare system is pivotal for navigating the intricate interplay between privacy and functionality. By addressing security concerns and implementing robust authentication mechanisms, stakeholders play a crucial role in upholding the confidentiality of patient data and fostering trust in the digital healthcare ecosystem, empowering them with a sense of responsibility.

1.5.2 PRIVACY PRESERVATION IN ENVIRONMENTAL MONITORING

WSN sensor nodes are positioned in an open space for environmental monitoring. Active and passive adversaries may attack these. Active and passive enemies are the two different categories of adversaries. An engaged adversary can physically capture and hack the sensor nodes [22]. A passive attacker cannot harm sensor nodes by destroying them, infiltrating specific nodes, or just hearing the message [23]. The study contrasts relevant privacy protection research with passive adversary research. The authors investigate various methods for protecting the privacy of the source node. Methods like random walks, multipath routing, ring routing, and phantom routing have all been used in the literature. The authors offer recommendations for models in WSNs that pertain to privacy [22]. In their 2014 paper, [22]

describe how random behaviour in packet forwarding to the sink can preserve location privacy. Several researchers do not consider preserving event privacy near the BS [23].

1.5.3 PRIVACY PRESERVATION IN MACHINE LEARNING

Recently, it has become more important to mine data sets dispersed across several parties without releasing further private information [1]. The authors cover the large data life cycle [2]. Data anonymization and encryption are essential techniques for protecting big data users' privacy [3], [4], [5]. Generalization and suppression are two k-anonymity techniques. By over-anonymizing the data, recipients may find it less useful [2]. To compute a function without revealing its private inputs, use secure multiparty computation [5], [6], [7], [8].

Health systems may divide the data vertically or horizontally [9]. The author suggests a training scenario where learners get horizontally split records. The authors provide a training dataset that is vertically partitioned and distributes characteristics to learners [1], [9]. Powerful analytical tools are being used more often, and as a result, more data is being produced faster, raising concerns about data privacy. As a result, several privacy-preserving machine learning algorithms are created for use in healthcare applications.

1.6 WSN-BASED HEALTHCARE SYSTEM

Body sensor network systems can assist people by providing medical services, including clinical observation, memory enhancement, clinical information access, and communication with the medical services provider in an emergency by SMS or GPRS. Continuous health monitoring using wearable or clothing-installed transducers and implanted body sensor groups will increase recognition of emergencies in at-risk individuals. They will benefit the sufferer and their relatives [7]. Also, these frameworks offer helpful methods for remotely obtaining and monitoring physiological signals without interfering with the patient's daily activities, thereby improving life quality.

It is crucial to continuously monitor the patient's physiological boundaries in an emergency clinic medical services observation framework. For instance, a pregnant woman's parameters, such as her blood pressure (BP), pulse, and foetal growth, are important for managing her medical condition [8].

A facilitator hub connected to an understanding body collects all of the signals from the remote sensors and transmits them to the base station, which is the suggested framework. The attached

sensors on the person's body structure are part of a remote body sensor network (WBSN) and can measure the pulse, circulatory strain, and other bodily functions. This system can recognize unusual situations, warn the patient, and email or text the doctor [20].

The proposed architecture also includes several distant transfer hubs responsible for sending and transferring the data supplied by the facilitator hub to the base station. This framework's main advantage over previous frameworks is that it uses less energy to prolong an organization's lifespan and accelerates and broadens correspondence inclusion to increase opportunities for improving patient happiness [19]. This framework was developed for emergency clinic medical services in multi-patient engineering.

In recent years, we have seen the growth of remote sensor organization in medical services, driven by innovation advancements in clinical sensors and low-power organized frameworks. These WSNs represent a commitment to unquestionably enhancing and broadening the scope of care across a wide range of contexts and for distinct segments of the population [20].

For instance, early framework models have demonstrated the capacity of WSNs to enable early identification of clinical disintegration through continuous patient observation in emergency clinics, improve the capacity of specialists on call to provide crisis care in enormous disasters through programmed electronic emergencies, further develop the existing nature of the older through brilliant conditions and enable enormous scope field investigations of clinical disintegration [21].

Recent years have witnessed the expansion of wireless sensor networks (WSNs) in the medical field, fueled by technological advances in clinical sensors and low-power organized frameworks. These WSNs are dedicated to undeniably strengthening and widening the breadth of care across various situations and for certain demographic groups [22].

Early framework models, for instance, have shown that WSNs have the potential to enable early identification of clinical disintegration through continuous patient observation in emergency clinics, improve the ability of specialists on call to provide crisis care in enormous disasters through a programmed electronic emergency, further develop the existence nature of the elderly through brilliant conditions, and enable enormous-scope field investigation.

1.6.1 PRIVACY PRESERVATION IN ENVIRONMENTAL MONITORING APPLICATION

There is growing interest in ecological observation for various applications, with remarkable effects on regular asset executives and safeguarding the economy and people's lives and

wellness. Common uses include, for example, Earth perception, meteorology, routine asset watching, horticulture and monitoring of timberlands, contamination management, perception and anticipation of catastrophic events, and fundamental foundation observing [20]. While these structures serve a vital function in our society, their acceptance can also give rise to several security and protection problems, which could hinder the development of future ecological applications. In this section, we set out the essential security and protection concerns about ecological data and the underlying frameworks for natural observation [23].

Frameworks for ecological observation allow for the examination of real anomalies and the design of tools for anticipation and reaction to dangerous situations [116][111]. A particular number of sensors designed to measure various real amounts, at least one handling hub, and a correspondence organisation make up an observation framework in general. The analogue signals that the sensors produce are modified and transferred into the computerized environment. The digital signals are sent to the registration devices at that moment, combining the information they collected to determine the intentional oddity [25].

These frameworks are becoming increasingly important for tracking the state of the climate in our cutting-edge civilization.

They have a crucial, indispensable role in identifying fresh ecological difficulties and providing assurances that can aid in concentrating on the ecological arrangements. These frameworks are also useful for understanding the relationships between the environment, conservative activities, daily life, and human health. For instance, climate change affects human wellness. It degrades the quality of water, land, developments, and forestry, while weather conditions influence agriculture's success and the forest industry's profitability [128]. So, there is much interest in monitoring the climate to relate prospective effects to observed idiosyncrasies and foresee fundamental but potentially dangerous events. For instance, we now understand a direct link between vascular diseases and a person's susceptibility to PM10 and PM2,5 [26].

1.6.2 PRIVACY PRESERVATION IN MACHINE LEARNING

A method of preventing information leakage in AI computations is gradually increasing AI protection. PPML makes several protection-improving approaches possible, allowing various information sources to agreeably create ML models without disclosing their private information [15].

The benefits of AI applications come with a risk to information security. For instance, suppose we consider apps for medical services or interruption identification. Information leaks and

cyberattacks are becoming increasingly frequent and expensive to contain. Large collections of data stored for planning purposes are attractive to cybercriminals because they may extract information that can be used by different persons or other valuable data that can be sold [16].
In healthcare applications, protecting personally identifiable information (PII), or information that could be used to identify a person specifically, is paramount. Digital attacks put the end clients to whom the information refers, as well as the businesses collecting the information, at risk of legal, financial, and reputational repercussions [131][132]. Because additional semi-identifiers may be used to identify a specific individual in the collection, it wouldn't be sufficient to merely remove PII from a dataset, such as names and addresses. In a study conducted by Latanya Sweeney, William Weld, the legislative leader of Massachusetts, was re-distinguished using anonymised health information data revealing just his date of arrival to the world, orientation, and postal division [3]. ML attempts to overcome these challenges by enhancing them with various information protection protecting approaches.
The protection protecting ML strategy was developed in response to the current cloud-based AI scenario, various association resources, and information security. There won't be a single solution to handle this PPML technique for all application types [17]. Different applications demand various types of considerations for protection. Furthermore, we need to find a way to reconcile situational concerns with the need to promote strong, independent-stage ideologies. Although recent research on protection-safeguarding AI has exploded, there is still a gap between theories and their applicability to verifiable circumstances [112 - 115].

1.6.3 PRIVACY PRESERVATION IN KNOWLEDGE GRAPH

Applications for the medical industry often use information diagrams to the benefit of all parties involved. For a specific case or to provide advice across many emergency clinics, the specialists potentially challenge the patients' historical context information diagram. Also, patients can query the information diagram using a Chabot framework or a clear-cut interface for the location of the closest expert with a high evaluation. The doctors might question the information diagram for a traditional drug family or the distinguishing salt of a particular prescription [20]. In medical services, the study suggests engineering for an information diagram. Protection is also a crucial requirement for a healthcare area. The report suggests ideas for protecting information diagrams used in applications for medical services [141-146].
Like a conventional chart, hubs and edges are addressed by an Information Diagram (KG). Yet, it differs from a traditional diagram in that it organises relationships, semantics, and elements

from the real world into a coherent whole [117-120]. The asset depiction structure (RDF) and the property diagram are examples of the portrayal plans in information diagrams. Diagram of Neo4j address information using a marked property diagram. A material with zero or more qualities is called a hub [21]. Similarly, a relationship between two hubs might have at least zero attributes. The number of triples stored in the RDF structure's subject-predicate-object structure has greatly increased.

For instance, the triple (abc, name of, patient) is used. For applications in medical care, RDF supports a variety of tiered information models and social information models. A connected chart with content and connections improved with semantics to the point where mind-boggling medical service decisions are made persuasively is known as a medical services information diagram [122-127]. Also, the Choice Emotionally Supportive Network (DSS), which is based on information diagrams, supports workers in the medical field, including specialists and physicians. Around the world, including in India, the need for DSS was felt during the coronavirus in 2019 to 2021 [22]. In India, notably, there is a severe labour shortage in the medical services sector due to a disparity between the population and the number of medical services specialists. Clinical information diagrams or information charts for medical services support the framework for such services in this way. Clinical information diagrams or information charts for medical services support the framework for such services in this way. The categorization on the information diagram. An information diagram represents the categorization [129-130].

1.7 MOTIVATION AND CONTRIBUTION

Nowadays, people are paying greater attention to the privacy issue in WSNs. In the healthcare industry, security and privacy problems are also significant. Privacy concerns regarding data gathered, sent, and analysed by WSNs, such as health and environmental monitoring. The topic of privacy in e-health is more complicated [15]. Health insurance and research, for example, frequently share the obtained data, privacy versus utilities [133-137]. Depending on the user's gender, nationality, and cultural background, different people have different perspectives, interests, and privacy needs. The research community has given the WSN much consideration for issues ranging from theoretical analysis to implementation. Additionally, as widespread (ubiquitous) computing spreads and gains acceptance, individual privacy is gradually eroding. Privacy concerns were raised in the early days of WSN as a secondary worry, but today, they are of more importance [138-140].

The Internet is crucial in today's technological world. Internet users are multiplying quickly [163]. General activities over the Internet include selling and purchasing products, online shopping, playing online games, social networks, hotel reservations, and train and flight tickets. In the modern era, people pay more attention to privacy issues in IoT. All of the above motivated us to work in this field [23].

Many application domains make use of data collection and data mining techniques. The management and frequent publication of sensitive personal data in some fields (such as medical records in the healthcare industry) causes privacy concerns. Healthcare data is sensitive as it contains the patient's personal information. Privacy is, therefore, a major concern for healthcare apps [20, 21]. For better outcomes, the knowledge graph for healthcare can be constructed using unstructured, semi-structured, and structured datasets. Due to privacy concerns, Multiple parties' medical knowledge graphs are not shared with other organisations. The notion of privacy preservation knowledge graphs is helpful when using medical knowledge graphs from several sources [26].

The contribution of our work can be summarized as follows:

For WSN-based healthcare applications, privacy preservation is accomplished by secret sharing and multipath routing. There are n components in the message. These components were then transmitted to servers via multipath routing [158] [160][162]. The maximum number of discontinuous routes between the source and destination nodes is computed to allow multipath routing. Also, each component's hash functions are calculated and provided to the server. The n servers are queried for the n components required to recreate the medical data [152] [157] [164].

A privacy preservation plan is considered for important and nominal occurrences with different privacy requirements and energy consumption levels. The source location privacy for event detection (SLP ED) is established using the grid-based source location privacy (GB SLP) and a chessboard alteration pattern (SLP ED CBA) technique. A security study has been completed in the work [148-151].

For the healthcare dataset, we provide privacy preservation approaches. Support Vector Machine (SVM) outperforms other methods, according to our comparison of six machine learning classification algorithms. We offer privacy protection for learning multiple classifiers based on PCA and horizontal data [153-156].

1.8 RESEARCH GAPS AND OBJECTIVES

The Research Gaps identified are as follows:

a) Privacy preservation framework for healthcare applications [2][8][31][43].

b) Current source location privacy preservation schemes require the same privacy level for all events. Schemes for handling heterogeneous monitoring environments (nominal event, critical event) [35][46][47] need to be developed.

c) Privacy enhancing techniques during data sharing for machine learning applications [59][60][61].

The following objectives have been articulated, based on the above problem statement and accomplished in this research work. The first objective is to develop a privacy preservation framework for IoT-enabled WSNs-based healthcare applications. The thesis' second goal is to create source location privacy preservation in event-driven WSNs that are IoT-enabled. The third objective is to develop privacy preservation machine learning in healthcare applications.

1.9 ORGANIZATION OF THE THESIS

For organizational purposes, the thesis is divided into six segments. Chapter 1 introduces the topic. The issue statement, the reason for doing the job, and its goal are as follows. Chapter 1 summarises the complete study project and its motivation and goal. Chapter 1 also includes the contribution to the work completed for the thesis.

Chapter 2 presents the literature review on security and privacy protection and covers privacy protection in health care and monitoring in more detail.

Chapter 3 suggests an architecture for healthcare apps that protects privacy. Secret splitting and multipath routing schemes ensure privacy. In addition, user authentication is proposed for healthcare applications.

For IoT-enabled Event-Driven WSNs, source location privacy preservation is discussed in Chapter 4. For applications involving environmental monitoring, three strategies are suggested.

Chapter 5 presents privacy-preserving methods for medical datasets. Machine learning techniques are used for healthcare data sets, and privacy-preserving schemes are proposed.

With a brief consideration of the potential continuation of our work in the future, Chapter 6 wraps up the thesis.

CHAPTER 2

LITERATURE SURVEY

This chapter presents the related literature on privacy preservation, emphasising privacy preservation algorithms. The related work is divided into three categories for healthcare applications: existing secure and privacy frameworks, protection of source location privacy in IoT-enabled event-driven WSNs, and data privacy preservation schemes.

2.1 SECURE AND PRIVACY FRAMEWORKS FOR HEALTHCARE APPLICATIONS

The ideal aim of the medical/health care framework is to combine data with information and communication technology (ICT) to enhance the limitations of existing schemes. It considers distant patient evaluation and makes it feasible for patients to monitor their clinical records from remote locations [32]. Various surveys have been published in [28], [29], [30], [31], [32], [33], [34] and table 2.1 which highlighted the privacy issues for healthcare applications. E-medical service is a new idea in medical care and clinical sciences of the 21st century. In a perfect world, e-medical services, while utilizing ICT, consider total patient confidentiality because patients have the power to permit/deny anybody having access to their records. E-Medical services dream of a medical care enterprise that considers modern advances in innovation and social restrictions. However, because current studies exclusively concentrate on their specific fields, there is a huge void in e-Medical research. This brings in proposed arrangements that, while sufficient to address specific issues, ignored work by and large as a piece of the broader enterprise. Recent advancements in communication technology have made e-Medical care an impending reality. Various issues still need to be handled [33], [34]. The preconditions of data security should be met in these frameworks, as they contain data of an incredibly confidential nature for patients. Security and monitoring measures should be installed in healthcare services to acquire patient trust. Using WSN for patient checking by making a body area network (BAN) is a moderately new phenomenon [33].

Table 2.1: Comparison of existing privacy preservation schemes for healthcare applications

S. No.	Title	Author's Name and year	Published in	Proposed Work	Limitations
1.	A medical healthcare system for privacy protection based on IoT [26].	Tianhe Gong, Haiping Huang, Pengfei Li, Kai zhang, Hao Jiang and 2015.	Seventh International Symposium on Parallel Architectures, Algorithms and Programming.	By taking into account the characteristics of the Internet of Things and privacy protection, a compact private homomorphism algorithm and an encryption method enhanced from DES are proposed.	Computation & communication are high.
2.	Hybrid Logical Security Framework for Privacy Preservation in the Green Internet of Things [110].	Isha Batra, Sahil Verma, Arun Malik, Kavita, Uttam Ghosh, Joel J. P. C. Rodrigues and 2020.	Sustainability.	The constrained application protocol (CoAP) and object security architecture for IoT are two widely used security protocols that are compared with HLSF (OSCAR).	Need to extend the authentication.
3.	A Privacy-Preserving Re-authentication Scheme for Mobile Wireless Sensor Networks [111].	Shunrong Jiang, Jiapeng Zhang, JingJun Miao, and Conghua Zhou and 2013	International Journal of Distributed Sensor Networks.	For mobile nodes, they create a simple mobile re-authentication mechanism.	Because of the extra-long pathways, it increases energy consumption and introduces packet delivery delay.
4.	Privacy-Preserving Wireless Medical Sensor Network [27].	Xun Yi, Athman Bouguettaya, , Andy Song and Jan Willemson and 2013.	IEEE Transactions on Dependable and Secure Computing.	They suggest employing numerous data servers to hold patient data as a viable strategy to thwart an insider assault.	It suffers from various issues, which results in increased overhead and communication cost.

Existing examinations have highlighted privacy-preserving prerequisites while exploring and investigating current research. However, there seems to be a gap, as sufficient studies assess research based on the security needs of e-Medical services. An architectural understanding of e-Medical services frameworks is fundamental to more readily figuring out the e-Medical services space and the security contemplations therein. There exist various functional e-Medical care ventures. However, there is no single arrangement of principles or structural plan for e-Medical care frameworks. Significant contrasts exist in taking care of patient electronic health records (EHR) [35]. EHR shows restraint control in a few functional undertakings, while other enterprises have dedicated medical care screens to managing EHR [35]. E-Medical services frameworks should be safeguarded from threats at each point. BAN and its correspondence communication link to smart devices have alarming message climate and safety efforts that are novel to that specific segment of the e-medical care venture. Cell phones are liable for gathering sensor information, pre-handling it, and sending it to an e-Medical

services centre organization [36].

These weaknesses are made worse because a patient uses a mobile device for personal usage and health data monitoring, making it a shared resource [37]. According to research, the number of smartphones and specialized applications is increasing and will soon play a crucial role in e-healthcare systems [38]. Prominent social media programs have also been introduced and used for e-healthcare social networking [39]. Communication links join all remote users to the core network and send all data from a mobile device to it. These weaknesses are intensified by a mobile device being a shared resource, and the patient is involved in their day-to-day activities in expansion to observe the monitoring data [40]. This is alongside the presentation and utilization of social media applications for e-medical care long-range informal communication [41]. These communication interfaces aggregate all the data from a cell phone to the server and interfaces generally remote clients to it.

Recent studies on healthcare have disregarded various crucial security factors in favour of access control data. Recent research also exhibited the requirement for smart devices and PDA interfaces for patients predicting the planning and development phase [42]. Ongoing investigations have shown that the lack of standardized security strategies has caused disruptions in e-medical services ventures. A detailed and extended focus on hypothetical prerequisites and executions has caused unintentional loss of data availability, workflow interruptions, and operational feasibility [43].

In [44], authors have examined various access control schemes to control and monitor users' access. In checking their advantages and disadvantages, no single method is sufficient for our ideal degree of access control. They have carried out a role identity-based access scheme. However, there are still some disadvantages that need to be addressed.

The unpublished multipath aggregation approach is provided for wireless sensor networks [45]. The authors note that in wireless sensor networks, secret-sharing multipath aggregation offers great secrecy [45], [46]. Data manipulation and packet insertion could cause the information in the packet to change. The attacker may also introduce false messages. Moreover, the intruder may modify such signals and merge data from sensor nodes into an aggregate message. [47] investigates the use of Body Sensor Networks (BSN) to collect vital body metrics (blood pressure, electroencephalography, electromyography, and so on). With the aid of a linked network, the obtained data is transmitted to the server. The BSN server and BSN nodes are connected via a local processing unit that also serves as a router as part of the design. Also, the BSN server care's Local Processing Unit (LPU) identifies and records alterations in the body. The situation of wearable clinical sensor nodes is given [48].

The authors [49] describe a new solution for message security in WSN that uses genetic-based biometric cryptography. The authors in [50] utilize the Share-mind system to do calculations on input data while maintaining the privacy of that data, providing a workable method to avoid an inside attack. The lightweight scheme protects patient data privacy and supports medical research. The authors suggest two private-preserving data aggregation strategies—CPDA and SMART—concentrating on additive data aggregation functions [51].

Some threats target transmitting physiological data online, including eavesdropping and data falsification [50, 54]. The authors have proposed two components to ensure data security and privacy. The proposed scheme generates a distraction matrix before the server session, followed by a procedure for transmitting encrypted data. This approach ensures that even if attackers can access the encrypted or scrambled data, they cannot decipher it into plaintext.

In reference [48], the authors present an algorithm that guarantees data correctness and discuss two privacy protection strategies. The first method employs homomorphic encryption theory, while the second uses a Data Encryption Standard-based encryption algorithm (DES). Both techniques ensure data privacy and are difficult for attackers to detect. Based on current parameters, homomorphic encryption is utilized to compute patient-related parameters. These algorithms have the advantage of small-scale code, lightweight computation, and minimal computational requirements [61].

Protecting privacy assets from an attacker is vital because the attacker may be able to monitor the course of the message and possibly take complete control of or seize the method of the asset. Several routing strategies provide depressingly random paths between sink and origin sites [54]. Their method enables the system to achieve more anonymity for the source region without impacting the network's longevity.

The developers of the alleged security [52] flaws have shown that they are not immune to assaults, including impersonation, sensor node theft, and leaked verification table information. They also demonstrate that it does not guarantee untraceability, safe mutual authentication, or anonymity. They suggest LAKS-NVT for the medical Internet of Things to address these security flaws, which does not call for a server verification table. In addition to invisibility, safe authentication mechanism, and unlikability, LAKS-NVT protects stolen sensor nodes, impersonation, and replay [53]. Moreover, LAKS-NVT is safe even if the server verification table is compromised since it does not save sensitive information and the user's authentication settings in the server's database [54]. The researchers conducted a rigorous security analysis of LAKS-NVT using a mathematical model called Real-or-random (ROR) to prove it can provide secure session key protection. They also used a widely accepted method called ‘Burrows-

Abadi-Needham' logic to accept LAKS-NVT, which can ensure secure message authentication [55][56]. They also conducted a thorough security evaluation of the planned system, LAKS-NVT, using the popular generally-approved "Automated Validation of Internet Security Protocols and Applications (AVISPA)" software tool to demonstrate it is safe [99]. The researchers evaluated the effectiveness of their scheme by comparing it to other current state-of-the-art schemes and analyzed its performance. They then used an NS2 simulator to conduct simulation tests and further evaluate their scheme [53].

Servers are typically viewed as reliable nodes. However, an attacker can access all the parameters in the server's database except for Kser, the server's master key. Additionally, an attacker can intercept, erase, substitute, inject, or replay information transmitted over public channels [55]. The researchers use a threat model known as the 'Dolev-Yao (DY) Threat Model', which assumes that the sensor node is untrustworthy. If an attacker gains control of the SN, they can use a power analysis attack to retrieve and access information stored in the server node. The attacker can then potentially use this information to launch other attacks [51]. The approach proposed by Xu et al. is divided into three stages: (a) Initialization Phase: During this phase, the system administrator configures the system by producing the server master key, Kser, and then saving it in the server memory [57]. The Sensor Node and Server Node verify each other and produce the current session key to obtain valuable medical services [58]. The proposed scheme's formal security is verified (LAKS-NVT). Using the "Automated Validation of Internet Security Protocols and Apps (AVISPA)" tool [59].

2.2 PRIVACY OF SOURCE LOCATION IN EVENT-DRIVEN WSNS

The authors investigate the phantom routing method for maintaining source location anonymity [62]. The strategy is divided into two stages. The first step is a random walk of h hops from the source node to an intermediate node. In the second step, the intermediary node floods the packet to the target node [62].

The researchers developed phantom routing, a unique technique that protects the position of the source node in sensor networks without significantly increasing energy usage. The approach entails picking 2 phantom nodes and employing a methodology to identify neighbors. For every source node, two randomly picked nodes act as phantom nodes. These nodes and the sink are not located in a straight line, resulting in distinct paths for two data packets. This protocol can confuse an attacker attempting to locate the source node in sensor networks [62]. Researchers investigate reference coordinates-based routing, which makes decisions and

forwards the packet to a sink. A node stores two unique groups of reference coordinates in its caches. To transmit a packet, the node selects one element from the candidate node set randomly and then chooses a coordinate from the pool at random. The node transmits a packet to its closest neighbor node. Path diversity is used in this strategy. Path diversity makes discovering an event-detecting node difficult for the attacker [27]. The authors advocate employing Phantom Routing based on the Annular Zone technique. The technique provides balanced energy usage and reasonable safety [62]. The entire network is separated into various levels with this technology depending on the distance of the nodes from the base station (NEAR and FAR layers).

There are two situations in which the choice of phantom nodes can occur. The first situation occurs when the SINK node is far from the event-detecting node. A phantom node is chosen randomly from FAR levels. In this scenario, FAR layers are employed to find the phantom node. Upon selecting the phantom node, the message is routed via the network from the source node to the phantom node. The authors propose a routing strategy [63] that employs a phantom source and the idea of angle anonymity in a routing algorithm to safeguard the efficient source's location privacy. The authors recommended using random paths to preserve the position of a node that detects an event [64]. Additionally, the comparison of existing schemes based on source location privacy preservation is explained in Table 2.2.

The suggested solution is divided into three sections. During the initial phase, each node divides its neighbors into three groups according to their distance from the access point: nearer set, equivalent set, and distant set. The node chooses a neighbor from the further or equivalent set at random and passes the packet to it for h1 hops, wherein h1 is a positive integer smaller than the network width. This resulted in a random stroll far from the base station. As in the

Table 2.2: Comparison of existing schemes based on Source Location Privacy Preservation

S. no.	Title	Author's Name and Year	Published in	Proposed Work	Limitations
1.	Preserving source location privacy in monitoring-based wireless sensor networks [112].	Xi Y, Schwiebert L, Shi W, and 2006.	International Parallel and Distributed Processing Symposium.	They incorporate the bloom filter within the packets so that they can avoid recently visited nodes.	The information about previously visited nodes may be obtained by the eavesdropper, posing a risk to the privacy of a person's location.
2.	Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records [113].	Beata Strack, Jonathan P. DeShazo, Chris Gennings and 2014	Hindawi Publishing Corporation.	The HbA1c assessment of diabetes may help to improve patient outcomes andreduce the cost of inpatient treatment.	Readmission rates remained the highest for patients with circulatory diagnoses.
3.	Protecting source location privacy in a clustered wireless sensor network against local eavesdroppers [114].	Al-Mistarihi MF, Tanash IM, Yaseen FS, Darabkh KA, and 2020.	Mobile Network and Application (Springer).	Clustered-based WSN DSP, DT, and Hybrid schemes are used.	Due to the extra-long pathways used in this technology, packet delivery latency and energy usage are increased.
4.	Energy-efficient source location privacy protection for network lifetime maximization against local eavesdropper in wireless sensor network (EeSP) [115].	Naveed Jan and Samadullah Khan, and 2022.	Transaction on Emerging Tel Tech. (Wiley).	Dynamic cluster head selection is used.	Lacking in providing efficient communication, improvement for the up-gradation of service.
5.	A source location protection protocol based on dynamic routing in WSNs for the social internet of things [117].	Han G, Zhou L, Wang H, Zhang W, Chan S, and 2018.	FuturGenerComput Syst. (Elsevier).	Route the packets subsequently sent in the sink's direction.	Due to the extra-long pathways used in this technology, packet delivery latency and energy usage are increased.

second phase, the packet is routed to the equivalent set for a certain quantity of hops and then to the nearer set till it reaches the base station. The authors propose a method based on inclination angles to select intermediary nodes while preserving source-location privacy. The phantom single-path routing strategy is compared to the 'angle-based routing scheme' [64][65]. According to the simulation findings, the ADRS system improves packet latency and network safety. The author achieves this by increasing the variety and unpredictability of route placements of phantom source nodes throughout the network [66].

The candidate node's remaining energy is considered while choosing the next hop node [67]. The authors examine privacy-aware protocols regarding routing protocols, parameters and trade-offs in WSNs under practical scenarios [68].

The authors study source location privacy preservation approaches based on clustering algorithms [69]. [70] published new research on privacy preservation techniques and solutions

for diverse data. The authors claim that employing a dynamic ring maintains the source location's secrecy [71]. The approach is divided into three stages. The intermediate node is selected in the first stage. The decision depends on where the destination and event-detecting nodes are deployed. The source node sends the information gathered to the intervening node. The second step transfers the packet from the intermediate node to the mixing ring. The packet in the active mixing ring is subsequently forwarded to the sink node. The authors [71] extend the source location privacy preservation schemes to multiple sinks. To improve source location privacy, packets in this system are separated and routed through numerous sink nodes using dynamic routing paths.

The authors divide the networks into cluster areas. As a result, the network's energy consumption is reduced [72]. The sensor nodes within the network rotate between alert and rest states in every zone. As a result, the network's energy consumption is reduced [72]. The technique takes the shortest route between the event-detecting and destination nodes. Nevertheless, if the attacker adopts the tactic of not visiting previously visited nodes and instead waits until a new node is discovered, the scheme is reduced to merely the shortest-way scheme [72]. The authors suggested that the packets from SN be sent to placed mediate and diversion nodes carefully. Due to the extra-long pathways used in this technology, packet delivery latency and energy usage are increased [116].

In [73], the authors have shown a solution to safeguarding the source's position in this study by appropriately altering the sensor routing protocols, preventing a threat from tracking the sensor signal back to its source. Giving special attention to flooding protocols, they consider location privacy and sensor network energy consumption while creating and testing our privacy-aware routing techniques. In response to these discoveries, they offered phantom routing, a versatile routing strategy that masks the location of the source. Phantom routing is a two-stage method that begins with a guided walk along a random path and ends with routing from the phantom origin to the sink. According to their findings, phantom routing effectively safeguards the source's position during sensor broadcasts.

The authors have offered a technique for safeguarding the anonymity of the place of origin based on random routing algorithms [74]. Packets are arbitrarily routed from the origin to the sink node via strategically placed media or diversions nodes to guarantee maximum privacy. Mediate, or diversion nodes are picked at random based on their location. Depending on the location of the originating node, packets are directed through several network zones. The suggested method ensures that the following packets are routed across a wide range of routing patterns, making it challenging for attackers to track them back to the location of the originating

node. The simulation findings demonstrate that the suggested technique outperforms earlier routing-based source location privacy solutions by effectively confusing the opponent while enhancing source location privacy [74]. To improve SLP in WSNs, the authors suggested in [75] a trace cost-based source location privacy protection system (TCSLP)' in WSNs for smart cities. They start by creating a phantom region where phantom source nodes are placed far away from the real source node. Secondly, they send packets using a combination of random and shortest path routing to boost the location secrecy of the actual node. The next design is a specialized trace cost zone, consisting of numerous sensor nodes with varied weights spread throughout mountains, plains, and woods to impede the adversary's trace speed. Eventually, a ring of packets forms around the sink node. The ring is made up of a large number of nodes drawn from neighboring grids. In contrast to the restricted flooding-based SLP and the improved SLP protocol based on SLP (SLP-E), which enable packets to be delivered in either direction but neglect overloaded nodes, our suggested technique may avoid node overuse and decrease route overlaps. According to simulation results, the proposed TCSLP system can increase safety time and improve SLP in wireless sensor networks (WSNs) for smart cities.

The authors in [76] evaluated the trajectory privacy research that has already been done about wireless sensor networks, location-based services, and geosocial networks. They categorized and defined the key tactics in each situation according to their preferences. They also looked at future trajectory privacy issues and approaches. They underline the importance of in-network computing, which is present in all three situations. When operations and algorithms are carried out on data streams as they are being delivered between network nodes, this is referred to as in-network computing. They concentrate on TPP methods and protocols and the safeguards they provide against privacy infractions that occur before data is transmitted to an offline database.

A strategy that provides source-area protection and material categorization through a two-stage directed interaction is suggested [77]. The message source randomly selects middle-of-the-road hubs during the first directing stage in the sensor space. Then, it sends the information packet to the haphazardly selected middle-of-the-road hub before being pointed to a ring hub. The local source area is protected at this level. The informational bundle will be combined with other bundles during the second acting stage using an organizational blending ring (NMR). This step provides global source-area security at the network level. Our reproduction findings further show that the suggested scheme is incredibly productive and can be applied to practical applications, even though it has the option to provide source-area protection to WSNs [77]. The research [78] suggests using Annular Zone (AZR)--based Phantom Routing as a routing

approach to guarantee acceptable source-location privacy while preserving balanced energy usage. This method separates the network into many levels depending on how close every node is to the SINK node. Whenever the node is remote from the SINK node, it chooses a phantom source randomly from the same layer. Otherwise, the FAR layers pick a phantom source at random. The message will then be sent along the annular routing channel from the source node to the phantom source. Our simulation results demonstrate that the suggested solution outperforms existing systems in terms of performance while ensuring source-location anonymity.

The work [79] gives a conventional design for the source-area protection problem and investigates the safeguard properties of several sensor direction conventions to handle source-area protection for sensor organizations. This paper suggests that security period and catch probability are two metrics for assessing source-area protection in sensor companies. Their analysis of popular directing techniques used in current sensor organizations also considered crucial framework difficulties, including energy use. They discovered that most conventions can't provide effective source-area protection. They suggest new practices that strengthen these guiding norms to enhance source-area protection. This protection enhancement should not have any substantial drawbacks related to a considerable increase in asset use. With minimal energy expansion above, we have created a technique known as "phantom routing," which has shown to be adaptable and ready to prevent the opponent from chasing the source location.

In contrast to traditional guiding plans, this study's directing strategy offers more grounded source area protection [80]. By providing incredibly arbitrary directing routes between the source and sink hubs, the study responds to some limits of four current schemes. To make the enemy's courses more overpowering, the strategy arbitrarily distributes bundles to the sink hub through intermediary hubs that are carefully placed in key locations. The suggested method employs a randomizing component that creates an irregular path for each advancing parcel to provide excellent protection. Reproductive findings show that the suggested scheme provides a larger window of well-being and a more robust defence against other schemes. Moreover, the strategy offers a stronger defence against patient and thoughtful adversary models [79, 80].

The Phantom Routing with Location Angle (PRLA) is a novel source-safeguarded WSN protocol proposed in [81]. In the PRLA, propensity points are known and used with direct arbitrary strolls, which attempts to avoid choosing routes harmful to preserving the source region. Reproduction findings reveal that PRLA reduces the wellness period by up to half with a slight increase in energy overhead, compared to the Apparition Single-Way Directing convention suggested in the text. This study proposes two phantom routing-based strategies for

providing source location privacy in the multisource/asset scenario, which has gotten little attention in the literature. Phantom routing's goal confuses the attacker by randomly relaying packets to remote nodes [81]. The first method is a 'phantom routing-based backward random walk' (PRBRW). PRBRW routes packets to the base station (BS) by incorporating greedy forwarding and a backward random walk (RW) strategy. The first method reduces network lifetime and has a low entropy measure despite having higher performance increases in the capture percentage and the safety period. To fix this, an upgraded phantom routing system is being created. PRLPRW ('phantom routing-based L-path RW') is recommended. The second approach is broken down into three steps: 1) the L-walk, 2) the greedy walk, and 3) the pure RW [79, 80 and 81]. This method performs better than competing methods in terms of capture ratio, safety duration, and entropy. There is a 477-fold rise in entropy and a tenfold rise in network longevity compared to PRBRW. The effectiveness of the developed analytical models is evaluated and contrasted with the standard shortest path routing protection-less technique (SPR). Compared to SPR, PRBRW and PRLPRW provide gains in capture ratio of 60 and 73 times, respectively. At the same time, previous pure RW and forward RW approaches relying on phantom routing only achieve gains in capture ratio of 54 and 34 times, respectively [81].

2.3 DATA PRIVACY PRESERVATION SCHEMES

The emerging healthcare Industrial Internet of Things (Health-IIoT) must address many basic security and privacy challenges, including secure fine-grained data transfer, privacy-preserving keyword-based cypher text retrieval, and malicious key delegation. We've divided it into two sections: Privacy preservation in Machine Learning and Knowledge Graphs.

2.3.1 PROTECTION OF PRIVACY IN MACHINE LEARNING

The information set is available online from UCI's machine learning repository in the .csv format [82]. Preliminary data analysis and preparation have retained only those features containing adequate information [82]. The dataset taken contains 101767 rows and 50 columns. The parameters selected for investigation include age, number of diagnoses, emergency visits, number of inpatients, number of lab tests, number of medications, number of outpatient visits, number of procedures, and length of stay in the hospital. We now present the visualization of the selected dataset. Data visualization is how data is in pictorial or graphic format. Data visualization helps people understand the importance of information in a simple, easy-to-

understand format by summarizing and presenting large quantities of data to communicate the information clearly and effectively. A heat map is a method to view the data showing the magnitude of a two-dimensional colour phenomenon.

The colour difference can alter with hue or intensity, providing unique visual clues as to how the phenomenon is dispersed or changed throughout space. The smaller dark grey needs to extend the authentication.

Designing effective methods for privacy-preserving deep learning requires a lot of work. In their assessment of several approaches, Zhang et al. [83] focused on collaborative learning and considered the two crucial stages of deep learning: training and inference. During the training and inference phases, Chang and Li [84] concentrated on privacy concerns, including assaults on learned models and their accompanying dangers and solutions. Tanuwidjaja et al. recently explored several privacy-preserving techniques [85] based on homomorphic encryption, multiparty computation, and differential privacy. The paper also included a comparison of the solutions that were looked at for each concept. Similarly, Riazi et al. [86] explored privacy-preserving deep learning methods focusing on cryptographic techniques. Together with descriptions of the solutions and performance comparisons, the paper also included key attacks on deep neural networks (DNNs). In addition, Boulemtafes et al. [87] emphasized open research and made recommendations after presenting a current study of the standardized confidentiality-preserving “deep learning” solutions and their evaluation outcomes. The aforementioned articles, however, only addressed the privacy issue in a generic framework and did not consider particular target environment restrictions.

When focusing on the IoT environment, Zheng et al. [88] presented a taxonomy of various privacy-preserving machine learning algorithms for the training and inference phases before discussing their drawbacks on IoT end devices. The authors also presented an obfuscation-based inference technique that protects privacy. The authors describe the framework to address the privacy issue in a diverse network of several clinical institutions while retaining the data's usefulness and the patient's privacy. The suggested framework's authentication and authorization elements must be expanded [119].

In [89], the authors expanded on their solution. Unfortunately, the assessment briefly summarizes the shortcomings of several privacy-preserving methods rather than providing a comprehensive overview of the existing alternatives. Furthermore, the restrictions are not assessed using a predetermined set of standards. Moreover, no distinction is made in the evaluation between training local and remote models. Table 2.3 describes the comparison of existing schemes based on machine learning applications.

Table 2.3: Comparison of existing schemes based on Machine Learning applications

S. no.	Title	Author's Name and Year	Published in	Proposed Work	Limitations
1.	Reauthentication scheme for mobile wireless sensor networks [118].	Vandana Mohindru, Ravindara Bhatt, Yashwant Singh, and 2019	Sustainable Computing Informatics and Systems.	They suggest a mobile node authentication mechanism with low energy consumption.	Because sensor nodes frequently travel from one place to another and reconnect to other sensor nodes, their mobility is a serious problem.
2.	A Two-tier Strategy for Priority-based Critical Event Surveillance with Wireless Multimedia Sensors [120].	Bhatt R, Datta R, and 2016	Wireless Networks. (Springer).	A two-tier technique that uses cheap audio tier nodes that are distributed densely has been proposed by the authors.	With the aid of WMSN, they explore the issue of monitoring significant events in an area of interest.
3.	Architecture for Preserving Privacy During Data Mining by Hybridization of Partitioning on Medical Data [121].	Asha Khatri, Swati Kabra, Shamsher Singh, and 2010	Mathematical/Analytical Modelling and Computer Simulation.	The research demonstrates that accuracy may be increased by using both vertical and horizontal division.	Increasing classification accuracy in both local and international mining can safeguard privacy.
4.	Classification of Crime Data for Crime Control Using C4.5 and Naïve Bayes Techniques [122].	Obuandike Georgina N., John Alhasan, and 2017	IJMAO.	C4.5 performed better with higher accuracy on the three dataset against Naïve Bayes.	Though it is a lazy classifier, it can compete effectively among other classifiers.
5.	A novel framework for preserving privacy of data using correlation analysis [123].	Animesh Tripathy, Matrubhumi Pradhan, and 2012	Knowledge-Based Systems.	This improved technique minimizes the number of 1's that must be detected in the database.	Randomization Method, Anonymization Method, Encryption Method.

Sun et al. [90] suggested a "Privacy-aware and Traceable Fine-grained System" that permits safe granular data transmission, retrieving information while protecting privacy, and effective encryption and decryption operations. Role-based encryption keys are the essential building block for the storage of sensitive data in cloud settings, and Sathya and Raja [91] introduced a Euclidean L3P-based Multi-objective Successive Approximation (MOSA) method, a strong privacy safeguard in the smart healthcare setting.

2.3.2 PRIVACY PRESERVATION IN KNOWLEDGE GRAPH

The ability to locate hospitals and healthcare facilities is crucial, particularly in cases where a patient requires a specialist with good reviews or government approval. In emergencies, it's important to be able to quickly locate a medical facility or trauma center with the necessary specialization. [92]. Clinicians can leverage knowledge graphs to identify diseases and connect

their symptoms to these diseases. This is accomplished through visualization and querying of the knowledge graph, which can improve doctors' decision-making process [93]. In the healthcare industry, knowledge graphs identify the most commonly prescribed medication, generic salt, or manufacturer for a specific drug. Today, medical practitioners give therapy by taking notes, evaluating a patient's medical history, or comparing it to previous case studies. Integrating Electronic Health Records (EHR) and knowledge graphs can assist in speeding up this process. Knowledge graphs can be used as a recommendation system in the healthcare area. The display and querying of knowledge graphs can also help clinicians make better decisions [94]. A knowledge graph model of the patient journey incorporates all contacts with healthcare stakeholders. The occurrences involve hospital visits, treatment, and many days in the hospital, among other things. This gives insight into the human body and may be utilised to frame medical operations by researchers, clinicians, and government authorities.KG aids in discovering medications with multiple uses, as well as communities and diseases associated with them. To help clinicians comprehend a patient's history or illness better, healthcare data from different facilities must be communicated. Information like ailment name, among other things, age, postal code, phone number, pin code, and religion, is included in the large-scale healthcare information. In big data analytics, several confidentiality-preserving strategies are implemented to preserve the data. Individual data privacy is at risk since it is shared throughout hospitals, insurance providers, etc. The graph is changed during the anonymization process to protect privacy. Data distribution strategies are frequently used to protect privacy, including dividing information into horizontal and vertical spaces available from diverse origins. As opposed to vertical distribution, which distributes the attributes or columns across several locations, partitioning horizontally keeps the columns constant in different locations [95]. Moreover, Ogundoyin et al. [96] and other research teams suggested the PAASH method, lightweight, fine-grained access control and privacy-preserving authentication system, for smart health [60]. The issues of smart healthcare in smart cities in terms of security, effectiveness, and privacy are discussed in this paper. Moreover, Vineela et al. [97] proposed an authentication technique that uses mutual authentication, and encryption is done between the user and the cloud environment to protect the confidentiality of data in the environment. Zhou et al. [98] created a human-in-the-loop-aided (HitL-aided) strategy to safeguard privacy in smart healthcare. They used a block design technique to obscure numerous health statistics gathered from hospitals and smart gadgets. To enable private access to health records from the smart healthcare platform, they also incorporated a human-in-the-loop (HitL). Also, a novel strategy for protecting privacy within the context of predictive modelling was put forth in a

paper by Krall et al. [98]. This approach reduces the risk of model inversion while satisfying the need for differentiated privacy. Hussain Seh et al. [100] defined an effective framework that uses machine learning techniques to preventatively preserve the security and confidentiality of clinical data to detect erroneous user access against Electronic Health records and to safeguard privacy in healthcare data. On the other hand, he et al. [101] proposed a password strength metre that considers the user's data. Users can choose passwords with a higher level of security with its assistance. Moreover, Ibaida et al. [102] proposed a unique privacy-preserving method that creates a lightweight neural network to lighten the load on the network while maintaining the privacy of the electrocardiogram signals (ECG). El Zouka et alrecent.'s study [103] defined a safe healthcare surveillance system that uses fuzzy logic-based decision support (FBIS) systems to determine the patient's status for the same reason. The suggested model includes a trusted setting for gathering verified physiological data. Moreover, Ma et al. [104] defined the SCF-CLSPE scheme as a secure certificate-less searchable public key encryption (SPE) technique for SHS that can withstand chosen keyword attacks (CKA) as well as keyword guessing attacks (KGA). This plan has also been evaluated, and it has been shown to have lower communication and computation costs. Jayaram and Prabakaran introduced a privacy-preserving additive homomorphic encryption for edge-layer processing data and removing non-sensitive data [105]. Also, a cloud layer adaptive weighted probabilistic classifier model is suggested for onboard disease prediction and remote rehabilitation of patients. Also, there are a lot of healthcare-related solutions, like those in Refs. [107–109], employ machine learning, deep learning, or combining the two to forecast serious diseases. These projects directly assess and track patient health to avoid serious health issues. Even so, these works offer little thought to patient privacy and instead place a greater emphasis on patient data. In opposition to these claims, Ge et al. work [106] ensured the data deletion approach by the data owner to restrict access to their health data while simultaneously seeking to detect disease using deep learning.

