2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM) | 979-8-3503-6509-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICONSTEM60960.2024.10568845

Cybersecurity Enabled Improved BigData Privacy Management Measures to Preserve Information with Privacy Concerns

[1]Sankar. R, [2]Alwin Christopher. T, [3]Daden Wilson. A, [4]Miruthula. M, [5]Vaishali. M, and [6]Kirthy Rajan. G

[1]Assistant Professor, Department of Master of Computer Applications, S.A.Engineering College, Chennai.
[2to6]PG Scholar, Department of Master of Computer Applications, S.A.Engineering College, Chennai

E-mail : sankar@saec.ac.in, 2299005@saec.ac.in, 2299010@saec.ac.in, 2299022@saec.ac.in, 2299059@saec.ac.in, 2299016@saec.ac.in

Abstract- In today's data-driven landscape, preserving the privacy and security of sensitive information within Big Data environments, particularly in sectors like healthcare, is of paramount importance. This paper introduces a comprehensive framework for Cybersecurity Enabled Improved Big Data Privacy Management Measures (IBPMM) tailored to address the challenges of safeguarding privacy while harnessing the potential of vast datasets. The framework integrates advanced cybersecurity measures with privacy management strategies, incorporating key components such as data encryption, access controls, anonymization techniques, threat detection, and incident response protocols. Regulatory compliance requirements, including GDPR and HIPAA, are carefully considered throughout the framework's development. Additionally, methodologies for data classification, risk assessment, and continuous monitoring ensure that privacy risks are systematically identified and mitigated. Privacy-preserving technologies such as differential privacy and homomorphic encryption enable secure data analysis while protecting individual privacy. Employee training and awareness initiatives foster a culture of privacy and security within organizations. The framework also includes robust incident response procedures to address security breaches promptly. By implementing the IBPMM framework, organizations can effectively balance the benefits of Big Data analytics with the imperative to preserve information with privacy concerns.

***Index Terms*—Cybersecurity, Big Data, Privacy Management, Framework, Data Encryption, Access Controls, Anonymization Techniques, Threat Detection, Incident Response, Regulatory Compliance.**

I. Introduction

In the rapidly evolving digital landscape, the proliferation of big data has revolutionized the way information is generated, collected, and analyzed across various sectors. Big data analytics has unlocked unprecedented opportunities for organizations to extract valuable insights, enhance decision-making processes, and drive innovation. However, amidst the wealth of data lies a pressing concern – the protection of individual privacy [1]. As the volume, velocity, and variety of data continue to expand exponentially, ensuring robust privacy management measures becomes paramount to safeguarding sensitive information from unauthorized access, misuse, and exploitation.

The intersection of big data and cybersecurity presents a complex landscape where the pursuit of data-driven insights must be balanced with the imperative to uphold privacy rights and regulatory compliance [2]. Cybersecurity, traditionally focused on safeguarding networks, systems, and data from cyber threats, has emerged as a critical enabler of enhanced privacy management practices within the realm of big data analytics. This convergence underscores the indispensable role of cybersecurity in fortifying the resilience of data ecosystems and fostering trust among stakeholders.

The advent of digital technologies, coupled with the proliferation of internet-connected devices, has precipitated an unprecedented data deluge, giving rise to the era of big data. Characterized by the generation of vast volumes of structured and unstructured data from diverse sources, including social media, IoT devices, sensors, and transactional records, big data has emerged as a potent resource for driving innovation and unlocking new business opportunities [3].

However, this proliferation of data has brought forth profound concerns regarding privacy infringement and data misuse. Individuals are increasingly apprehensive about the collection, storage, and analysis of their personal information, as evidenced by mounting public scrutiny and regulatory interventions such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). These regulatory frameworks underscore the imperative for organizations to implement robust privacy management measures and uphold stringent data protection standards to mitigate privacy risks and preserve individual rights.