2.4 SUMMARY

This chapter shows the current situation of the existing schemes. Moreover, it also identifies the research gaps in the existing literature, as shown in section 1.8. The next chapters of the thesis present the proposed objectives.

CHAPTER 3

FRAMEWORK FOR HEALTHCARE APPLICATIONS

Healthcare systems built on Wireless Sensor Networks (WSN) are becoming more prevalent daily in informing people about their health and living conditions. WSN-based healthcare applications, however, have privacy and security issues. The vulnerability of WSN-based healthcare applications to attacks and security concerns is intriguing and difficult. For WSN-based healthcare applications, the chapter introduces a multipath routing, secret sharing, and hashing-based privacy preservation solution. The wireless sensor network's healthcare data collection is divided into *n* components. Also, each component's hash value is calculated using a well-known hashing method. To identify changes in the message, look for changes in the hash value. Multipath routing is then used to send these *n* components to *n* servers. In-depth simulations are provided in this paper to support the novel technique. Findings demonstrate that multipath routing and secret splitting preserve privacy in a WSN-based healthcare system.
Healthcare applications are considered promising fields for wireless sensor networks, where patients can be monitored in hospitals and at home using Wireless Medical Sensor Networks (WMSNs). Mobile body networks have wireless sensor devices worn by patients that provide physiological sensing. While the patient data is transmitted to the physician, an adversary may capture the physiological data from the wireless channels and alter the physiological data. After the attacked data (i.e., altered data) is sent to the physician, it could endanger the patient.
The threat of sophisticated attackers intercepting data in the current network landscape is a constant concern. Traditional single-path data transmission methods are susceptible to attacks like eavesdropping, man-in-the-middle, and tampering. Using a multipath secret sharing-based scheme is beneficial and essential to combat these vulnerabilities proactively.
The multipath secret sharing-based scheme is a comprehensive strategy that involves dividing the secret data into multiple shares and transmitting each share over different network paths. This comprehensive approach significantly enhances security, as an attacker must intercept all the separate paths to reconstruct the original secret. For instance, if a message is split into n shares, each sent via a distinct route, an attacker must intercept all n shares to reconstruct the original document. This comprehensive nature of the multipath strategy ensures robust protection against potential data breaches, making it a vital mechanism for secure communications in hostile environments.

3.1 WSN-BASED HEALTHCARE SYSTEM

The autonomous sensor nodes that make up a WSN-based healthcare system communicate with one another via wireless technology. These nodes collect physical data from the region of interest, including motion, temperature, pressure, etc. The medical system helps keep track of each patient's condition and monitors their disease. Doctors assess the patients and urge them to take certain safeguards for a specific time. As a result, the healthcare monitoring system offers in-home help for patients who are elderly or have special needs [165]. Information about the patient should never be made public since it might be used inappropriately, or privacy concerns might prevent people from fully utilising technology. E-healthcare refers to information and services that stakeholders can share or change using appropriate technology [167]. With the help of such a healthcare system, medical staff members and/or doctors can discuss the data and offer patients the best treatment option. Patient data for the study is abundant in electronic health records. The issue is that society will learn facts about patient privacy. M-health [172] refers to gathering patient or individual health data via a mobile device. The development of sensor network technologies and the usage of these technologies in healthcare applications are expanding quickly [168]. Today, many applications, including those for monitoring blood pressure and heart rate, are in use. WBANs are a new industry that has emerged to address the growing use of sensor technology [177].

In a multipath secret sharing scheme, privacy is measured by the extremely low probability of an attacker intercepting all paths simultaneously and the minimal information leakage from any single intercepted share. Path independence and the redundancy of shares play a significant role in enhancing privacy, ensuring the secret remains protected even if some paths are compromised.

Shamir’s Secret Sharing Scheme splits the message (secret) into n shares. These n shares are transmitted via n multi-path routes using a routing algorithm. They are then reconstructed at the receiver side using Shamir’s Secret Sharing Scheme.

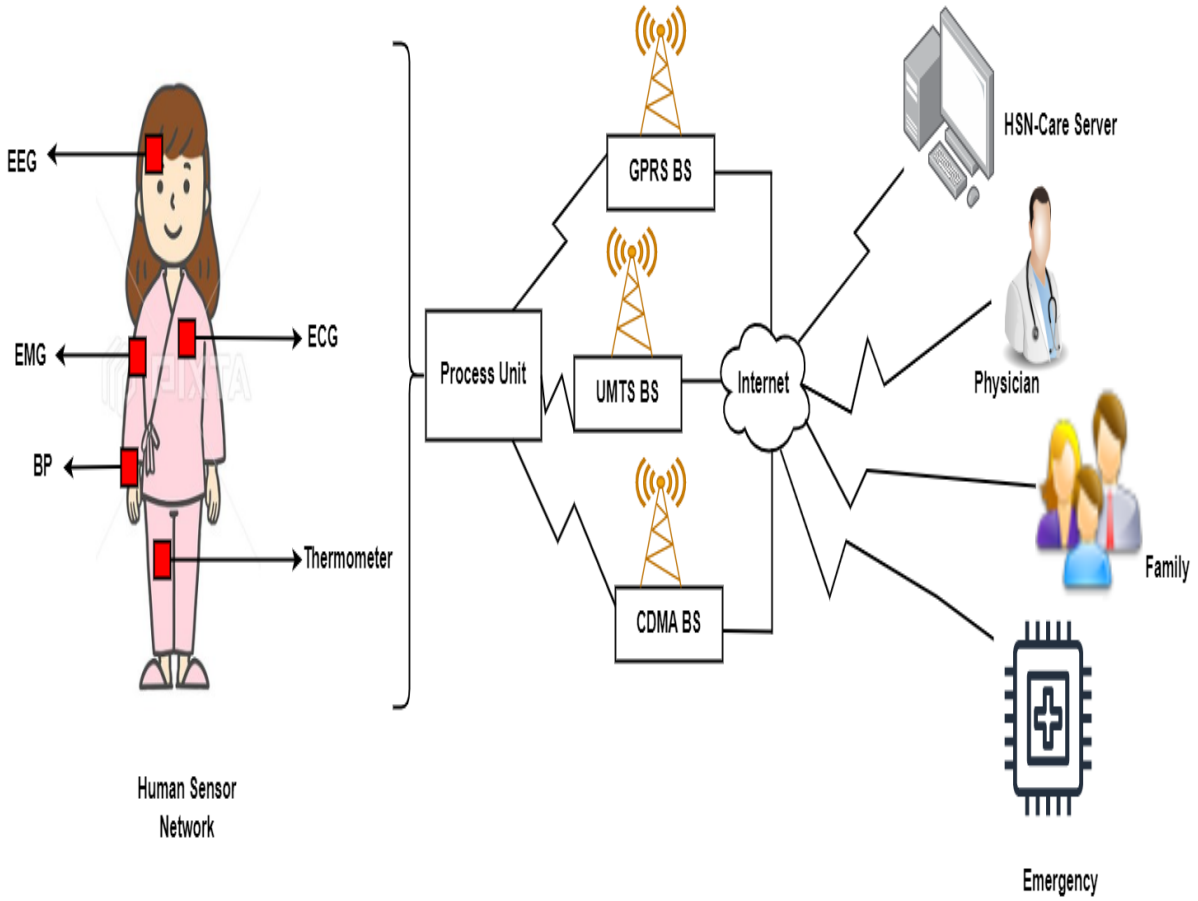


Figure 3.1: WSN Architecture

Figure 3.1 represents the primary architecture of the healthcare system, which is based on WSN. The sensing field regularly monitors parameters like blood pressure, brain signals, heartbeats, and patient temperature. The processing unit then processes the information collected and analyses the parameters [181]. The systems' stakeholders are provided with the information processed by the base station through the Internet. Health department professionals monitor the patient based on the information gathered from the base station. Nevertheless, attackers using a high-end receiver can catch hold of the data collected from the medical sensors and further may circulate the same on different social networking sites and the dark net. The authors discuss a technique for preserving privacy that protects against inside attacks [174]. As wireless communication allows for easy eavesdropping, hackers can quickly introduce harmful messages into the network. Many lightweight encryptions and a MAC generation technique are proposed to provide a secure connection between data servers and medical sensors [169]. One needs an identity-based signature or digital signature to protect the information. In WSNs, multipath routing is described using various methods [166]. The system authenticates the sender (x) and the receiver (y) via a multipath routing protocol.

An attacker capturing and modifying data packets in transit can gain sensitive information, disrupt communications, inject malware, and manipulate transactions for financial gain. The potential consequences of these actions are severe, including identity theft, financial fraud, corporate espionage, and a competitive advantage, significantly benefiting the attacker.

3.1.1 SECURE MUTUAL AUTHENTICATION AND KEY AGREEMENT

A secure mutual authentication and key agreement (MAKA) system for the Internet of Medical Things is an important security aspect for protecting users' health information while providing efficient healthcare services [9]. Many MAKA schemes have been introduced in recent decades to ensure user privacy. Many of his subsequent MAKA schemes were developed to overcome these security flaws using the smart card and biometrics [170]. However, these schemes store sensitive user data in the server database. So, if the data stored on the server is disclosed to an attacker, the entire system will collapse.

Several authentication protocols have been developed for the Medical Internet of Things to ensure user privacy. However, these protocols do not protect against verification theft or table leak attacks nor provide secure mutual authentication, anonymity, or untraceability [11].

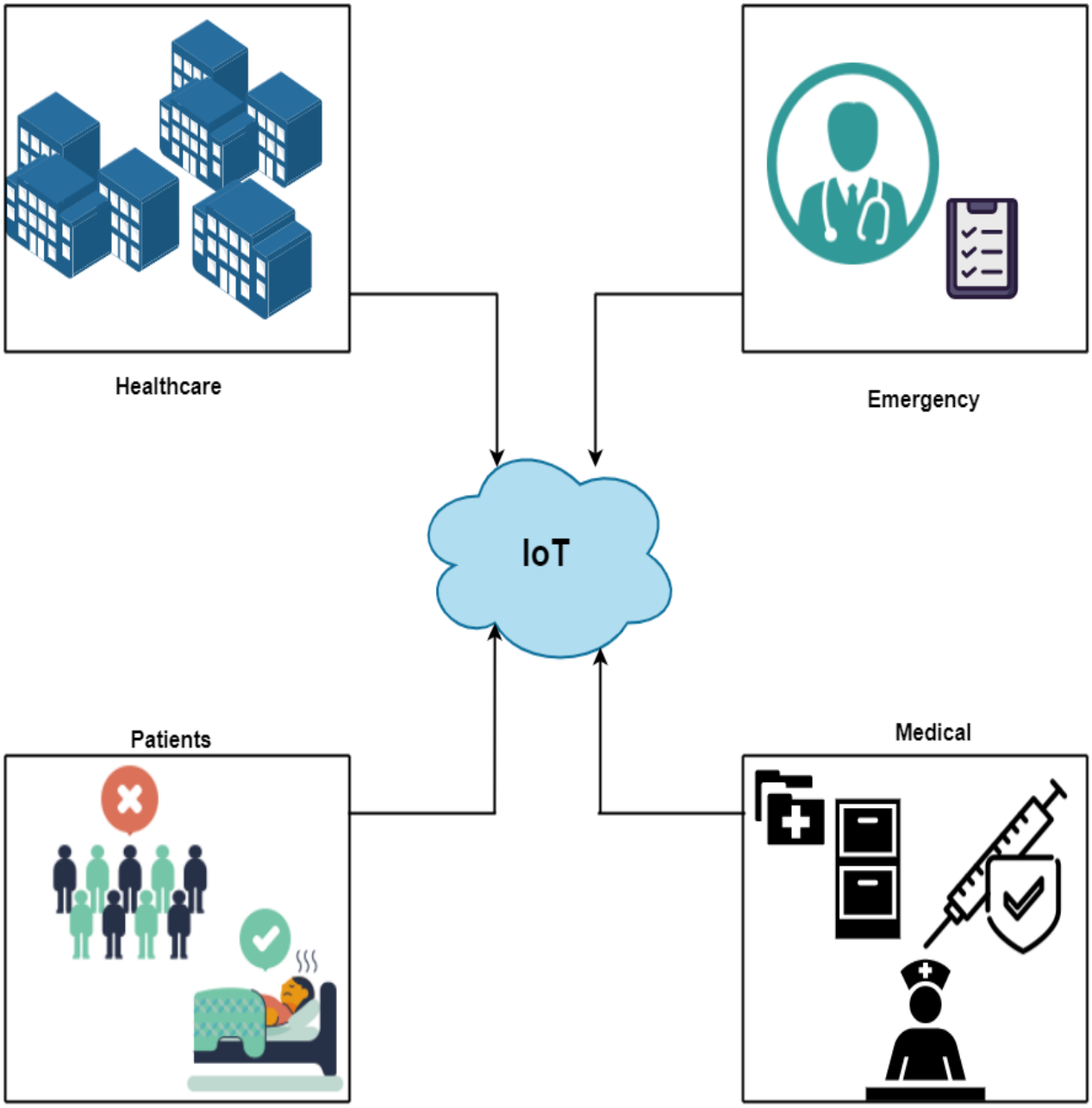


Figure 3.2: Medical-IoT System

The Internet of Things (IoT) allows objects to connect and communicate over the Internet. Through various new technologies, such as radio frequency identification (RFID), sensor technology and embedded system technology, the IoT will bring the concept of intelligent identification and management to life. Wireless sensor networks (WSNs) are an important technological backbone of the Internet of Things, providing data sources for Internet of Things applications [12]. A WSN typically consists of multiple sensor nodes that communicate with each other over a wireless network. It is typically used to monitor environmental conditions in

a specific area based on information gathered from sensor nodes. The overall WSN architecture has three participants: users, gateway nodes, and sensor nodes [13]. To collect environmental parameters, deploy a sensor node in the region of interest and transmit these parameters to the gateway node via a wireless channel. Authorized users can access this data, and the combination and analysis of this data help managers make the right decisions. WSN has important applications in many industries, such as health monitoring, intelligent transportation, and environmental monitoring [14].

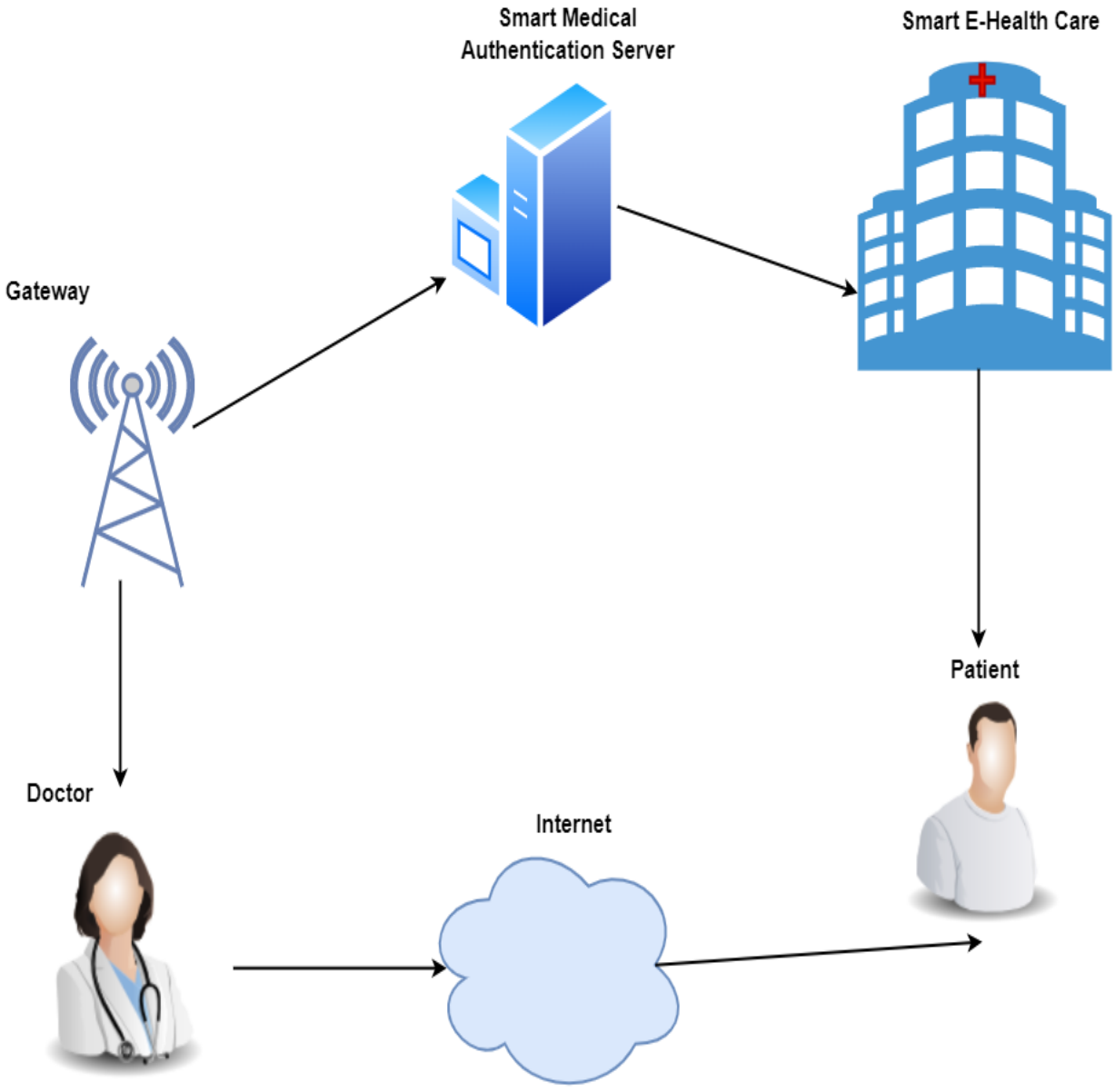


Figure3.3: System Model

3.1.2 SYSTEM MODEL AND THREAT MODEL

Figure 3.3 depicts the system model, which consists of three entities: the Doctor (D_i), the Medical Authentication server, and the Patient (P_i).

TABLE 3.1: NOTATIONS AND THEIR SIGNIFICANCE

Symbol	Significance
PWD_i	Passwordofthedoctor
BIO_i	BiometricidentityoftheDoctor
randN	RandomNumber
Keysecret	SecretKeyoftheDoctorend
$DID_{Gateway}$	Doctorgatewayidentity
T_{id}	Time-stampid
$MAS_{gatewayid}$	Gatewayidentityofthemedical authenticationserver
Tokens	*A short*authenticationidentity
||	Concatenationoperation
h ()	Hashfunction

(i) Doctor: The doctor's device must prove its authenticity on the network before communicating with the medical authentication server to access the patient's health data.

(ii) Smart Medical Gateway (SMG): The gateway acts as an interface between the authentication server and the doctor. The smart medical gateway (SMG) unit involves various smart medical appliances. The data generated from these medical appliances are aggregated through the SMG over an insecure channel. Meanwhile, the medical authentication server verifies and authenticates every medical staff request.

(iii) Patients: Patients are persons who are admitted to the hospital for treatment. The patients are equipped with various smart IoT sensor nodes that gather and communicate the real-time health data of patients. These IoT-enabled sensor nodes are low-power devices with limited battery constraints.

A. Threat Model

The threat model's security is important when sending data over insecure wireless channels. The user's authentication and validation may be vulnerable to various threats. The threat model generally consists of active and passive attacks. The attacker can perform such attacks and affect the communication and processing of the medical IoT network. The attacker can alter use-full data generated from the smart medical IoT nodes. Moreover, they can also impersonate the doctor's device and forge secret information from it.

Active attacks involve the attacker altering or disrupting communications (e.g., message modification, impersonation, and denial of service). Passive attacks involve eavesdropping or monitoring communications without changing them (e.g., traffic analysis and interception).

The work primarily considers one type of attack model: passive attack. Rest assured, the work is focused on passive attacks. With the help of these [115] references, the attacker is passive, which means that he will not harm the network's everyday workings. The attacker will not destroy any network equipment but will use his powers to locate the sensor nodes.

Improving and overcoming such issues requires a substantial technique that protects the network and authenticates every user's device. In this paper, we have assumed the Dolev-Yao (DY) threat model [23] [19]. In the DY threat model, attackers can intercept and capture important secret identity exchanges between the trusted doctor/user and the smart medical centre. The attacker can also perform inside threats to forge the device data.

3.2 PRELIMINARIES

This section covered Security enrichment and how Shamir's secret splitting can improve it. The fundamentals of a secret sharing scheme are now presented.

3.2.1 Secret Sharing

A process known as secret sharing defines secret distribution. This is done by a group of contributors with access to the data. Only when the various components of a secret are

combined again can it make sense; on their own, individual shares are meaningless. An individual shareholder who tries to get data can never update the data.

3.2.2 Working on Secret Sharing

The Shamir secret-sharing technique has several variations. Here, we employ the Shamir secret-sharing system based on a threshold. Assume that *r* and *p* are both positive integers, with $r < p$ and *p* being the number of participants required for *r* participants to compute the value of secret value *S*. Let us now assume that (*h*, *j*), scheme. When ($h = j$), all the contributors are required simultaneously to reconstruct the secret (S). On the other hand, when ($h < j$), only the (*h*) contributors are required simultaneously for the reconstruction of the secret (S).

Equation 3.1 considers a polynomial for the Shamir secret-sharing scheme.

$$f(y)=a_0+a_1y+a_2y^2+\cdots+a_{k-1}y^{k-1} \quad (3.1)$$

Let's use an illustration to describe Shamir's secret to make the concepts easier to understand. Let S, the medical information gathered from the node of the wireless medical sensor networks, be a secret. Let S have a value of 1121. In Equation (3.1), the polynomial is displayed. For producing frequent integers with j = 6 and h = 3, where *h* is the number of participants required to reconstruct the secret S. The values of the coefficients a_1 and a_2 are 150 and 60, respectively. Equation 3.1 can now be written as shown in Equation (3.2).

$$f(y)=1121+150y+60y^2 \quad (3.2)$$

The points of secret-sharing for all six participants are as follows: (1, 1331), (2, 1661), (3, 2111), (4, 2681), (5, 3371), (6, 4181). Each participant's share is a tuple. The shares are distributed among six participants. Only three (h<j) participants must reconstruct the secret, as shown in the next paragraph.

The tuple values (x_0,y_0) = (2, 1661), (x_1,y_1) = (4, 2681), (x_2,y_2) = (5, 3371) can now reconstruct the secret S. The process of generating the secret S can be explained by considering the Lagrange Polynomials.

$$l_o = \frac{y - y_1}{y_0 - y_1} X \frac{y - y_2}{y_0 - y_2} = \frac{y - 4}{2 - 4} \text{X} \frac{y - 5}{2 - 5} = \frac{1}{6}y^2 - \frac{3}{2}\text{y} + \frac{10}{3}$$

$$l_1 = \frac{y - y_0}{y - y_0} \text{X} \frac{y - y_2}{y_1 - y_2} = \frac{y - 2}{4 - 2} \text{X} \frac{y - 5}{4 - 5} = -\frac{1}{2}y^2 + \frac{7}{2}\text{y} - 5$$

$$l_2 = \frac{y - y_0}{y_2 - y_0} \text{X} \frac{y - y_1}{y_2 - y_1} = \frac{y - 2}{5 - 2} \text{X} \frac{y - 4}{5 - 4} = \frac{1}{3}y^2 - 2\text{y} + \frac{8}{3}$$

$$\text{f}(y) = \sum_{j=0}^{2} \text{y}_\text{j}\, \text{Xl}_\text{j}\,(y) = 1661\,(\,\tfrac{1}{6}y^2 - \tfrac{3}{2}y + \tfrac{10}{3}) + 2681(-\tfrac{1}{2}y^2 + \tfrac{7}{2}y - 5) + 3371(\tfrac{1}{3}y^2 - 2y + \tfrac{8}{3}) \quad (3.3)$$

As shown in Equation 3.4, the reconstructed function is the same as the original function.

$$f(y) = 1121 + 150y + 60y^2 \quad (3.4)$$

3.2.3 Multipath Routing

A source can use several routing algorithms to find different paths for the secure transmission of messages from the source to the sink node [3, 18].

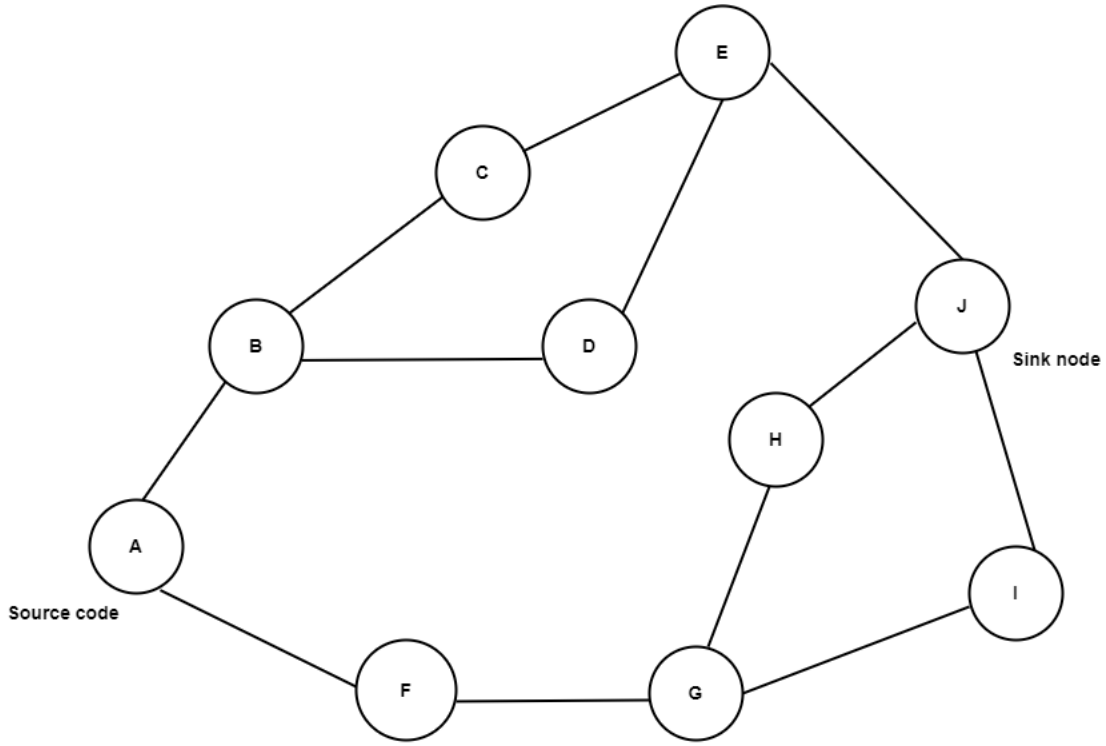


Figure 3.4 (a): Multipath Routing

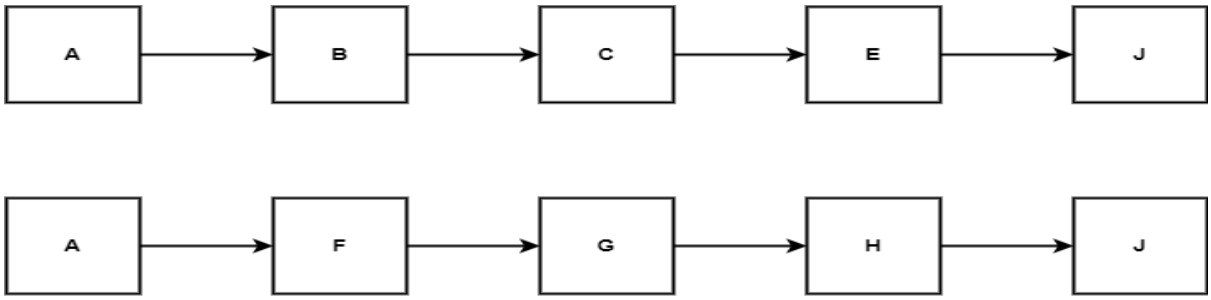


Figure 3.4 (b): Different routing paths

Figures 3.4(a) and 3.4(b) depict multipath routing with node A as a source node and node J as a sink node. Numerous other methods exist for transporting data from source node A to sink node J. The total discontinuous pathways between the source and destination are depicted in Figure 3.4 (a), using node A as a source and node J as a sink node. For multipath routing, computing several routes from a source node to a destination node is required [19]. For the network to be more reliable, some paths must be disconnected. Now, consider a healthcare

system based on a WSN as a collection of connected nodes and represent it as a graph G = (V, E) as shown in Figure 3.5 (a), Figure 3.5 (b), and Figure 3.5 (c). The two nodes, x and y, are separate vertices in a graph G where G = (V, E), so they are not next to each other, according to the well-known Menger's theorem.

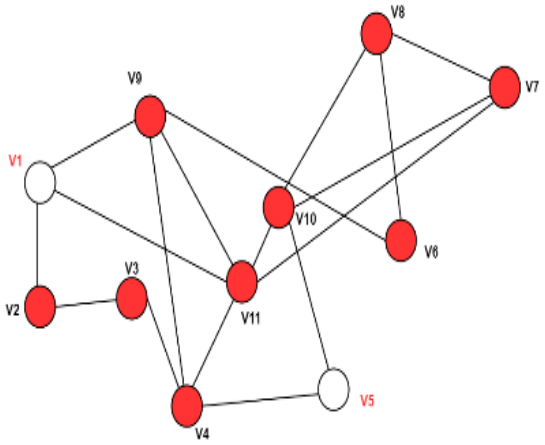


Figure3.5 (a): Example of Menger's Theorem

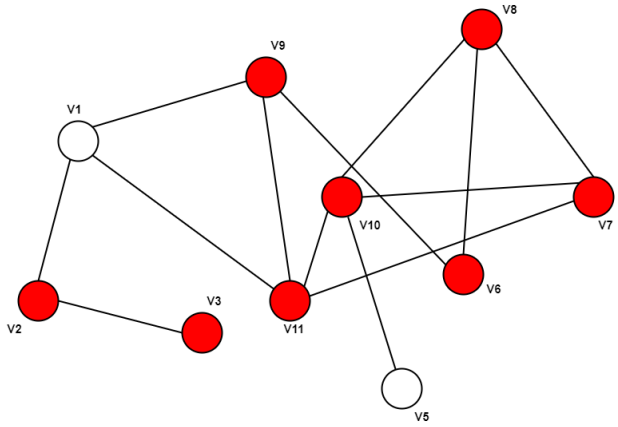


Figure 3.5 (b): Deletion of V4

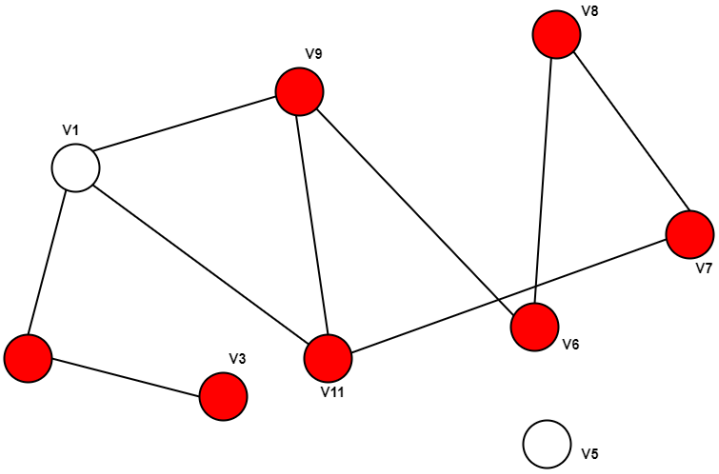


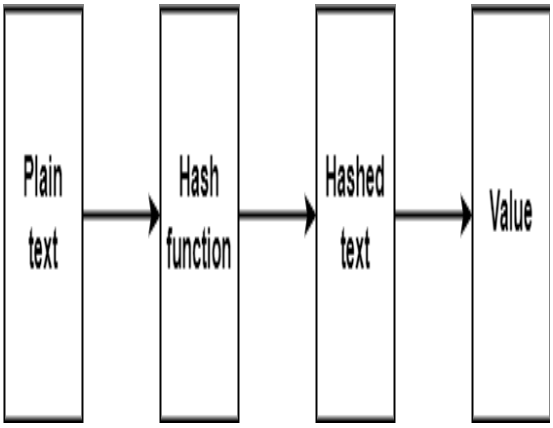
Figure 3.5 (c): Deletion of V10

The maximum number of pairwise vertex disjoint pathways from the x toy is λ (x,y), while the minimum vertex cut size is K(x, y) (K(x,y) =λ (x,y)). A separating set is a group of vertices in a graph G for any two vertices, U and V, leaving the graph with U and V in separate components when removed. A U-V separating set with minimum cordinality is a minimum U-V separating set. Let U and V be the non-adjacent vertices in graph G; this indicates no direct route between U and V. This is by Menger's theorem. The number of disconnected U-V

pathways in graph G equals the size of the minimal U-V separating set. Figure 3.5(a) shows how to locate the internally disjoint pathways from V1 to V5 in the graph using Menger's theorem. However, there are various feasible routes from node V1 to V5. The Menger Theorem states that all we need to consider is the smallest distance between sets V1 and V5. In this example, we can delete V4, and if we want to separate V5 from the rest of the graph, we can delete V10; the size of the minimum V1 - V5 separating set is 2. The remainder of the graph is now connected to V1, and V5 is a component. V4 and V10 have been deleted, as shown in Figures 3.5 (b) and 3.5 (c).

3.2.4 Hashing and Message Digest

The idea of a hash table is the foundation of the searching method known as hashing. A hash table is a data structure that links keys and values. Hashing speeds up searches and expedites insertion and removal, slowing down processing. A block diagram of hashing is shown in Figure 3.6 (a). The first text is the key and serves as the hash function's input. In essence, a hash function creates a function that accepts a hash table as an array. The indices will be 0 to L-1 if the hash table is L in size. The result is derived from the hash function and is either the address of the key or an index of an array of hash tables containing the key. We need a function H(k), where k represents the key and is a one-to-one mapping to integers in the range (0, L-1), where L is the size of the hash table, H is the hash function, and H(k) is the key's hash. Figures 3.6(a) and 3.6(b) depict the hash function's fundamental operation. A message multiple of a fixed length is often padded onto an integer, and the padding contains the value of the original message length, expressed in bits. Length fields offer security measures that make it more difficult for an outside attacker to create another message with the same hash value. The output of a hash function is expressed as h = H(M), where M is the size of the input data. The message digest is the term used to describe the hash function output when applied [14].



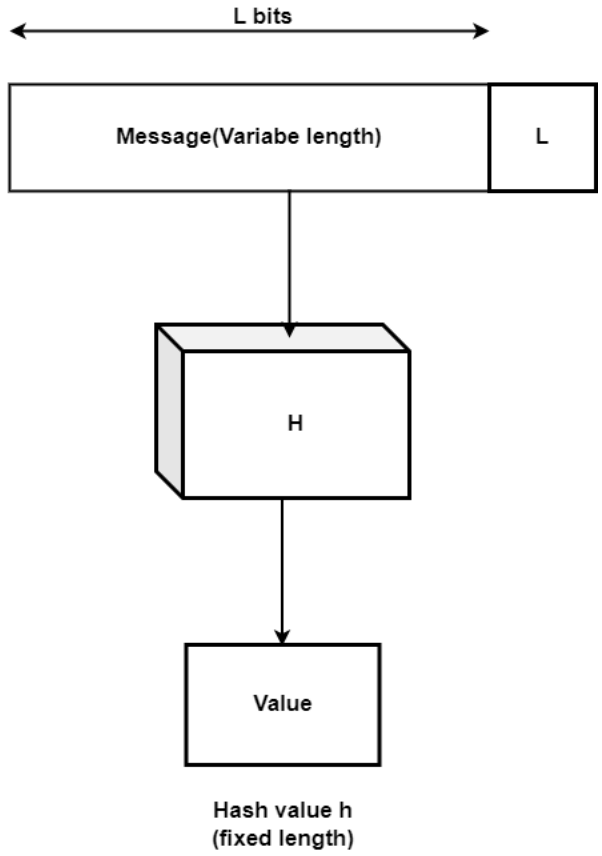


Figure 3.6 (a): Block Diagram of Hashing

Figure 3.6 (b): Hash Function

The network sends the message digest prepared at the sender's end. If the message digest obtained from the sender side matches the one obtained from the receiver side, the message is considered secure. Otherwise, it is deemed insecure.

3.2.5 Properties of Hashing

A mathematical operation known as hashing reduces a message's fluctuating size to a fixed, manageable size. Four distinct features of a hash function H include:

1. A message of any length entered into the hash function (H) generates an output known as a message digest.
2. H(x) = Y can be easily calculated, but it cannot be done in reverse, i.e., h(x) = Y cannot be calculated.
3. As finding y is not equal to x in this circumstance results in limited collision resistance, hashing both x and y h(y = x) is computationally inefficient.

4. In this case, the phrase "high collision resistance" appears, and it is possible to find a pair from (x, y) such that the hashing function of x equals the function of y h (x = y).

3.3 PROPOSED WORK

This section discusses the proposed architecture in sub-sections 3.3.1 and authentication protocols in sub-sections 3.3.2 and 3.3.3, respectively.

3.3.1 Proposed Architecture

Figure 3.7 depicts the architecture of multipath splitting and reconstruction, with hashing used for integrity protection. M represents the initial message size transported from sender to receiver. Hashing is used to protect communication streams from outside attackers. This makes it clear that the multipath splitting and reconstruction are done using the multipath secret sharing scheme, and hashing is specifically used to ensure the messages' integrity, not for routing, as previously stated.

For this purpose, the original message M is split into three parts labelled M1, M2, and M3. These message streams and their hash values are sent to the D1, D3, and D4 servers in the receiver's direction, where the original message is finally combined or rebuilt. First, message M was separated into pieces M1, M2, and M3. These splitting boxes send the message streams and their associated hash values to the servers. The message M1 is then sent to the D1 server along with its hash value. From the sender node to the server D1, H(M1) is sent. From the source node to the destination server D1, the message stream (M1) is transferred. The source nodes send messages M2 and M3 and hash H(M2) and H(M3) to servers D3 and D4, respectively. On the server side, the incoming stream of messages is processed by these servers, D1, D3, and D4, and the hashed input message is converted. The M1, HR(M1), M2, HR(M2), and M3, HR(M3) outputs from servers D1, D3, and D4 are then sent to the receiver for reconstruction. The servers forward data for message reconstruction at the message reconstruction point.

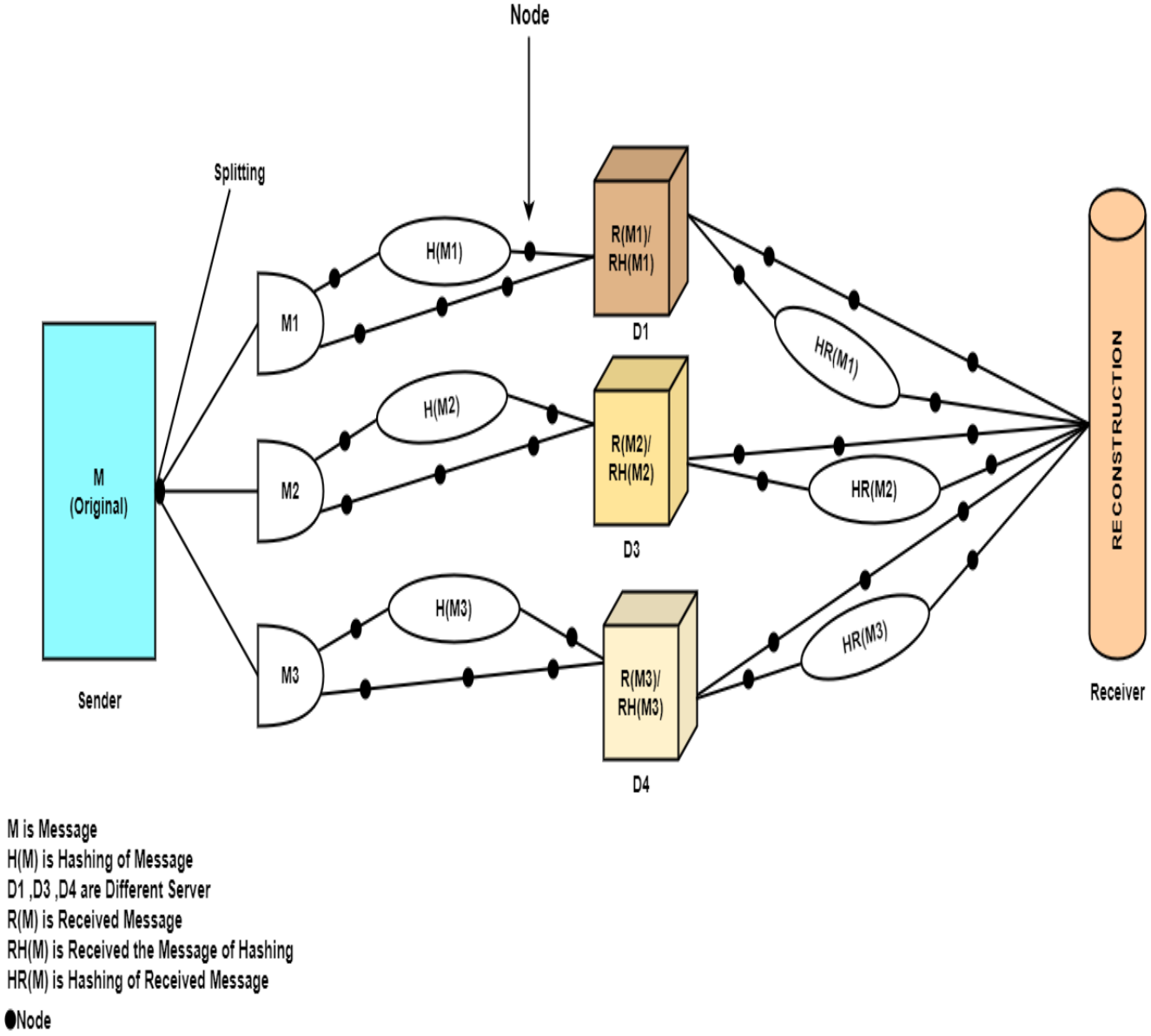


Figure 3.7: Privacy Preservation Framework

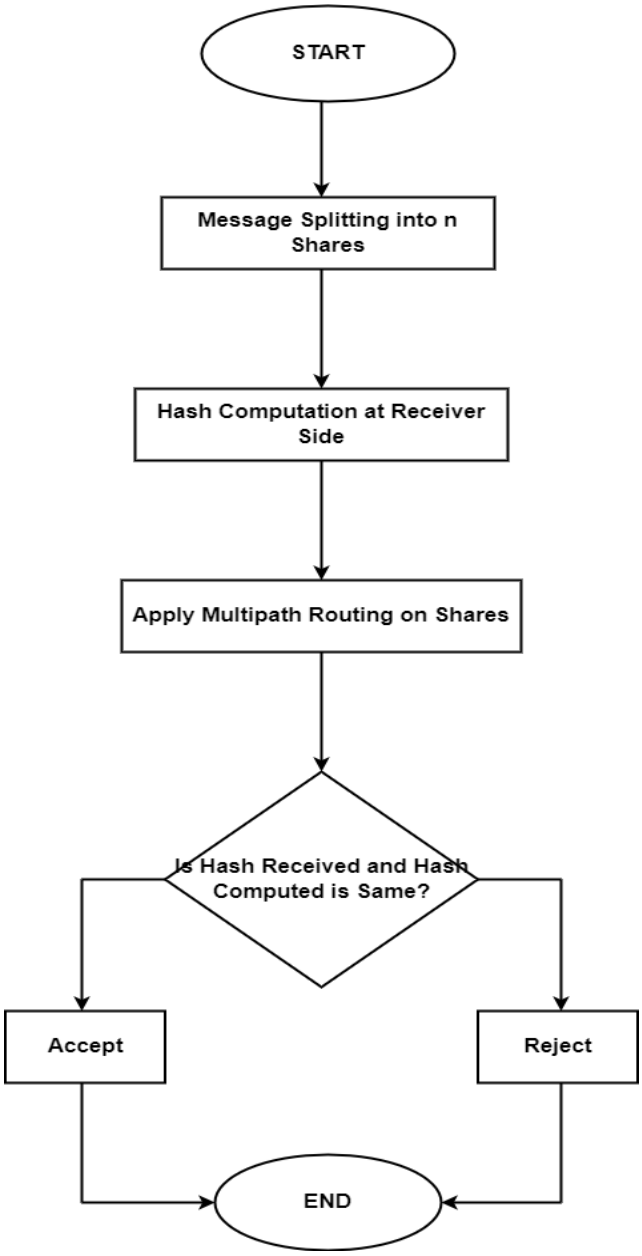


Figure 3.8: Flow Chart of Proposed Scheme

Figure 3.8 shows the proposed model's flowchart. Shares of the initial message separated. Multipath routing was used for these shares' subsequent transfer. The hash function is also constructed using multipath routing at the receiver end. A receiver receives the share and hash function applied to shares during the reconstruction step. Next, determine whether the received shares were the same with or without hashing. If so, the message is accepted; otherwise, it is rejected.

3.3.2 Proposed user authentication protocol

Authentication is important for users in Wireless Body Sensor Networks due to the nature of sensitive information. The authentication approach is presented for the user node in the network.

Proposed Protocol for User Authentication:

Step 1: Registration phase of the user node

The user node first registers at the server and says S_i. The user node is termed the patient node or doctor node. The user node obtains its identification (U_j), and the hash value $H(u_j)$ is generated from user biometric details, a random number (R_u), and a pairwise session key (Ks_iu_j) from $S_i$via offline distribution in a secure manner.

Step 2: Authentication Phase

1) The user node (U_j) starts the authentication process with the Si server. The user node (U_j) sends the *msg* to S_i. The *msg* includes the following details: $U_j|| S_i || t_u || MAC_u$. The details include the identification of the user node (U_j), identification of the server (S_i), timestamp (t_u) of the user node U_j, and the message authentication code of the *msg* (MAC_u). The MAC_u includes the following details: $MAC_u = (K_{siuj} , U_j || S_i || t_u || H(R_u))$.

$U_j \rightarrow S_i$: $U_j|| S_i || t_u || MAC_u$

2) The server checks the time stamp for the message's validity. If $t_u$is valid else exit. If the message is valid then the server computes MAC_{u*}, where $MAC_{u*} = (K_{siuj}, U_j || S_i || t_u || H(R_u))$.

If $MACu_* = MAC_u$then proceed to 3 else exits.

3) The Server node (S_i) send the *msg* to U_j. The *msg* includes the following details: $S_i || U_j|| t_{si} || MAC_{si}$. The details include the identification of the server (S_i), identification of the user node (U_j), timestamp (t_{si}) of the server node S_i, and the message authentication code of the *msg* (MAC_{si}). The MAC_{si} includes the following details: $MAC_{si} = (K_{siuj} ,S_i || U_j || t_{si}|| H(u_j))$.

$S_i \rightarrow U_j$: $S_i || U_j|| t_{si} || MAC_{si}$

4) Upon receiving the message from the server, the user node checks MACsi*, where MAC_{si*}

= (K_{siuj}, S_i || U_j|| t_{si} || $H(u_j)$).

If MAC_{si*} = MAC_{si}then authentication is successful, exit.

3.3.3 Proposed Mutual Authentication Scheme

Authenticating doctors' and patients' devices is still critical in medical IoT networks. We have proposed a mutual authentication session key protocol for smart IoT medical networks to enhance security. Our scheme uses multi-factor authentication to protect doctor and patient data and provide an authentication mechanism.

PROPOSED SCHEME

This work proposed a secure scheme for a smart IoT-based healthcare system to eradicate security issues. This scheme provides a secure key agreement and maintains privacy. The proposed scheme consists of multiple phases are follow as:

1) Registration phase
2) Login and Authentication Phase
3) Password Update Phase

1) Registration Phase (RP):

- ***R_1:*** The Doctor chooses the identity (D_{id}) and password PWD_i. Then it chooses the bio-metric id's Bio_i, with the timestamp value T_i into the smart devices as a bio-metric format shown in Table 3.2.
- ***R_2:*** Doctor's smart device computes these identities as $RP_1 = h(D_{id}||PWD_i||Bio_i)$, $RP_2 = h(R_1||r||T_i)$, and $RP_3=h(T_i||r_i||RP_2)$, then send this secret identity to the medical server via a secure channel.
- ***R_3:*** The medical authentication server receives the request sent by the Doctor's end. The medical authentication server first verifies the timestamp value $(T_2T_1) < T$, if true, it continues, otherwise, it discards the request. After timestamp verification, the medical server calculates $RP_4=h(BIOiPWD_iN\ T_i)$ and $RP_5 = RP_4(A_1PWDiT_i)$ and stores this identity in its key table.
- ***R_4:*** After verifying these identities, the medical authentication server grants a unique Token identity (TOKENid), RP6=h(PWDid||Ti||N||Tokenid), and sends this identity to the Doctor using the secure channel.

TABLE 3.2: SUMMARY OF ACCESS CONTROL PROCEDURE BETWEEN DOCTOR AND MEDICAL AUTHENTICATION SERVER

Doctor end	**Medical Authentication Server**
STEP1: Compute$RP_1=h(D_{id}\|\|PWD_i\|\|Bio_i)$,$RP_2=h(R_1\|\|r\|\|T_i)$,and$RP_3=h(T_i\|\|r_i\|\|RP_2)$,	**STEP2**: TheMedicalServerFirstCalculates $RP_4=h(BIOi\|\|PWD_i\|\|N\|\|T_i)andRP_5=RP_4\oplus(A_1\|\|PWDi\|\|T_i)$ and stores this identity in itskeytable. Where[T_{id}-$T_{current}<\Delta T$]
STEP 3: Medical authentication server granted an unique Token Identity (TOKENid),RP6=h(PWDid||Ti||N||Tokenid) and send this identity to the Doctor's end using the secure channel.	
STEP 4: The Doctor uses its smart device and enter its biometric, and PWD identity$h(D_{id}\|\|r\|\|PWD_i\|\|BIO_{i)}$,computes $L_1=h_{BIOi}\|(T_{id}\|ID_{gateway})$, $L_2=h(L_1\|r\|T\|Key_{secret})$,$L_3=h(Token_{id}\|Ti\|n\|L_2)$,then $L_4=HBIO_i\|D_{id}\|r\|PWDi\|L_2$	
	STEP5: (D_{id},PWD,BIO_i,$TOKEN_{id}$), $L_5=h(T_i\|\|R\|\|DID_{gateway}\|\|D_{id}\|\|Key_{secret})$
	STEP 6: Medical authentication server computes received password, Token identity and other secret identity,$L_6=h(PWD_iToken_{id}T_{id})$andverifies$TOKEN_{ID}=TOKEN_{ID}andPWD_{ID}=PWD_{ID}$andgrantedmessage$L_7=h(PWD_iGrantedT_{id}Key_{secret}MAS_{gatewayid})$. Access granted to user.
STEP 7: User access the service using secure channel.	

- ***R_5:*** After receiving the values from the medical server, the Doctor verifies the timestamp $(T_2 T_1) < T$, if true, it continues otherwise, discard the message. Then, the Doctor computes and stores these entities in his/her device smartcard.

2) ***Login and Authentication Phase (LP):*** This phase provides the login and authentication functionality of our proposed scheme. Table 3.2 demonstrates the steps of this phase.

- ***LP_1:*** The Doctor uses its smart device and enters its bio-metric, and PWD identity $h(D_{id}||r||PWD_i||BIO_i)$, then computes $L_1=h_{BIOi}||(T_{id}||ID_{gateway})$, $L_2=h(L_1||r||T||Key_{secret})$, $L_3=h(Token_{id}||Ti||n||L_2)$, and then $L_4=HBIO_i||D_{id}||r||PWDi||L_2$. The doctors end the $E[L_4||T_{id}$ to the medical server end for the further authentication process.
- ***LP_2:*** On receiving the message from the Doctor gateway end, the medical authentication server extracts and computes secret identities from the message L_4 and timestamp value to protect the secret identities from adversaries. After verification of the timestamp, the authentication server extracts some secrets id's from the message $(D_{id}, PWD, BIO_i, TOKEN_{id})$, $L_5=h(T_i||R||DID_{gateway}||D_{id}||Key_{secret})$.
- ***LP_3:*** After computing and verifying the secret identities, the Smart medical authentication server stores them in their key table. Then, the medical authentication server computes the received password, Token identity and another secret identity, $L_6=h(PWD_i||Token_{id}||T_{id})$ andverifies $TOKEN_{ID}=TOKEN_{ID}$ and $PWD_{ID}=PWD_{ID}$ if true, and it continues; otherwise, it discards the request.
- ***LP_4:*** After authenticating the identities, the medical authentication sends granted message $L_7=h(PWD_i||Granted||T_{id}||Key_{secret}||M$ to the Doctor gateway end by computing.
- ***LP_5:*** After getting the message, the Doctor end verifies the timestamp using $(T_2\text{-}T_1)<\Delta T$. *If true, it continues; otherwise, it* discards the packet. Now, the Doctor is capable of accessing the medical authentication server.

3) ***Password update Phase:*** The password update phase is initiated when the login and authentication phases are completed. The validation scheme can be given the element of updating the secret password PWDid. The password update features allow doctors and users to update their passwords regularly without affecting the network.

1: The Doctor enters the secret identities of this smart device terminal, provides the h(PWDold||Did), and enters the bio-metric identity PUP=h(BIOold||Dgateway||Tid||Did), and

this request is sent to the medical authentication server end.

2: After receiving the identities from the Doctorend, the authentication server extracts and verifies all secret identities h($BIO_{old}||D_{gateway}||T_{id}||D_{id}$), it also verifies *PUP=PIP'* if true, it continues otherwise, discard the request. Finally, the authentication server allows users to change their password and biometric identity.

4. RESULT AND SIMULATION

This section discusses the results of the proposed shamir scheme in sub-section 3.4.1 and authentication protocols in sub-section 3.4.2.

4.1 Shamir Scheme

Using Matlab 2010, we have modelled the Shamir sharing scenario when the secret key is 1121 (role of the polynomial) [6]. For various values of x, the originally transmitted signal or polynomial function f(x) is shown. For the same values of x, the reconstructed signal at the receiver is likewise displayed. The fact that the reconstructed signal crosses over with the initially transmitted signals attests to the success of the reconstruction technique.

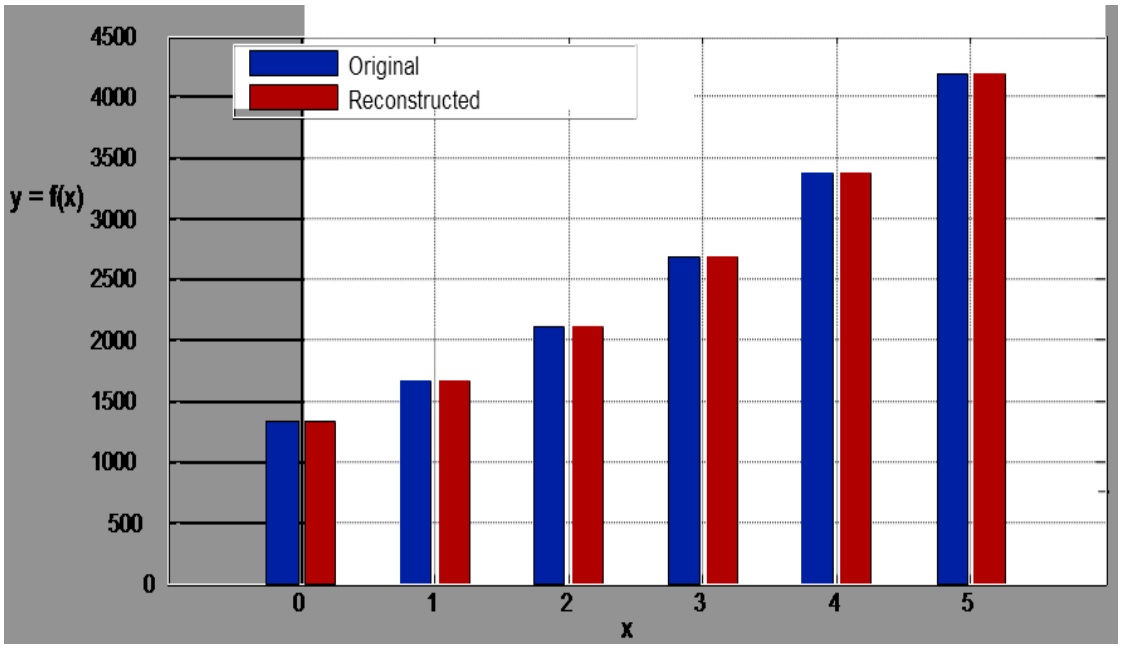


Figure 3.9: Splitting and Reconstruction of Shamir's Secret Sharing Scheme

Realistic channels refer to network conditions, including interference, noise, and varying signal strengths, which can lead to packet losses. "The significance of Figure 3.9 lies in its demonstration of the robustness of the multipath routing scheme. This figure shows the results obtained after executing the routing under realistic channel conditions, where packet losses are common due to interference, noise, and varying signal strengths. Despite these losses, reliable transmission protocols ensure that all shares are successfully delivered to their destinations, allowing for 100% accurate data recovery. Therefore, Figure 3.9 illustrates the effectiveness and reliability of the proposed routing and reconstruction method."

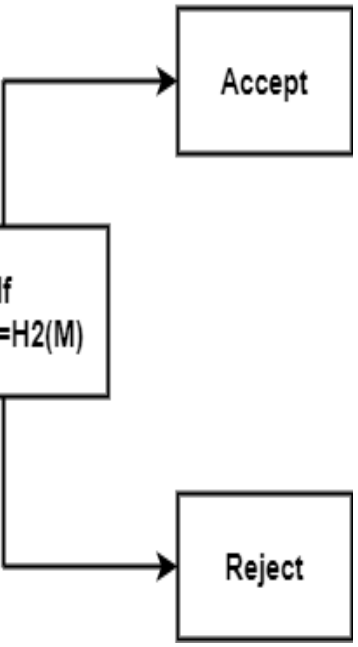
We have introduced privacy preservation for WSN-based healthcare applications utilizing secret sharing and multipath routing. The data collected from the wireless sensor network is split into components. These components are further transferred to servers with the help of multipath routing. A maximum number of disjoint paths is computed from the source node to the destination node to achieve multipath routing. Further, the hash functions are calculated for each component and sent to the server. These n components are retrieved from the n server to reconstruct the medical data. Results and analysis validate our approach. Figure 3.9 illustrates the splitting and reconstruction of the message (secret) using Shamir's secret-sharing scheme. This revised explanation clarifies the importance of Figure 3.9 and connects it to the system's overall reliability and robustness.

Table 3.3: Table of Splitting and Reconstructed

		Shamir's Secret Sharing Scheme		
Sr. No.	**D**	**Splitting X**	**Reconstruction f(x)**	
1	D0	1	1331	
2	D1	2	1661	D1
3	D2	3	2111	
4	D3	4	2681	D3
5	D4	5	3371	D4
6	D5	6	4181	

Table 3.4: Table of Splitting and Reconstructed with Hashing

Signal	Transmitted	Received	Receiver Operation	
M	H(M)	H(M)	H1(M)	If H1(M)=H2(M)
M	M	M	H2(M)	



4.2 Performance Analysis of the Authentication Scheme

This section presents a performance analysis of the proposed authentication scheme. The message overhead is as follows: the MAC size is 4 bytes, the timestamp is 8 bytes, a random number is 8 bytes, identification is 1 byte, and the key size is 16 bytes [1].

Table 3.5: Message overhead

User node to the server	14 bytes
Server to user node	14 bytes

Table 3.5 presents the message overhead of the user node and server node. The overhead includes 2 bytes of identification, 8 bytes of timestamp, and 4 bytes of MAC.

Table 3.6: Cryptographic operations

Number of encryption /decryptions by user node	2
Number of encryption /decryptions by server	2
Number of MAC generation by user node	2
Number of MAC generation by server node	2

Table 3.6 presents the number of cryptographic operations for the transmission and reception of messages for the authentication approach.

5. CONCLUSION

Although medical sensor networks used in healthcare applications have many advantages, they also present security and privacy problems for patients. The development of technology has made it possible to communicate electronic medical data for telemedicine through a network. Data may be intentionally misleading, though. Thus, it is essential to safeguard the information against abuse. Hashing is a fundamental technique for protecting the original message. The chapter describes a novel approach to improving health data securely transmitted across multiple channels. If a message is received, hashing can quickly determine whether it is authenticated. Hashing provides a high level of security for the privacy criteria. The original message has been divided into three parts using the proposed technique, and the split messages are then sent to various servers using multipath routing. The implementation of the multipath routing scheme improves transmission security. The findings of this technique outperform plain text communication in terms of privacy preservation for a WSN-based healthcare system.

CHAPTER 4

SOURCE LOCATION PRIVACY PRESERVATION IN IOT-ENABLED EVENT-DRIVEN WSNS

WSNs that utilize the Internet of Things (IoT) have become increasingly popular in healthcare and monitoring applications due to their small size, scalability, and cost-effectiveness. Privacy preservation in healthcare, a process crucial for protecting patient location, involves obfuscation. The patient's original location is perturbed by a shift to a new location, aided by random displacement along the x and y coordinates. A timestamp is added to prevent future value repetition. The weak adversary's role is limited to eavesdropping on the message (traffic) from the source to the healthcare provider, with the healthcare databases and the patient's source location remaining unknown to this weak adversary [182]. However, healthcare databases can be accessed by a strong adversary on the other hand.

Location privacy protection (LPP) mechanisms should dynamically adapt the privacy level based on the sensitivity of the visited locations. Thus, differential privacy-preserving schemes can help achieve the privacy of a patient visiting a hospital by adding random noise to the user's location [183].

In a smart health system, patients' location information is periodically sent to hospitals, which helps hospitals provide improved healthcare services. The location information and the time stamp alone can reveal a patient's private information. In dummy location generation-based privacy protection mechanisms, a fake or dummy location is created randomly corresponding to each actual location, and only the dummy location information is sent to the destination [184].

However, maintaining privacy for monitoring applications. To conserve energy, an event-driven system must minimize traffic when no events occur but handle a large volume of data when an event is detected [173]. To safeguard source location privacy in WSN monitoring applications, this chapter suggests a framework for privacy protection. The chapter provides a comprehensive security study of grid-based deployment in wireless sensor networks. It describes three schemes aimed at protecting the location privacy of the source of events during event detection: Source Location Privacy for Event Detection (SLP_ED), Chessboard

Alteration Pattern for Source Location Privacy (SLP_ED_CBA), and Grid-based Source Location Privacy (GB_SLP) [76]. These schemes are designed to ensure that the location of the source generating the event remains confidential, thereby enhancing the network's security and privacy. It's important to note that these schemes are not about event detection but protecting the source's location and privacy during detection, which is a crucial distinction.

Explanation

- **Event Detection Schemes**: Typically focus on identifying and responding to specific events within the network.
- **Protecting Location Privacy**: This policy ensures that the physical location of the source generating the event remains undisclosed to prevent attacks or unauthorized tracking.

It becomes evident that the chapter's contribution is pivotal. It focuses on enhancing source location privacy during event detection in grid-based deployments rather than using event detection methods.

The next section describes the proposed source location privacy preservation methodology in event-driven wireless sensor networks. A source node in environmental monitoring detects two types of events. These events can be critical or nominal. Upon detection of the event, gathered data is transmitted to the sink node. The algorithm selects a low-energy consumption path for nominal events [27]. For critical events, the algorithm chooses high-energy consumption paths. The main objective of an adversary is to reach the source node. The initial and final positions for an adversary are the sink and the source node, respectively. The adversary's initial position is at the sink node. From the sink node, the adversary backtracks to the source node.

4.1 THE PROPOSED MODEL IN AN EVENT-DRIVEN WSN

This section presents the network model for event-driven WSN. The following are the network model's underlying assumptions:

a) The Base Station (BS) is positioned in a rectangular area of interest.
b) All network nodes are aware of the position of the single destination node (BS).
c) The nodes are densely distributed and remain stationary after deployment.
d) The placement of the source is random in the region of interest.
e) The network’s nodes meet the coverage and connectivity threshold criteria.
f) The adversary's monitoring radius and the sensor node's communication range are

equal for simplicity.

g) A Passive attacker can only monitor the network's traffic and not destroy the sensor network [171].

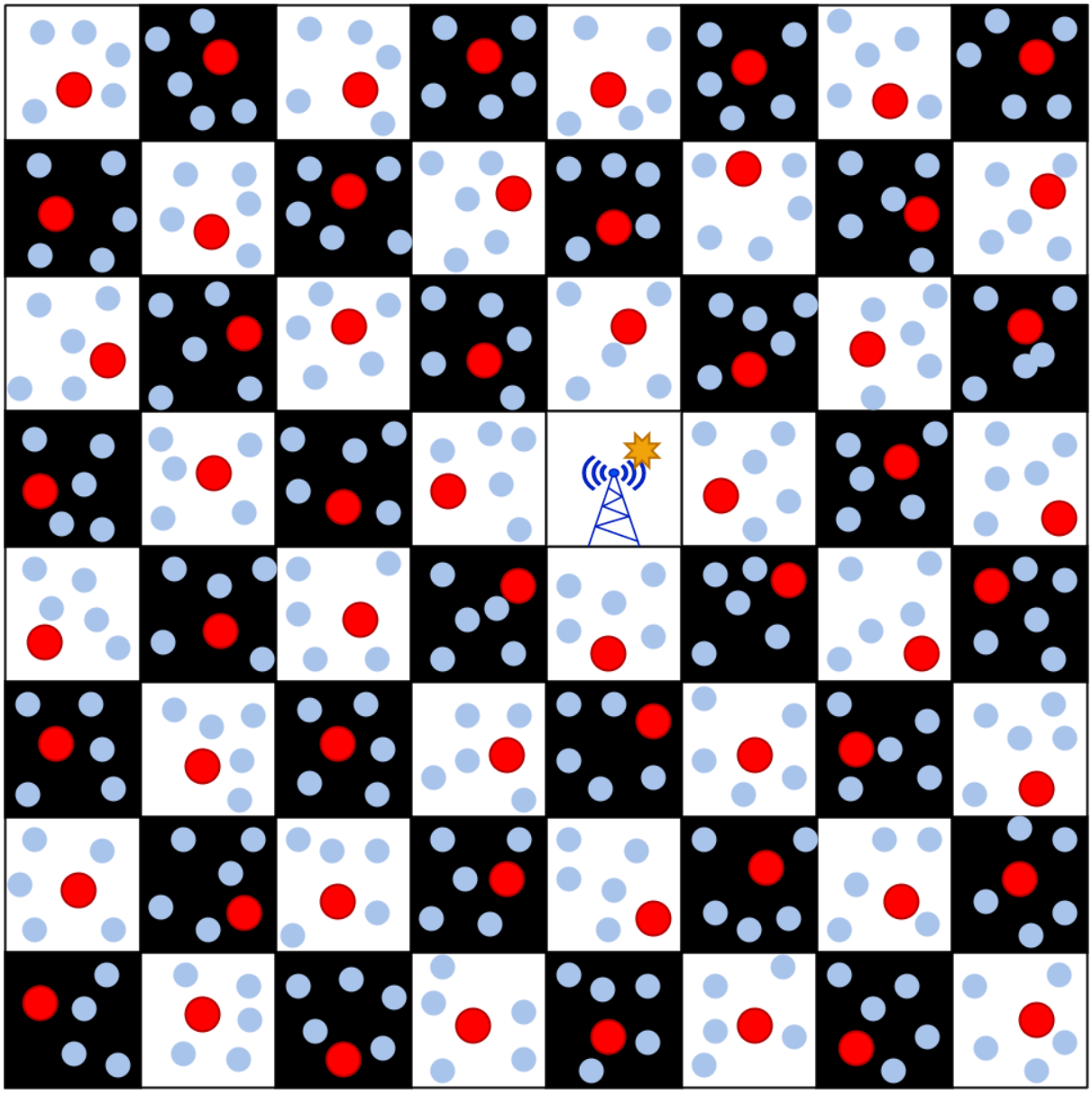


Figure 4.1: Chessboard Deployment plane for source location privacy

Figure 4.1 presents the chessboard deployment plane for source location privacy. Important

performance indicators for maintaining source location privacy include energy consumption, packet latency, and safety period. The following metrics are used in the work.

(i) Let the attacker reach from the designation node to the source node in time *t*. During this time *t*, the number of data packets the source node transfers is *p* packets to the BS. The value of *p* determines the safety level.

(ii) Delay: The average time a packet takes from its source to its sink node. The packet's typical hop count from source to sink is utilized to determine the delay [176].

(iii) Energy Consumption: Energy consumption is defined as the average number of packets sent from the source node to the sink node for the safety period. The packet transmission and reception energy is used in the routing process [179]. This means that each node along the routing channel only sends and receives the packet once (see equations (1) through (3)). On the other hand, the source and destination nodes only have energy for transmission (E_{trans}) and reception (E_{rec}), respectively. Transmission energy is proportional to the size of packets in (k) bits. Transmission energy is proportional to the transmitting and receiving node distance (d). The reception energy is proportional to the packet size (k). The packet size and distance are measured in bits and metres, respectively. The constants E_{elec} and ε_{amp} are taken as 50 nJ/bit and 100 pJ/bit/$m^{2,}$ respectively. Equations 4.1 to 4.2 are the energy consumed per hop in transmission and reception. Equation 4.3 is the total energy consumed per hop of transmission and reception by the sender and receiver node.

$$E_{trans} = (E_{elec} * k) + (\varepsilon_{amp} * k * d^2) \quad (4.1)$$

$$E_{rec} = (E_{elec} * k) \quad (4.2)$$

$$E_{total} = E_{trans} + E_{rec} \quad (4.3)$$

In real-life settings, the occurrences are either nominal or significant. A source node is capable of detecting both nominal and critical events. The suggested algorithm chooses a high-energy path for critical events, whereas, for nominal events, it chooses a low-energy path. Similarly, the proposed algorithm selects a high safety level for a critical event and a low safety level for the nominal path.

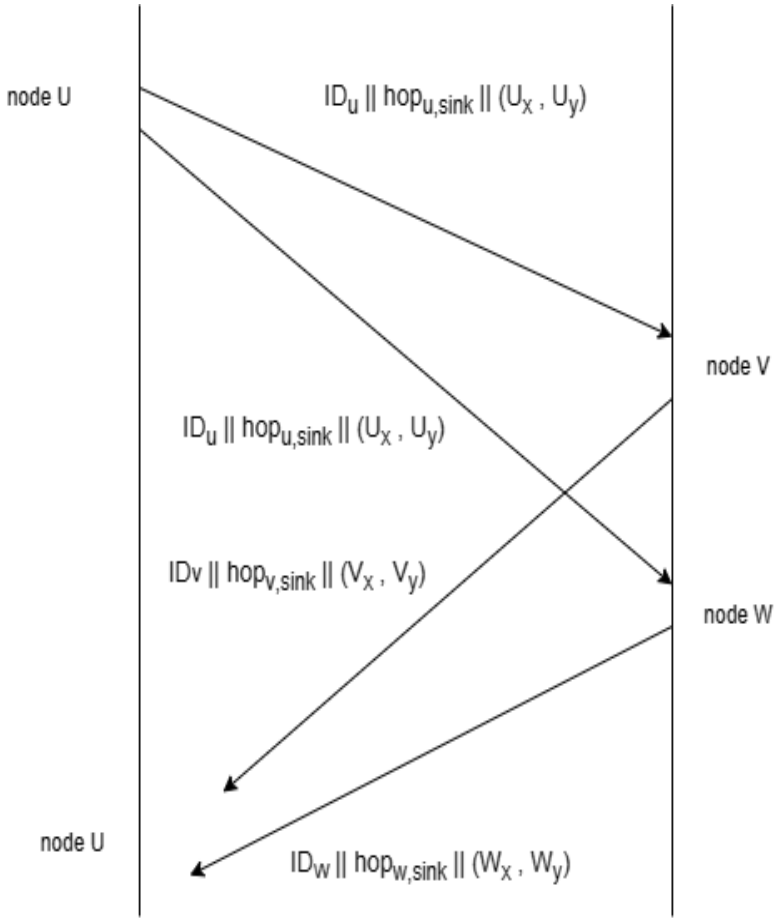


Figure 4.2: Initialization phase

Figure 4.2 shows the initialization stage of the suggested approach. Each node broadcasts its identification, location, and hop count to the sink during this phase to its neighbors. Each node sorts its neighbors into one of three groups: farther set, equivalent set, or closer set after the startup phase. A node's closer set consists of neighbors closer to the base station than the node itself. The farther set includes all of a node's neighbors farther away from the base station than the node. The third group, known as counterparts, includes the node's neighbors that are at the same depth as itself about the base station. Each node assigns its neighbour to one of three categories after completing the initialization step successfully: a closed set, an equivalent set, or a further set. The source Location Privacy for Event Detection (SLP_ED) algorithm is used for nominal and critical events.

Proposed Algorithm (*SLP_ED*)

StepI:

//The random walk phase.

//User-defined *h* hops

if (*message_priority*is*critical_event*)

{

for hop_count = 1: *h* hops

{

if ($current_{node}$ is not the $boundary_{node}$)

{

$temp_{node}$← (get-farthest-neighbor ($current_{node}$));

send($temp_{node}$);

$current_{node}$←$temp_{node}$;

}

}

}

If (*message_priority* is*normal_event*)

goto:StepII

StepII:*dynamic-shortest-path* routing

Select *dynamic-shortest-path* routing from $current_{node}$ to sink

The SLP_ED algorithm takes the source node, destination node, priority of an event, and the number of hops as input. The algorithm chooses the neighbour in the farthest set from the current node in each iteration. The process continues for h hops. The algorithm moves into step II if *hop_count* equals h or a boundary node is encountered.

The energy model is the standard model used in recent papers (Jan et al., 2019) and (Al-Mistarihi et al., 2020). The two recent works are compared with the proposed approach. The energy comparison is only one parameter; however, the significant parameter is the safety level in our work for source location privacy preservation schemes.

The privacy level or safety period is the number of packets sent successfully by the event-detecting node before an attacker finds the node. Our work has two types of events: critical and normal. The safety level for a critical event is higher than that for a normal event. However,

normal events consume less energy.

Two simulation environments have been created similar to custom-built simulators, and these have been available in recent research papers. Simulation environment I is a 250 m × 250 m rectangular grid, and simulation environment II is a 400 m × 400 m rectangular grid. The network is scalable with various deployment strategies, with different deployment areas in rectangular or circular deployment strategies.

Two suggested nominal and critical event identification methods for source privacy preservation are *SLP_ED* and *SLP_ED_CBA*. If the message priority is normal_event, go to Step II. In Step II, select the dynamic_shortest_path routing from the current node to the sink. In medical applications, the potential risks of an attacker finding the source of events are significant. This could lead to privacy breaches, targeted attacks, or unauthorized surveillance. Therefore, it's crucial to understand what constitutes critical and nominal events, as it highlights the importance of protecting the source location.

Motives for Attacking Source Location:

1. **Patient Privacy**: Attackers might seek to discover the location of patients who generate specific health data, such as those undergoing treatment for sensitive conditions, to exploit their personal.

The thesis discusses the source location privacy protection for IoT-enabled WSNs for monitoring applications. Three algorithms are proposed in the thesis. The chapter gives a security study of the grid-based deployment and describes three event detection source location privacy protection schemes for unattended asset/event-driven monitoring in hostile environments. The first algorithm, the SLP_ED algorithm, takes the source node, destination node, priority of an event, and the number of user-defined h hops as input. For regular events, privacy preservation of the source node is less, and the packet is forwarded from the source node to the sink node using *dynamic-shortest-path* routing. However, for critical events (such as habitat monitoring applications for endangered species), the privacy preservation of the source node is required, and the proposed algorithm works in two steps. The first step proposes a backward random walk-up to user-defined h-hops. In the second step, min-hop routing is used to forward the packet from the phantom node to the sink node.

The first algorithm is source location privacy for event-driven (SLP_ED). It works on the idea of phantom routing to relay the packet to a distant node randomly to increase the network's safety level. The algorithm works in two steps: in step one, the packet is forwarded to the nest

hops (up to h hops) using backward random routing and then to the sink node with the help of a minimum-hop routing scheme in the second step.

The second algorithm is an energy-efficient version of the source location privacy for event-driven (SLP_ED) and works on a chessboard alteration pattern (SLP_CBA). The sensor periodically changes the state from active to sleep in a chessboard pattern. Only half of the nodes are active at any given time, and the rest are in sleep mode. The active node changes its state and is in sleep mode. Similarly, the asleep node changes its state and is in active mode, similar to [115].

The third algorithm selects a subset of the sensor nodes (using a well-known sampling algorithm) outside the coverage area. The nodes in the triangle coverage area are not selected to work as phantom nodes. However, the nodes outside the coverage are the candidate set for the phantom node. To compute the triangle coverage, the user-defined input specifies the sink node's location and the coverage angle.

Privacy preservation for IoT-enabled medical applications is discussed in Chapter 5. IoT-enabled medical devices or smartphones collect data from the sensors and send the raw data to the application server, which stores them. A task of the utmost importance is to develop methods and tools for sharing data in a more hostile environment so that the shared data remains practically useful while individual privacy is preserved. This undertaking is called privacy-preserving data publishing (PPDP). Chapter 5 discusses k-anonymity, randomness and other methods to preserve data privacy in medical applications.
Chapter 5 also presents the framework for the knowledge graph for sharing the data across medical organizations.
The next paragraph presents the Source Location Privacy for Event Detection with a Chess alteration pattern (*SLP_ED_CBA).*
SLP_ED_CBA algorithm takes the source node, destination node, priority of an event, number of hops, and active nodes as input. The algorithm chooses the neighbour farthest away from the current node in each iteration. The process continues for h hops. The algorithm moves into step II if *hop_count* equals h or a boundary node is encountered. For both step 1 and step 2, the active nodes alter in a chessboard manner. Only half of the nodes are active at any given time, and the rest are in sleep mode. The active node changes its state and is in sleep mode, and the asleep node changes its state and is in active mode.

Proposed Algorithm (SLP_ED_CBA)

Step 1:

//The random walk phase.

//User-defined *h* hops

If (*message_priority* is *critical_event*)

{

for hop_count = 1: *h* hops

{

if (current node is not the boundary node)

{

temp node← active_node ((get-farthest-neighbor ($current_{node}$)));

send(temp node);

$current_{node}$←$temp_{node}$;

}

}

}

If (*message_priority* is*normal_event*)

goto: Step II

Step II: dynamic-shortest-path routing

Select dynamic-shortest-path routing from current_node to sink

Step 1 in the source location privacy leads to an increase in safety level. Figure 4.3 shows the chessboard alteration (CBA), consisting of active and sleeping sensors and alternating stages. As shown in Figure 4.3, the deployment plane is divided into a chessboard pattern with black and white nodes or sleep and active nodes. The adversary traces back the route from the sink to the BS. An opponent's initial and final locations are the base station and source node.

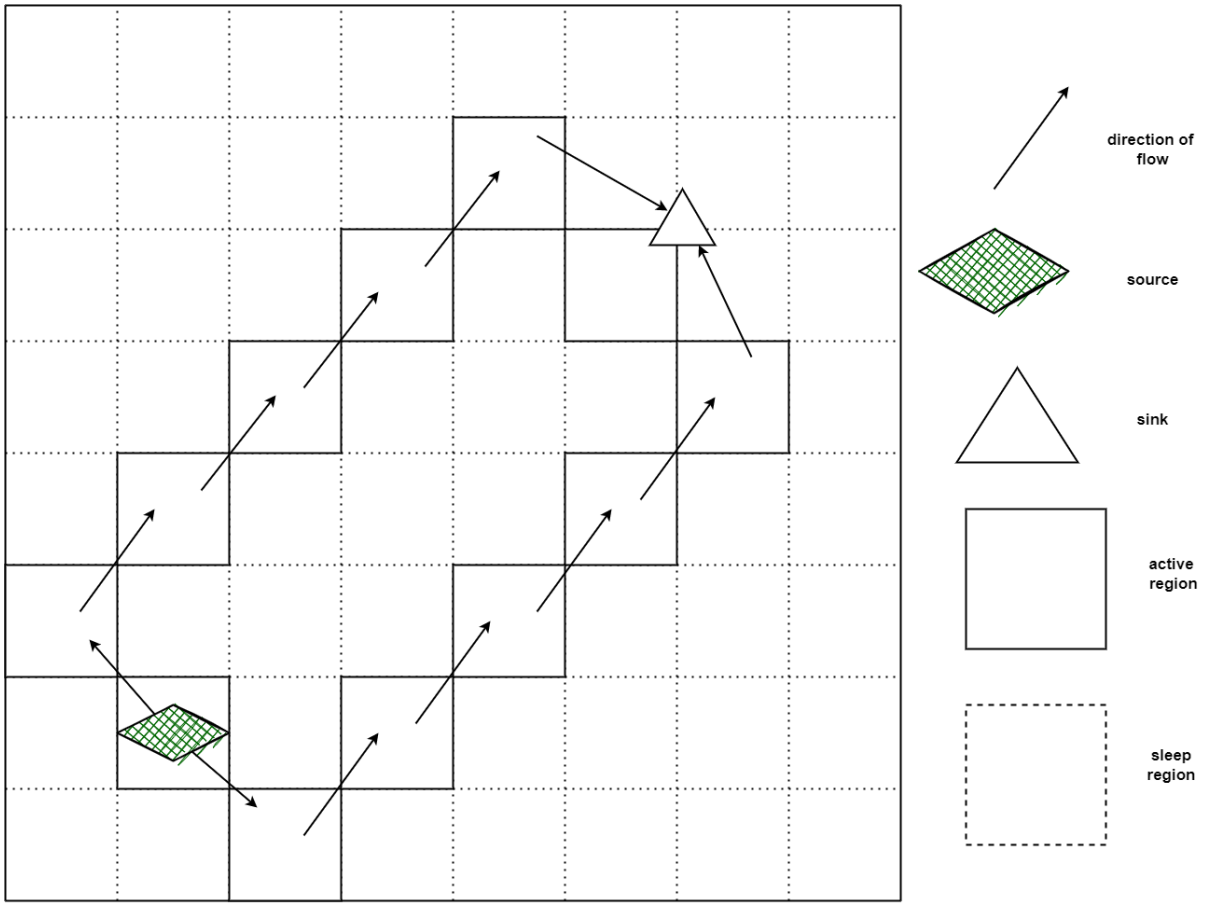


Figure 4.3: SLP_ED_CBA with chessboard alteration (CBA) pattern

In the first step of the algorithm, a sensor node sends an information packet to a remote node referred to as the virtual sensor node. Each node then forwards this packet to a designated neighbour on its farther list, selected by a specific process that repeats for a set number of hops, denoted as h1. The second phase involves the packet moving towards the base station with fewer nodes involved. The next paragraph presents the Adversary strategy for backtracking from the sink node to the source node. Two strategies, strategy I and strategy II, are discussed in the following paragraph.

An adversary's starting location is a sink node or base station. The attacker observes packet transfer to the sink node while waiting near it. The attacker locates the immediate sender node in the event of packet transmission. The sender node performs the one-hop transfer between the sink and attacker nodes. The immediate circumstance is where an adversary is if they use either Adversary strategy I or II, as shown in Figure 4.4. The attacker observes the incoming packet transmission. The attacker moves hop-by-hop, starting from the sink node and reaching the event location. When the attacker reaches the source node, the process ends.

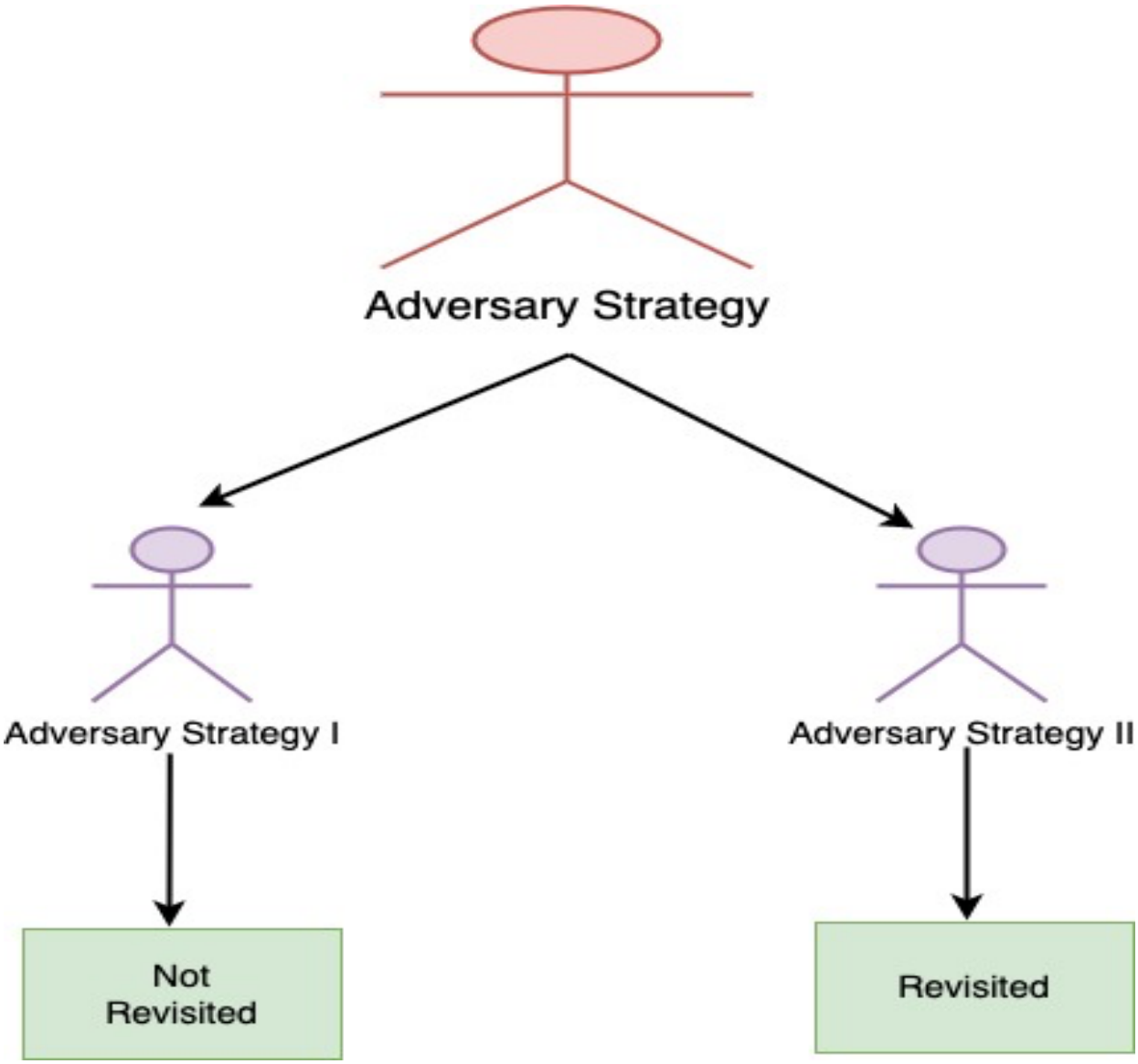


Figure 4.4: Types of Adversary Strategy

(i) Adversary strategy—I: If an adversary has not visited the node, the adversary goes to the immediate place. The nodes are not visited again. This is done to avoid switching between nodes that have previously been visited.