Despite the recognition of privacy as a fundamental right and the enactment of regulatory frameworks aimed at safeguarding personal data, organizations encounter numerous challenges in effectively managing privacy within the realm of big data analytics. One of the primary challenges lies in reconciling the tension between data utility and privacy preservation [4] [5]. While organizations seek to derive actionable insights from vast datasets to enhance operational efficiency and gain competitive advantage, they must navigate the inherent trade-offs between data anonymization and data utility. The process of anonymizing data to protect individual identities often entails sacrificing certain attributes or reducing data granularity, thereby diminishing its analytical value.Figure 1 illustrates the cybersecurity measures implemented to safeguard and preserve sensitive information [6].

Figure 1. Cybersecurity for Preserve Information

Additionally, the decentralized nature of big data ecosystems poses challenges in ensuring end-to-end data protection and governance [7] [8]. With data being generated, stored, and processed across disparate systems and cloud environments, organizations face complexities in implementing cohesive privacy controls and maintaining visibility into data flows. The lack of interoperability standards further exacerbates these challenges, hindering seamless data governance and compliance across heterogeneous IT landscapes. Recognizing the symbiotic relationship between cybersecurity and privacy, organizations are increasingly leveraging cybersecurity capabilities to bolster their big data privacy management practices [9] [10]. By implementing robust access controls, encryption mechanisms, and threat detection solutions, organizations can mitigate the risk of data breaches and unauthorized disclosures, thereby enhancing privacy protection.

II.RELATED STUDY

It was common for real-world datasets to have associated data. Unfortunately, data correlation was a feature of datasets that had previously been ignored by existing data privacy methods. Privacy breaches were triggered by this data linkage, which went undiscovered by the majority of researchers. A higher degree of connection among data raised the likelihood of such assaults, which in turn enhanced the likelihood of homogeneity, background knowledge, and linkage breaches, all of which led to such privacy leaks. The enormous number of real-world datasets—what was called "Big Data"—only served to amplify this issue. A number of academics had put forward techniques that used data privacy algorithms, correlation analysis, and machine learning models to stop privacy leaks in massive datasets caused by correlation. The association among data was initially examined in the present proposed effort. Two methods for analyzing correlations in data were looked at: distance correlation analysis and mutual information correlation analysis [11]. For data with several dimensions, it was discovered that distance correlation analysis provided better results. Following the computation of the correlation, the data was partitioned into blocks. To guarantee the data privacy demands, the differential privacy technique was applied. Data usefulness, mean average error, change with data size, and privacy budget values were among the many parameters that were used to obtain the findings. When compared to previous studies, the results demonstrated that the suggested technique yielded more useful data. The suggested technique also offered data privacy assurances that were on par with the other outcomes. As a result, the suggested approach improved data usefulness while still meeting the necessary data privacy obligations.

Since big data and AI enabled the prediction of third parties using the anonymous data of numerous individuals, they provided a new problem for data protection. Predicted data included things like income, gender, age, health, sexual orientation, race/ethnicity, and so forth. Such "predictive analytics" uses were based on comparing the user's behavior data (such as use, tracking, or activity data) with data from many other users, sometimes anonymized, that had been processed using ML. The essay began by pointing out that there was a great risk of predictive analytics being utilized in a way that caused social exclusion, discrimination, and inequality. Indeed, the use of anonymized bulk data occurred in a mostly uncontrolled domain, and existing data protection legislation in the EU did not govern these possibilities. The phrase "predictive privacy" described a strategy for protecting personal information from the potential misuse of predictive analytics [12]. When large amounts of people's data were used to make predictions about individuals without their consent or awareness, it violated their predictive privacy.