(ii) Adversary strategy - II: The opponent may return to the nodes and move to a more proximate location.

The grid-based Source Location Privacy (GB_SLP) scheme is presented below.

Proposed Algorithm:

Step 1:

Wait for an event ()

if (event ())

send*alarm_packet*to *sink_node* in secure manner

send *source_id* to *sink_node* in secure manner

upon receipt of event information by *sink_node*

sink_node computes the coverage_area

select a set S′ (phantom nodes) outside coverage_area // use Equation 4.4 to 4.9

select a set S″ (phantom nodes) using reservoir sampling

source_node receives the set S″

Step 2:

select *phantom_node* from S″

delete the selected *phantom_node* from S″

send the packets from *source_node* to *phantom_node*

Step 3:

Route the packet using dynamic shortest path routing from *phantom_node* to *sink_node*

if(S″ != φ)

goto: **Step 2**

else

goto: **Step 1**

One simple sampling method with an O(n) time complexity is reservoir sampling. Reservoir sampling can now be described as follows: The set S′ initially contains *n* elements. The reservoir sampling selects (sample) a subset S″. The sampling is random in nature, as presented below.

```
// reservoir_sampling(phantom_nodes)
for count=1 uptocount =c
reservoir[count]: =S′[count]

for count= k+1 upto count = n
temp: =random_integer(1, count)
if(temp<= c)
reservoir [temp]:=S′ [count]
```

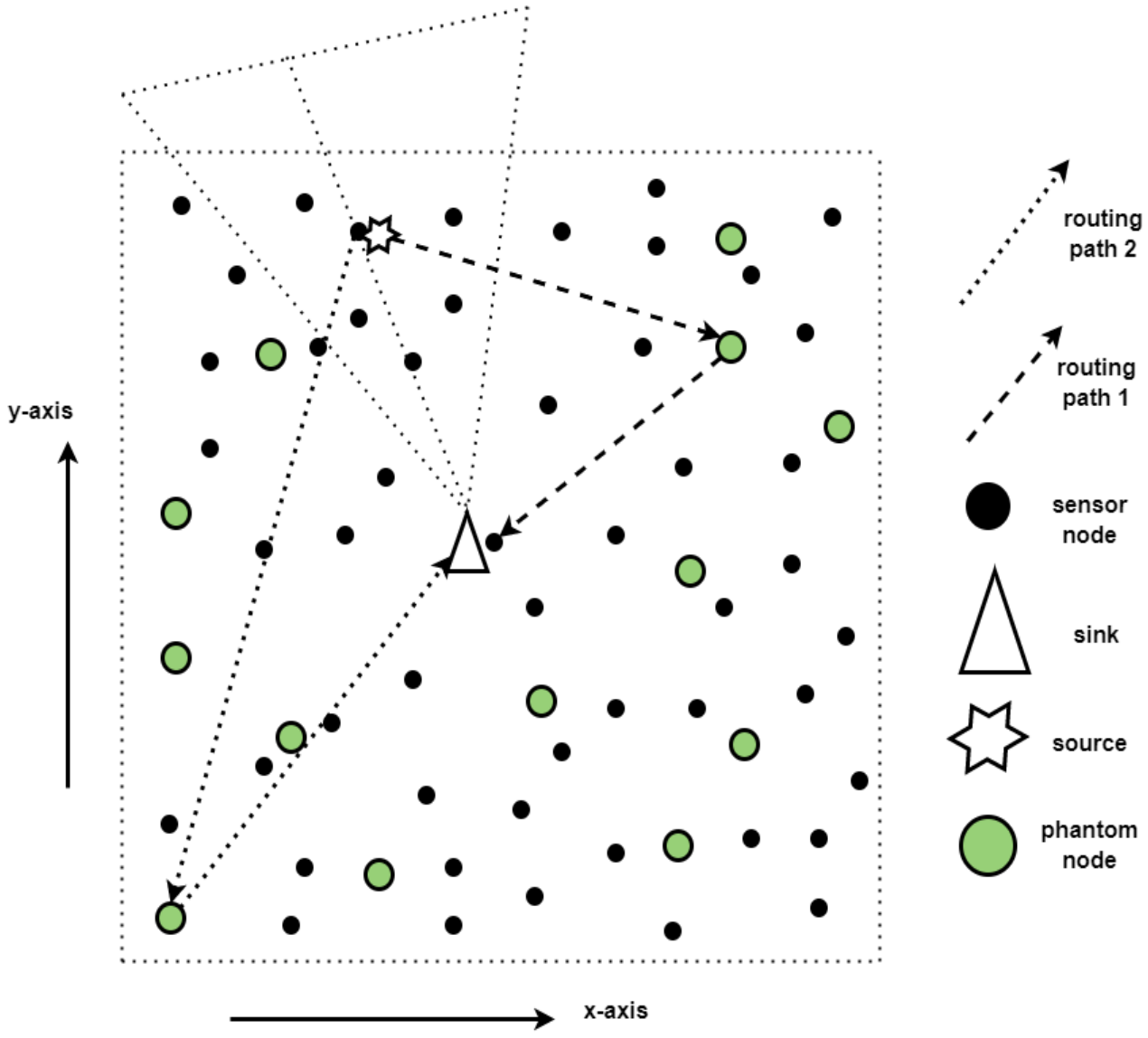


Figure 4.5: Grid_SLP (Selection of phantom nodes)

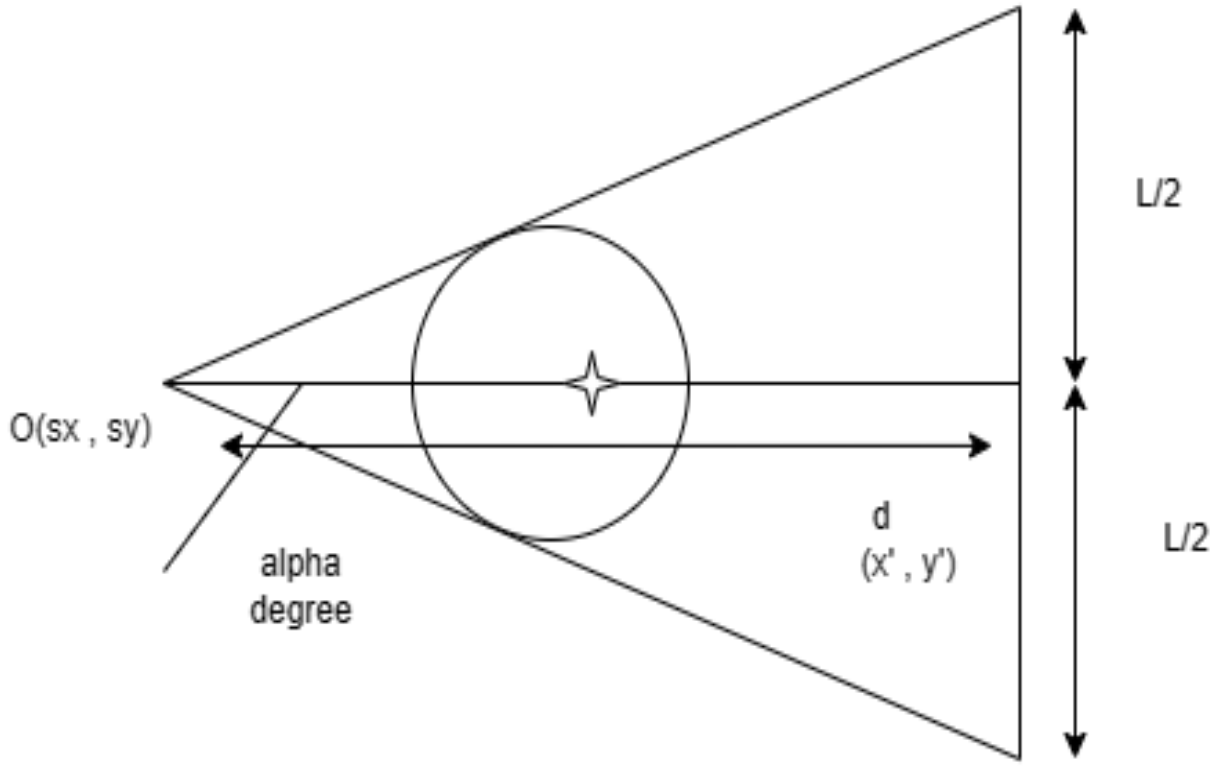


Figure 4.6:*GB_SLP* deployment area with triangle coverage

Using a coverage algorithm and sampling, a grid-based source location scheme chooses the phantom nodes. The nodes in the triangle coverage area are not selected to work as phantom nodes. However, the nodes outside the coverage are the candidate set for the phantom node (refer to Figure 4.5). To compute the triangle coverage, the user-defined input specifies the sink node's location and the coverage angle. The distance, denoted by d, between the sink and source nodes, is shown in Figure 4.6. The triangle covers the point if both conditions (4.8 and 4.9) are satisfied and the sink is placed at (0, 0). These candidate nodes are sampled using a well-known reservoir sampling algorithm. These sampled nodes are the candidates for the phantom nodes in the grid-based source location privacy scheme.

$$x_1 = x - s_x \quad (4.4)$$

$$y_1 = y - s_y \quad (4.5)$$

$$x' = \cos\alpha \times x_1 + \sin\alpha \times y_1 \quad (4.6)$$

$$y' = -\sin\alpha \times x_1 - \cos\alpha \times y_1 \quad (4.7)$$

$$x' \leq d \qquad (4.8)$$

$$(L/2d) \times x' \leq y' \leq (L/2d) \times x' \qquad (4.9)$$

4.2 PERFORMANCE ANALYSIS FOR SECURITY

Let shortest_path take p_{hops} from the *source_node* to the *sink_node*. Assume τ_{hop} is the time required by the attacker to backtrack one hop. Thus, the total time an attacker requires can easily be calculated as the product of $sp_{hops}$and τ (time unit).

The total number of hops required from the *source_node* to the *phantom_node* in our suggested method *GB_SLP* is $H_{src,\,ph}$ hops. Further, assume that the total number of hops required from phantom_nodeto the *sink_node* is given by $H_{ph,snk}$. Equation 4.10 now gives the total time to trace back.

$$(H_{src,ph} + H_{ph,snk}) \times \tau_{hop} \qquad (4.10)$$

$$sp_{hops} \times \tau_{hop} \leq (H_{src,ph} + H_{ph,snk}) \times \tau_{hop} \qquad (4.11)$$

Equation 4.11 compares the time required by the shortest path and routing via phantom_node. As shown in Figure 4.6, the phantom_node is positioned beyond the triangle's coverage. Compared to the shortest_path method, the GB_SLP technique offers a higher safety level. The safety level can be further improved by replacing the phantom nodes after every Tps duration, as shown in Equation 4.12.

$$T_{ps} \leq MIN(H_{src,ph} \times \tau_{hop} , H_{ph,snk} \times \tau_{hop}) \qquad (4.12)$$

Additionally, let's assume that there are N_{ph} *phantom_nodes* available. In that case, the overall level of security can be improved by (Refer to Equation 4.13):

$$sp_{hops} \times \tau_{hop} \leq N_{ph} \times MIN (H_{src,ph} \times \tau_{hop} , H_{ph,snk} \times \tau_{hop}) \qquad (4.13)$$

Equation 4.14 presents the minimum value from the *source_node* to the *phantom_node*.

$$SP_{hops} \times \tau_{hop} \leq N_{ph} \times (H_{src,ph} \times \tau_{hop}) \qquad (4.14)$$

Equation 4.15 presents the minimum value from *phantom_node* to the *sink_node*.

$$SP_{hops} \times \tau_{hop} \leq N_{ph} \times (H_{ph,snk} \times \tau_{hop}) \quad (4.15)$$

Assume T_{active} is the time the source_node is sending the data. If the following condition is met, the attacker cannot locate the source node (Refer to Equation 4.16).

$$T_{active} \leq N_{ph} \times MIN(H_{src,ph} \times \tau_{hop} , H_{ph,snk} \times \tau_{hop}) \quad (4.16)$$

4.3 LIMITATIONS

(i) First Algorithm: SLP_ED

SLP_ED (Source Location Privacy Event Driven)

- **Objective**: Provide location privacy using event-driven mechanisms to obscure the source location.
- **Limitations**:
    - **Event Overhead**: Event-driven mechanisms can introduce significant overhead, especially in environments with frequent events.
    - **Predictability**: If the event-driven patterns become predictable, attackers may still be able to infer source locations.

(ii) Second Algorithm: SLP_ED_CBA

SLP_ED_CBA (Source Location Privacy Event Driven with Chess Board Alteration)

- **Objective**: Improve upon SLP_ED by introducing the Chess Board Alteration (CBA) method to enhance privacy and reduce predictability.
- **Improvements**:
    - **Chess Board Alteration**: This technique involves altering the communication pattern to resemble a chessboard, which helps obscure the source location more effectively.

- **Unpredictability**: The CBA method introduces higher unpredictability, making it more difficult for attackers to deduce the source location based on event-driven patterns.

- **Limitations**:
  - **Complexity**: The introduction of CBA increases the algorithm's complexity and may require more sophisticated management.

(iii) Third Algorithm: Grid-Based

Grid-Based Approach

- **Objective**: Enhance location privacy by organizing the network into a grid structure, allowing for more efficient routing and energy management while maintaining the benefits of event-driven privacy mechanisms.
- **Improvements**:
  - **Structured Routing**: A grid-based structure allows for more systematic and efficient routing, reducing energy consumption and improving scalability.
  - **Enhanced Privacy**: The grid structure can integrate advanced privacy mechanisms more effectively, offering better protection against location-based attacks.
- **Limitations**:
- **Initial Configuration**: Setting up a grid-based structure may require substantial initial configuration, which can be challenging in dynamic or large-scale environments.
  - **Maintenance**: Maintaining the grid structure can be complex and resource-intensive, especially with changing network conditions (e.g., node mobility varying energy levels).

4.4 SIMULATION AND RESULTS

The section compares the proposed approach to current techniques by presenting two simulation environments and parameter settings.

The simulator is a custom-built simulator. The simulation results are obtained for our proposed approach and compared with existing approaches (Al-Mistarihi et al., 2020) and (Jan et al., 2022). Two simulation environments have been created similar to the built simulators available in recent research papers [114][115]. Simulation environment I is a 250 m × 250 m rectangular

grid. Simulation environment II is a 400 m × 400 m rectangular grid. The network in each scenario is modelled using our custom simulator written in C with the Windows platform.

Simulation Environment I

A 250 m × 250 m rectangular grid serves as the simulation setting. In the grid-based deployment, the nodes are uniform. The simulation assumes the sensor node, base station, and adversary operational communication range are ten metres. Nodes are classified into two categories. Type-1 nodes may detect critical events, whereas type-2 nodes can detect nominal events.

Moreover, the proposed system requires high confidentiality for critical events and low confidentiality for nominal events. The nodes are distributed densely to ensure extensive network coverage and connectivity. The size of each packet in this study is 2048 bits. The forward-random walk and shortest-path algorithms are compared to the SLP_ED methods. The shortest path algorithm determines the optimal route between the source and sink nodes. On the other hand, the forward-random walk randomly chooses the next node from a list of neighbour nodes in each iteration to reach the sink node. The forward-random walk and shortest path algorithms provide a similar route for nominal and critical events.

The following are the energy transmission and reception criteria in the simulation environment. The values for E_{elec} and ε_{amp} are taken as $50 \times (10^{-9})$ Joules/bit and 100 pico-Joules/bit/ m^2 respectively.

The suggested method (SLP_ED) operates in two steps for critical event detection. The algorithm chooses the subsequent node randomly from the list of neighbours of the current node. The neighbour list contains nodes that are in the farther set. The packet travels for *h* hops in step 1. However, as soon as the packet reaches the boundary node, step 1 is terminated, and step 2 is initiated. When a significant event occurs, the value of *h* is greater than one. This is reasonable as critical events require higher safety levels. The nominal event is *zero* when no privacy restrictions are present. The packet travels the shortest route to the sink during the second phase. In step 2, the SLP ED and SLP ED CBA algorithms use a dynamic shortest path to reach the sink node.

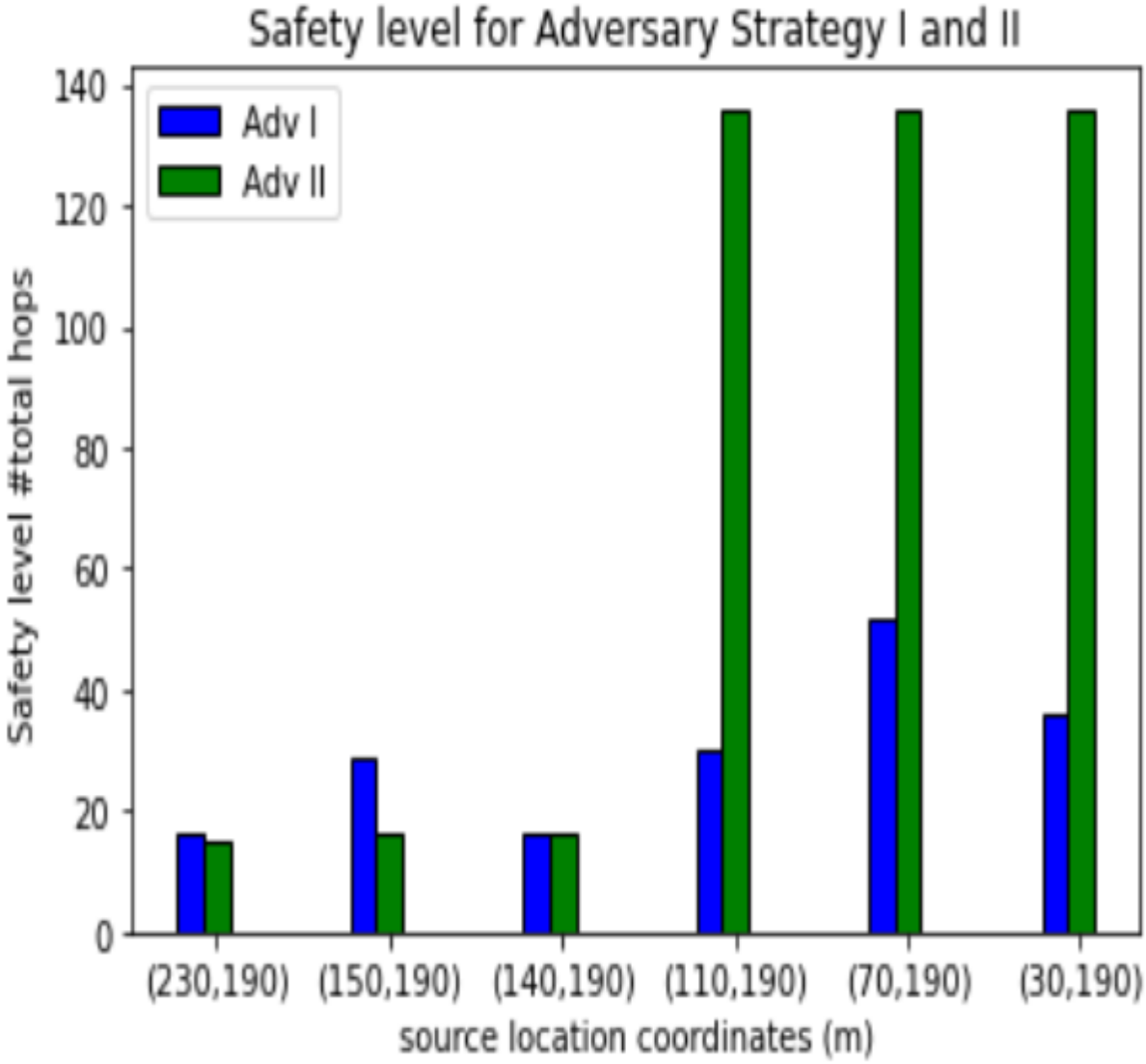


Figure 4.7: Safety level for adversary strategy

The sink node is the adversary's starting location. The adversary uses strategy-I or strategy II to backtrack from the sink node to the source node hop-by-hop. Figure 4.7 shows the safety level for attacker strategies. Adversary Strategy II's safety level rises in the suggested system. As shown in Figure 4.7, the safety level increases for adversary strategy II whenever the distance between the source and sink increases. The sink node's position is (240 m, 240 m).

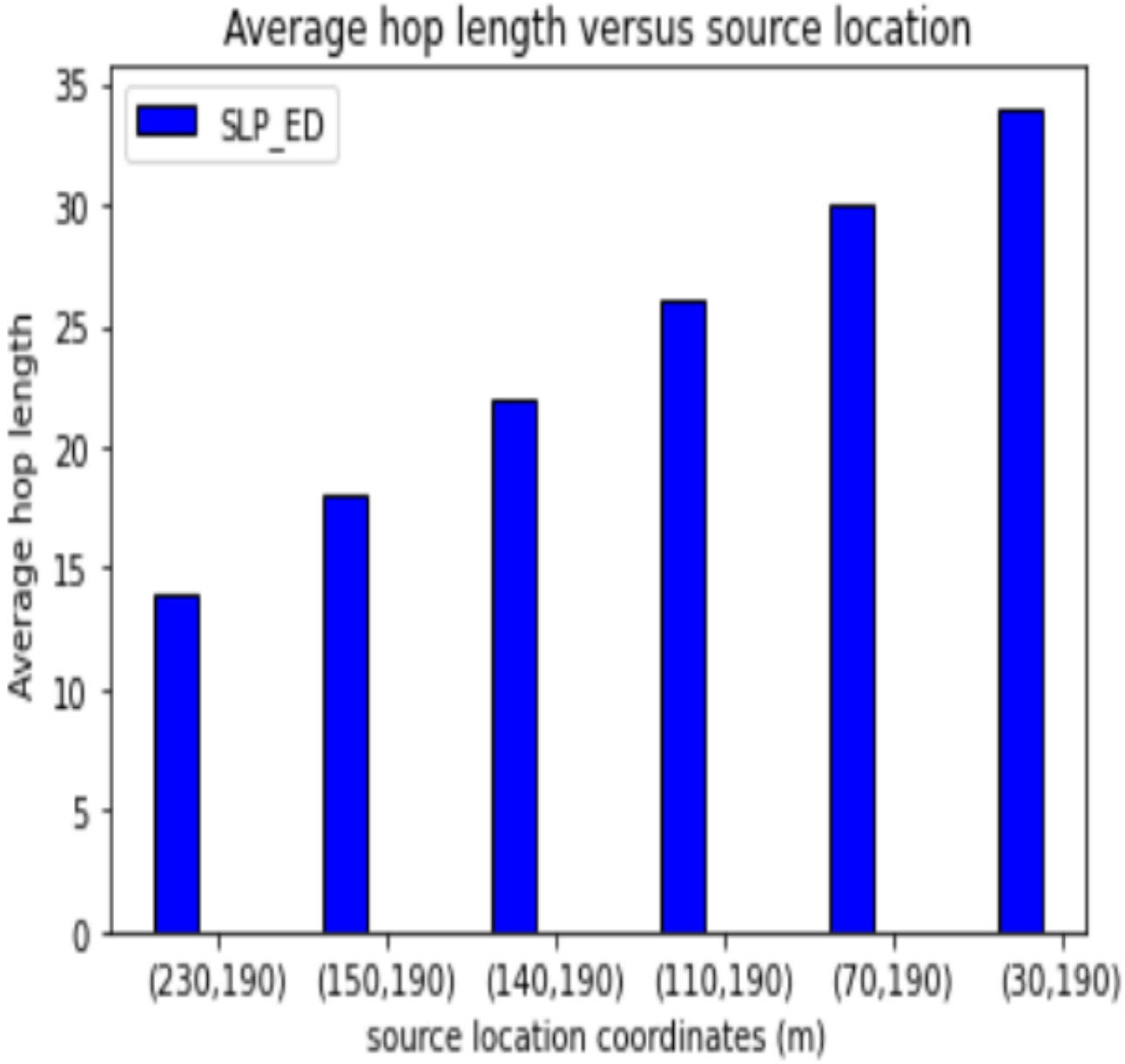


Figure 4.8: Plot of average hop length

More packets can be sent from the source node to the sink node in Strategy II. This is because the adversary position alternates between unvisited and visited nodes. Figure 4.7 depicts the SLP_ED critical event scheme. The coordinates of the source location sites are: (30 m, 190 m), (70 m, 190 m), (110 m, 190 m), (150 m, 190 m) to (230 m, 190 m) in 40 m increments. The results suggest that SLP_ED delivers a higher safety level for critical events. The average hop length plot for critical events for the SLP_ED scheme is shown in Figure 4.8.

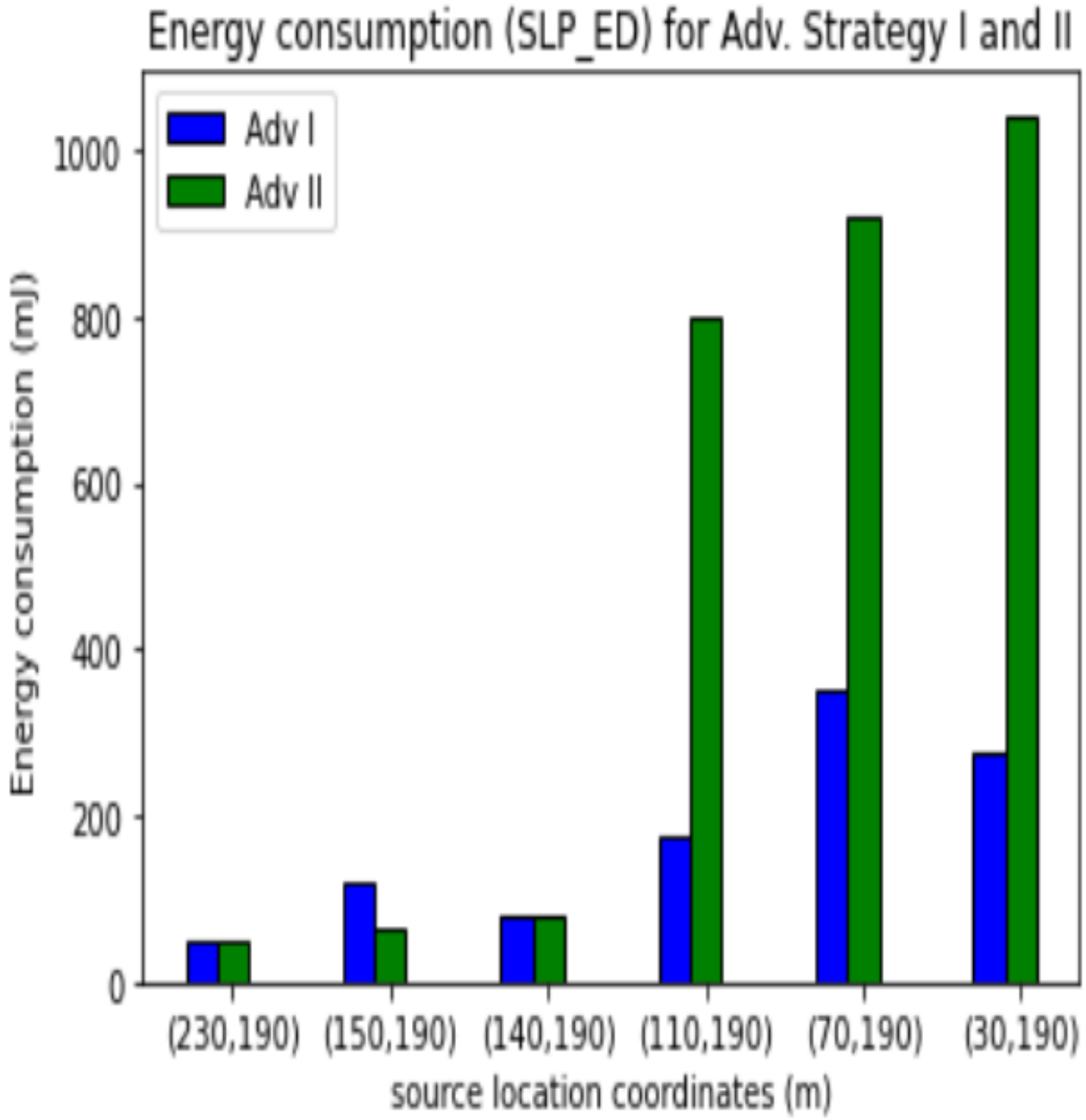


Figure 4.9: Energy consumption (mJ) for (SLP_ED)

The energy consumption in a network versus the source location for the SLP_ED technique is shown in Figure 4.9. The energy is measured in milli-Joules. The six source sites are spaced 40 m apart and range from (30 m to 190 m) to (230 m to 190 m). Sink and opponent coordinates are at (240m, 240 m). The user-defined value for *h* hops is taken as *four* for step I. The algorithm uses the shortest path for step II. The suggested plan offers a better safety level. The SLP_ED approach, on the other hand, consumes more energy than forward routing.

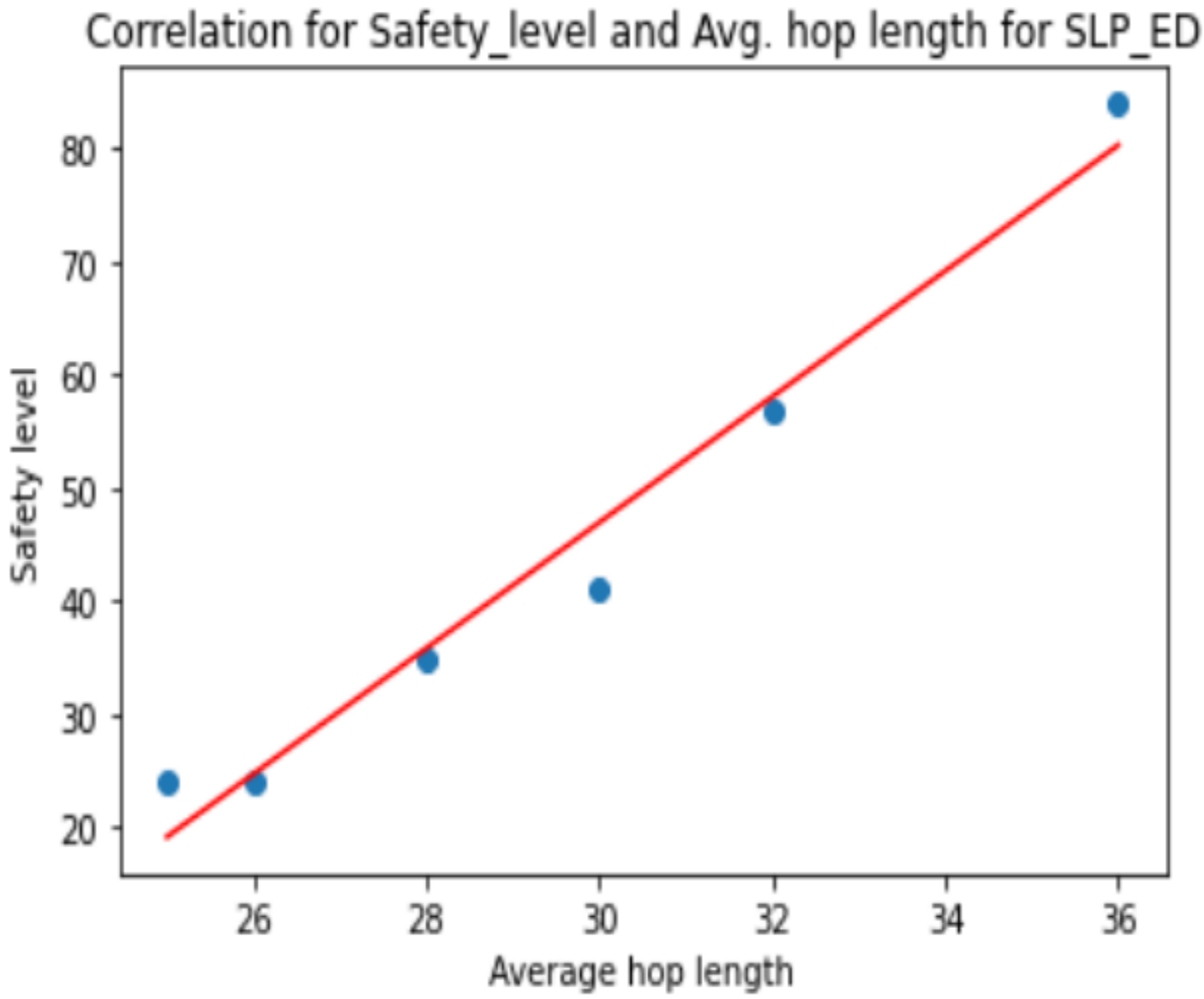


Figure 4.10: Safety level and Average hop length Correlation

Figure 4.10 for (SLP_ED) shows the relationship between the average hop length and the safety level. As the average hop length and safety level value increase, this leads to path diversity.

Simulation Environment II

A 400 m × 400 m rectangular grid serves as the simulation setting. In the grid-based deployment, the nodes are uniform. The deployment area is divided into 64 clusters, and 441 sensor nodes are deployed in the network. The node changes its state between active and sleep nodes periodically. The active and sleep nodes represent the chessboard deployment pattern. The nodes on the white portion of the chessboard deployment pattern are active, and the black portion is in sleep mode. On changing state, the active nodes are in the sleep node and vice-versa. All sensor nodes, base station, and attacker have a communication range of 100 meters, and the opponent employs Strategy II. The source node coordinates are (140, 120), (160, 120), and (200, 120). Simulation environment II takes the sink node coordinates as (300 300) meters.

The researchers suggest the Dynamic Shortest Path Scheme (DSP) algorithm, which chooses the next_hop node based on the current_node location. The selection strategy for the *next_hop* node in the dynamic-shortest-path algorithm considers the two factors. The first is the distance of the node from the base station. The second is the residual energy among the candidate nodes. The DSP technique deploys in a checkerboard pattern, creating a fresh cluster head after each round. The active and sleep cluster zones are switched off according to the chessboard deployment pattern. To address the privacy preservation challenge, the researchers (Jan et al., 2019) propose arranging sensors in a checkerboard pattern called Energy-efficient Source Location Privacy Protection (EeSP). Sensor nodes in the white and black regions alternate between the awake and sleep states to save energy across the network.

Two packets are sent to the sink in the first epoch over separate routes. These packets use the active zones only, as half of the nodes are in inactive or sleep mode. In the next epoch, the node changes state, and the sleep node becomes active. Two packets are delivered via separate routes in the second epoch. Thus, only half of the nodes are active in the chessboard deployment pattern.

In both epochs, the algorithm chooses nodes with higher residual energy than nodes in other regions and are closer to the sink.

The proposed method, called source location privacy for event detection (SLP_ED), is executed in two ways: with or without the Chess Board Alternation (CBA) pattern. In each of its iterations, the algorithm chooses the farthest neighbor from the current node. The process continues for h hops. The algorithm moves into step II if *hop_count* equals *h* or a boundary node is encountered. For both step 1 and step 2, all the nodes are in active mode without a CBA pattern. However, in SLP_ED_CBA, half of the nodes are in an active state, and the rest half is in a sleep state. For both step 1 and step 2, the active nodes alter in a chessboard manner.

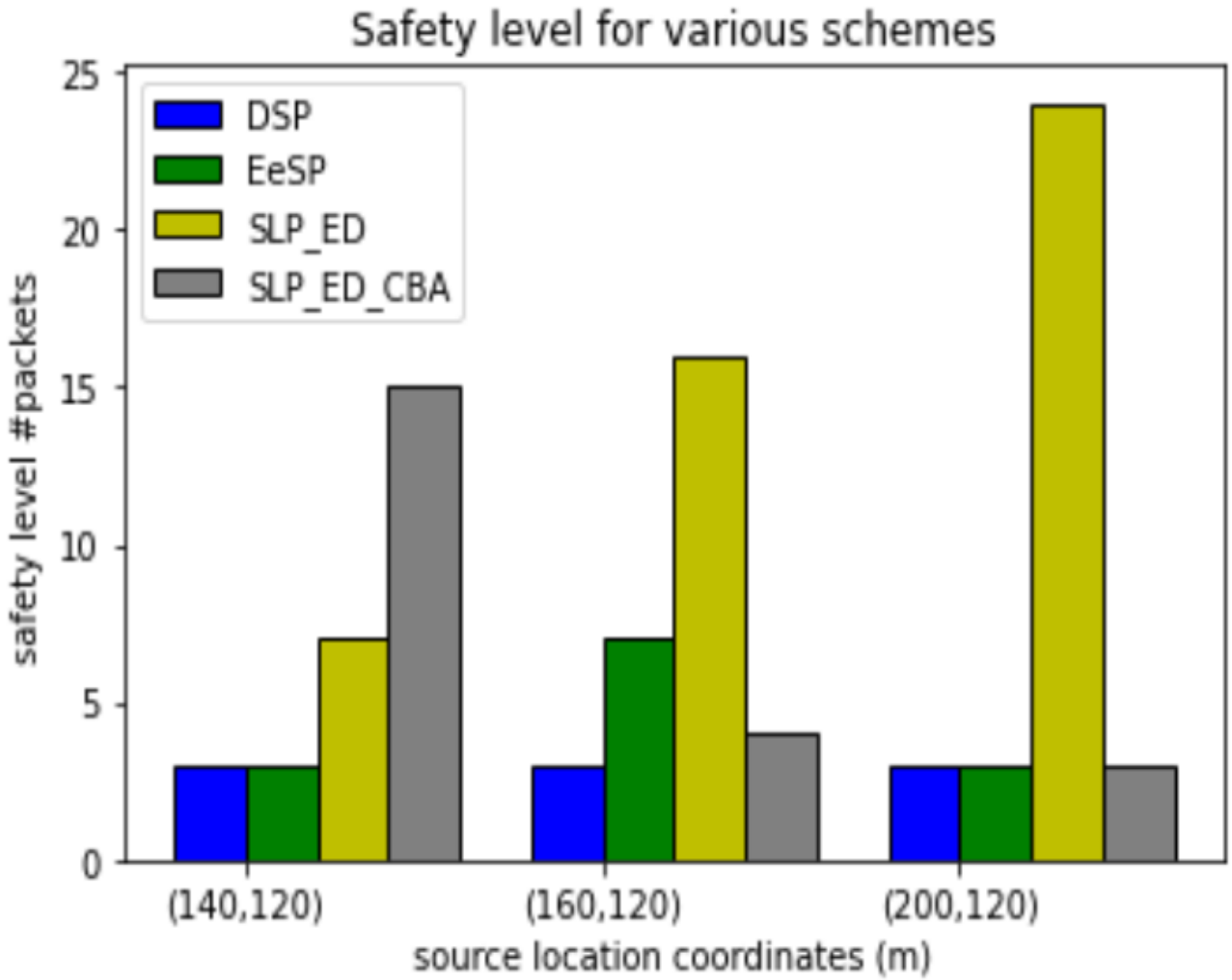


Figure 4.11: Safety level for the algorithms

Figure 4.11 shows the safety levels for four algorithms. SLP_ED and SLP_ED_CBA are the two proposed algorithms. The safety levels of the proposed algorithms are compared with the EeSP and DSP algorithms. As shown in the figure, the safety levels of the proposed algorithms are higher than those of EeSP and DSP schemes. A random walk in the opposite direction caused an increase in safety levels.

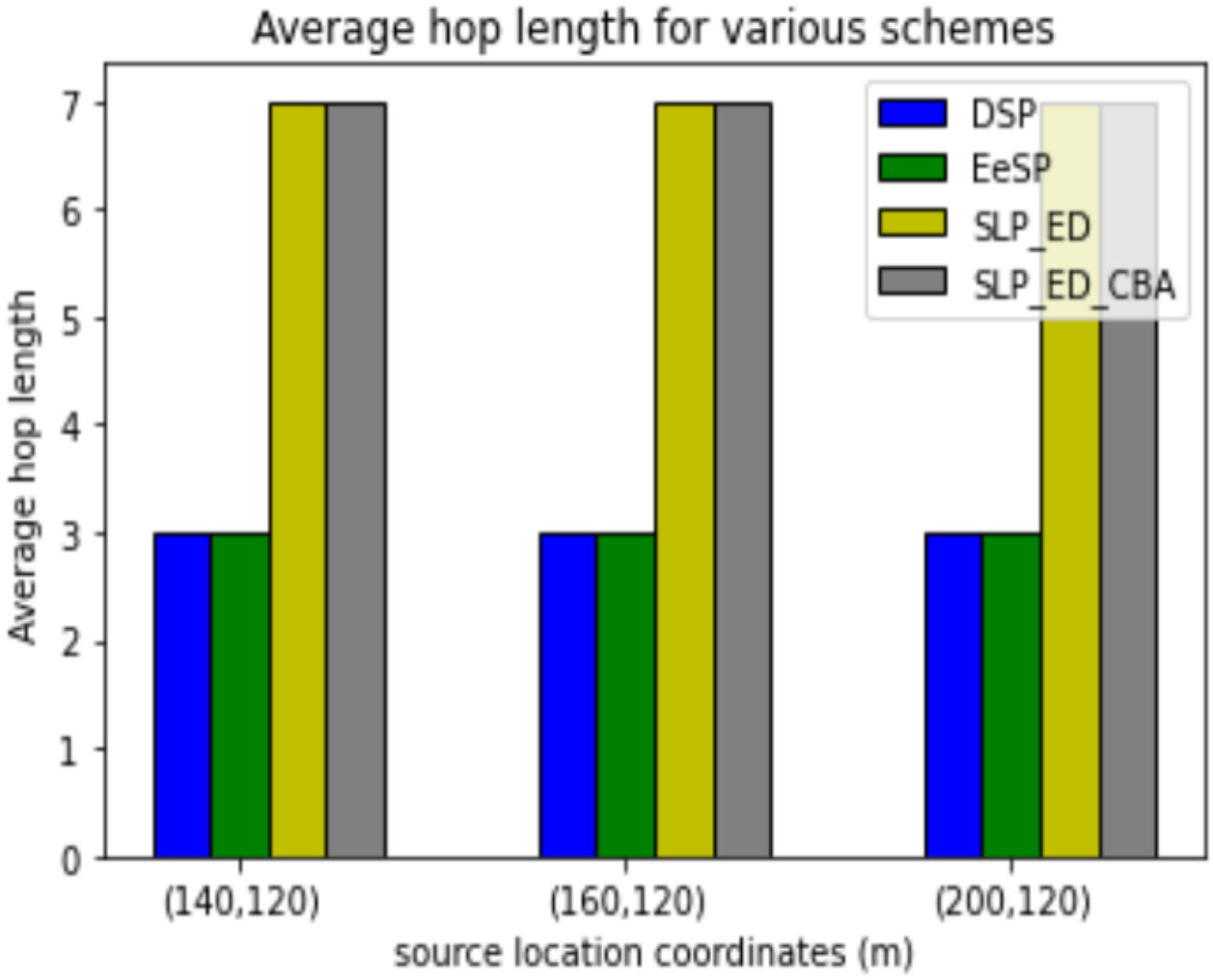


Figure 4.12: Average hop length for algorithms

Figure 4.12 shows the average hop length for four algorithms. The two proposed algorithms are SLP_ED_CBA and SLP_ED. The parameters of the proposed algorithms are compared with those of the EeSP and DSP algorithms. The average hop length of the proposed algorithms is higher when compared to EeSP and DSP schemes, as shown in the figure. A random walk in the opposite direction caused the increase in the average hop length. However, the increase in average hop length leads to an increase in the energy consumption in the network, as shown in Figure 4.13.

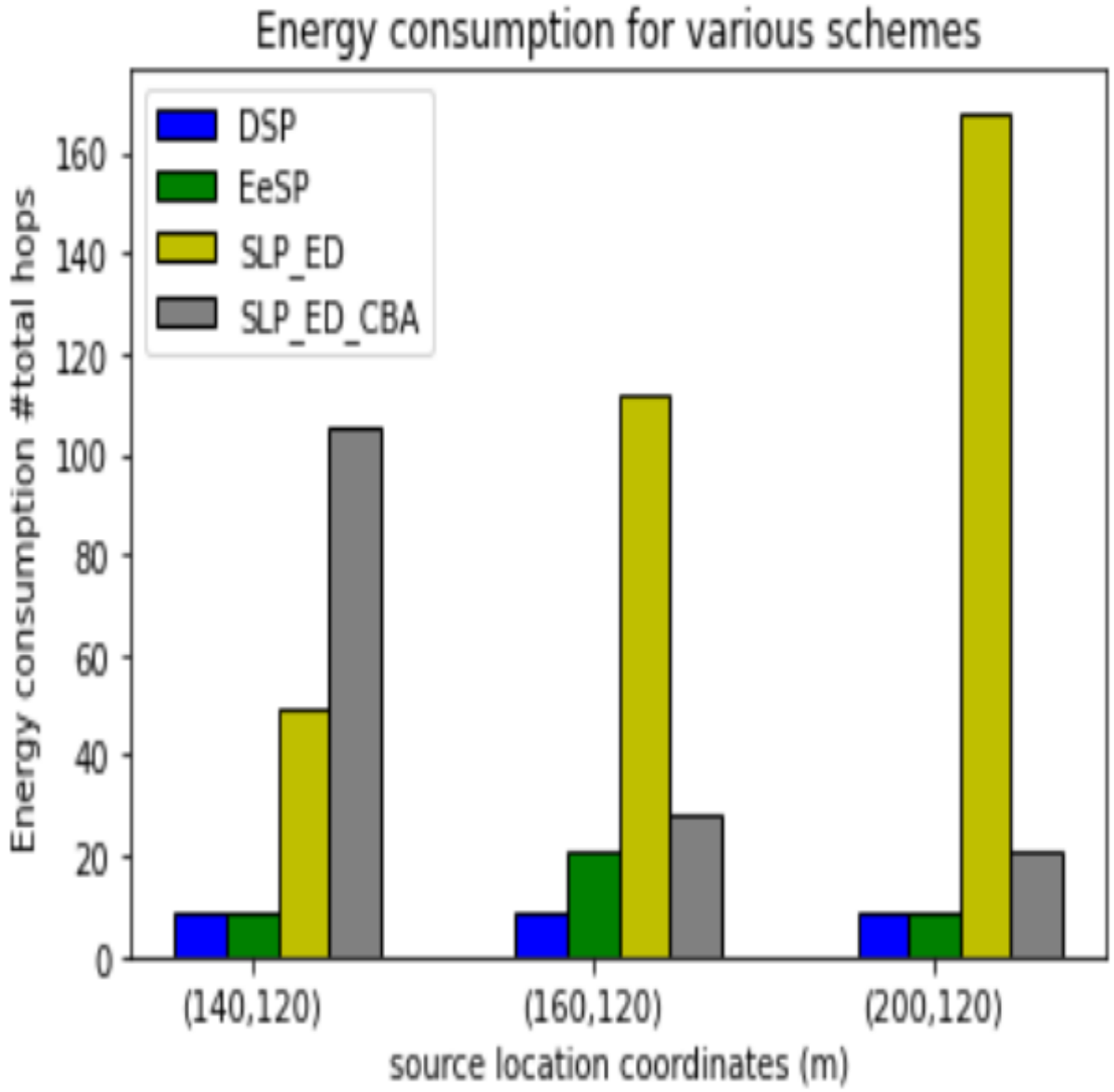


Figure 4.13: Energy consumption for the algorithms

The above-mentioned Figure 4.13 depicts the overall energy consumption at three different source locations in hops for four algorithms. The results demonstrate that our approach uses more energy than the suggested scheme. On the other hand, the proposed strategy provides a better safety level. Our proposed approach and comparison with existing approaches (Al-Mistarihi et al., 2020) and (Jan et al., 2022).

4.5 CONCLUSION

Privacy preservation is a crucial concern in Wireless Sensor Networks driven by events. The study contrasts SLP_ED with Forward Random Walk (FRW), Dynamic Shortest Path (DSP), and EeSP for various deployment scenarios. Nominal events require shorter hop paths with lower delay and less energy consumption. On the other hand, critical events require higher hop length with high delay and more energy consumption in the network. EeSP and DSP algorithms are compared with SLP_ED_CBA and SLP_ED on three parameters: energy consumption, average hop length, and safety levels for two deployment scenarios. Compared to DSP and EeSP systems, SLP_ED_CBA and SLP_ED offer higher degrees of safety.

Nonetheless, SLP_ED_CBA and SLP_ED consume more energy and have longer hops on average. The work can easily be extended to mobile and heterogeneous node deployment scenarios. The main idea of monitoring applications is to control the trade-off between location privacy and energy efficiency. In contrast, healthcare monitoring is to control the trade-off between location privacy and service utility.

The future scope of the work can be summarized as follows: for a real-world environment, the two-deployment plane can be extended into a three-dimensional plane. In addition, the work can be extended to Location Privacy Preservation with multiple source locations for healthcare and monitoring applications.

CHAPTER 5

DATA PRIVACY PRESERVATION SCHEME

Medical cyber-physical systems integrate a network of medical equipment, which is essential to healthcare. The security and privacy of medical data are, without any doubt, one of the main issues in the design of a Medical cyber-physical system. Significant progress in processing power over the last decade has allowed numerous academics to successfully apply several machine intelligence health applications. Machine Learning applications can provide valuable information to all stakeholders in the healthcare system. The data can facilitate patient care and diagnose diseases initially [180]. Better treatment is made possible by early disease identification. For example, it would be helpful if a doctor knows a patient's risk for a particular disease based on lab test results and family history. However, machine learning algorithms for healthcare applications are privacy-sensitive. The issue of data privacy is gaining importance in healthcare applications. Several privacy preservation algorithms for healthcare applications are available in the literature. Machine learning algorithms require large quantities of training data. The challenges for data privacy in healthcare applications are associated with machine learning algorithms. This chapter presents various Machine Learning Classification Techniques for a healthcare dataset. Further, we give privacy preservation techniques for the healthcare dataset. We compare six machine learning classification algorithms and observe that Support Vector Machine (SVM) performs better than other techniques. The original dataset is preserved by applying privacy-preservation techniques to the data. The work also explores data preservation techniques to secure machine learning models from leaking sensitive information. We observe that employing a single privacy-preserving technique could not provide optimal results.

Applications for medical services often use "knowledge graphs" to benefit all parties involved. For certain cases, the professionals might consult with other emergency rooms or consult knowledge graphs to learn more about the patients' past. Patients can essentially use a simple interface or a Chabot framework to query the "knowledge graph" for the location of the closest

expert with a high evaluation [76]. Physicians may question the information diagram for a family of conventional drugs, and they can also tell one drug from another by its salt. The work makes an engineering suggestion for a "Knowledge Graph" in the context of medical services. A crucial requirement for a clinical area is protection and conservation. The work offers protective protection approaches for "Knowledge graphs" used in applications for medical services. This chapter has focused heavily on machine learning algorithms for healthcare data analytics.

Privacy-preserving in healthcare data analytics refers to the techniques and methods used to protect sensitive patient information while allowing for effective data analysis.

Knowledge graph construction for better understanding of the data or insights into data.

5.1 INTRODUCTION

Recently, machine learning has become very popular for gathering beneficial data for business and scientific research applications in the healthcare sector. Healthcare data is sensitive as it contains the patient's personal information. Thus, privacy is a significant issue for healthcare applications. Machine learning helps doctors notice disease at the initial stage of the disease, leading to necessary medication. Figure 5.1 illustrates the various machine-learning techniques. Supervised, unsupervised, and reinforcement learning are the three main subfields of machine learning. The phrase "predictive learning" relates to supervised learning. Using data from similar items that correspond to the class of the unknown item, a machine can predict the class of the item. Descriptive learning is another name for unsupervised learning [175]. By combining related objects, a system can identify patterns in unidentified objects. In reinforcement learning, a computer learns to act independently to accomplish objectives.

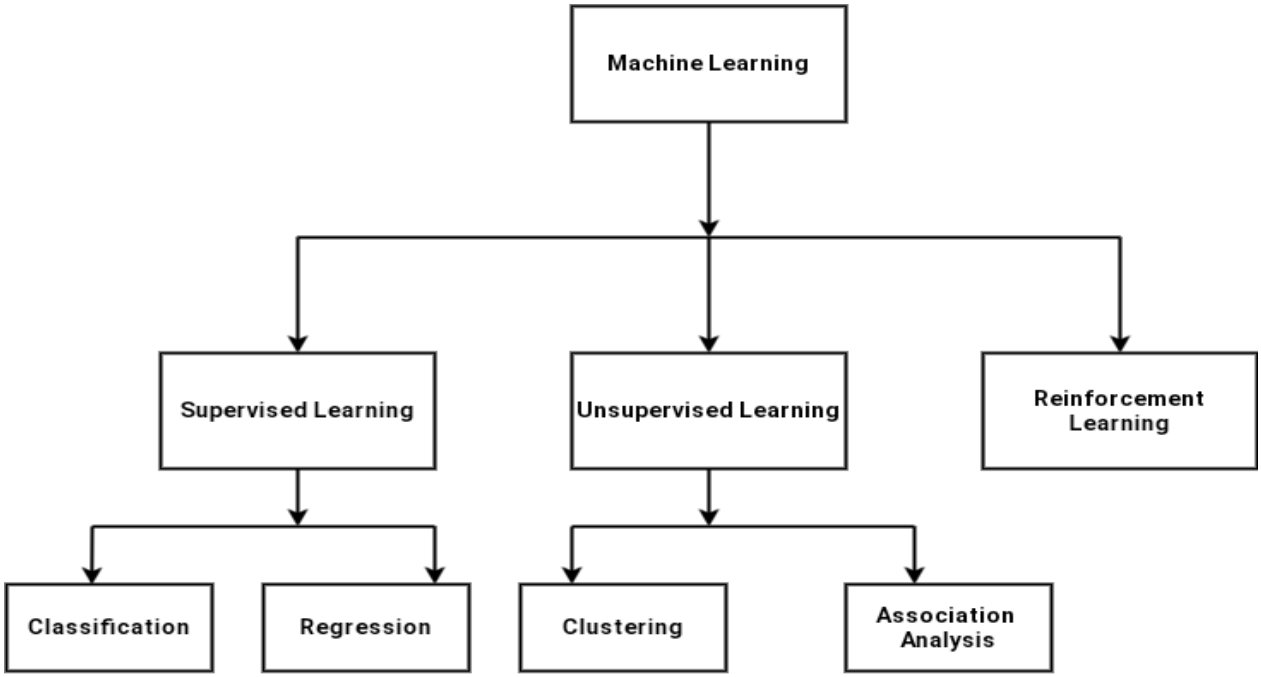


Figure 5.1: Types of Machine Learning

People have recently become increasingly concerned about privacy issues in healthcare and other applications. Current e-healthcare systems, on the other hand, lack privacy and user trust. Sharing a patient's information would severely threaten data privacy. Further, a breach of privacy leads to moral, legal, and social problems. Summarization, data separation, and data obfuscation are popular privacy-preserving techniques. We now present some popular machine learning algorithms and privacy preservation techniques.

Mining data sets spread across many parties has recently become essential without disclosing further private information [1]. The authors discuss the big data life cycle [2]. Data anonymization and encryption are imperative approaches in the privacy preservation of big data [3][4][5]. Suppression and generalization are two k-anonymity strategies. The provided data may become less helpful to receivers of excessive anonymization [2]. Secure multiparty computation is a mechanism to calculate a function without disclosing its private inputs [5] [6][7][8].

Health systems share the data horizontally or vertically [9]. The author proposes a horizontally separated training scenario with records provided to learners and presents a vertically partitioned training dataset with attributes distributed to learners [1] [9].

The authors suggest an architecture that combines horizontal and vertical data distribution for medical applications [10]. The authors propose a method combining vertical and horizontal

partitioning [11]. The objective of vertical partitioning is to induce diversity among learners. The primary purpose of horizontal partitioning is to generate diversity among several classifiers. To achieve a trade-off between data privacy and data utility, the authors propose maintaining privacy by rearranging various properties of big data [12]. The authors present a survey on multiple classifier systems [13]. Results show that multiple-classifier systems outperform single-classifier systems [13]. The authors discuss strategies for horizontally partitioned data to carry out PCA computing [14] [15] [16]. Data contributed by various participants are horizontally divided in the data space [15]. The authors investigate secret sharing framework-based PCA computation for horizontally partitioned data [15]. Privacy maintaining Knowledge Graph (KG) becomes a key research question to tackle [17].

The data set is accessible online in .csv format from UCI's machine learning repository [18]. Preliminary data analysis and preparation have retained only those features containing adequate information [18].

The description of the chosen dataset is now shown. The dataset used has 50 columns and 101767 rows. Age, Number of Diagnoses, Number of Emergency Visits, Number of Inpatients, Number of Lab Tests, Number of Medicines, Number of Outpatient Visits, Number of Procedures, and Duration in Hospital are the features selected for the study. The visualisation of the chosen dataset is now being presented.

Data visualisation is the presentation of data in pictorial or graphic format. It helps people understand the importance of information in a simple, easy-to-understand format by summarizing and presenting large quantities of data to communicate it clearly and effectively.

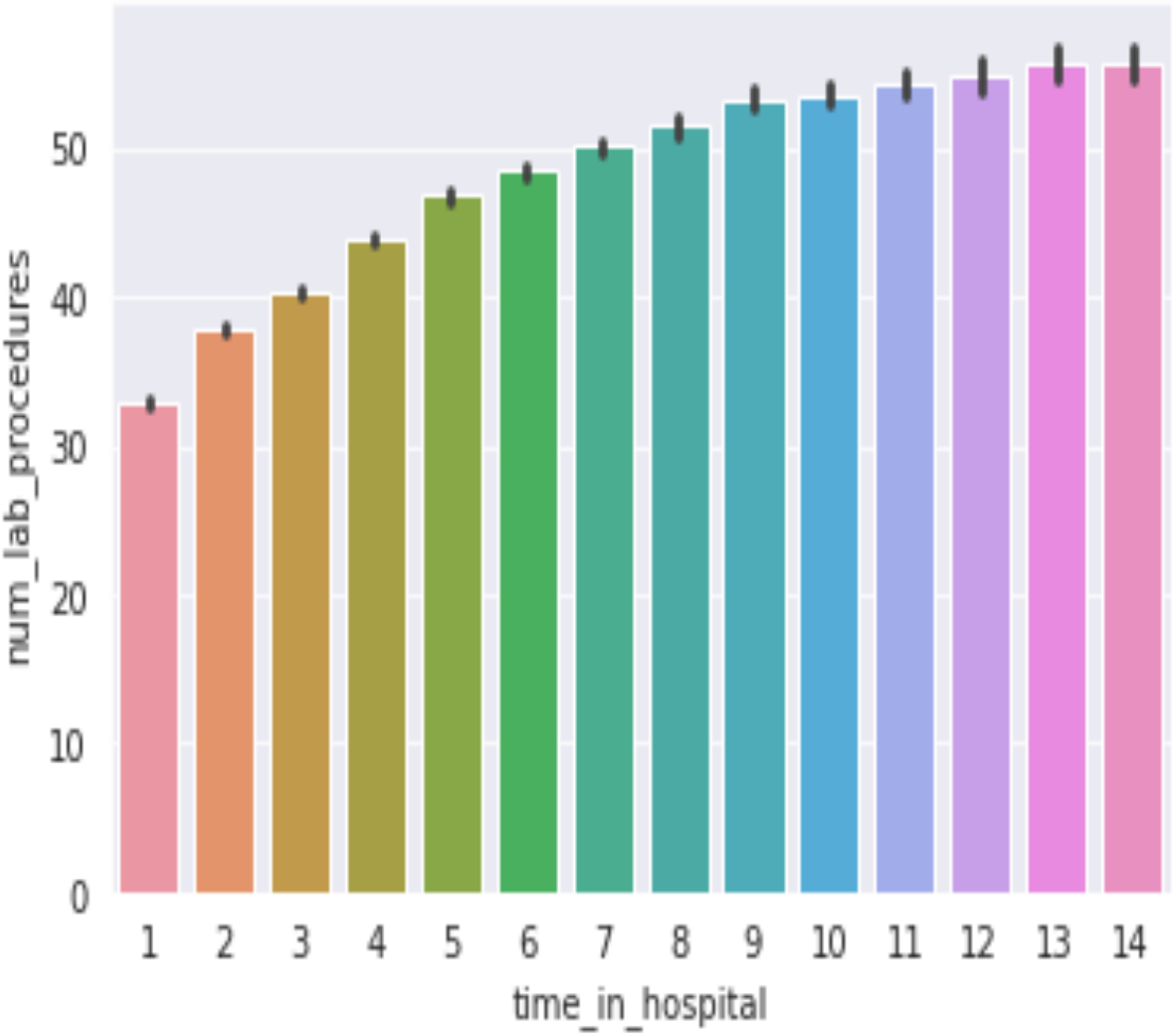


Figure 5.2 (a): Plot of the number of lab procedures versus time in the hospital

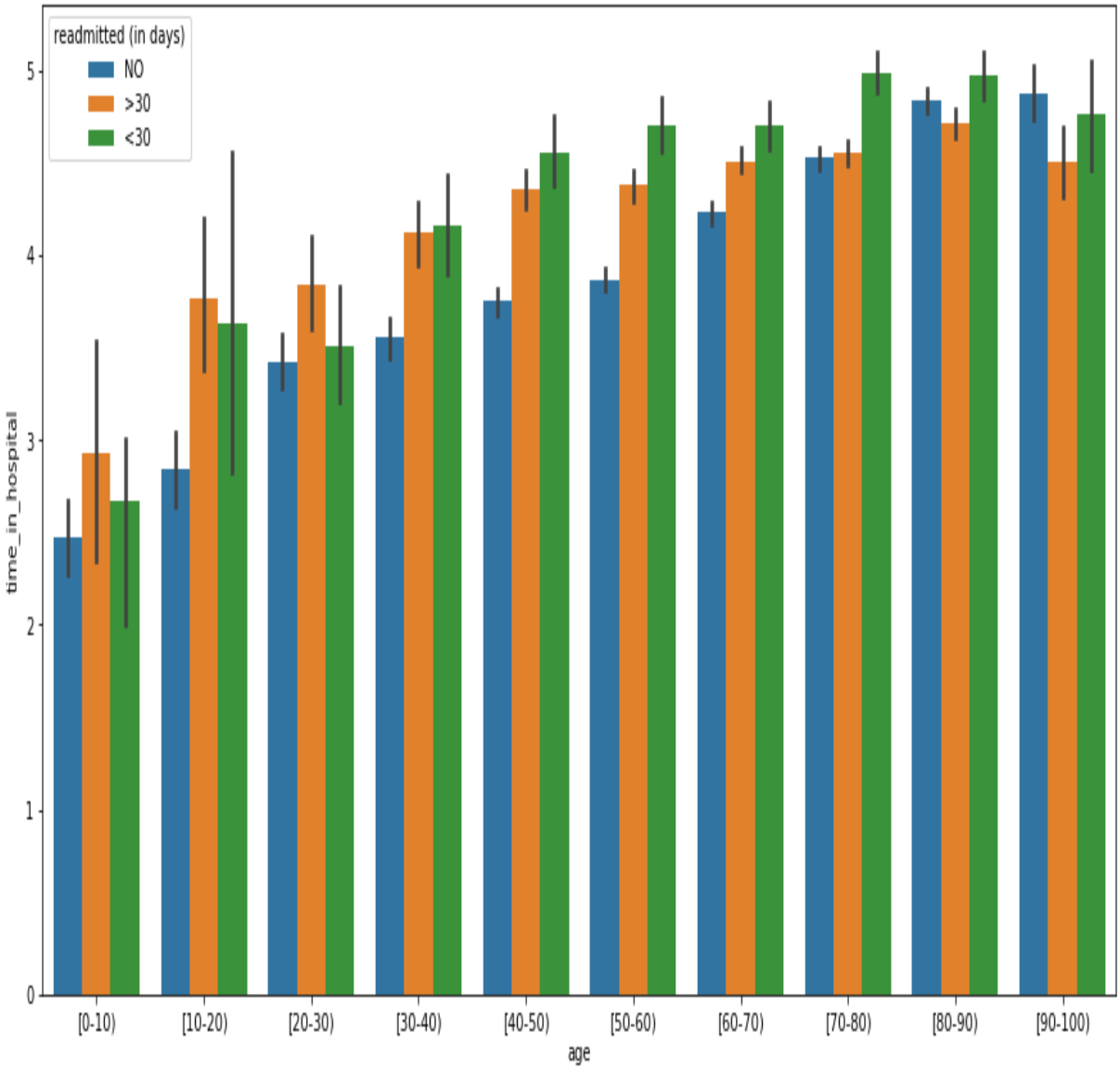


Figure 5.2 (b): Plot of time in the hospital versus age

Figure 5.2 (a) plots a bar chart between the numbers of lab procedures versus time in the hospital. The number of lab procedures is directly proportional to the time in the hospital. Figure 5.2 (b) plots time in the hospital versus age. The plot for the time in hospital for three cases: case 1 is no readmission, whereas case 2 and case 3 are readmission greater than 30 days and less than 30 days, respectively. As shown in Figure 5.2 (b), the time in the hospital increases with age.

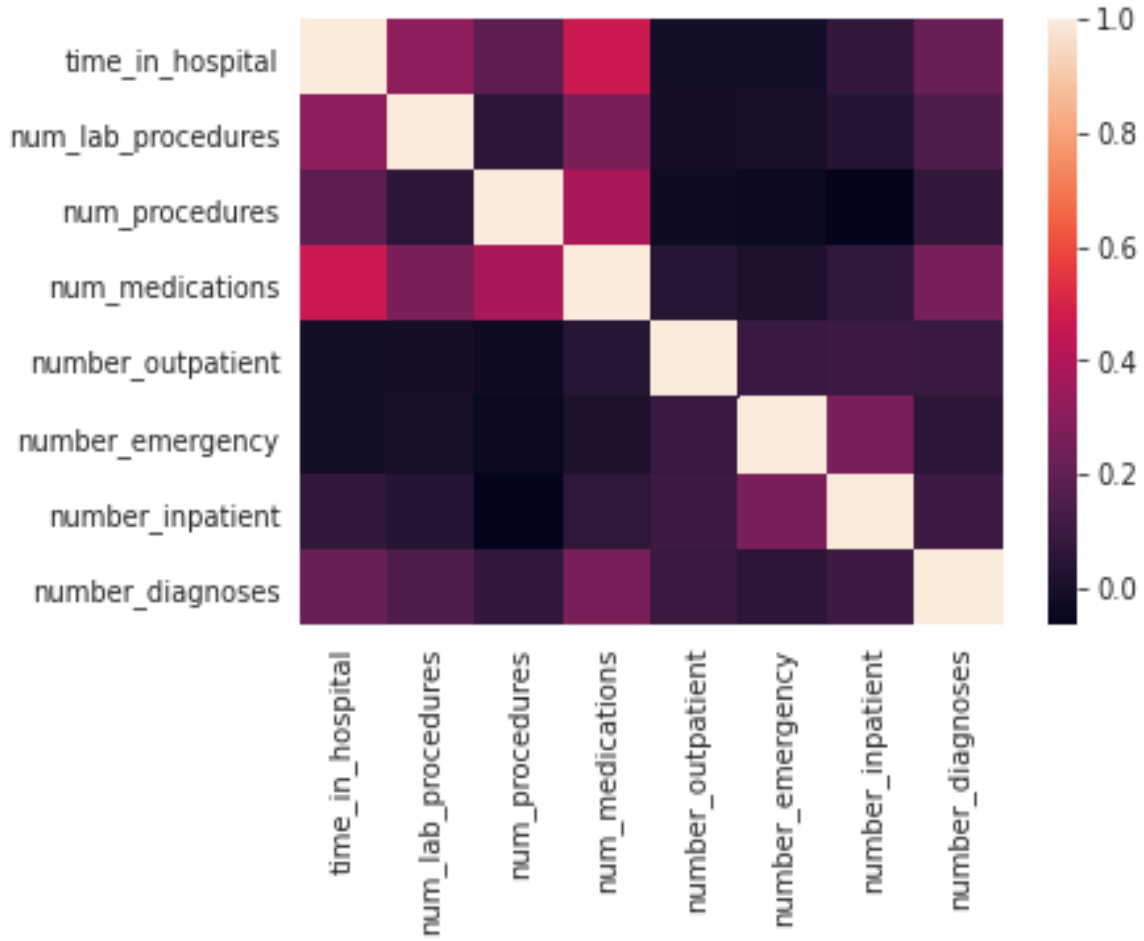


Figure 5.2 (c): Plot of heat map

Figure 5.2 (c) plots the heatmap for the diabetes dataset. A heat map is a method to view the data showing the magnitude of a two-dimensional colour phenomenon. The colour variation can vary with hue or intensity, providing clear visual indications of how the phenomenon is clustered or varies across space. The smaller dark grey or black pixels represent the larger values, and lighter squares represent smaller values.

5.1.1 Classification Techniques: The classification technique predicts the dependent variable using independent features. The classification method consists of two steps: building the classification model (model training) comes first, and then applying the model to forecast class labels comes second. The test Set size is 25%, whereas the Training Set size is 75%. Age, Number of Diagnoses, Number of Emergency Visits, Number of Inpatients, Number of Lab Procedures, Number of Medicines, Number of Outpatient Visits, Number of Procedures, and Duration in Hospital are the chosen features (independent variables). The independent

variables are represented by a matrix "X"; the dependent variable, i.e., readmission of patients dependent on these independent variables, is denoted by a column vector, "Y".

LOGISTIC REGRESSION:

Logistic Regression follows a probabilistic approach to determining whether a particular event is going to occur or not.

$$g(z) = \frac{1}{(1+e^{-z})} \qquad (5.1)$$

It uses the sigmoid function g(z), where z is the real value and output g(z) is between 0 and 1. The dependent variable Y is binary (0, 1) in the logistic regression, while the independent variables (X) are continuous. The logistic regression predicts the probability of Y being 1 given some X values. The logistic formulas for the probability of Y =1 are known as P. The chances of Y being 0 are (1 - P).

$$\ln(P / (1 – P)) = a + b X \qquad (5.2)$$

The 'ln' symbolises a natural logarithm, and $a+b$ **X** represents the regression line equation. The coefficients *a* and *b* are used in the equation. To prevent overfitting, we have selected only nine significant features. Since we need to make sure that features are on a similar scale, we perform feature scaling on our selected features, using the formula given below, x_{min} is the minimum value and x_{max} is the maximum value of that feature respectively. χ_i is the i^{th} feature of the selected features.

$$x_{i=\frac{x_i}{xmax-xmin}} \qquad (5.3)$$

RANDOM FOREST CLASSIFIER:

An ensemble classifier is Random Forest. Ensemble learning is how several models, such as classifiers or experts, are systematically developed to address a particular issue in artificial intelligence. Random forest uses and combines many decision tree classifiers. The reason for using many trees in random forests is to sufficiently train the trees so that several models contribute to each feature. Following the creation of the random forest by combining the trees, the output of the several trees is combined using a majority vote. The simulation is done in

Python. The parameter "n-estimators" required in Python for this algorithm denotes the number of trees taken as 10 in this work. The second parameter used by this algorithm in Python is "Criterion". The algorithm criterion is set to "entropy".

Since we need to ensure that features are on a similar scale, we perform feature scaling on our selected features using the formula below: xmin is the minimum value, and xmax is the maximum value of that feature, respectively, and χ_i is the i^{th} feature of the selected features.

$$x_{i=\frac{x_i}{xmax-xmin}} \quad (5.4)$$

DECISION TREE CLASSIFIER:

One of the most popular methods of classification is the decision tree method. The purpose is to create a model based on the functionality of the input vector variable that will predict the value of the output variable. Each node of a decision tree corresponds to one of the feature vectors.

The decision tree classifier separates the dataset by different split lines. These split lines try to divide the dataset into various sections with different dimensions so that each section contains data points mainly of only one particular category and minimizes the entropy. This optimal split creates a tree-like structure in the background, which involves different decision (if-else) conditions. After creating these sections in our training set, we finally use them for our test set. We use each data point to identify which category or section they fall under and assign them their respective category.

The predictions for all data points are 0 and 1 in a "speed" column vector, which we compare with our test set data of the "y" column vector. This helps us determine the accuracy of our results with the help of the confusion matrix created. The algorithm is simulated in Python. Further, the Criterion is a parameter we must set in Python for the decision tree algorithm. The criterion is set to "entropy" for this algorithm.

K-NEAREST NEIGHBORS CLASSIFIER:

The training data set, test data set, and several neighbours are the inputs for the kNN algorithm. The algorithm selects the number of neighbour as five. The algorithm computes the closest 'k'

training data points whose distances are the least from the test data points. Following this, the test data point maps to the majority class label in the training data. The distance between data points is computed based on Euclidean distance. The distance between two points is called the Euclidean metric in Euclidean space. Manhattan, Minkowski, Chebyshev, etc., are other metrics used in work. Since we need to ensure that features are on a similar scale, we perform feature scaling on our selected features.

We present a kNN demonstration via a diagram. The plot of the training data set with the two independent variables x1 and x2 is shown in Figure 5.3 (a). The training data set's dependent variable, displayed in black and red as 1 or 0, is 1 or 0. For the prediction of the test data set, we select the number of neighbours as five, as shown in Figure 5.3 (b). As shown in the figure, three of five neighbors belong to the red class; test data is either red or 0.

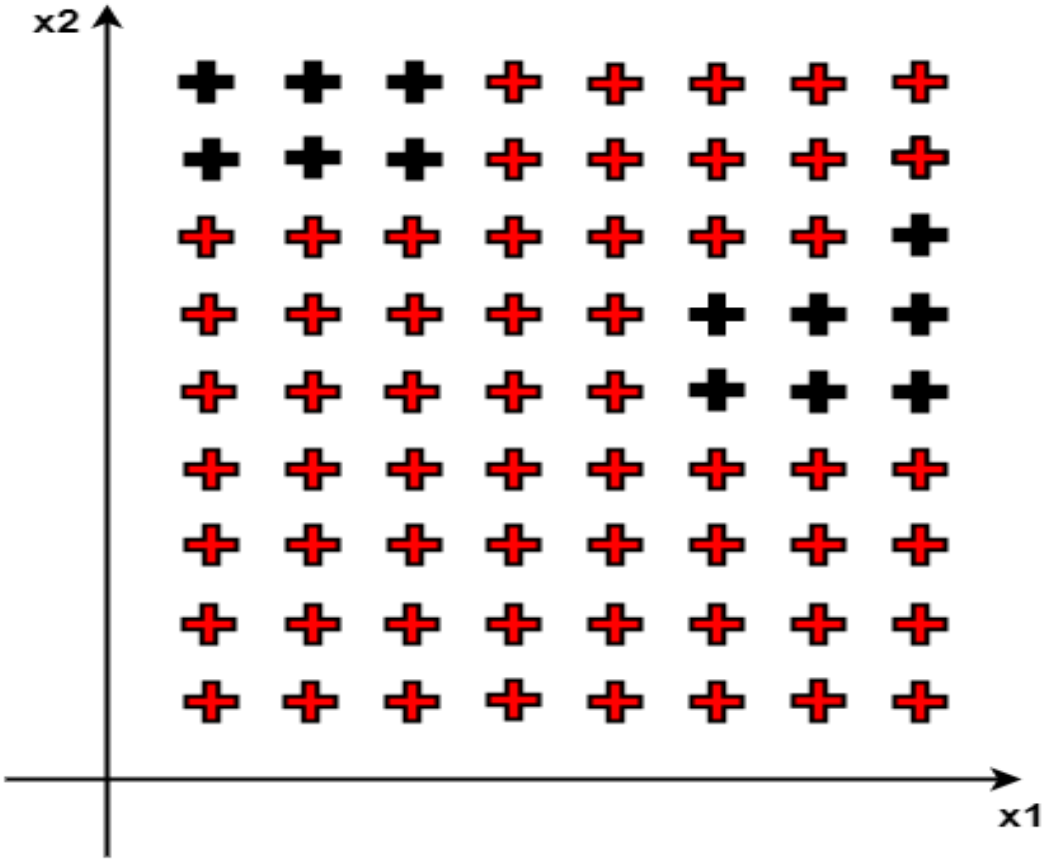


Figure 5.3 (a): Training data

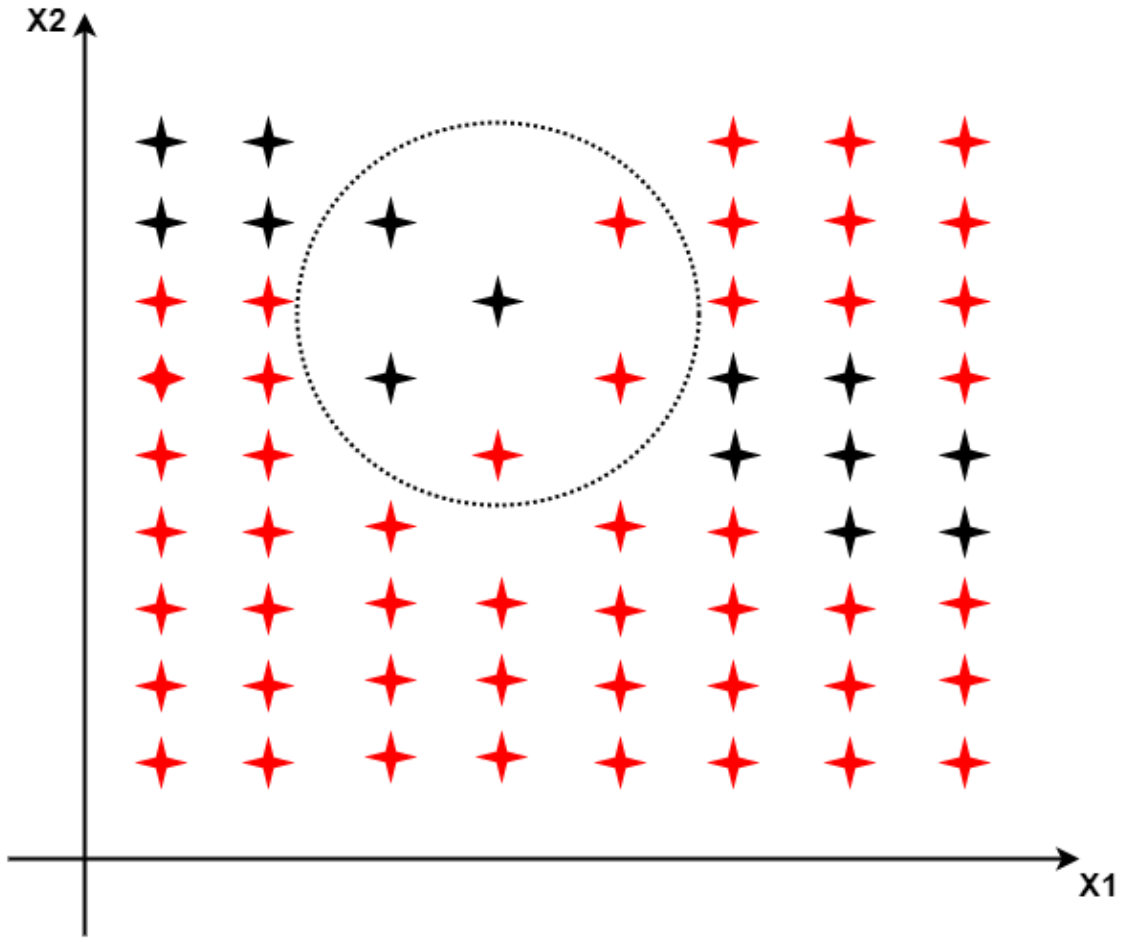


Figure 5.3 (b): kNN Illustration

SUPPORT VECTOR MACHINE (SVM) CLASSIFIER:

A Support Vector Machine classifier is used for both regression and classification. Support Vector Regression (SVR) is the name for solving a regression problem.

When we have a linearly separable group of points of two different classes, the goal of the SVM is to find the separation between classes or hyper-planning. Separating two classes ensures that the perpendicular distance between the nearest points in each class is maximum.

Figure 5.3 (c) shows the plot of two independent variables, x1 and x2, for different data points. The figure illustrates a hyperplane separating data points into two classes: red and black colour. Since the test data points fall in the red region, red is allocated to test data points.

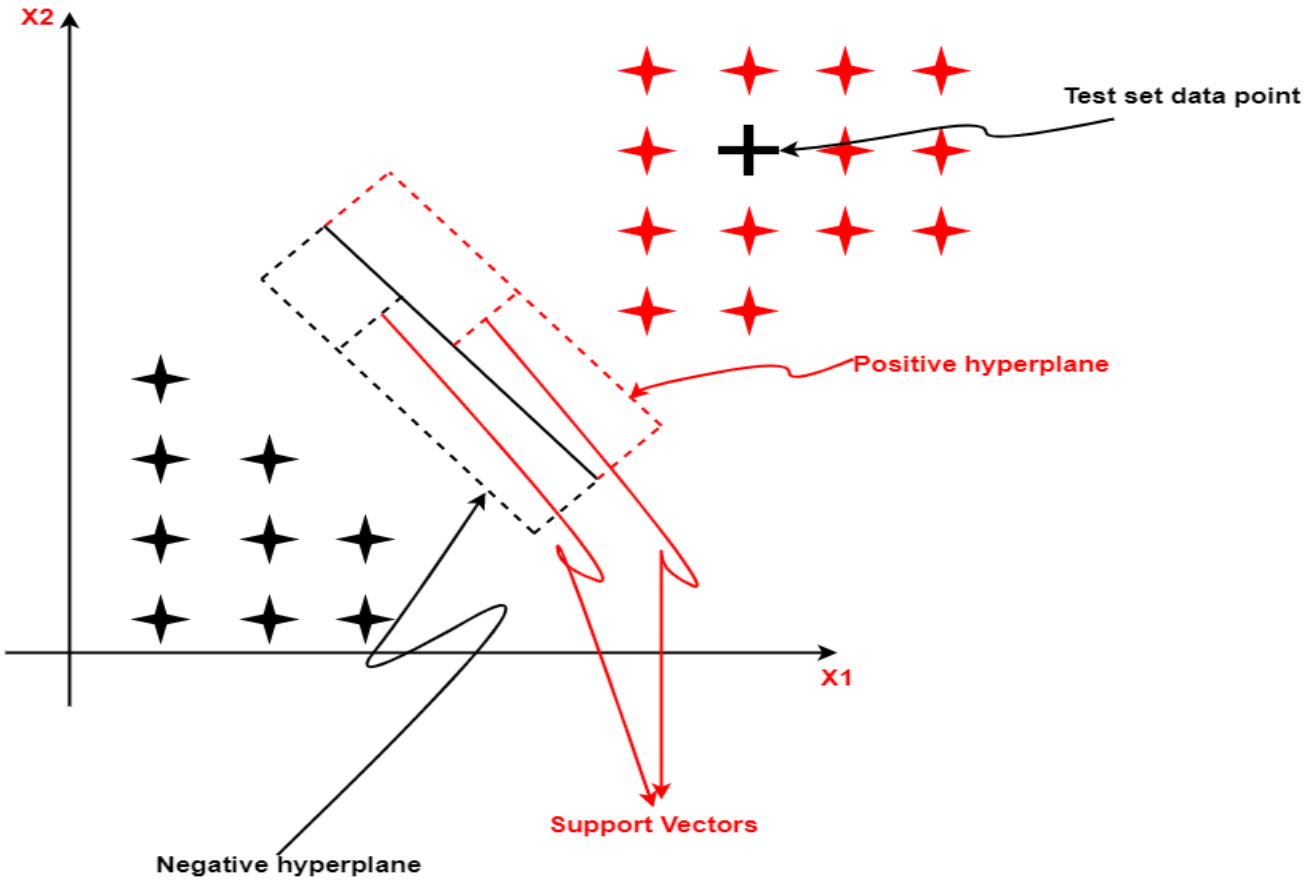


Figure 5.3 (c): SVM illustrations

NAIVE BAYES CLASSIFIER:

Within the algorithm of the Naive Bayes classifier, Bayes' probability theorem calculates conditional probability, i.e. the likelihood of an event supported by previous data accessible on the events.

The naive Bayes Classifier relies on Bayes' theorem for its predictions. Bayes' theorem talks about conditional probability as shown, $P(B|A)$ to represent the probability of event B, given that it has already occurred, and represents the probability of event A, given that B has already occurred. Naive Bayes' Classifier uses Bayes' theorem as shown below $P(y|x) = \frac{P(x|y)P(y)}{P(x)}$

Here, P(y) denotes Prior Probability, P(x) denotes Marginal Likelihood, P ($x|y$) denotes Likelihood, P ($y|x$) denotes Posterior Probability, x denotes matrix of independent variables, and y denotes column vector of the dependent variable. These probabilities need to be calculated in order.

To calculate the marginal likelihood, we plot the dataset in n-dimensional space and choose a circle with a radius of our choice around the data point whose prediction we must make.
To calculate the marginal likelihood, we plot the dataset in n-dimensional space and choose a circle with a radius of our choice around the data point whose prediction we must make.

$$P(x) = \frac{No.of\ similar\ observations(no.of\ points\ lying\ in\ circle)}{Total\ Observations(total\ datapoints)} \quad (5.5)$$

Similarly, for calculating likelihood, we choose the numerator as several similar observations favourable to y and the denominator as total observations favourable to y. The posterior probability is calculated as one of the results is greater than or equal to 0.5. Since we must ensure features are similar, we perform feature scaling on the dataset.

5.1.2 KNOWLEDGE GRAPH (KG)

A knowledge graph (KG) handles hubs and edges like a traditional chart. Yet, it differs from the "standard graph" since the components of this current reality, connections, and meanings are structured [1]. A couple of the representational schemes used in the "knowledge graph" are the "Property graph" and the "Resource Description Framework (RDF)". Neo4j uses a designated "property graph" to address a "knowledge graph." A material having zero or more qualities is referred to as a hub [2]. Similarly, a relationship between two hubs might have at least zero attributes. The subject-predicate-object triples are stored away substantially more in the RDF structure. For instance, the triple (abc, name of, patient) is used. For applications in medical care, RDF supports a variety of tiered information models and social information models. An associated graph with a substance, connections enriched with semantics to such a degree that mind-boggling decisions are successfully made for medical services, is known as a "knowledge graph" for medical services. Moreover, the Knowledge Graph-based Decision Support System (DSS) supports professionals in the medical field, such as specialists and doctors [3]. The need for DSS was realised worldwide, including in India, during the coronavirus pandemic from 2019 to 2021. In India, in particular, the medical services sector is struggling with a severe labor shortage due to the disparity between the population and the number of medical services specialists. Clinical "knowledge Graphs" or "knowledge Graphs" for medical services support the framework for medical services in this way. The "knowledge graph" categorization is shown in Figure 5.4 [1].