Big data analysis relied heavily on deep learning. Since the cloud provided powerful computing resources and storage locations, it was commonly used to perform big data analytic jobs. However, the need for data owners to preserve their privacy ran counter to the open cloud model. A workable solution was required to utilize cloud resources for data training while protecting user privacy. In order to tackle this preservation challenge, the research proposed a privacy-preserving deep learning model

(PPDLM) [13]. A homomorphic encryption (HE) method was first used to encrypt the data to protect its privacy. On top of that, unlike homomorphic encryption, the sigmoid function, which was used by the deep learning algorithm as its activation function, processed non-addition and non-multiplication operations using the least-squares approach. In conclusion, the experimental results demonstrated that PPDLM significantly impacted the safeguarding of data privacy information. The computational efficiency of PPDLM was higher than that of NPPDLM, the Non-Privacy Preserving Deep Learning Model.

Machine learning, artificial intelligence, and the internet made many things easier to do, but they also exposed users' personal information to substantial security threats. However, problems existed with the conventional approach to BD privacy and security. For instance, typical models' protection effects weren't appropriate, and the privacy protection pace was too sluggish when confronted with data privacy attacks and large data information intrusion attacks. It was challenging to safeguard existing data privacy with the BD security and privacy protection architecture due to its limitations. A thriving area of study was the development of privacy and security models for BDs that used picture encryption algorithms. Incorporating an image encryption method into the BD security and privacy protection paradigm was possible due to its benefits in data encryption. A chaotic system-based technique for picture encryption was presented in this study [14]. This approach achieved an improved reduction of the picture's pixel correlation coefficient and enhanced encryption of the user's image privacy and data. When tested against tuple assaults, the conventional BD security and privacy protection model had a success rate of 75.15 percent, while the model built using an image encryption method based on a chaotic system had an even higher success rate of 82.25%.

Several issues with network security stemmed from the green light for information retrieval on the Internet. To detect intrusions or other forms of unauthorized access to protected networks, intrusion recognition was an important area of network security research. In those days, intrusion detection had established itself as a serious field. In order to increase system accuracy and decrease the false-positive fraction, research had focused on several datasets. To solve some of the problems with abuse and anomaly detection, that paper suggested a novel intrusion detection system that used deep learning and big data analytics. To detect hostile or unauthorized activity and enable a reaction during a breach of confidentiality, the suggested approach could detect any unusual network activity [15]. The distributed and parallel big data analytics platform was utilized by the suggested system. In addition to enhancing accuracy, parallel and distributed systems shortened training times. The data was correctly labeled as normal or abnormal according to the experiment. With a recognition percentage of 96.11%, the suggested technique significantly improved overall recognition accuracy compared to current tactics.

III. Methodology

Framework Development:
The development of the comprehensive framework for IBPMM integrating cybersecurity and privacy management began with a structured approach, utilizing methodologies such as the NIST Cybersecurity Framework and guidelines from the HIPAA Privacy Rule. The objective was to establish a framework that addressed the unique challenges faced by healthcare organizations in preserving the privacy and security of patient data within Big Data environments while ensuring compliance with regulatory requirements. Key components were identified, including data encryption, access controls, anonymization techniques, threat detection, and incident response protocols. These components were carefully tailored to the healthcare sector, taking into account industry best practices and specific regulatory frameworks such as HIPAA. The framework provided a structured roadmap for healthcare organizations to follow in implementing robust cybersecurity and privacy management measures to protect sensitive patient information. Figure 2 shows the block diagram of suggested approach.

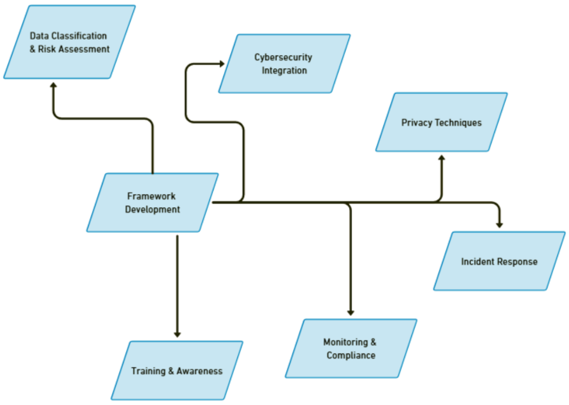

Figure 2. Block Diagram of Proposed Model

Data Classification and Risk Assessment:
A methodology was developed for classifying healthcare data based on its sensitivity and potential privacy risks. This involved categorizing data types such as demographics, medical history, and treatment records, and assigning risk levels accordingly. Thorough risk assessments were conducted to identify potential vulnerabilities and threats to patient data privacy, considering factors such as data storage, transmission, and processing. High-risk areas were prioritized for data protection measures, ensuring that resources were allocated effectively to mitigate the most significant threats. By systematically classifying and assessing data risks, healthcare organizations could better understand their data landscape and implement targeted measures to

safeguard sensitive information.

Cybersecurity Integration:
Advanced cybersecurity measures were integrated into Big Data processing pipelines to protect patient data from unauthorized access and breaches. Encryption mechanisms such as AES-256 were implemented to secure data both at rest and in transit, ensuring that patient information remained protected even if systems were compromised. Robust access controls, including role-based access control (RBAC) and multifactor authentication (MFA), were put in place to restrict access to authorized personnel only. Intrusion detection and prevention systems (IDPS) were deployed to monitor network traffic and detect anomalies indicative of potential security breaches. By integrating these cybersecurity measures into Big Data environments, healthcare organizations could significantly reduce the risk of data breaches and protect patient privacy.

Privacy-Preserving Techniques:
Privacy-enhancing technologies (PETs) were implemented to enable data analysis while preserving patient privacy. Differential privacy techniques were utilized to anonymize aggregate data, ensuring that individual patient information remained protected. Homomorphic encryption was employed to perform computations on encrypted data without decrypting it, allowing for secure data analysis while maintaining confidentiality. Secure multi-party computation (MPC) protocols were explored to enable collaborative data analysis across multiple parties without exposing raw data. Additionally, data anonymization methods such as k-anonymity or generalization were used to de-identify individual patient records while retaining their utility for analysis purposes. These privacy-preserving techniques ensured that patient data could be leveraged for insights without compromising individual privacy.

Continuous Monitoring and Compliance:
Mechanisms were established for continuous monitoring of data usage and access patterns, with real-time monitoring of access logs and data flows within Big Data platforms. Logging and auditing capabilities were implemented to track data access and processing activities, ensuring accountability and transparency. Regular security assessments and vulnerability scans were conducted to identify potential weaknesses in the infrastructure and address them promptly. Compliance with HIPAA and other regulatory requirements was ensured through periodic audits and assessments, with mechanisms in place to address any non-compliance issues. By continuously monitoring data usage and ensuring compliance with regulations, healthcare organizations could maintain the integrity and confidentiality of patient information.

Employee Training and Awareness:
Comprehensive training programs were developed to educate healthcare staff on data privacy best practices, cybersecurity protocols, and compliance requirements under HIPAA. Training sessions covered topics such as patient data privacy regulations, cybersecurity threats, and the organization's policies and procedures for safeguarding sensitive information. Role-specific training was provided to different staff members, including healthcare providers, administrative staff, and IT personnel, ensuring that each employee understood their responsibilities in protecting patient data. A culture of privacy and security awareness was fostered throughout the organization, with open communication channels for reporting security incidents or concerns. By investing in employee training and awareness, healthcare organizations could empower their workforce to play an active role in safeguarding patient information.

Incident Response and Remediation:
An incident response plan (IRP) was developed to address security incidents related to patient data breaches promptly. The plan outlined clear procedures for identifying, containing, and mitigating security incidents, with an incident response team tasked with coordinating response efforts. Communication protocols were established for notifying affected parties, including patients, regulatory authorities, and business partners, in accordance with HIPAA breach notification requirements. Post-incident reviews were conducted to analyze the root causes of security incidents and identify areas for improvement in the incident response process.