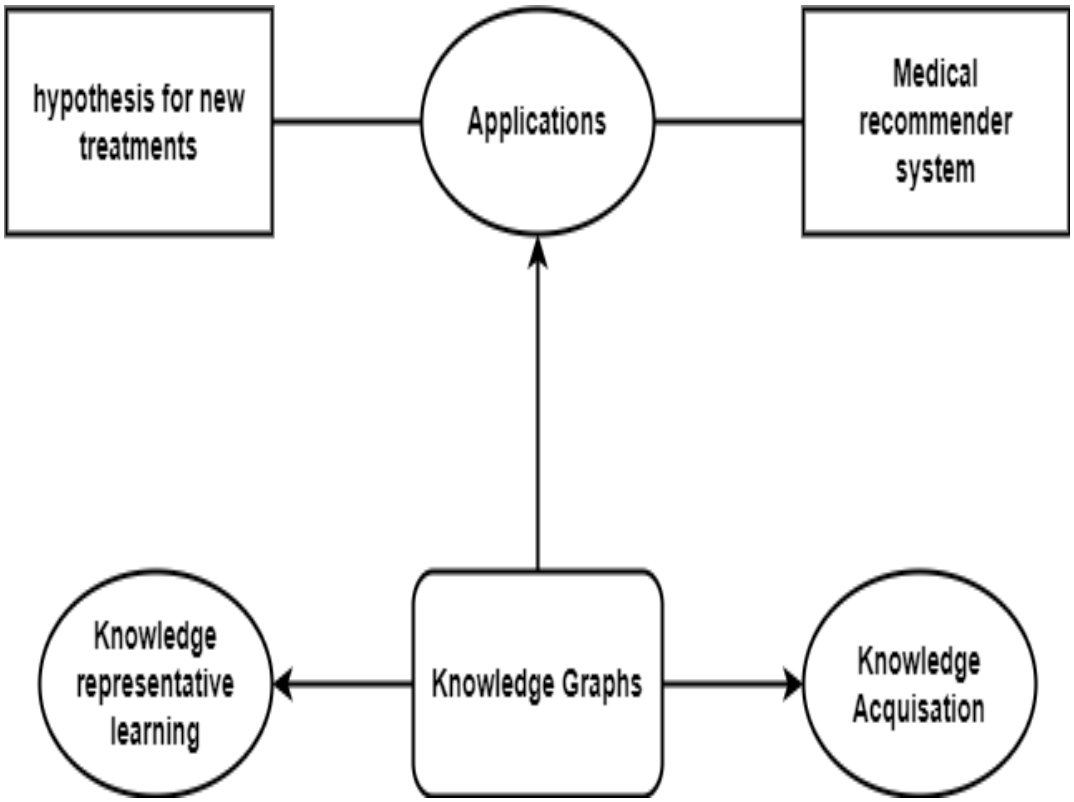


Figure 5.4: KG Categorization

A thorough classification of the current study's clinical "knowledge graph" into three groups: "knowledge-representative learning, knowledge acquisition, and applications." The study suggests the creation of a "knowledge graph" for the medical services field in light of engineering and metaphysics. Additionally, the therapeutic "knowledge graph" engineering incorporates protective practices. Given the vast amount of information available for the treatment of diabetes, the designers discuss the enhancement of a data architecture [4]. The framework gathers various data from many sources, preprocesses the data, and stores it as a data set. A "knowledge graph" can be added to the architecture to aid collaborators in the clinical field. The remainder of the work is divided into sections: Segment 3 follows section 2, which addresses the linked work. Segment 2 discusses the usage of information diagrams in the medical services sector and KG protection and conservation. The diagram is suggested for applications in medical services in segment three of information engineering.

APPLICATIONS OF KNOWLEDGE GRAPH

KG is used to find the manufacturer of a certain medication, the standard salt, or the drug that is most often prescribed. The medical services sector uses a manual process to create summary notes and review patients' clinical histories or connections to earlier contextual investigations to provide therapy. The electronic health record (EHR) reconciliation using the information diagram can help speed up the conversation. Information diagrams can be a foundation for suggestions in the medical services industry. All partnerships with providers of medical services are included in the information diagram that illustrates the patient excursion. The occurrences include treatment, several days spent in the emergency clinic, trips to emergency clinics, etc. This information on the human body could be applied in procedures by professionals, specialists, and governmental organisations. KG helps to identify between multifunctional drugs, recognise networks, and locate the linked illness. The infection and side effects diagram made using Neo4j is shown in Figure 5.5.

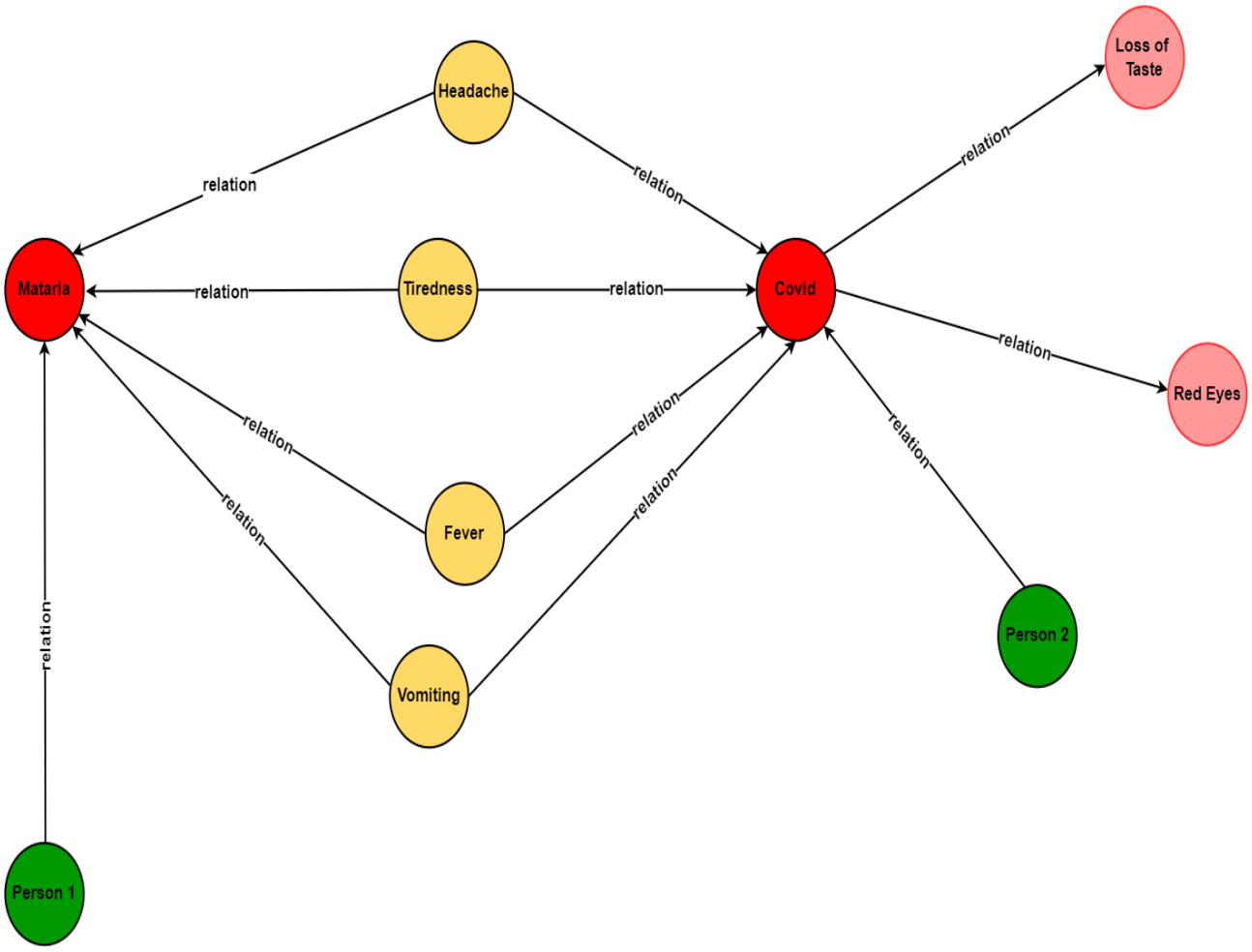


Figure 5.5: Property Graph

PRIVACY PRESERVATION FOR HEALTHCARE

Information on medical services from different emergency clinics should be exchanged for experts' improved comprehension of an infection or the patient history. The huge array of medical services data includes sensitive information, including infection name, age, PIN code, postal address, phone number, religion, and other details. In large-scale information analysis, a few protection-saving strategies are implemented to preserve the information. Due to information sharing between insurance companies, emergency rooms, and other organizations, there is a risk to the privacy of individual information. An interaction known as anonymization modifies the diagram to maintain privacy [5-7]. It is common practice to use information transmission strategies like the level and vertical section of the information available from multiple sources for preservation and conservation. When the properties or segments are distributed over many locations in vertical appropriation, level apportioning maintains the sections or characteristics at various locations. Tables 5.1(a) and 5.1(b) outline the scenario when anonymization is applied to a dataset of medical services.

Table 5.1 (a): Before applying anonymization on a healthcare dataset

Zip	Age	Disease
247***	2*	heart
247***	2*	skin
247***	2*	eye
247***	>40	heart
247***	>40	eye

Table 5.1 (b): After applying anonymization on a healthcare dataset

Zip	Age	Disease
247667	29	heart
247589	23	skin
247888	27	eye
247456	58	heart
247874	63	eye

5.2 PROPOSED WORK

The proposed study predicts hospital readmission rates for different individuals using their HbA1c results and previous responses to diabetes diseases. The work includes a framework for privacy preservation in medical data.

Machine learning healthcare applications extract useful information from data. However, privacy should be preserved when mining sensitive data. For example, medical research represents an important application that must both extract useful information and protect patient privacy.

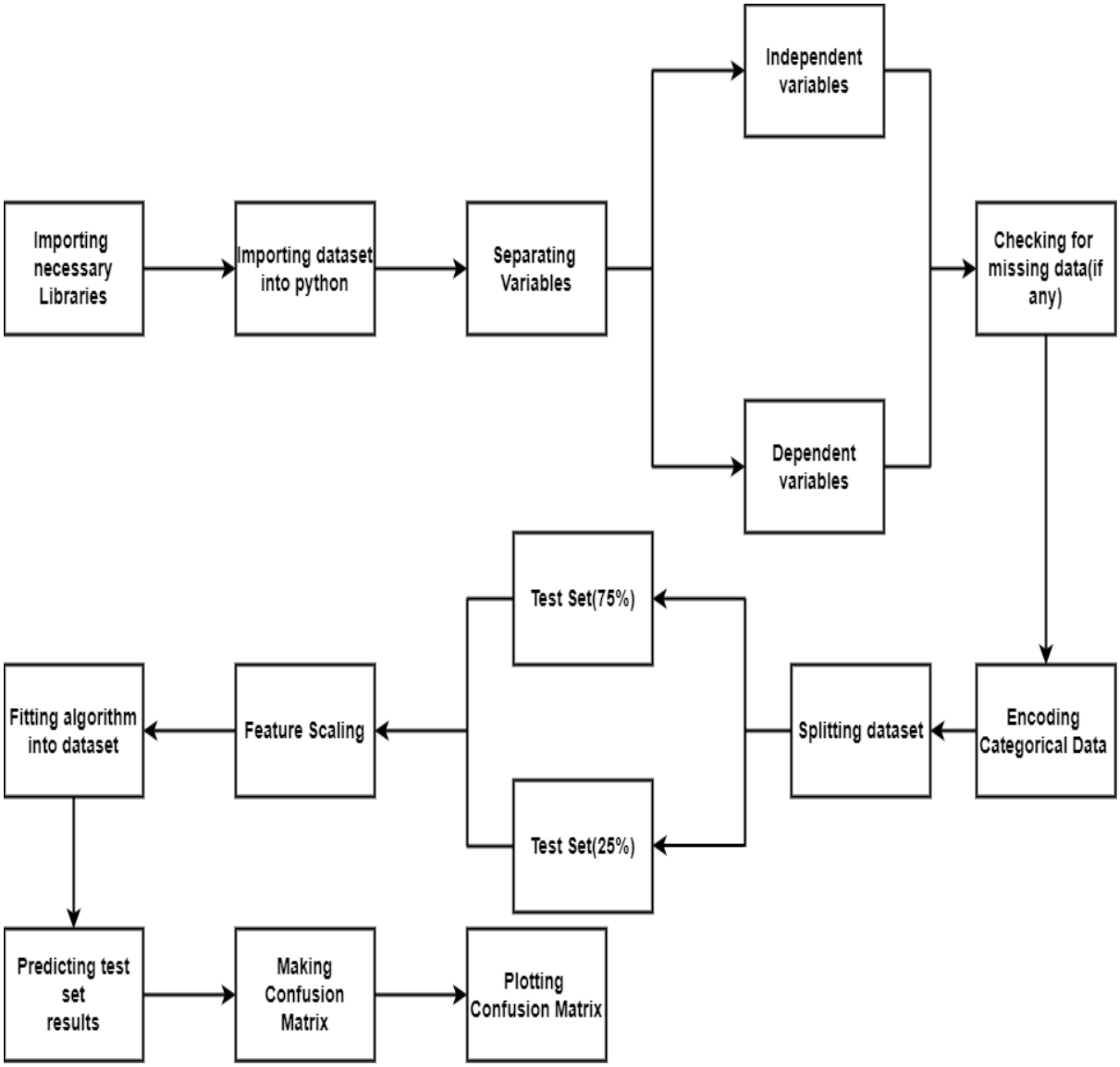


Figure 5.6: Flowchart for the machine learning process

5.2.1 EVALUATION OF MODEL PERFORMANCE

Evaluation of the proposed ML model shows the findings in the binary classification problem, letting TP, TN, FP, and FN have true positive, true negative, false positive, and false negative class labels. These metrics can be computed as the following:

$$\text{Sensitivity} = TP/(TP + F\,N) \tag{5.6}$$

$$\text{Precision} = TP/(TP + F\,P) \tag{5.7}$$

$$\text{Specificity} = TP/(FP+TN) \tag{5.8}$$

$$F1-\text{Score} = TP/\{TP+1/2(FP+FN)\} \tag{5.9}$$

$$\text{Accuracy} = TP+TN/\{(TP+FP)+(TN+FN)\} \tag{5.10}$$

The statistical evaluation of the classification performance of the ML model is formulated and plotted the relationship between the observed values and predicted values by using the confusion matrix. The performance of the model is obtained by using the four categories of the confusion matrix. We have obtained the classification Information from the confusion matrix. Various metrics (measures) can be used to evaluate a classifier's accuracy. True Positive (TP) and True Negative (TN) classifications are correct predictions. If an effect is wrongly forecast to be positive, if it is negative, it is called False Positive (FP). A result incorrectly predicted as negative when positive is called False Negative (FN). Euclidean distance, Manhattan distance, Minkowski, and Chebyshev metrics for different kNN classifiers are used in the thesis.
Section 5.1.1 discusses the six classification techniques used in the work in detail. It evaluates model performance using a diverse range of six classifications discussed in Section 5.1.1. We run six algorithms on the selected dataset. These include the Decision Tree Classifier, K-Nearest Neighbors Classifier (KNN), Logistic Regression, Naive Bayes Classifier, Random Forest Classifier, and Support Vector Machine (SVM). The rows selected for the training and test sets are 75% and 25%, respectively. We consider 101767 rows and nine columns for the training and testing set.

Figure 5.6 presents the flowchart of the proposed work. The three cases are as follows: the first is readmission in less than 30 days, and the second is greater than 30 days. The third case is the patient's no readmission. We run six algorithms on the selected dataset. Numerous algorithms include the Decision Tree Classifier, K-Nearest Neighbors Classifier (KNN), Logistic Regression, Naive Bayes Classifier, Random Forest Classifier, and Support Vector Machine (SVM). The rows selected for the Training Set and Test Set are 75% and 25%, respectively. We consider 101767 rows and nine columns for the training and testing set. The rows selected for the Training Set and test set are 76325 and 25442, respectively, in random order. Further, we group age features in [0, 10), [10, 20), ..., [90,100). The three prediction cases are as follows: Readmission in less than 30 days is (1, 0, 0) depending upon whether readmission in less than 30 days is in a particular row. Similarly, readmission in greater than 30 days is (0, 1, 0) depending on whether readmission in greater than 30 days is in a particular row. No Readmission is (0, 0, 1) depending on whether No Readmission is in a particular row. We denote 'rg' as patient readmission in greater than 30 days, 'rl' as patient readmission in less than 30 days, and 'rn' as no readmission of the patient. Based on the above criteria, the three

equations are as follows:

$$rg = 1 - (rn + rl) \quad (5.11)$$

$$rl = 1 - (rn + rg) \quad (5.12)$$

$$rn = 1 - (rg + rl) \quad (5.13)$$

Most of the time, 70 to 80 percent of the input data for supervised learning is used to train the model. The remaining 20–30% of the data are utilised as test data to confirm the model's effectiveness. The holdout method partitions the input data into training and test data and holds back some input data to validate the trained model. Cross-validation, a repeated holdout, is a particular variant of the holdout method. Cross-validation employs a non-replacement sampling strategy. There is a limit to the number of samples drawn using the cross-validation method. On the other hand, bootstrapping employs a sampling with replacement approach and allows for unlimited samples. We now present the metrics for the evaluation of the classification approach.

5.2.2 METRICS FOR CLASSIFICATION

The accuracy of a classifier can be assessed using a variety of metrics (measures). The number of accurate categories divided by the total number of categories determines how accurate each categorization model is. We are interested in the positive class, while the negative classes are those we don't care about (which may be combined into one negative class). Table 5.2 shows the four possible outcomes of one prediction on the test set. True Positive (TP) and True Negative (TN) classifications are correct predictions. If an effect is wrongly forecast to be positive, if it is negative, a False Positive is called (FP). False Negative is when a result is wrongly projected as negative when it is positive (FN). Table 5.2 displays the truth in the rows and the algorithm's decisions in the columns. The goal is to reduce the number of false positives and false negatives in the outcome. The term "confusion matrix" in Table 5.2 shows the prediction categories.

Table 5.2: Confusion matrix

Actual class (observation)	Hypothesized class (prediction)	
	Positive class	Negative class
Positive	True Positive	False Negative
Negative	False Positive	True Negative

Let the data D = {x, y}. Let h(x,w) be a binary classification, and the hypothesized class is $\hat{y} = h(x, w)$. In the confusion matrix, the pair of labels (y, $\hat{y}$) indicates each observation coordinate; the first mark indicates the matrix's row, and the second marks the column.

The correct number of categories divided by their total number is the success rate.

$$Success\ rate = \frac{TP+TN}{TP+TN+FP+FN} \qquad (5.14)$$

A feature can provide information that resembles what is jointly offered by one or more other features. The relevance of features should be measured by how much information a feature offers. Distance-based similarity measures are the most common distance measure.

Let us use d_{ij} to depict a *distance metric* or *dissimilarity measure* between patterns *i* and *j*. For patterns *i* and *j*, we have the vector of *n* measurements (, and (, respectively.

Euclidean distance: Euclidean distance $d_{ij}$is the most prevalent distance measurement. d_{ij} is the distance between two patterns *i* and *j*.

$$the\ the\ d_{il} = \sqrt{\left(x_1^{(i)} - x_1^{(j)}\right)^2 + \left(x_2^{(i)} - x_2^{(j)}\right)^2 + \cdots + \left(x_n^{(i)} - x_n^{(j)}\right)^2} \qquad (5.15)$$

Manhattan distance: is the absolute difference and is given as:

$$d_{ij} = \sum_{k=1}^{n} \left|x_k^{(i)} - x_k^{(j)}\right| \qquad (5.16)$$

Minkowski metric: The Minkowski metric (sometimes called alternative to the L_p norm) is a generic type of metric for n-dimensional patterns.

$$L_P\left(x^{(i)}, x^{(j)}\right) = \left(\sum_{k=1}^{n} \left|x_k^{(i)} - x_k^{(j)}\right|^P\right)^{1/P} \qquad (5.17)$$

Where $p \geq 1$ is a selectable parameter. Euclidean distance results when p equals 2 and settings p=1 allow the distance from Manhattan.

$$L_P\left(x^{(i)}, x^{(j)}\right) = \left\|x^{(i)} - x^{(j)}\right\|_P = \left(\sum_{k=1}^{n} \left|x_k^{(i)} - x_k^{(j)}\right|^P\right)^{1/P} \qquad (5.18)$$

(Minkowski norm)

$$\left\|x^{(i)} - x^{(j)}\right\|_2 = \left[\left(x_1^{(i)} - x_1^{(j)}\right)^2 + \left(x_2^{(i)} - x_2^{(j)}\right)^2 + \cdots + \left(x_n^{(i)} - x_n^{(j)}\right)^2\right]^{1/2} \qquad (5.19)$$

(Euclidean norm)

$$\left\|x^{(i)} - x^{(j}\right\|_l = \sum_{k=1}^{n} \left|x_{jk}^{(i)} - x_k^{(j)}\right| \qquad (5.20)$$

(Manhattan norm)

This section presents metrics for machine learning. Measurements from machine learning

statistics have proven beneficial in Null Hypothesis; ANOVA, Chi-Square Test; Confidence Interval; P-values; t-test; t/Distribution; Z-scores, and other metrics.

5.2.3 PRIVACY PRESERVATION IN HEALTHCARE AND CYBER-PHYSICAL SYSTEMS

After extraction, the features are stored in a separate database, to which the data preservation techniques are applied. The concept of privacy preservation began with disclosing the data to the users in the first stage. Each stage raises a new question: whether we want to hide the dataset's attributes or the rules to maintain privacy. Figure 5.7 shows the comparison with and without privacy preservation.

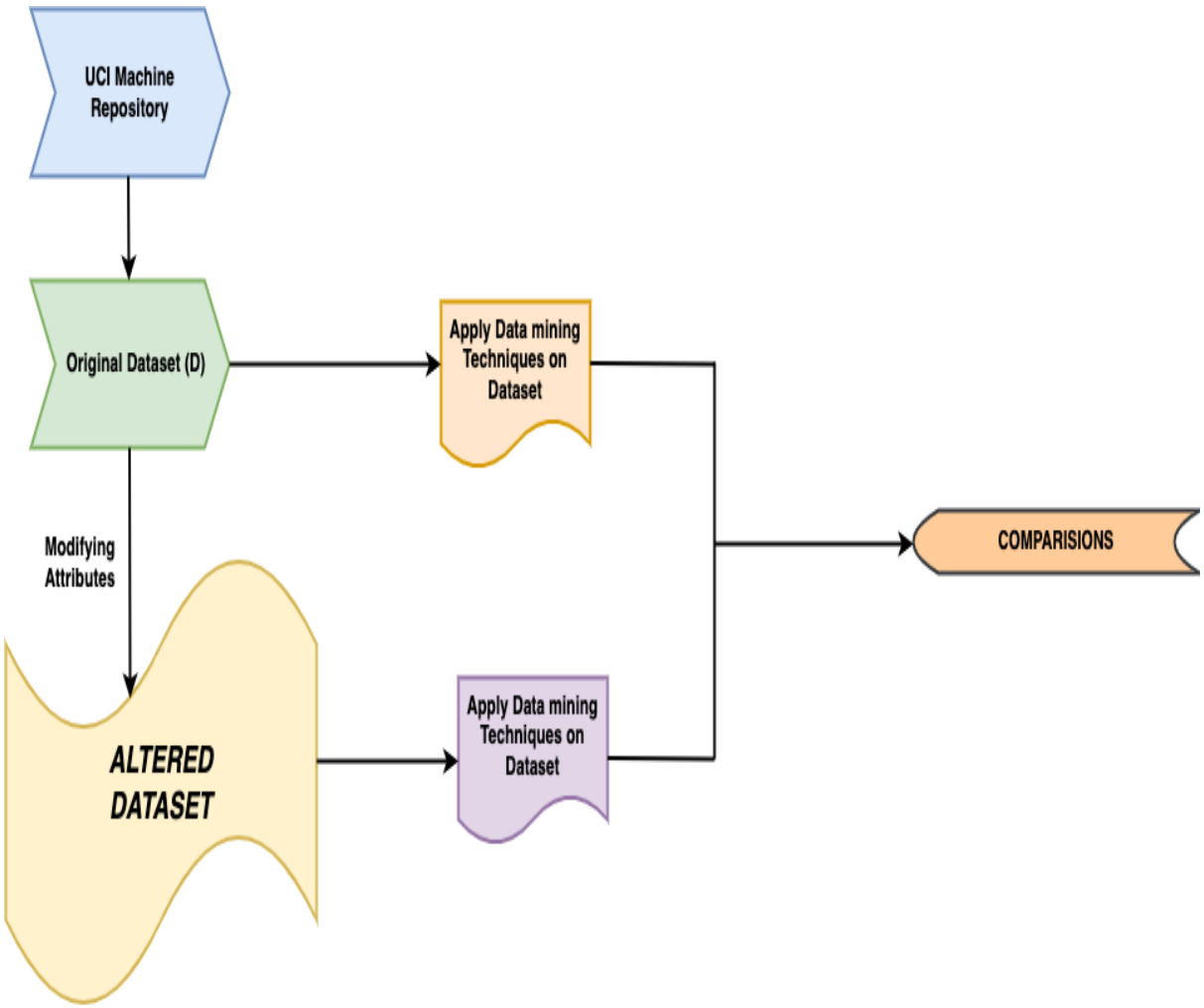


Figure 5.7: Comparison with and without privacy preservation

The hidden attribute actions include data modifications, randomization, swapping, aggregation and suppression. Anonymization is the process of data modification before the data is analysed. Pseudonymisation and anonymisation are popular privacy-enhancing techniques. Pseudonymisation changes the data value using randomization or encryption techniques. Anonymisation removes direct and indirect personal identifiers. Privacy laws differ from country to country. Privacy laws apply to pseudonymised data as indirect identifiers, combined with other identifiers, can reveal the person's identity. Randomization is the technique of adding noise to information through a probability distribution. The data is disseminated over several locations using horizontal and vertical distribution. Data distribution is horizontal when dispersed over multiple sites with the same attributes. The distribution is vertical if data is distributed over several sites with different attributes. Figure 5.8 presents the framework of privacy preservation.

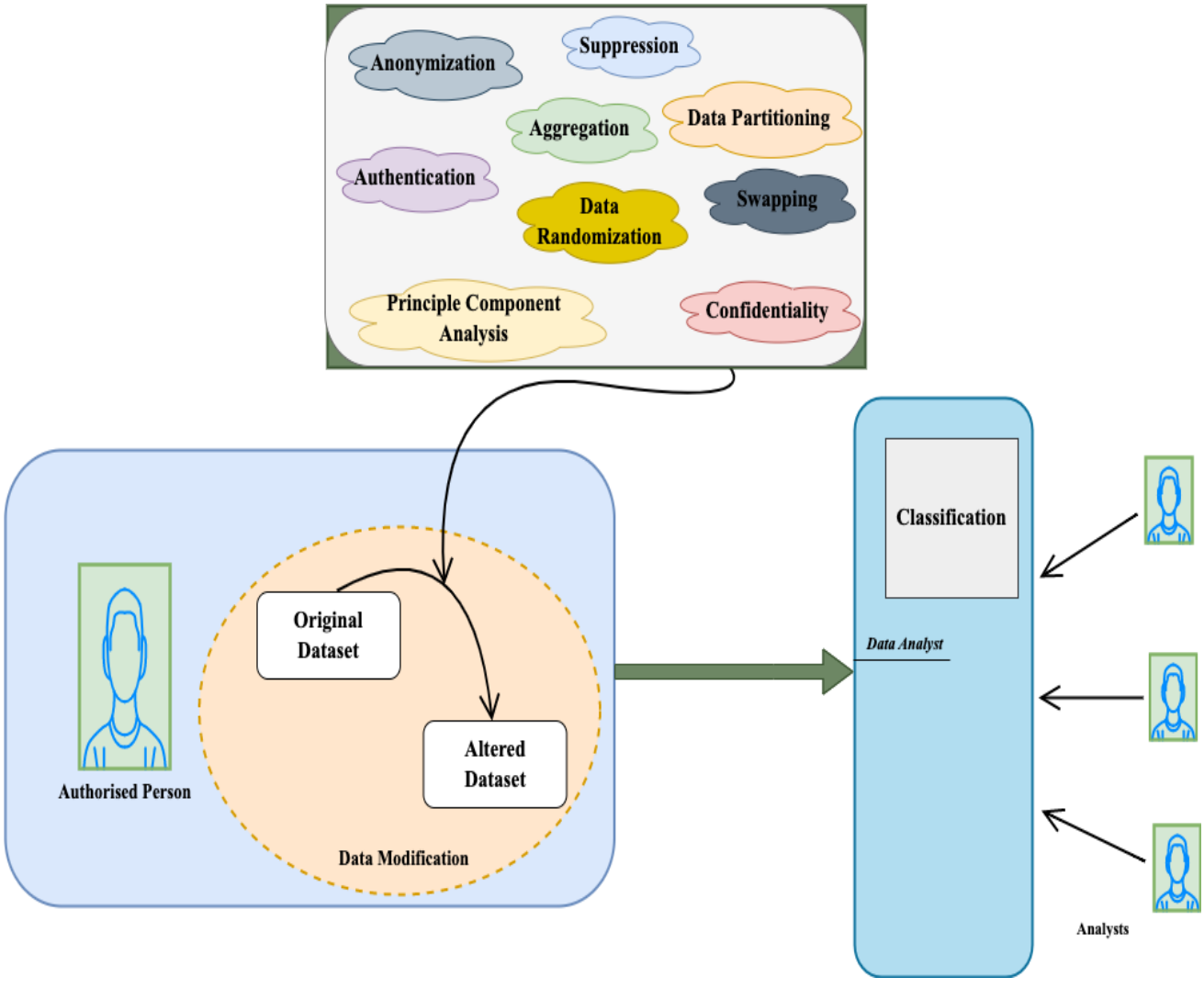


Figure 5.8: Framework of privacy preservation

Differential privacy promises to make it nearly impossible for anyone to identify private information about an individual from a dataset. This is particularly vital as large datasets are available today, including quasi-identifying information such as zip code, gender, birthdate, etc. Noise-adding Mechanisms in differential privacy: Laplace, Exponential, and Gaussian Mechanisms. The Laplace distribution adds noise from a symmetric continuous distribution to a true answer. Laplace could be used for numeric queries only, while exponential can be applied for numerical or categorical queries. The Gaussian mechanism is an alternative to the Laplace mechanism, which adds Gaussian noise instead of Laplacian noise.

5.2.4 KG ARCHITECTURE AND BUILDING OF HEALTHCARE

Semi-endlessly structured and unstructured datasets can be combined to create knowledge graphs for medical services. The triples subject, predicate, and article are the three elements of the triples used to build knowledge graphs.

KG ARCHITECTURE

Figure 5.9 shows the engineering for the knowledge graph. This section explains the engineering process before moving on to the basic steps in developing a clinical knowledge graph.

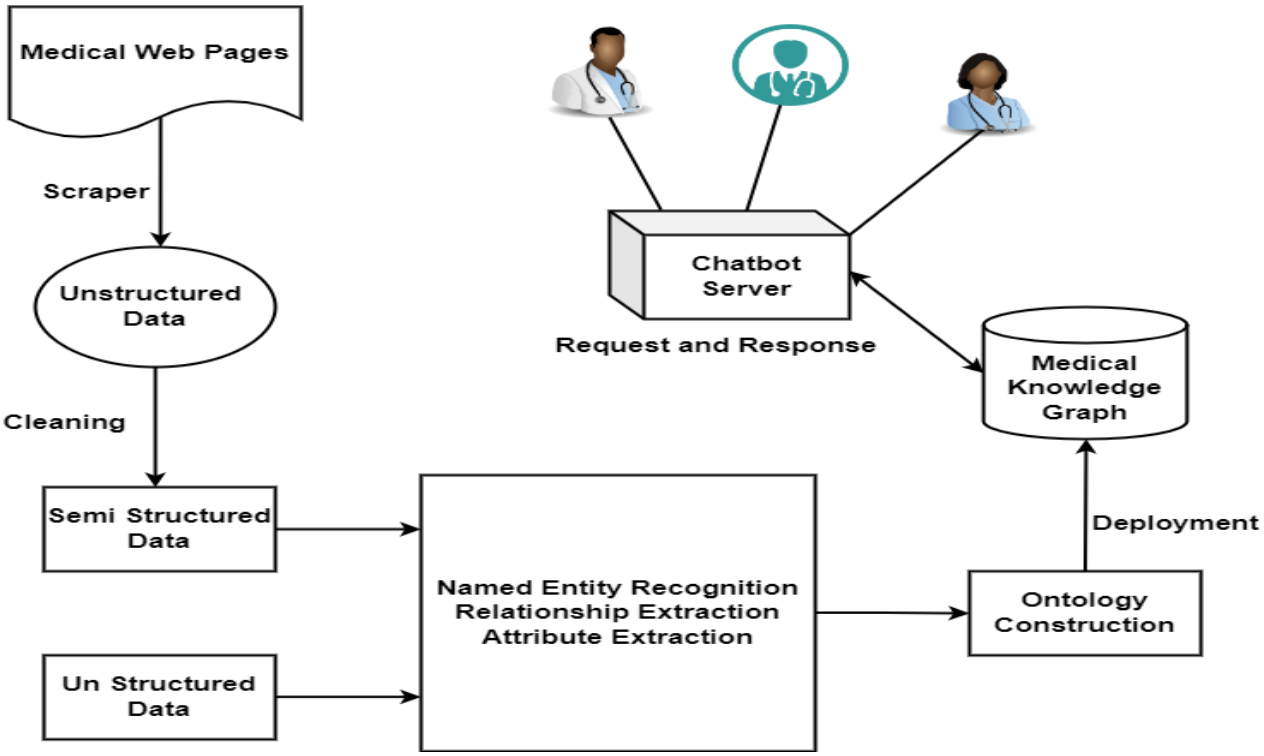


Figure 5.9: Medical KG Architecture

DATA

The organised, semi-structured, and unstructured data set is gathered from diverse sources. Information gathering in the therapeutic environment should be possible with the help of a web crawler or bot. Web crawlers use computations to extract information about medical services by connecting links across site pages.

PREPROCESSING OF DATA

Because it is now heterogeneous, the acquired data should be transformed into a sensible arrangement for building an input-to-knowledge graph.

ONTOLOGY MODELLING

The clinical information diagram gains semantic depth via ontology demonstration. Distinguishing classes and subclasses for the clinical space is one of the most crucial stages in ontology demonstration.

EXTRACTION

Three types of information extraction exist: substance, connection, and characteristic extraction [8]. Subject, predicate, and article are the three triple elements used to build knowledge graphs. The metaphysical extraction of subject, predicate, and item from handled information is the next step in creating an "information diagram." A named element is distinguished from the collected data using named entity recognition (NER). Three categories of NER techniques are described [5]. Rules-based, learning-based, and neural network-based techniques are the three main categories. Moreover, the standard-based method calls for appropriate clinical area information and clinical space-related highlights. Other classifications for learning-based systems include unsupported, essentially endlessly administered categories. Administered strategies need information that has been remarked on. Unaided techniques don't require an explanation. Clinical space-related ontologies are unnecessary for brain network-based techniques, although they need a large dataset. A key step in providing semantic data to clinical "knowledge graphs" is the development of information diagrams connection extraction.

EVALUATION

The two key techniques are manual evaluation by clinical professionals and question-and-answer-based assessment projects. Manual evaluation should be doable by randomly reviewing the knowledge graph and using chart examining processes to collect tests. Clinical

space experts can evaluate these cases. The question-answer-based framework can be implemented manually or automatically. The results may be compared to the space master's, and clinical knowledge graphs can accommodate assessment score measures.

The clinical knowledge graph uses three types of measurements: Valid Positive, Authentic Negative, and Misleading Negative. A positive event that the framework predicts as certain is genuinely positive, while a negative event that the framework predicts as definite is deceptively positive. A good event the framework expects to be negative is a deceptive negative.

"Precision= TP / (TP + FP) (5.21)

recall= TP / (TP + FN) (5.22)

F-measure= 2 × (precision × recall) / precision+recall) (5.23)

F measure is the harmonic mean of precision and recall" (5.24)

STORAGE

RDF-based plans and chart-based plans are used to complete knowledge graph storage. RDF uses subject, predicate, and article for KG. In addition, RDF leverages IRI/URI to create information diagrams. A well-known open-source tool called Neo4j provides chart-based designs with local graphic hoarding. Capability frameworks for knowledge graphs should be flexible and be able to manage massive amounts of clinical data for questioning and display. The clinical knowledge graphs use SPARKQL, a common inquiry language [9]. Clinical information diagrams are imagined using Neo4j representation tools. The developers present a utility-safeguarding anonymization scheme for EHR [10]. Knowledge graph storage is completed using RDF-based plans and chart-based plans. For KG, RDF employs subject, predicate, and article. In addition, RDF leverages IRI/URI to create information diagrams. A well-known open-source tool called Neo4j provides chart-based designs with local graphic hoarding. Capability frameworks for knowledge graphs should be flexible and be able to manage massive amounts of clinical data for questioning and display. The clinical knowledge graphs use SPARKQL, a common inquiry language [9]. Clinical information diagrams are imagined using Neo4j representation tools. Developers have presented a utility-safeguarding anonymization scheme for EHR [10].

MEDICAL KG

Figure 5.10 addresses the clinical knowledge graph toy model for a triple and contains

information regarding Tom Smith. The information diagram shows that Tom Smith also has coronary disease. The knowledge graph can determine whether Tom Smith is a patient.

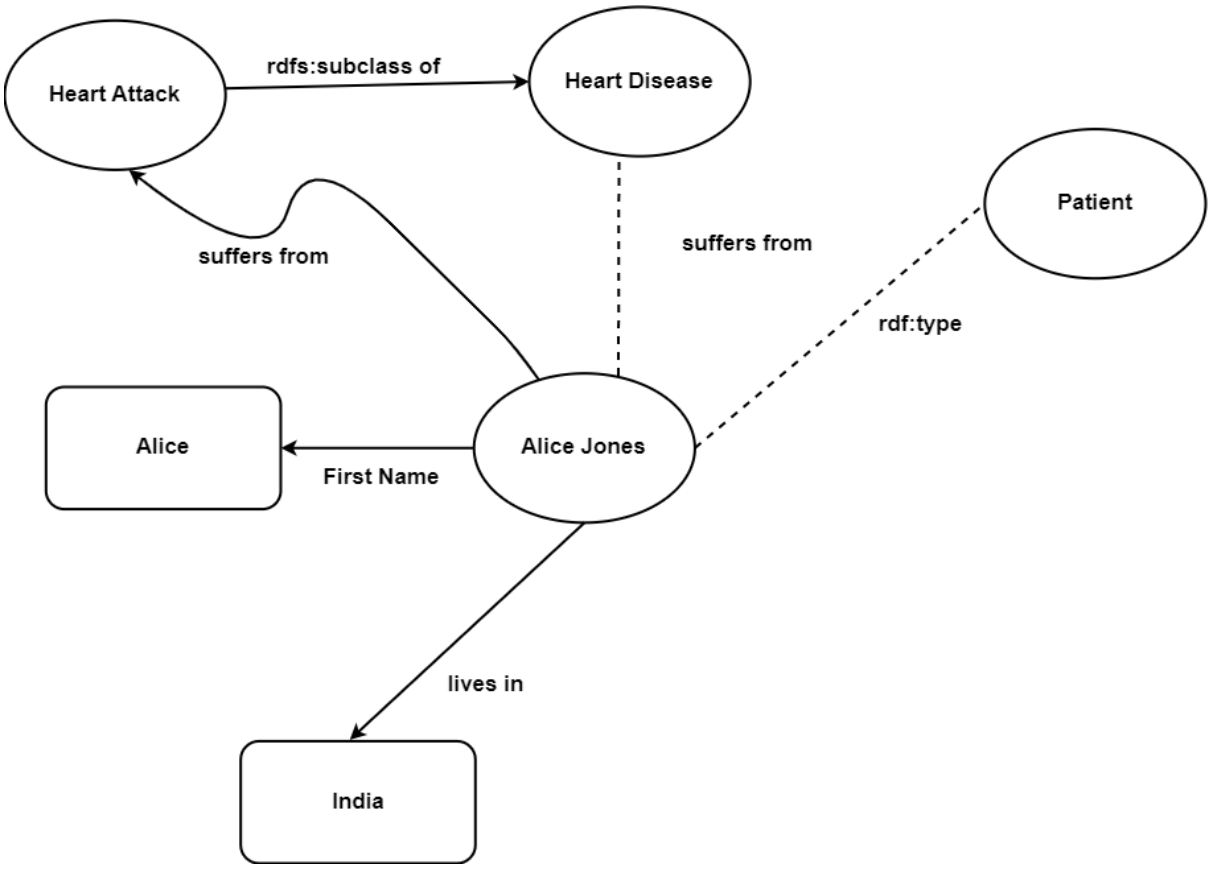


Figure 5.10: Toy Example of Medical KG

PRIVACY PRESERVATION IN MEDICAL GRAPH

Various groups' clinical knowledge graphs are their own and shouldn't be shared with other groups to maintain privacy. The protection protecting the KG notion may be used for clinical knowledge diagrams from various groups.

ORIGINAL GRAPH AND AFTER ANONYMIZATION

The customer names are changed since they are sensitive. Also, many preservation and conservation processes, such as supposing some qualities like age, have been completed. Also, a few links are clumsily made to ensure the security of the knowledge graph for the medical services area. The mathematical feature can be combined with other qualities or transferred into a larger category. As shown in Figure 5.11, the mathematical characteristic of age may be summarized. As seen in the scientific classification tree in Figure 5.11, age can also be turned into range credits. A scientific classification tree can also represent simple qualities

(allude to Figure 5.11). The original and anonymous diagram is shown in Figures 5.12 (a) and 5.12 (b).