In conclusion, the development and implementation of the comprehensive framework for IBPMM integrating cybersecurity and privacy management provided healthcare organizations with a structured approach to preserving the privacy and security of patient data within Big Data environments. By systematically classifying and assessing data risks, integrating advanced cybersecurity measures, implementing privacy-preserving techniques, continuously monitoring data usage, investing in employee training and awareness, and developing robust incident response protocols, healthcare organizations could effectively protect sensitive patient information while ensuring compliance with regulatory requirements such as HIPAA.

IV. RESULTS AND DISCUSSIONS

The working principle of the comprehensive framework for IBPMM integrating cybersecurity and privacy management revolves around a holistic approach to safeguarding sensitive data within Big Data environments, particularly within the healthcare sector. At its core, the framework operates on the principle of proactive risk management and continuous improvement. It begins with a thorough understanding of the data landscape, where data is classified based on its sensitivity and potential privacy risks. Through rigorous risk assessments, vulnerabilities and threats to data privacy are identified and prioritized, guiding the allocation of resources towards the most critical areas.

Table 1. Cybersecurity Integration

Security Measure	**Implementation Status**	**Completion Date**
Data Encryption	Implemented	5/15/2023
Access Controls	In Progress	-
IDS	Implemented	4/30/2023
Multi-factor Authentication	Planned	6/30/2023
Security Patch Management	Implemented	3/20/2023

Table 1 outlines the cybersecurity integration efforts within the organization. As depicted, several security measures have been strategically implemented or are currently in progress. Data Encryption, ensuring data confidentiality, was successfully implemented by May 15, 2023. Security Patch Management, vital for mitigating vulnerabilities, was completed by March 20, 2023. Intrusion Detection Systems (IDS) are operational since April 30, 2023, enhancing threat detection capabilities. Access Controls are currently being worked on, aiming to fortify system access security. Moreover, Multi-factor Authentication is slated for implementation by June 30, 2023, further bolstering authentication mechanisms for enhanced data protection.

Table 2. IDS Performance

Metric	**Value**
FPR	0.05
FNR	0.02
DR	0.98
Precision	0.95
Recall	0.98
Accuracy	0.97
MTTD	20 minutes
MTTR	45 minutes

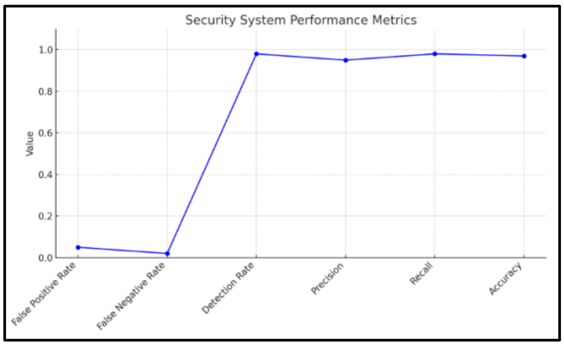

Figure 3. Security System Performance Metrics

In Table 2 and Figure 3 and 4, the performance metrics of the Intrusion Detection System (IDS) are meticulously detailed. With a False Positive Rate (FPR) of 0.05 and False Negative Rate (FNR) of 0.02, the IDS exhibits high precision and detection capabilities, evidenced by its Detection Rate (DR) of 0.98 and Precision of 0.95. Additionally, the system achieves a commendable Recall of 0.98 and an overall Accuracy of 0.97, signifying its reliability in identifying and mitigating security threats. Notably, the Mean Time to Detect (MTTD) stands at 20 minutes, demonstrating the system's efficiency in promptly identifying potential intrusions, while the Mean Time to Respond (MTTR) of 45 minutes highlights its effectiveness in swiftly addressing security incidents.