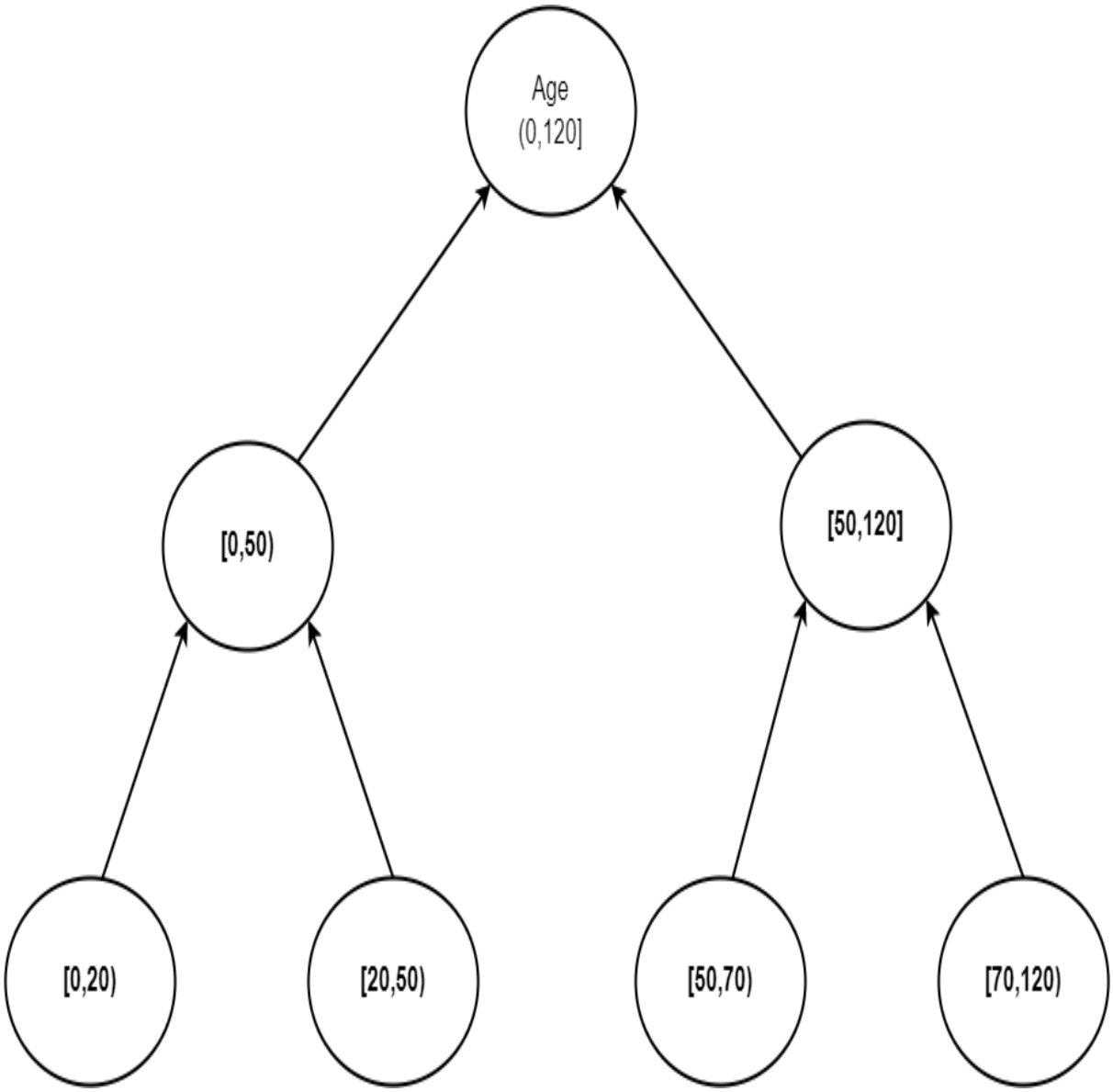


Figure 5.11: Age Taxonomy Tree

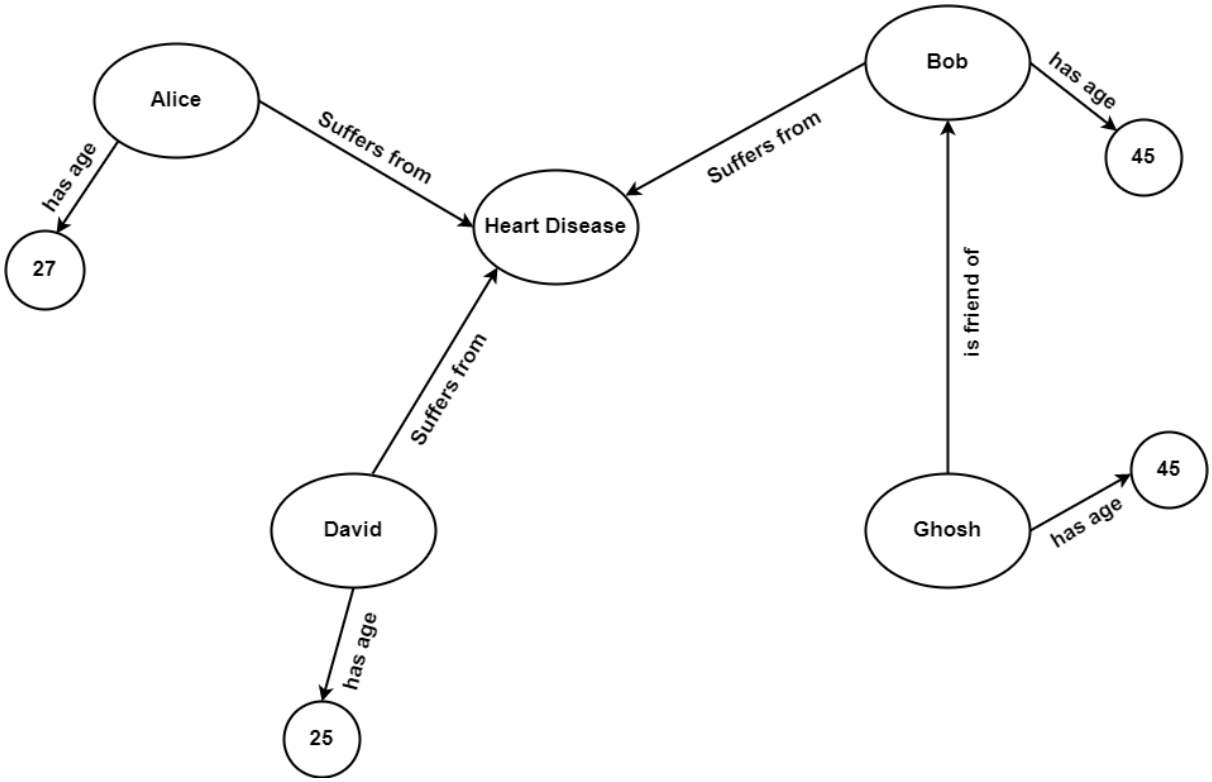


Figure 5.12 (a): Original Graph

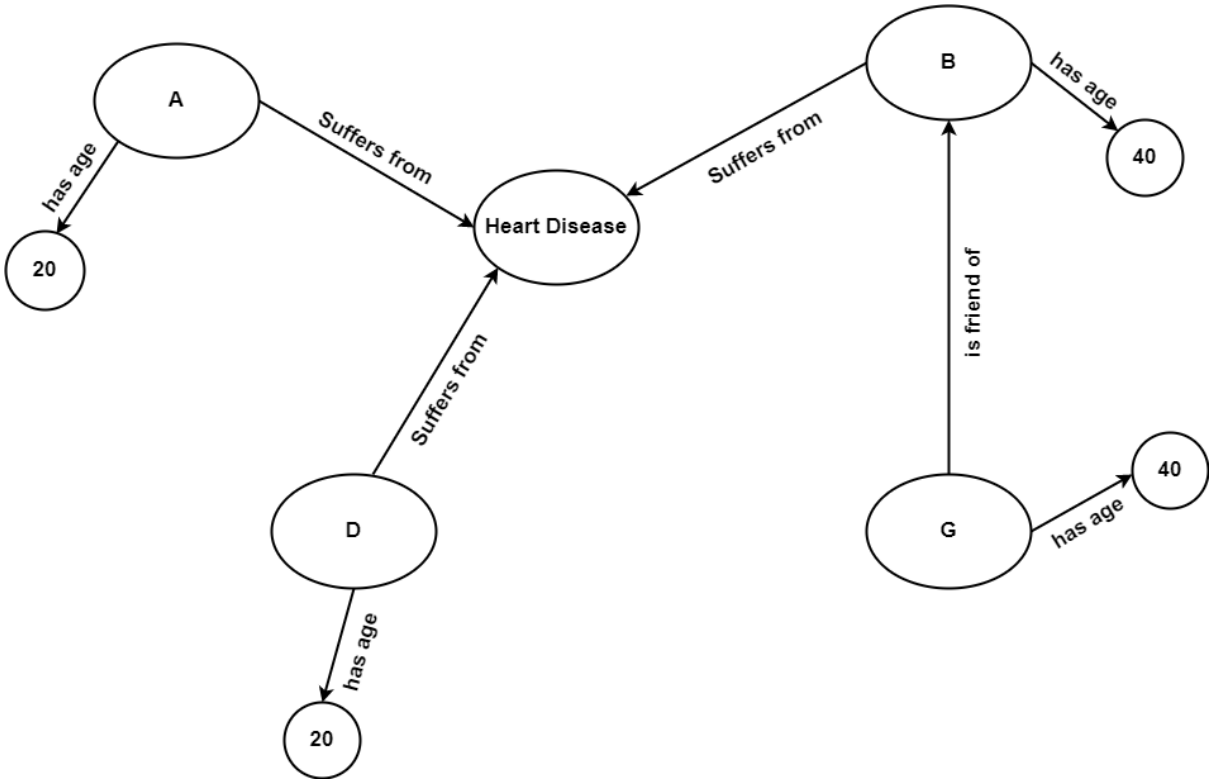


Figure 5.12 (b): Anonymized Graph

5.3 OBSERVATIONS AND RESULTS

Our results section explained the simulation part in total. Firstly, we selected the diabetic dataset. There are 101768 rows and 50 columns in the .csv file format in the diabetes dataset. In this diabetic dataset, we have chosen nine features from the dataset. We simulate six machine learning classification techniques: Decision Tree Classifier, Random Forest Classifier, Support Vector Machine (SVM) Classifier, K-nearest neighbors (KNN) Classifier, Logistic Regression, and Naive Bayes Classifier. These six techniques work on nine selected features of the diabetic dataset. The selected features are Age, Number of Diagnoses, Number of Emergency Visits, Number of inpatients, Number of Lab Procedures, Number of Medications, Number of Outpatient Visits, Number of Procedures, and Time in Hospital. The simulation work is done in the Python environment. The six techniques observed the independent features and predicted the patient's readmission rate with the help of three cases. The three cases are as follows: the first is readmission in less than 30 days, and the second is greater than 30 days. The third case is the patient's no readmission. The features selected in all techniques are our independent features, which predicted the dependent feature, the Patient Readmission rate. A training set and a test set were created from the dataset. The test Set size is 25%, whereas the Training Set size is 75%. Moreover, the training set is chosen using a Python random sampling feature. All the independent variables are represented by a matrix "X", and the dependent variable, i.e., the readmission of the patients, is denoted by a column vector, "Y".

HbA1c Test: The work talks about the HbA1c Test (glycated haemoglobin), a test of haemoglobin (Hb). The HbA1c test gauges blood glucose levels and reveals a person's susceptibility to diabetes. If the level of Hemoglobin A1c in the blood is more than 6%, it is considered abnormal. The dataset consists of HbA1c measurement records of 70,000 patients. We had a significant overview of how many patients were prone to diabetes, how they were medically treated, the major tests performed on them, and how many were readmitted in that period.

We used different techniques that could best fit our records and predict future readmissions after removing insignificant features from the dataset. The results show that the readmission was analyzed for less than and greater than 30 days.

Out of the six machine learning classification techniques applied, the Support Vector Machine

(SVM) outperformed all the other techniques and best fit the data. The prediction accuracy is 88.81% in case of patient readmission in less than 30 days, 65.5% in case of inpatient readmission in greater than 30 days, and 62.37% in case of no patient readmission. The results from the other five techniques are close to those obtained from SVM.

Privacy preservation in medical data within Wireless Sensor Networks (WSNs) is a critical concern due to the sensitive nature of medical information. Here are some key points and techniques used to ensure data privacy in this context:

1. **Anonymization Techniques**:
    - **K-Anonymity**: Anonymization is the process of modifying data before it is given for data analytics so that de-identification is not possible. If an attempt is made to de-identify by mapping the anonymized data with external data sources, K indistinguishable records will result.
    - **L-Diversity**: This technique extends k-anonymity by ensuring sensitive attributes have at least l well-represented values. Another method called L diversity has been proposed to address homogeneity attacks. As per L diversity, there must be L well-represented values for the sensitive attribute (disease) in each equivalence class. Implementing L diversity is not always possible because of the variety of data.
    - **T-Closeness**: Ensuring that the distribution of a sensitive attribute in any equivalence class is close to the distribution of the attribute in the overall table. Another improvement to L diversity is the T closeness measure, where an equivalence class is considered to have 'T closeness' if the distance between the distributions of sensitive attributes in the class is no more than a threshold and all equivalence classes have T closeness. T closeness can be calculated on every attribute concerning sensitive attributes.
    - **Randomization Technique:** Randomization adds noise to the data, generally done by the probability distribution. It is applied in surveys, sentiment analysis, etc. Randomization does not need knowledge of other records in the data. It can be applied during data collection and preprocessing time. There is no anonymization overhead in randomization.
    - **Data distribution technique:** The data is distributed across many sites. Distribution of the data can be done in two ways:
    - **Horizontal distribution of data:** When data is distributed across many sites with the same attributes, it is said to be horizontal. Horizontal distribution of data can be

applied only when some aggregate functions or operations are to be applied to the data without actually sharing it.

- **Vertical distribution of data:** When person-specific information is distributed across different sites under the custodianship of other organizations, the distribution is called vertical distribution.

5.3.1 RESULTS & ANALYSIS

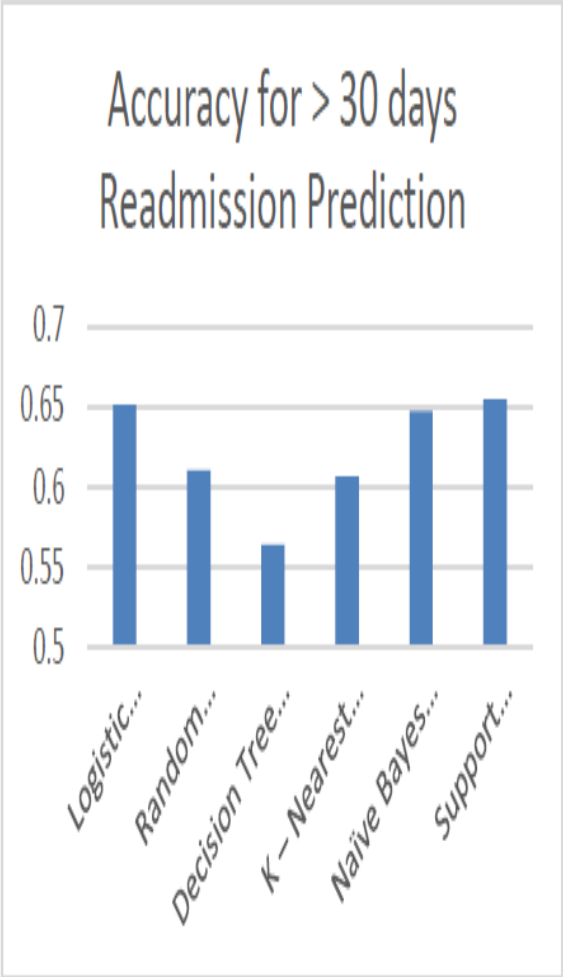


Figure 5.13 (a): Prediction for greater than 30 days

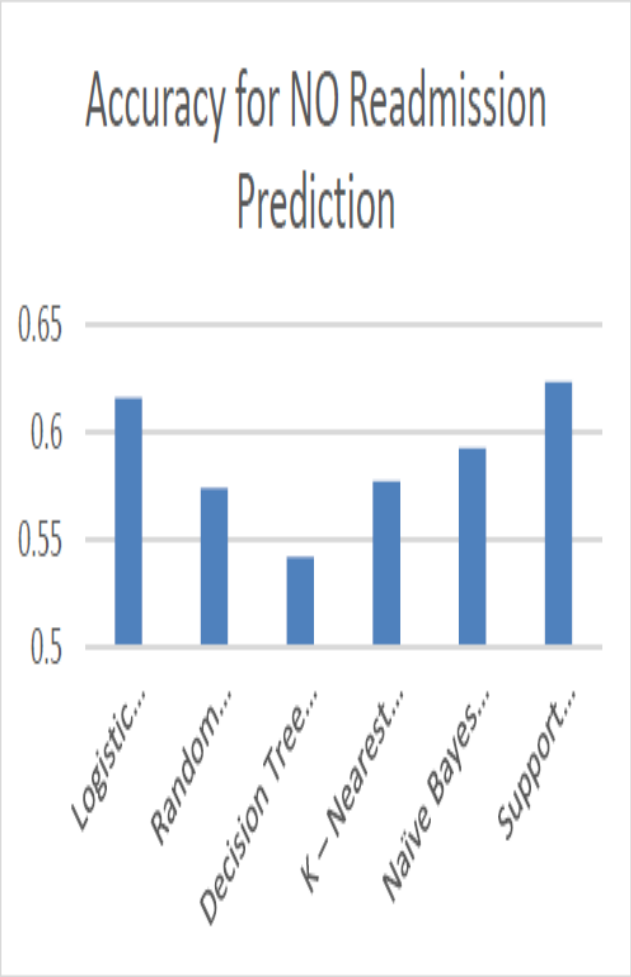


Figure 5.13 (b): Prediction for no readmission

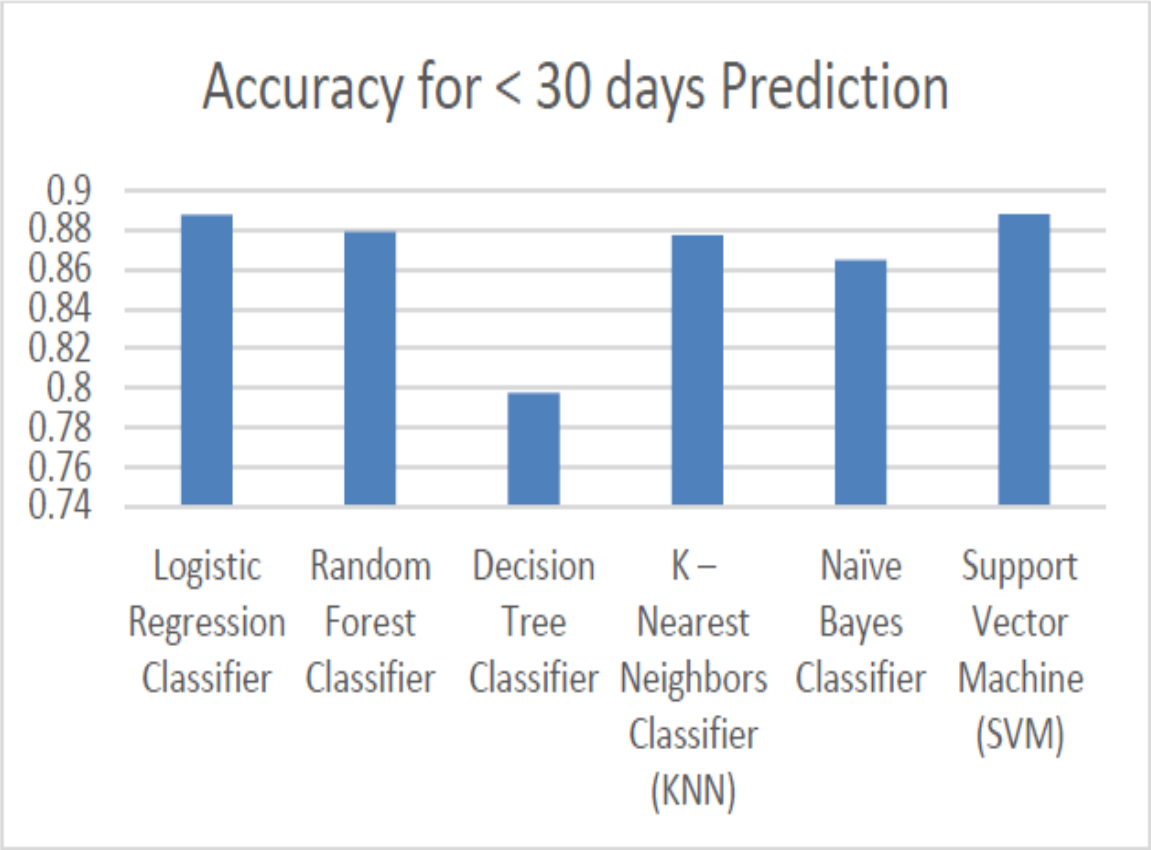


Figure 5.13(c): Prediction for readmission in less than 30 days

Figures 5.13 (a), 5.13 (b), and 5.13 (c) present six machine-learning classification algorithms. Figures show that Support Vector Machine (SVM) performs better than other techniques. Table 5.3 presents four distance metrics for the three cases.

Table 5.3: Distance Metrics Comparison

	Readmission in less than 30 days (Accuracy %)	Readmission in greater than 30 days (Accuracy %)	No Readmission (Accuracy%)
Euclidean	87.72	60.69	57.74
Manhattan	87.67	60.62	57.64
Minkowski	87.72	60.69	57.74
Chebyshev	87.70	60.65	57.68

The power parameter differs for different kNN classifier schemes for distance metrics comparison. The scheme with a value of p equals two for Minkowski is equivalent to Euclidean distance.

5.3.2 MACHINE LEARNING IN HEALTHCARE APPLICATIONS

The dataset taken contains 101767 rows and 50 columns. The features chosen for analysis are Age, Number of Diagnoses, Number of Emergency Visits, Number of inpatients, Number of Lab Procedures, Number of Medications, Number of Outpatient Visits, Number of Procedures, and Time in Hospital. This dataset is available online at http://dx.doi.org/10.1155/2014/781670.

Six algorithms, Decision Tree Classifier, K-Nearest Neighbors Classifier (KNN), Logistic Regression, Naive Bayes Classifier, Random Forest Classifier, and Support Vector Machine (SVM), were used in the work. The rows selected for the Training Set and Test Set are 75% and 25%, respectively. Table 5.4 presents the list of features and their descriptions in the initial dataset.

Table 5.4: List of features and their descriptions in the initial dataset

id	***age***	***time_in_hospital***	***num_lab_procedures***	***num_procedures***	***num_medications***
0	*[0-10)*	*1*	*41*	*0*	*1*
1	*[10-20)*	*3*	*59*	*0*	*18*
2	*[20-30)*	*2*	*11*	*5*	*13*
...	*...*	*...*	*...*	*...*	*...*
101764	*[80-90)*	*10*	*45*	*2*	*21*
101765	*[70-80)*	*6*	*13*	*3*	*3*

Out of the six machine learning classification algorithms applied, Support Vector Machine (SVM) outperformed and best fitted the data and gave an accuracy of 88.81% in case of patient readmission in less than 30 days, 65.5% in case of patient readmission in greater than 30 days and 62.37% in case of no patient readmission. The results from the other five algorithms, too, were appreciable and were close to what the support vector machine algorithm predicted.

After extraction, the features are stored in a separate database to which the data preservation techniques are applied. The concept of privacy preservation began with disclosing the data to the users in the first stage. Each stage raises a new question: whether we want to hide the dataset's attributes or the rules to maintain privacy. Table 5.5 compares six machine learning algorithms.

Table 5.5: Comparison of Machine Learning Techniques

Sr. No.	***Techniques***	***Readmission>30 Days***	***No Readmission***	***Readmission<30Days***
1.	*Logistic Regression*	*0.6514*	*0.6165*	*0.8877*
2.	*Random Forest*	*0.6102*	*0.5744*	*0.879*
3.	*Decision Tree*	*0.5640*	*0.5422*	*0.7975*
4.	*K-Nearest Neighbor*	*0.6069*	*0.5774*	*0.8772*
5.	*Naïve Bayes*	*0.6477*	*0.5929*	*0.8649*
6.	*Support Vector Machine*	*0.655*	*0.6237*	*0.8881*

5.3.3 PRIVACY PRESERVATION RESULTS

This section implements privacy preservation using Python Libraries and Google Colab. Table 5.6 presents swapping *age* and *time_in_hospital* columns for the dataset.

Table 5.6: Swapping for data preservation: age and time_in_hospital columns

Id	***age***	***time_in_hospital***	***num_lab_procedures***	***num_procedures***	***num_medications***
0	*1*	*[0-10)*	*41*	*0*	*1*
1	*3*	*[10-20)*	*59*	*0*	*18*
2	*2*	*[20-30)*	*11*	*5*	*13*
3	*2*	*[30-40)*	*44*	*1*	*16*
4	*1*	*[40-50)*	*51*	*0*	*8*
…	*…*	*…*	*…*	*…*	*…*
101761	*3*	*[70-80)*	*51*	*0*	*16*
101762	*5*	*[80-90)*	*33*	*3*	*18*
101763	*1*	*[70-80)*	*53*	*0*	*9*
101764	*10*	*[80-90)*	*45*	*2*	*21*

Table 5.7: Data randomization

id	***age***	***time_in_hospital***	***num_lab_procedures***	***num_procedures***	***num_medications***
0	*1*	*[0-10)*	*59*	*1*	*40*
1	*3*	*[10-20)*	*78*	*45*	*40*
2	*2*	*[20-30)*	*54*	*35*	*91*
...	*...*	*...*	*...*	*...*	*...*
101763	*1*	*[70-80)*	*81*	*38*	*82*
101764	*10*	*[80-90)*	*12*	*95*	*62*
101765	*6*	*[70-80)*	*70*	*60*	*34*

The table describes the dataset using the Googlecolab environment. While applying the data suppression technique, substitute each numerical value with its first digit or a combination of asterisks. However, Semi-Suppression replaces the percentage of characters with an asterisk. Table 5.7 presents the data randomization.

Table 5.8: Suppression

id	*age*	*time_in_hospital*	*num_lab_procedures*	*num_procedures*	*num_medications*
0	*1*	*[0-10)*	*9**	*1*	*40*
1	*3*	*[10-20)*	*8**	*45*	*40*
2	*2*	*[20-30)*	*4**	*35*	*91*
3	*2*	*[30-40)*	*3**	*49*	*41*
4	*1*	*[40-50)*	*7**	*48*	*91*
...	*...*	*...*	*...*	*...*	*...*
101761	*3*	*[70-80)*	*8**	*20*	*17*
101762	*5*	*[80-90)*	*9**	*34*	*69*
101763	*1*	*[70-80)*	*1**	*38*	*82*
101764	*10*	*[80-90)*	*2**	*95*	*62*
101765	*6*	*[70-80)*	*0**	*60*	*34*

Table 5.9: Aggregation on num_medications (low, medium, high)

id	*age*	*time_in_hospital*	*num_lab_procedures*	*num_procedures*	*num_medications*
0	*1*	*[0-10)*	*9**	*1*	*MEDIUM*
1	*3*	*[10-20)*	*8**	*45*	*MEDIUM*
2	*2*	*[20-30)*	*4**	*35*	*HIGH*
3	*2*	*[30-40)*	*3**	*49*	*MEDIUM*
4	*1*	*[40-50)*	*7**	*48*	*HIGH*
...	*...*	*...*	*...*	*...*	*...*
101761	*3*	*[70-80)*	*8**	*20*	*LOW*
101762	*5*	*[80-90)*	*9**	*34*	*HIGH*
101763	*1*	*[70-80)*	*1**	*38*	*HIGH*
101764	*10*	*[80-90)*	*2**	*95*	*HIGH*
101765	*6*	*[70-80)*	*0**	*60*	*MEDIUM*

Tables 5.8 and 5.9 describe the suppression and the aggregation on num_medications (low, medium, high). Data distribution is horizontal when dispersed over multiple sites with the same attributes. The distribution is vertical if data is distributed over several sites with different attributes. The original data set consists of 50 features or columns and 101767 rows. We apply PCA and horizontal partition techniques to the dataset shown in Table 5.10. Each model receives samples from the original dataset in multiple classifier systems. Model 1 receives nine columns and [0 to 12720] rows in our proposed system. Similarly, model 2 receives nine columns and [12720, 25442) rows, Model 3 receives nine columns and [25442, 38163) rows, … and so on.

Table 5.10: Six classifiers with PCA applied to the dataset

Sr. No.	***Techniques***	***Readmission>30 Days***	***No Readmission***	***Readmission<30Days***
1.	*Logistic Regression*	*0.6514*	*0.6165*	*0.8877*
2.	*Random Forest*	*0.6102*	*0.5744*	*0.879*
3.	*Decision Tree*	*0.5640*	*0.5422*	*0.7975*
4.	*K-Nearest Neighbor*	*0.6069*	*0.5774*	*0.8772*
5.	*Naïve Bayes*	*0.6477*	*0.5929*	*0.8649*
6.	*Support Vector Machine*	*0.655*	*0.6237*	*0.8881*

Table 5.11: Differential privacy: Randomize all numerical values through Laplace distribution: Randomize

id	***age***	***time_in_hospital***	***num_lab_procedures***	***num_procedures***	***num_medications***
0	*5.000000*	*[0-10)*	*9**	*1*	*MEDIUM*
1	*5.000000*	*[10-20)*	*8**	*45*	*MEDIUM*
2	*17.602735*	*[20-30)*	*4**	*35*	*HIGH*
3	*19.111837*	*[30-40)*	*3**	*49*	*MEDIUM*
4	*5.000000*	*[40-50)*	*7**	*48*	*HIGH*
...	*...*	*...*	*...*	*...*	*...*
995	*15.756559*	*[70-80)*	*6**	*30*	*LOW*
996	*5.000000*	*[0-10)*	*5**	*85*	*MEDIUM*
997	*8.514236*	*[60-70)*	*5**	*34*	*HIGH*
998	*26.261881*	*[70-80)*	*1**	*81*	*HIGH*
999	*17.666898*	*[50-60)*	*9**	*87*	*HIGH*

Table 5.12: Differential privacy: Randomize categorical values through Exponential distribution:Randomize num_medications

id	*age*	*time_in_hospital*	*num_lab_procedures*	*num_procedures*	*num_medications*
0	*1*	*[0-10)*	*7**	*12*	*LOW*
1	*3*	*[10-20)*	*6**	*3*	*HIGH*
2	*2*	*[20-30)*	*6**	*77*	*HIGH*
3	*2*	*[30-40)*	*6**	*24*	*LOW*
4	*1*	*[40-50)*	*7**	*81*	*HIGH*
...	*...*	*...*	*...*	*...*	*...*
995	*13*	*[70-80)*	*8**	*80*	*LOW*
996	*1*	*[0-10)*	*1**	*88*	*MEDIUM*
997	*11*	*[60-70)*	*8**	*60*	*HIGH*
998	*12*	*[70-80)*	*6**	*18*	*HIGH*
999	*9*	*[50-60)*	*3**	*75*	*LOW*

Table 5.11 presents the differential privacy: Randomize all numerical values through Laplace distribution: Randomize, and table 5.12 describes the differential privacy: Randomize categorical values through Exponential distribution: Randomize num_medications. Logistic Regression on the dataset was performed without and with differential privacy, resulting in an accuracy of 88.96% and 88.97%, respectively.

5.4 Data Preservation Explanation

After extraction, the above features are stored as a separate database on which the data preservation techniques are applied. The concept of privacy preservation began with disclosing the data to the users in the first stage. Each stage raises a new question: whether we want to hide the dataset's attributes or the rules to maintain privacy. The hidden attribute actions include data modifications, randomization, swapping, aggregation and suppression. The alternate versions of the dataset help preserve data. The original dataset hides from the public to preserve the data using attribute hiding techniques.

Before data is analysed, anonymization is changing data [19] [20]. Randomization is adding noise to information through probability distribution [19]. The data is disseminated over several locations using horizontal and vertical distribution [19]. Data distribution is horizontal when dispersed over multiple sites with the same attributes. The distribution is vertical if data is dispersed among multiple sites with various properties [19] [21].

Table 5.13: List of features and their descriptions in the initial dataset

id	***age***	***time_in_hospital***	***num_lab_procedures***	***num_procedures***	***num_medications***
0	*[0-10)*	*1*	*41*	*0*	*1*
1	*[10-20)*	*3*	*59*	*0*	*18*
2	*[20-30)*	*2*	*11*	*5*	*13*
3	*[30-40)*	*2*	*44*	*1*	*16*
4	*[40-50)*	*1*	*51*	*0*	*8*
...	...	...	...	...	...
101761	*[70-80)*	*3*	*51*	*0*	*16*
101762	*[80-90)*	*5*	*33*	*3*	*18*
101763	*[70-80)*	*1*	*53*	*0*	*9*
101764	*[80-90)*	*10*	*45*	*2*	*21*
101765	*[70-80)*	*6*	*13*	*3*	*3*

Table 5.13 describes the dataset with the help of the googlecolab environment. We substitute each numerical value with its first digit or a combination of asterisks while applying the data suppression technique. Data Randomization is the process of making something random. Since we want the valid values hidden in the dataset, we substitute for some altered value for a particular column (between 1-100). Using data aggregation techniques, we try to group and replace the data with the group representative. For example, the number of a patient's medicines is displayed as low (less than 30), medium (30-60), or high (>60) instead of defining any number.

Table 5.14: After applying attribute hiding techniques with the help of *googlecolab* environment

```
X['num_lab_procedures ''] [0:5]
```

```
0     1*
1     9*
2     1*
3     4*
4     1*
Name: num_lab_procedures, dtype: object
```

```
X['num_procedures ''] [0:5]
```

```
0     67
1     33
2     14
3     31
4     63
Name: num_procedures, dtype: int64
```

```
X['num_medications ''] [0:5]
```

```
0     LOW
1     LOW
2     LOW
3     LOW
4     LOW
Name: num_medications, dtype: object
```

Table 5.14 presents the dataset after applying attribute-hiding techniques. Multiple classifier systems have demonstrated advantages over individual classifier systems. The ensemble systems rely on a pool of classifiers to help classify the new pattern instead of employing a single classifier approach.

Classification aims to assign a particular object in a predetermined category, sometimes named labels, described in its x attributes. We have got a collection of six different classifiers. The majority vote combines the binary output of six classifiers and the maximum number of votes for the ensemble output. The weighted majority voting model shows that some classifiers perform more effectively than others [22].

The original data set consists of 50 features or columns and 101767 rows. We apply PCA and horizontal partition techniques to the dataset. Each model receives samples from the original dataset in multiple classifier systems. Model 1 receives nine columns and [0 to 12720] rows in our proposed system. Similarly, model 2 receives nine columns and [12720, 25442) rows, Model 3 receives nine columns and [25442, 38163) rows; similarly, model 4 receives nine columns and [38163, 50884), Model 5 receives nine columns and [50884, 63605) rows, Model 6 receives nine columns and [63605, 76324), That is out of 101767 rows 75 % were extracted randomly for training purposes. The 75 % was divided into six parts and independently given to each classifier. The rest were taken for prediction. Each expert is trained, and all these models are tested against the test dataset for accurate computations. Model 1 is trained for the logistic regression classifier, and model 2 is trained for the Random Forest classifier. Model 3 and model 4 are trained for the decision tree and kNN classifier, respectively. Further, model 5 and model 6 are trained for Naive Bayes and SVM classifiers. The results are compared using majority voting techniques. Results are shown in Table 5.15. Accuracy results show a slight improvement for the single classifier.

Table 5.15: Result of the six classifiers with PCA applied to the dataset

	RG	RN	RL
Logistic Regression	0.6514	0.6165	0.8877
Random forest	0.6102	0.5744	0.879
Decision	0.564	0.5422	0.7975
kNN	0.6069	0.5774	0.8772
Nieve	0.6477	0.5929	0.8649
SVM	0.655	0.6237	0.8881

Pseudonymisation and anonymisation are popular privacy-enhancing techniques [23]. Pseudonymisation changes the data value using randomization or encryption techniques. Anonymisation removes direct and indirect personal identifiers. Privacy laws differ from country to country. Privacy laws apply to pseudonymised data as indirect identifiers, in combination with other identifiers, can reveal the person's identity. However, Privacy laws do not apply to anonymised data. We now present the framework for selecting a particular privacy-enhancing scheme based on a weighted approach similar to [23]. However, compared to [23], our approach takes only two inputs: privacy and utility values.

$$\text{score} = (w1 * \text{privacy} + w2 * \textit{utility}) / (w1 + w2) \quad (5.25)$$

The choice of weights depends upon the healthcare records to enhance privacy. The privacy and utility values are discrete in the range of (1, 2, 3), where one signifies not essential and 3 essentials. We now plot the score for two cases:

Case 1: Data privacy is essential, but utility is not. score = (0.5 * 3 + 0.5 * 1)/ (0.5 + 0.5) = 2

Case 2: Data privacy is not essential, but utility is essential. score = (0.5 * 1 + 0.5 * 3)/ (0.5 + 0.5) = 2

Case 3: Data privacy and utility are not essential. score = (0.5 * 1 + 0.5 * 1)/ (0.5 + 0.5) = 1

Case 4: Privacy of data and utility are essential. score = (0.5 * 3 + 0.5 * 3)/ (0.5 + 0.5) = 3

Figure 5.14 (a): Score for different parameters settings for privacy and trust

Figure 5.14 (a) presents the score for different parameter settings. The privacy and trust weights are kept at 0.5 each. Figure 5.14 (b) illustrates the degree of privacy versus utility [24]. The security and privacy of medical data are undoubtedly significant considerations for developing

Cyber-Physical Systems [25].

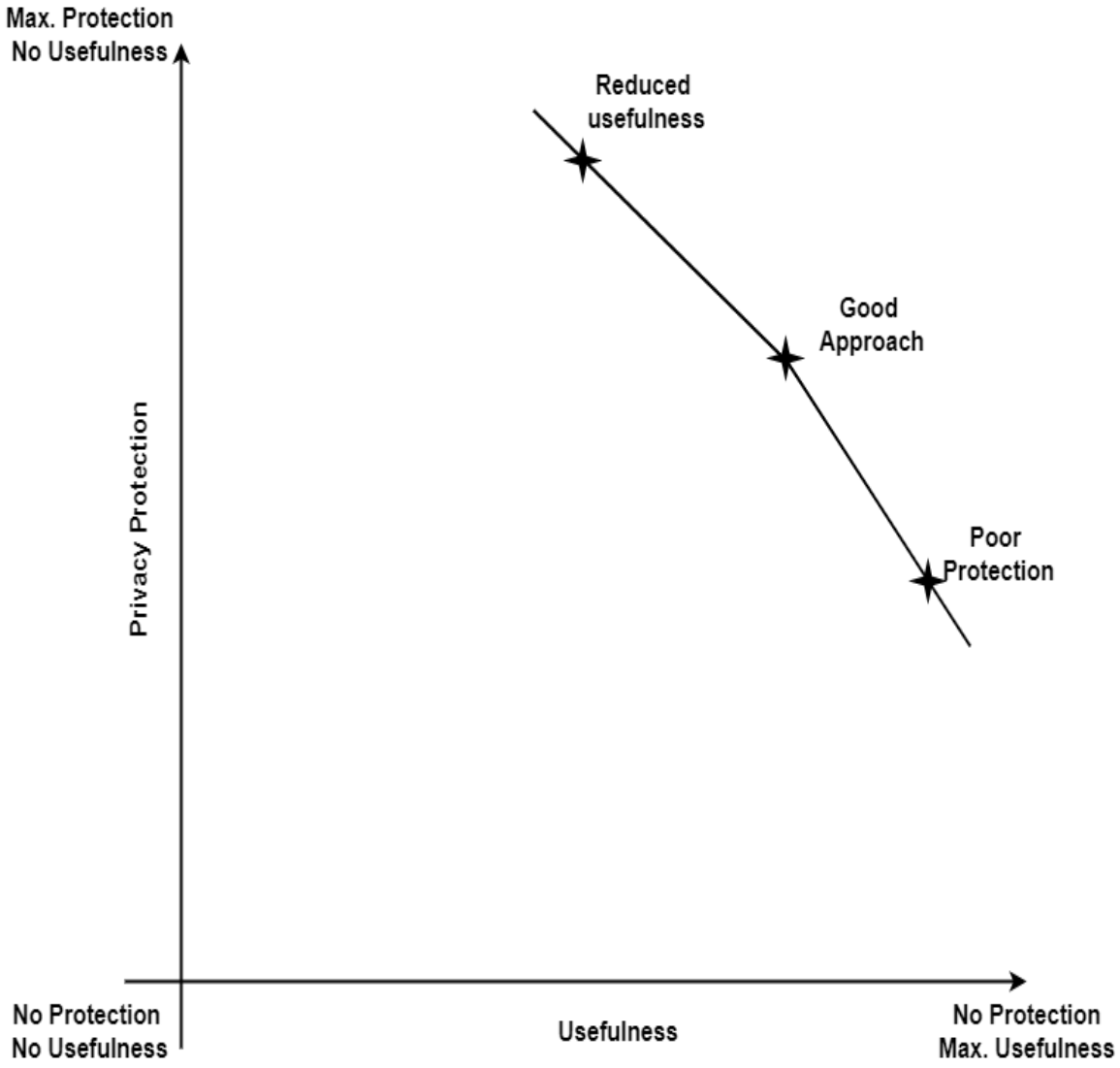


Figure 5.14 (b): degree of privacy protection versus utility

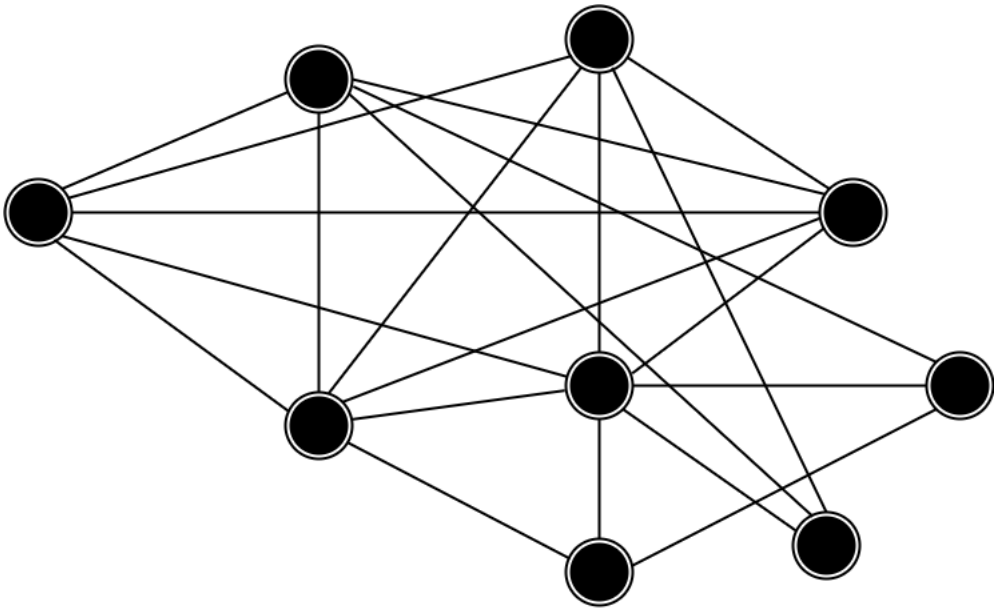


Figure 5.14 (c): Healthcare data converted into a graph using the neo4j tool [26]

Figure 5.14 (c) presents healthcare data converted into a graph using the neo4j tool. This graph implements well-known anonymization algorithms, Random Add/Delete and Random Walk, to analyze the preservation levels of the graph [26].

These fundamental strategies for data alteration used to safeguard data privacy can lead to diverse effects under different conditions.
These basic data modification techniques to preserve data privacy can produce different results under different circumstances. The single method of privacy preservation cannot deliver ideal outcomes. Different approaches may perform better under different conditions. It is unnecessary to produce optimal results with the procedures used for data preservation.

HbA1c measurement can help develop approaches for minimizing readmission rates and prices for treating people living with diabetes as a predictor for readmission rates for people with diabetes diseases. The support Vector Machine (SVM) classifier gave us a better accuracy of 88.81% beneath the case of patient readmission in less than or < 30 days, 65.5% just in case of patient readmission in higher than or > 30 days, and 62.37% just in case of no patient readmission. There is no loss of statistics because of Privacy. The preservation of data will increase the utility of information in the future.

5.5 CONCLUSION

We present privacy preservation techniques for the healthcare dataset. We compared six machine learning classification algorithms and observed that Support Vector Machine (SVM) performs better than other techniques. We observed that employing a single privacy-preserving technique could not provide optimal results. Different techniques perform better under different conditions. We present privacy preservation based on PCA and horizontal data for multiple classifier learning.

The construction of a clinical knowledge graph and engineering methods are discussed in the study. This article provides a framework for a clinical knowledge graph that specialists, physicians, and patients with various illnesses can use. The anonymization or ejection of patient-sensitive information from knowledge graphs is complete. Developing a data framework for clinical knowledge graphs can also expand the framework. Combining cutting-edge data storage technologies with the information architecture for chronic infections and lifestyle disorders will benefit specialists, staff, and patients. The partners will benefit from enhancing a knowledge graph user interface in light of the Android platform. Patients, clinical understudies, and expert consultants advising remote locations are anticipated consumers of the frameworks.

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

This chapter concludes the security and privacy preservation research in environmental monitoring and healthcare applications. A privacy preservation framework in wireless medical sensor data is proposed to hold patient data, and numerous data servers are used. Hashing is an essential technique for safeguarding the original message. The thesis presents a novel approach to improving health data security across several channels. If a message is received, hashing can quickly determine whether it is authenticated. Hashing provides a high level of protection for the privacy criteria. The original message has been divided into three parts using the proposed technique, and the split messages, along with the hash value, are then sent to various servers using multipath routing. The thesis presents a performance analysis of the proposed authentication scheme.

A source location privacy-preserving scheme is proposed for crucial and nominal occurrences for environmental monitoring applications. The work introduces two proposed event detection algorithms. EeSP and DSP algorithms are compared with 'SLP_ED and SLP_ED_CBA' regarding energy consumption, average hop duration, and safety standards for two deployment scenarios. Compared to DSP and EeSP systems, 'SLP_ED and SLP_ED_CBA' offer higher degrees of safety. Nonetheless, SLP_ED and SLP_ED_CBA consume more energy and have longer hops on average. After deployment, the nodes are static. Work can easily be extended to the mobile node and heterogeneous node deployment scenarios. The design may be expanded to three dimensions to reflect the real-world environment. The main idea of monitoring applications is to control the trade-off between location privacy and energy efficiency. In contrast, healthcare monitoring is to control the trade-off between location privacy and service utility. In addition, the work can be extended to Location Privacy Preservation with multiple source locations for healthcare and monitoring applications.

With the increased use of technology, a massive amount of data is being generated in healthcare. Machine learning healthcare applications extract useful information from data. However, when mining sensitive data, privacy must be preserved. 'Privacy-Preservation in

Data Mining (PPDM)' solves this problem. We compare six machine learning classification algorithms and observe that Support Vector Machine (SVM) performs better than other techniques. The work also explores data preservation techniques to secure machine learning models from leaking sensitive information. We observe that employing a single privacy-preserving technique could not provide optimal results. Logistic Regression on the dataset was performed without and with differential privacy, resulting in an accuracy of 88.96% and 88.97%, respectively.

It is suggested that a framework for medical knowledge graphs that assists medical professionals and patients with various disorders be used. Knowledge graphs are anonymized or have patient-sensitive data removed for privacy preservation. Creating an information system for medical knowledge graphs can further expand the system. Incorporation of big information collection technologies and knowledge graphs for lifestyle-associated disorders or chronic diseases would also benefit physicians, staff, and patients. The stakeholders will benefit from creating an Android-based knowledge graph user interface in an emergency. Medical students, patients, and specialised doctors who provide consultation in far-flung areas could be potential consumers of the systems.

REFERENCES

[1] P. Neves, M. Stachyra, J. Rodrigues, "Application of wireless sensor networks to health care promotion", Journal of communications software and systems, Vol. 4, No. 3, pp. 181-190, September 2008.

[2] P. Kumar and H.-J. Lee, "Security issues in healthcare applications using wireless medical sensor networks: A survey, " 12(1): 55-91, Sensors. 2011.

[3] M.A. Sahi,H. Abbas, K.Saleem,X. Yang, A.Derhab, M.Orgun,W. Iqbal,I. Rashid,A. Yaseen,"PrivacyPreservationineHealthcareEnvironments:AReview".IEEEAccess.2017 Oct. 30.

[4] Y. Xi, L. Schwiebert, and W. Shi, "Preserving source location privacy in monitoring-based wireless sensor networks," Proceedings 20th IEEE International Parallel & Distributed Processing Symposium. IEEE, 2006.

[5] J. KangJ, D. NyangD, "A Privacy-Preserving Mobile Payment System for MassT ransit". IEEE Transactions on Intelligent Transportation Systems. 2017 Aug;18(8):2192-205.

[6] E. Stavrou,A. PitsillidesA,"A survey on secure multipath routing protocols in WSNs". Computer Networks. 2010 Sep15; 54(13):2215-38.

[7] A. Anjum,K.K. Choo, A.Khan, A.Haroon,S. Khan, S.U.Khan, N.Ahmad, B.Raza,"An efficient privacy mechanism for electronic health records". Computers&Security.2018 Jan1;72:196-211.

[8] W. Lou, W. Liu and Y. Fang, "Spread: improving network security by multipath routing," IEEE Military Communications Conference, 2003. MILCOM 2003., 2003, pp. 808-813 Vol.2, doi: 10.1109/MILCOM.2003.1290216.

[9] T. Claveirole, M. D. De Amorim, M. Abdalla and Y. Viniotis, "Securing wireless sensor networks against aggregator compromises," in IEEE Communications Magazine, vol. 46, no. 4, pp. 134-141, April 2008, doi: 10.1109/MCOM.2008.4481352.

[10] P. Gope and T. Hwang," BSN-Care: A secure IoT-based modern healthcare system using body sensor network," IEEE Sensors Journal, " 16(5):1368-76, Mar 1, 2016.

[11] B. Shanthini, S. Swamynathan," Genetic-based biometric security system for wireless sensor-based health care systems," 2012 International Conference on in

Recent Advances in Computing and Software Systems (RACSS), pp. 180-184, IEEE, Apr 25, 2012.

[12] S. Williams,"Cryptography and network security:Principles and practices".ed:PearsonEducation.2006;17.

[13] Z.Y. Wu, Y.C. Lee, F. Lai, H.C. Lee, Y. Chung,"A secure authentication scheme for telecare medicine information systems". Journal ofmedicalsystems.2012 Jun1;36(3):1529-35.

[14] R.R. Selmic,V.V. Phoha,A. Serwadda,"WSN Platforms''.In Wireless Sensor Networks 2016 (pp.197-215), Springer, Cham.

[15] W. He, X. Liu, H. Nguyen, K. Nahrstedt and T. Abdelzaher, "PDA: Privacy-Preserving Data Aggregation in Wireless Sensor Networks," IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications, 2007, pp. 2045-2053, doi: 10.1109/INFCOM.2007.237.

[16] Raja M, Datta R. An Enhanced Source Location Privacy Protection Technique for Wireless Sensor Networks using Randomized Routes. IETE Journal of Research. 2017 Sep 1:1-3.

[17] R. Bhatt,R. Datta,"ATwo-tier Strategy for Priority based Critical Event Surveillance with Wireless Multimedia Sensors".Wireless Networks, 2017 Jan22:267-284.

[18] N. Li, , N. Zhang, , S.K. Das, , and B. Thuraisingham," Privacy preservation in wireless sensor networks: A state-of-the-art survey," Ad Hoc Networks, 7, 1501–1514, 2009.

[19] Y. Xi,L. Schwiebert,W. Shi,"Preserving source location privacy in monitoring-based wireless sensor networks", In Parallel and Distributed Processing Symposium,IPDPS2006.20th Internationalpp.8 IEEE.

[20] S. Kraijak and P. Tuwanut,"A survey on iot architectures, protocols, applications, security, privacy, real world implementation and future trends," in *11th International Conference on Wireless Communications, Networking and Mobile Computing(WiCOM2015)*,Sept2015,pp.1–6.

[21] L.Chen, S.Thombre,K.Järvinen,E.S.Lohan,A.Alénsavikko,H.Leppäkoski M.Z.H.Bhuiyan, S.Bu-Pasha, G.N. Ferrara, S.Honkala, J.Lindqvist,

[22] L. Ruotsalainen, P. Korpisaari, and H.Kuusniemi,"Robustness, security and privacy in location-based services for future iot:A survey,"*IEEEAccess*,vol.5, pp.8956–8977,2017.

[23] F.K. Santosoand N.C.Vun,“Securing iot for smart home system,”in *International Symposiumon Consumer Electronics(ISCE)*.IEEE,2015,pp.1–2.

[24] X.Zhu,Y.Badr,J.Pacheco,andS.Hariri,“Autonomic identity framework for the internet of things,”in *2017 International Conference on Cloud and Autonomic Computing(ICCAC)*,Sept 2017,pp.69–79.

[25] G. Kortuem,F.Kawsar,V.Sundramoorthy,D.Fitton*etal.*,“Smart object sas building blocks for the internet of things,”*IEEE Internet Computing*,vol.14,no.1,pp.44–51,2009.

[26] T. Gong, H . Huang, P. Li, K. Zhang, H. Jiang,"A medical healthcare system for privacy protection based on IoT," In Parallel Architectures, Algorithms and Programming (PAAP), Seventh International Symposium on 2015 Dec 12 (pp. 217-222). IEEE, 2015.

[27] X . Yi, J . Willemson, F. Nait-Abdesselam," Privacy-preserving wireless medical sensor network," 2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pp. 118-125, IEEE, Jul 16, 2013.

[28] R. P. Nia, R. P. Mganga, ``Enhancing information security in cloud computing services using SLA based metrics," Ph.D. dissertation, Dept. Comput. Sci., Blekinge Inst. Technol., Karlskrona, Sweden, 2011.

[29] Verizone 2014 Data Breach Investigation Report, Figure 19; Accessed: Aug. 27, 2016. [Online]. Available: http://www.verizonenterprise. com/resources/reports/rp_Verizon-DBIR-2014_en_xg.pdf

[30] Accessed: Jan. 4, 2016. [Online]. Available: http://www.nytimes. com/2015/02/06/business/experts-suspect-lax-security-left-anthemvulnerable-to-hackers.html.

[31] J.-J. Yang, J.-Q. Li, and Y. Niu, ``A hybrid solution for privacy-preserving medical data sharing in the cloud environment," Future Generat. Comput. Syst., vols. 4344, pp. 7486, Feb. 2015.

[32] C. R. Baker et al., ``Wireless sensor networks for home health care," in Proc. 21st Int. Conf. Adv. Inf. Netw. Appl. Workshops (AINAW), vol. 2. May 2007, pp. 832837.

[33] M. Al Ameen, J. Liu, and K. Kwak, ``Security and privacy issues in wireless sensor networks for healthcare applications," J. Med. Syst., vol. 36, no. 1, pp. 93101, 2012.

[34] T. Giannetsos, T. Dimitriou, and N. R. Prasad, ``People-centric sensing in assistive healthcare: Privacy challenges and directions," Secur. Commun. Netw., vol. 4, no. 11, pp. 12951307, 2011.

[35] S. Avancha, A. Baxi, and D. Kotz, ``Privacy in mobile technology for personal healthcare," ACM Comput. Surv., vol. 45, no. 1, p. 3, 2012.

[36] V. Stanford, ``Pervasive health care applications face tough security challenges," IEEE Pervasive Comput., vol. 1, no. 2, pp. 812, Apr. 2002.

[37] P. Kulkarni, and Y. Öztürk, ``Requirements and design spaces of mobile medical care," ACM SIGMOBILE Mobile Comput. Commun. Rev., vol. 11, no. 3, pp. 1230, 2007.

[38] N. Lee and O. Kwon, ``A privacy-aware feature selection method for solving the personalization-privacy paradox in mobile wellness healthcare services," Expert Syst. Appl., vol. 42, no. 5, pp. 27642771, 2015.

[39] M. N. K. Boulos, D. M. Giustini, and S. Wheeler, ``Instagram and WhatsApp in health and healthcare: An overview," Future Internet, vol. 8, no. 3, p. 37, 2016.

[40] M. Henze, L. Hermerschmidt, D. Kerpen, R. Häuÿling, B. Rumpe, and K. Wehrle, ``A comprehensive approach to privacy in the cloud-based Internet of Things," Future Generat. Comput. Syst., vol. 56, pp. 70-71, Mar. 2016.

[41] A. Sajid and H. Abbas, ``Data privacy in cloud-assisted healthcare systems: State of the art and future challenges," J. Med. Syst., vol. 40, no. 6, pp. 116, 2016.

[42] N. H. Hassan and Z. Ismail, ``Information security culture in healthcare informatics: A preliminary investigation," J. Theor. Appl. Inf. Technol., vol. 88, no. 2, p. 202, 2016.

[43] R. Parks, H. Xu, C.-H. Chu, and P. B. Lowry, ``Examining the intended and unintended consequences of organizational privacy safeguards enactment in healthcare: A grounded theory investigation," Eur. J. Inf. Syst., vol. 26, no. 1, pp. 3765, May 2016.

[44] H. A. J. Narayanan and M. H. Güne³, ``Ensuring access control in cloud provisioned healthcare systems," in Proc. IEEE Consum. Commun. Netw.Conf. (CCNC), Jan. 2011, pp. 247-251.

[45] J. Bugeja, A. Jacobsson, and P. Davidsson,“On privacy and security challenges in smart connected homes,” in *European Intelligence and Security Informatics Conference*. IEEE, 2016, pp. 172–175.

[46] H. Khemissa and D. T and jaoui, “A light weight authentication scheme for e-

health applications in the context of internet of things," in *2015 9th International Conference on Next Generation Mobile Applications, Services and Technologies*.IEEE,2015,pp.90–95.

[47] A.J.J.Valera,M. A.Zamora,andA. F.Skarmeta,"An architecture based on internet of things to support mobility and security in medical environments," in *2010 7th IEEE consumer communications and networking conference*. IEEE,2010,pp.1–5.

[48] H.Boyes,B.Hallaq,J.Cunningham,andT.Watson,"The industrial internet of things(iiot):Ananalysis framework,"*Computersinindustry*,vol.101,pp.1–12,2018.

[49] M.S.HossainandG.Muhammad,"Cloud assisted industrial internet of things(iiot)enabled framework for health monitoring,"*ComputerNetworks*,vol.101,pp.192–202,2016.

[50] J.Sengupta,S.Ruj,andS.D.Bit,"A comprehensive survey on attacks, security issues and blockchain solutions for iot and iiot,"*Journal of Network and ComputerApplications*,vol.149,p.102481,2020.

[51] K. A.AbuhaselandM.A.Khan,"A secure industrial internet of things(iiot) framework for resource management in smart manufacturing,"*IEEEAccess*,vol.8,pp.117354–117364,2020.

[52] S.Qi,Y.Lu,W.Wei,andX.Chen,"Efficient data access control with fine-grained data protection in cloud-assisted iiot,"*IEEE Internet of Things Journal*,vol.8,no.4,pp.2886–2899,2020.

[53] A.Karati,S.H.Islam,andM.Karuppiah,"Provably secure and lightweight certificate less signature scheme for iiot environments,"*IEEE Transactionson Industrial Informatics*,vol.14,no.8,pp.3701–3711,2018.

[54] D.Liu,A.Alahmadi,J.Ni,X.Lin,andX.Shen,"Anonymous reputation system for iiot-enabled retail marketing at oppos blockchain," *IEEE Transactionson Industrial Informatics*,vol.15,no.6,pp.3527–3537,2019.

[55] Q.Jiang,S.Zeadally,J.Ma,andD.He,"Lightweight three-factor authentication and key agreement protocol for internet-integrated wireless sensor networks,"*IEEE Access*,vol.5,pp.3376–3392,2017.

[56] Y.S.DabbaghandW.Saad,"Authentication of wireless devices in the internet of things:Learning and environmental effects,"*IEEE Internet of Things Journal*,vol.6,no.4,pp.6692–6705,2019.

[57] N.Li,D.Liu,andS.Nepal,"Lightweight mutual authentication for iot and its applications,"*IEEE Transactions on Sustainable Computing*,vol.2,no.4,pp.359–

370,2017.

[58] L.Viganò,“Automated security protocol analysis with the avis patool,”Electronic Notes in Theoretical Computer Science,vol.155,pp.61–86,2006.

[59] Z.Ali,S.A.Chaudhry,M.S.Ramzan,andF.Al-Turjman,“Securing smart city surveillance:A lightweight authentication mechanism for unmanned vehicles,”*IEEEAccess*,vol.8,pp.43711–43724,2020.

[60] H.KhemissaandD.Tandjaoui,“An ovel lightweight authentication scheme for heterogeneous wireless sensor networks in the context of internet of things,”in *2016 Wireless Telecommunications Symposium(WTS)*.IEEE,2016,pp.1–6.

[61] A.Adeel,M.Ali,A.N.Khan,T.Khalid,F.Rehman,Y.Jararweh,andJ.Shuja,“Amulti -attack resilient lightweight iot authentication scheme,”*Transactionson Emerging Telecommunications Technologies*,p.e3676,2019.

[62] A.Armando,Basin*etal.*,“The avispatool for the automated validation of internet security protocols and applications,”in *International conference on computer aided verification*.Springer,2005,pp.281–285.

[63] S.Li,L.DaXu,andS.Zhao,“The internet of things: asurvey,”Information Systems Frontiers,17(2),243-259,2015.

[64] S. Misra, M. Maheswaran, and S. Hashmi, “Security Challenges and Approaches in Internet of Things,” Springer Briefs in electrical and computer engineering,2017.

[65] L.Atzori,A.Iera,andG.Morabito,“The internet of things:Asurvey,” Computer networks, 54(15),2787-2805,2010.

[66] Q. Jing, A. V. Vasilakos, J. Wan, J. Lu, and D. Qiu, “Security of the Internet of Things: perspectives and challenges,” Wireless Networks, 20(8),2481-2501,2014.

[67] C. W. Tsai, C. F. Lai, and A. V. Vasilakos, “Future Internet of Things:open issues and challenges,” Wireless Networks, 20(8),2201-2217,2014.

[68] M. Nawir, A. Amir, N. Yaakob, and O. B. Lynn, “Internet of Things(IoT): Taxonomy of security attacks,” in 3rd International Conference on Electronic Design(ICED),IEEE, pp.321-326,2016.

[69] R. Roman, P. Najera, and J. Lopez, “Securing the internet of things,”Computer,44(9),51-58,2011.

[70] X. Wei, N. C. Valler, H. V. Madhyastha, I. Neamtiu, and M. Faloutsos,“Characterizing the Behavior of Hand held Devices and It

sImplications,”ComputerNetworks,114,1-12,2017.

[71] K.T.Nguyen,M.Laurent,andN.Oualha,“Survey on secure communication protocols for the Internet of Things,” AdHocNetworks, 32,17-31,2015.

[72] P.Gope,andT.Hwang,“BSN-Care:AsecureIoT-based modern health care system using body sensor network,” IEEE sensors journal, 16(5),pp.1368-1376,2015.

[73] C. Ozturk et al., “Source-location privacy in energy-constrained sensor network routing” in Proc. 2nd ACM Workshop on Security of Ad hoc and Sensor Networks (SASN ’04). New York, NY, USA: Association for Computing Machinery, 2004, pp. 88-93,doi:10.1145/1029102.1029117.

[74] L. C. Mutalemwa and S. Shin, “Strategic location-based random routing for source location privacy in wireless sensor networks,” Sensors (Basel), vol. 18, no. 7, p. 2291, 2018, doi:10.3390/s18072291.

[75] H. Wang et al., “TCSLP: A trace cost based source location privacy protection scheme in WSNs for smart cities,” Future Gener. Comput. Syst., vol. 107, pp. 965-974, 2020, [doi:10.1016/j.future.2017.07.051.

[76] M. Guo and N. XinyuJinPissinou, Sebastian Zanlongo, “Bogdan Carbunar, and S. S. Iyengar,” ACM Comput. Surv., vol. 48, p. 2, article 23, 2015. In-Network Trajectory Privacy Preservation.

[77] Y. Li and J. Ren, "Preserving Source-Location Privacy in Wireless Sensor Networks," 2009 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, Rome, Italy, 2009, pp. 1-9, doi: 10.1109/SAHCN.2009.5168962.

[78] Z. Zeng et al., “Source-location privacy protection in wireless sensor networks using AZR routing” in Proc. International Conference on Wireless Networks (ICWN), 2014, (p. 1). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

[79] P. Kamat et al., “Enhancing source location privacy in sensor network routing” in ICDCS, Proc. 25th International Conference on Distributed Computing Systems, OH, USA, June 2005, ’05.

[80] L.Mutalemwa and S. Shin, “Achieving Source Location Privacy Protection in Monitoring Wireless Sensor Networks through Proxy Node Routing,” Sensors, vol. 19, no. 5, p. 1037, Feb. 2019, doi: 10.3390/s19051037.

[81] W.P. Wang, L. Chen and J. . -X. Wang, "A Source-Location Privacy Protocol

in WSN Based on Locational Angle," 2008 IEEE International Conference on Communications, Beijing, China, 2008, pp. 1630-1634, doi: 10.1109/ICC.2008.315.

[82] R. Manjula, T. Koduru and R. Datta, "Protecting Source Location Privacy in IoT-Enabled Wireless Sensor Networks: The Case of Multiple Assets," in IEEE Internet of Things Journal, vol. 9, no. 13, pp. 10807-10820, 1 July1, 2022, doi: 10.1109/JIOT.2021.3126171.

[83] D. Zhang et al., "A survey on collaborative deep learning and privacy-preserving" in IEEE Third International Conference on Data Science in Cyberspace (DSC), vol. 2018, 2018, doi:10.1109/DSC.2018.00104.

[84] S. Chang and C. Li, "Privacy in neural network learning: Threats and countermeasures," IEEE Netw., vol. 32, no. 4, pp. 61-67, 2018, doi:10.1109/MNET.2018.1700447.

[85] H. C. Tanuwidjaja et al., "A survey on deep learning techniques for privacy-preserving" in International Conference on Machine Learning for Cyber Security. Cham: Springer, 2019 Sept. 19, pp. 29-46.

[86] M. S. Riazi et al., "Deep learning on private data," IEEE Secur. Privacy, vol. 17, no. 6, 54-63, 2019, doi:10.1109/MSEC.2019.2935666.

[87] A. Boulemtafes et al., "A review of privacy-preserving techniques for deep learning," Neurocomputing, vol. 384, pp. 21-45, 2020 [doi:10.1016/j.neucom.2019.11.041].

[88] M. Zheng et al., "Challenges of privacy-preserving machine learning in IoT" in Proc. First International Workshop on Challenges in Artificial Intelligence and Machine Learning for Internet of Things 2019 Nov 10, 2019, pp. 1-7, doi:10.1145/3363347.3363357.

[89] D. Xu et al., "Lightweight and unobtrusive data obfuscation at IoT edge for remote inference," IEEE Internet Things J., vol. 7, no. 10, 9540-9551, 2020. arXiv preprint, doi:10.1109/JIOT.2020.2983278.

[90] M. M. Dhanvijay, and S. C. Patil, "Internet of Things: A survey of enabling technologies in healthcare and its applications," Computer Networks, 153,113–131,2019.

[91] S.He,B.Cheng,H.Wang,Y.HuangandJ.Chen,"Proactive personalized services through fog-cloud computing in large-scale IoT-based healthcare application," China Communications, 14(11),1-16,2017.

[92] F. Andriopoulou, T. Dagiuklas, and T. Orphanoudakis, "Integrating IoT and

fog computing for healthcare service delivery," in Components and services for IoT platforms(pp.213-232). Springer, Cham,2017.

[93] S. K. Sood, and I. Mahajan, "A fog-based healthcare framework for chikungunya," IEEE Internet of Things Journal,5(2),794-801,2017.

[94] P. Verma, S. K. Sood, and S. Kalra, "Cloud-centric IoT based student healthcare monitoring framework," Journal of Ambient Intelligence andHumanizedComputing,9(5),pp.1293-1309,2018.

[95] R. Amin, S. H. Islam, G. P. Biswas, M. K. Khan, and M. S. Obaidat,"Design and analysis of an enhanced patient-server mutual authentication protocol for telecare medical information system,"Journal of medical systems,39(11),p.137,2015.

[96] Y. K. Chen, "Challenges and opportunities of internet of things," In 17thAsia and South Pacific design automation conference, IEEE, pp. 383-388,2012.

[97] J.Gubbi,R.Buyya,S.Marusic,andM.Palaniswami,"Internet of Things (IoT): A vision, architectural elements, and future directions," Future generation computer systems, 29(7),1645-1660,2013.

[98] G. P. Hancke, K. Markantonakis, and K. E. Mayes, "Security challenges for user-oriented RFID applications within the Internet of things," Journal of Internet Technology, 11(3),307-313,2010.

[99] Y. Harel, I. B. Gal, and Y. Elovici, "Cyber security and the role of intelligent systems in addressing its challenges,"2017.

[100] Z. Yan, P. Zhang, and A. V. Vasilakos, "A survey on trust management for Internet of Things," Journalofnetworkandcomputerapplications,42,120-134,2014.

[101] S. Winsen, "Threat modelling for future vehicles: on identifying and analyzing threats for future autonomous and connected vehicles," Master's thesis,University of Twente,2017.

[102] L.Zhou, and H. C. Chao, "Multimedia traffic security architecture for the internet of things," IEEE Network, 25(3),35-40,2011.

[103] M. F. Al-Mistarihi1, I. M. Tanash, F.S. Yaseen, K.A. Darabkh, "Protecting Source Location Privacy in a Clustered Wireless Sensor Networks against Local Eavesdroppers, " Mobile NetwAppl, 25, 42–54, 2020.

[104] A. Arivarasi,&P. Ramesh,P, "An improved source location privacy protection using adaptive trust sector- based authentication with honey encryption algorithm in WSN".Journal of Ambient Intelligence and Humanized Computing,1-13. 2011.

[105] R. Bhatt,R., R. Datt, “Redeployment strategies for Wireless Sensor Networks under random node failure sand budget constraints. In 2012 2nd IEEE International Conference on Parallel, Distributed and Grid Computing(pp.767-772).IEEE. 2012.

[106] R. Bhatt,R.Datta,“Costmodelsfor3-DdeploymentofWirelessMultimediaSensor Networks’’.In*2016IEEERegion10Conference(TENCON)*(pp. 3485-3489).IEEE.

[107] M. Cunha,R. Mendes,J P . Vilela,“ A survey of privacy-preserving mechanisms for heterogeneous datatypes,”ComputerScienceReview,Volume41,100403.2021.

[108] J.M. deFuentes,L. González-Manzano,&Mirzaei,O, “Privacy models in wireless sensor networks:A survey”.JournalofSensors,2016.

[109] H. Farman,B. Jan,Z. Khan,&A. Koubaa, “A smart energy-based source location privacy preservation model for Internet of Things-based vehicular adhoc networks. *Transactions on Emerging Telecommunications Technologies*,*33*(2),e3973. 2022.

[110] I. Batra et al., “Hybrid logical security framework for privacy preservation in the green internet of things,” Sustainability, vol. 12, no. 14, p. 5542, 2020, doi:10.3390/su12145542.

[111] S. Jiang et al., “A privacy-preserving reauthentication scheme for mobile wireless sensor networks,” Int. J. Distrib. Sens. Netw., vol. 9, no. 5, p. 913782, 2013, doi:10.1155/2013/913782.

[112] Y. Xi et al., “Preserving source location privacy in monitoring-based wireless sensor networks” in Proc. 20th IEEE International Parallel & Distributed Processing Symposium. IEEE, 2006, Apr., (p. 8-pp)..

[113] B. Strack et al., “Impact of HbA1c measurement on hospital readmission rates: Analysis of 70,000 clinical database patient records,” BioMed Res. Int., vol. 2014, 781670, 2014, doi:10.1155/2014/781670.

[114] M. F. Al-Mistarihi et al., “Protecting source location privacy in a clustered wireless sensor network against local eavesdroppers,” Mob. Netw. Appl., vol. 25, no. 1, pp. 42-54, 2020, doi:10.1007/s11036-018-1189-6.

[115] N. Jan and S. Khan, “Energy-efficient source location privacy protection for network lifetime maximization against local eavesdropper in wireless sensor network (EeSP),” Trans. Emerg. Telecommun. Technol., vol. 33, no. 2, p. e3703, 2022, doi:10.1002/ett.3703.

[116] L. C. Mutalemwa and S. Shin, “Strategic location-based random routing for source location privacy in wireless sensor networks,” Sensors (Basel), vol. 18, no. 7,

p. 2291, 2018, doi:10.3390/s18072291.

[117] G. Han et al., "A source location protection protocol based on dynamic routing in WSNs for the Social Internet of Things," Future Gener. Comput. Syst., vol. 82, pp. 689-697, 2018, doi:10.1016/j.future.2017.08.044.

[118] V. Mohindru et al., "Reauthentication scheme for mobile wireless sensor networks," Sustain. Comput. Inform. Syst., vol. 23, pp. 158-166, 2019, doi:10.1016/j.suscom.2019.07.010.

[119] F. Jabeen et al., "Enhanced architecture for privacy preserving data integration in a medical research environment," IEEE Access, vol. 5, pp. 13308-13326, 2017, doi:10.1109/ACCESS.2017.2707584].

[120] R. Bhatt and R. Datta, "A two-tier strategy for priority based critical event surveillance with wireless multimedia sensors," Wirel. Netw., vol. 22, no. 1, pp. 267-284, 2016, doi:10.1007/s11276-015-0971-7.

[121] A. Khatri et al., "Architecture for preserving privacy during data mining by hybridization of partitioning on medical data" in Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation, vol. 2010. IEEE, 2010, May, pp. 93-97, doi:10.1109/AMS.2010.31.

[122] G. N. Obuandike, J. Alhasan, and M. B. Abdullahi, IJMSO, vol. 2017, pp. 139 - 153, Apr. 2018.

[123] A. Tripathy and M. Pradhan, "A novel framework for preserving privacy of data using correlation analysis" in Proc. International Conference on Advances in Computing, Communications and Informatics, 2012, Aug., pp. 650-655, doi:10.1145/2345396.2345502.

[124] P. Gautam,M.D. Ansari,&S.K. Sharma,S.K' "Enhanced security for electronic health care information using obfuscationand RSA algorithm in cloudcomputing. International Journal of Information Security and Privacy(IJISP),13(1),59-69.

[125] C. Gu,"Source Location Privacy in Wireless Sensor Networks Under Practical Scenarios:Routing Protocols,ParameterisationsandTrade-Offs(Doctoraldissertation,UniversityofWarwick). 2018.

[126] G. Han,M. Xu,Y. HEJiang,J.,Ansere,J.A.,&Zhang,"A dynamicring-based routing scheme for source location privacy in wireless sensor networks."Information Sciences,504,308-323. 2019.

[127] G. Han, H.Wang, X.Miao,L. Liu,&Y. Peng, "A Dynamic Multipath Scheme for ProtectingSource-

LocationPrivacyUsingMultipleSinksinWSNsIntendedforIIoT.”IEEE Transactions on IndustrialInformatics,Vol.16, No.8,pp.5527-5538. 2020.

[128] N. Jan,&S. Khan,“Energy-efficient source location privacy protection for network lifetime maximization against local eavesdropper in wireless sensor network(EeSP)”.Transactions on Emerging Telecommunications Technologies,e3703. 2019.

[129] N. Jan,S.Khan,AH. Al-Bayatti,M.O, Alassafi,M.O., &M.A. Alqarni, “C2S2-LOOP: Circular Chessboard-Based Secure Source Location Privacy Model Using ECC-ALO in WSN” Wireless Communications and Mobile Computing,2021.

[130] Z. Jia,X. Wei, H.Guo,W. Peng,W.,&Song,“A privacy protection strategy for source location in WSN based onangle and dynamical adjustment of node emission radius”.Chinese Journal of Electronics,26(5), 1064-1072. 2017.

[131] J. Jiang,G. Han,H. Wang,& M. Guizani,M.“A survey on location privacy protection in wireless sensor networks. Journal of Network and Computer Applications,125,93-114. 2019.

[132] P. Kamat,Y. Zhang,W. Trappe,&Ozturk,C.“Enhancing source-location privacy in sensor network routing. In25th IEEE international conference on distributed computing systems (ICDCS'05)(pp.599- 608). IEEE. 2005.

[133] S. Kawai, & T. Asaka,T.“Event-driven wireless sensor networks using energy-saving data collection”.In201218thAsia-Pacific Conference on Communications(APCC)(pp.300-305).IEEE.

[134] P. Kumar, et al. "Source location privacy using multiple-phantom nodes in WSN. "*TENCON 2015-2015 IEEE Region 10 Conference*. IEEE, 2015.

[135] N. Li,N. Zhang,SK. Das,& B. Thuraisingham, “Privacy preservation in wireless sensor networks:A state-of-the-artsurvey.”*AdHocNetworks*,*7*(8),1501-1514. 2009.

[136] R. Manjula, & R. Datta,“An energy-efficient routing technique for privacy preservation assets monitored with WSN.In *Proceedings ofthe2014 IEEE Students' Technology Symposium*(pp. 325-330). IEEE. 2014.

[137] R. Manjula,T. Koduru,andDatta, “Protecting Source Location Privacy in It Enabled Wireless Sensor Networks:the Case of Multiple Assets”.IEEE Internet of Things Journal, DOI 10.1109/JIOT.2021.3126171. 2017.

[138] F. Mukamanzi,M. Raja,T. Koduru,,&Datta,“Position-independent and Section-based Source Location Privacy Protection in WSN”.IEEE Transactions on Industrial Informatics. 2022.

[139] R.K. Nirala, & M.D. Ansari,"Performance evaluation of loss packet percentage for asymmetric key cryptography in VANET." In 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC)(pp.70-74).I EEE.

[140] C. Ozturk,Y. Zhang&W. Trappe,W."Source-location privacy in energy-constrained sensor network routing."In Proceedings of the 2nd ACM workshop on Security of Adhocand Sensor Networks(pp. 88-93). 2004.

[141] A. Pudasaini,N.D. Devi,S.S. Hwang,&S. Shin,"Directional random routing for enhancing source location privacy in wireless sensor networks."In *2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN)*(pp.344-348).IEEE.

[142] M. Raja, & R. Datta, "An Enhanced Source Location Privacy Protection Technique for Wireless Sensor Networks using Randomized Routes".IETE Journal of Research,64(6),764-776.

[143] P.K. Roy,A.K. Tripathy,SK. Singh,&K.C. Li,K.C."Recent advancements in privacy-aware protocols of source location privacy in wireless sensor networks:Asurvey".Computer Science and Information Systems, (00), 7-7. 2022.

[144] R. Tandon,&P.K. Gupta,"A Novel Pseudonym Assignment and Encryption Scheme for Preserving the PrivacyofMilitaryVehicles."DefenceScienceJournal,71(2). 2021.

[145] P. Spachos,D. Toumpakari,& D. Hatzinakos,"Angle-based dynamic routing scheme for source

[146] Location privacy in wireless sensor networks."In *2014 IEEE 79th Vehicular Technology Conference(VTC Spring)* (pp. 1-5). IEEE. 2014.

[147] L. Ting,M. Khan,A. Sharma,&M.D. Ansari,"A secure framework for IoT-based smart climate agriculture system:Toward blockchain and edge computing."Journal of Intelligent Systems,31(1),221-236. 2022.

[148] H. Wang,L. Wu,L, &H. Jiang,"Energy balanced source location privacy scheme using multibranch path in WSNs for IoT."Wireless Communications and Mobile Computing, 2021.

[149] H. Wang,G. Han,C. Zhu, S. Chan,&W. Zhang,W."TCSLP: A trace cost based source location privacy protection scheme in WSNs for smart cities".Future Generation Computer Systems. doi: 10.1016/j.future.2017.07.051

[150] X. Fan, G. Wang, K. Chen, X. He, and W. Xu, "PPCA: Privacy-preserving

Principal Component Analysis Using Secure Multiparty Computation (MPC)," 2021, [Online]. Available: http://arxiv.org/abs/2105.07612

[151] A. Cuzzocrea, "Privacy-preserving big data stream mining: Opportunities, challenges, directions," *IEEE Int. Conf. Data Min. Work. ICDMW*, vol. 2017-November, pp. 992–994, 2017, doi: 10.1109/ICDMW.2017.140.

[152] Z. Zeng, M. Zeng, H. Liu, and U. States, "Source - Location Privacy Protection in Wireless Sensor Networks using AZR Routing," pp. 1–6.

[153] S. Networks, "SPS and DPS: Two New Grid-Based Source Location," 2019.

[154] K. Abouelmehdi, A. Beni-Hessane, and H. Khaloufi, "Big healthcare data: preserving security and privacy," *J. Big Data*, vol. 5, no. 1, pp. 1–18, 2018, doi: 10.1186/s40537-017-0110-7.

[155] J. Qian, X. Y. Li, C. Zhang, and L. Chen, "De-anonymizing social networks and inferring private attributes using knowledge graphs," *Proc. - IEEE INFOCOM*, vol. 2016-July, 2016, doi: 10.1109/INFOCOM.2016.7524578.

[156] A. Mukasheva, T. Iliev, and G. Balbayev, "Development of the Information System Based on BigData Technology to Support Endocrinologist-Doctors," *2020 7th Int. Conf. Energy Effic. Agric. Eng. EE AE 2020 - Proc.*, no. November, pp. 12–15, 2020, doi: 10.1109/EEAE49144.2020.9278971.

[157] M. Rotmensch, Y. Halpern, A. Tlimat, S. Horng, and D. Sontag, "Learning a Health Knowledge Graph from Electronic Medical Records," *Sci. Rep.*, vol. 7, no. 1, pp. 1–11, 2017, doi: 10.1038/s41598-017-05778-z.

[158] S. Ji, S. Pan, E. Cambria, P. Marttinen, and P. S. Yu, "A Survey on Knowledge Graphs: Representation, Acquisition, and Applications," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 33, no. 2, pp. 494–514, 2022, doi: 10.1109/TNNLS.2021.3070843.

[159] C. Zhang, H. Jiang, X. Cheng, F. Zhao, Z. Cai, and Z. Tian, "Utility analysis on privacy-preservation algorithms for online social networks: an empirical study," *Pers. Ubiquitous Comput.*, vol. 25, no. 6, pp. 1063–1079, 2021, doi: 10.1007/s00779-019-01287-0.

[160] O. Kocabas, T. Soyata, and M. K. Aktas, "Emerging Security Mechanisms for Medical Cyber-Physical Systems," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 13, no. 3, pp. 401–416, 2016, doi: 10.1109/TCBB.2016.2520933.

[161] F. N. Wirth, T. Meurers, M. Johns, and F. Prasser, "Privacy-preserving data sharing infrastructures for medical research: systematization and comparison," *BMC*

Med. Inform. Decis. Mak., vol. 21, no. 1, pp. 1–13, 2021, doi: 10.1186/s12911-021-01602-x.

[162] S. Bamford, "Applications of privacy-enhancing technology to data sharing at a global pharmaceutical company," *J. Data Prot. Priv.*, vol. 3, no. 3, pp. 281–290, 2020.

[163] J. K. B and M. Wo, "Multi Sampling Random Subspace Ensemble for Imbalanced," vol. 3, pp. 360–369, 2020, doi: 10.1007/978-3-030-19738-4.

[164] C. C. Aggarwal and P. S. Yu, "A General Survey of Privacy-Preserving Data Mining Models and Algorithms," pp. 11–52, 2008, doi: 10.1007/978-0-387-70992-5_2.

[165] M. Al-Rubaie and J. M. Chang, "Privacy-Preserving Machine Learning: Threats and Solutions," *IEEE Secur. Priv.*, vol. 17, no. 2, pp. 49–58, 2019, doi: 10.1109/MSEC.2018.2888775.

[166] P. Ram Mohan Rao, S. Murali Krishna, and A. P. Siva Kumar, "Privacy preservation techniques in big data analytics: a survey," *J. Big Data*, vol. 5, no. 1, 2018, doi: 10.1186/s40537-018-0141-8.

[167] B. Strack *et al.*, "Impact of HbA1c measurement on hospital readmission rates: Analysis of 70,000 clinical database patient records," *Biomed Res. Int.*, vol. 2014, 2014, doi: 10.1155/2014/781670.

[168] C. Chen, J. Cui, G. Liu, J. Wu, and L. Wang, "Survey and Open Problems in Privacy Preserving Knowledge Graph: Merging, Query, Representation, Completion and Applications," pp. 1–24, 2020, [Online]. Available: http://arxiv.org/abs/2011.10180

[169] R. V. Banu and N. Nagaveni, "Preservation of data privacy using PCA based transformation," *ARTCom 2009 - Int. Conf. Adv. Recent Technol. Commun. Comput.*, pp. 439–443, 2009, doi: 10.1109/ARTCom.2009.159.

[170] Y. Liu *et al.*, "Privacy-Preserving PCA for Multiparty Modeling," 2020, [Online]. Available: http://arxiv.org/abs/2002.02091

[171] M. Al-rubaie, P. Wu, J. M. Chang, and S. Kung, "Privacy-Preserving PCA on Horizontally-Partitioned Data," pp. 280–287.

[172] M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Inf. Fusion*, vol. 16, no. 1, pp. 3–17, 2014, doi: 10.1016/j.inffus.2013.04.006.

[173] Y. Qu, J. Xu, and S. Yu, "Privacy-preserving in big data sets through multiple

shuffles," *ACM Int. Conf. Proceeding Ser.*, 2017, doi: 10.1145/3014812.3014886.

[174] J. P. Barddal, "Vertical and Horizontal Partitioning in Data Stream Regression Ensembles," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2019-July, no. July, pp. 1–8, 2019, doi: 10.1109/IJCNN.2019.8852244.

[175] A. Khatri, S. Kabra, S. Singh, and D. K. Mishra, "Architecture for preserving privacy during data mining by hybridization of partitioning on medical data," *AMS2010 Asia Model. Symp. 2010 - 4th Int. Conf. Math. Model. Comput. Simul.*, pp. 93–97, 2010, doi: 10.1109/AMS.2010.31.

[176] K. Xu, H. Yue, L. Guo, Y. Guo, and Y. Fang, "Privacy-Preserving Machine Learning Algorithms for Big Data Systems," *Proc. - Int. Conf. Distrib. Comput. Syst.*, vol. 2015-July, pp. 318–327, 2015, doi: 10.1109/ICDCS.2015.40.

[177] C. Chen *et al.*, "Secret Sharing based Secure Regressions with Applications," 2020, [Online]. Available: http://arxiv.org/abs/2004.04898

[178] N. Sharma and R. Bhatt, "Privacy Preservation in WSN for Healthcare Application," *Procedia Comput. Sci.*, vol. 132, pp. 1243–1252, 2018, doi: 10.1016/j.procs.2018.05.040.

[179] L. Lildholdt, "Privacy-Preserving Machine Learning in Healthcare," p. 114, 2021, [Online]. Available: https://projekter.au.dk/fileadmin/ingen_mappe_valgt/LasseLildholdt_MasterThesis.pdf

[180] S. Pati *et al.*, "Federated learning enables big data for rare cancer boundary detection," *Nat. Commun.*, vol. 13, no. 1, 2022, doi: 10.1038/s41467-022-33407-5.

[181] L. Xu, C. Jiang, J. Wang, J. Yuan, and A. Y. Ren, "Information Security in Big Data: Privacy and Data Mining," *IEEE Access*, vol. 2, pp. 1149–1176, 2014.

[182] M. A. Abdo, A. A. Abdel-Hamid and H. A. Elzouka, "A Cloud-based Mobile Healthcare Monitoring Framework with Location Privacy Preservation," 2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT), Sakheer, Bahrain, 2020, pp. 1-8, doi: 10.1109/3ICT51146.2020.9311999.

[183] M. Min *et al.*, "Semantic Adaptive Geo-Indistinguishability for Location Privacy Protection in Mobile Networks," in *IEEE Transactions on Vehicular Technology*, vol. 73, no. 6, pp. 9193-9198, June 2024, doi: 10.1109/TVT.2024.3354881.

[184] I. Natgunanathan, A. Mehmood, Y. Xiang, L. Gao and S. Yu, "Location Privacy Protection in Smart Health Care System," in *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3055-3069, April 2019, doi: 10.1109/JIOT.2018.2878917.

LIST OF PUBLICATIONS

Journal Publications

1. N. Sharma, and R. Bhatt, "Privacy Preservation in WSN for Healthcare Application", Procedia Computer Science, 132(2018), pp. 1243-1252, 2018. ISSN: 18770509. DOI: 10.1016/j.procs.2018.05.040.
2. N. Sharma, and R. Bhatt, "Privacy-preserving knowledge graph for healthcare applications", Journal of Physics: Conference Series, 2339 (1):012013. ISSN: 1742-6588 (print), 1742-6596 (web), (2022). DOI:10.1088/1742-6596/2339/1/012013.
3. N. Sharma, and R. Bhatt, "Source location privacy preservation in IoT-enabled event-driven WSNs", International Journal of Pervasive Computing and Communications (IJPCC), 19(5), pp. 782-798, 2023. ISSN: 1742-7371, DOI: https://doi.org/10.1108/IJPCC-05-2022-0214.
4. N. Sharma, and R. Bhatt, "Privacy-preserving Machine Learning in Healthcare Applications and Medical Cyber-Physical Systems", Informatica. ISSN: 0350-5596 (print edition), ISSN: 1854-387.

Presented in International Conferences

1. N. Sharma, and R. Bhatt, "FoG Computing based IoT in Healthcare Application", November. *In 2020 Sixth International Conference on Parallel, Distributed and Grid Computing* (PDGC), pp. 442-446, IEEE, 06-08 November 2020. DOI: 10.1109/PDGC50313.2020.9315745.
2. N. Sharma, and R. Bhatt, "Privacy Preservation with Machine Learning for Healthcare Applications", *3rd Himachal Pradesh Science Congress*, pp. 157, 2018.
3. N. Sharma, and R. Bhatt, "Secure Mutual Authentication Scheme IoT-Based Healthcare", August. *International Conference on Emerging Trends & Innovation in Science, Engineering and Education (IESEE-2022)*, pp. 253-261, Springer, 2022.