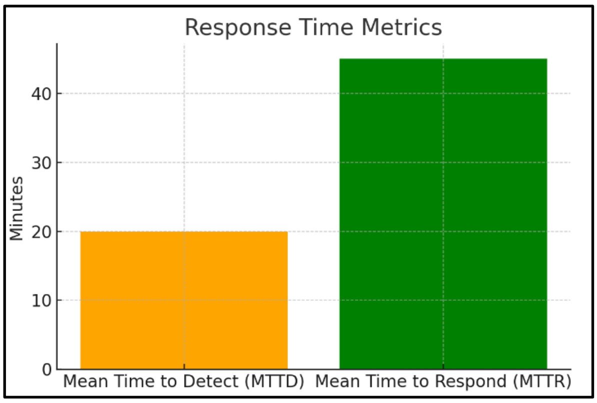

Figure 4. Response Time Metrics

Advanced cybersecurity measures, including encryption, access controls, and threat detection, are integrated into Big Data processing pipelines to protect data from unauthorized access and breaches. Privacy-preserving techniques such as differential privacy and data anonymization enable secure data analysis while preserving individual privacy. Continuous monitoring and compliance mechanisms ensure ongoing adherence to regulatory requirements, while employee training and awareness initiatives foster a culture of security throughout the organization.

Table 3. Performance Comparison

IDS	**FPR**	**FNR**	**DR**	**Throughput (Mbps)**
Snort	0.01	0.03	0.97	1000
Suricata	0.02	0.05	0.95	900
Bro/Zeek	0.03	0.04	0.96	800
Snort + Suricata	0.005	0.02	0.98	1100

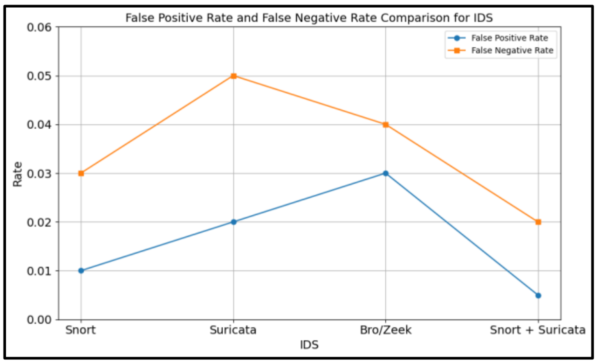

Figure 5. FPR & FNR Comparison

Table 3 and Figures 5 and 6 provides a comparative analysis of various IDS solutions, including Snort, Suricata, Bro/Zeek, and a combined Snort and Suricata setup. Each IDS is evaluated based on key performance metrics such as FPR, FNR, DR, and Throughput (Mbps). Snort demonstrates the lowest FPR of 0.01 and FNR of 0.03, along with a high DR of 0.97 and impressive Throughput of 1000 Mbps. Suricata and Bro/Zeek also exhibit strong performance, albeit with slightly higher FPR and FNR. The combined Snort + Suricata configuration showcases the best overall performance, achieving the lowest FPR of 0.005, the lowest FNR of 0.02, a high DR of 0.98, and the highest Throughput of 1100 Mbps, indicating its superior capability in effectively identifying and mitigating security threats with minimal impact on network performance.

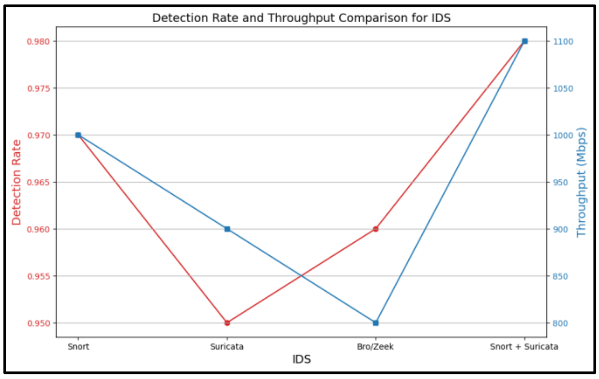

Figure 6. DR and Throughput Comparison

In the event of a security incident, a robust incident response plan is activated, facilitating prompt identification, containment, and mitigation of the breach, followed by thorough post-incident reviews to improve future response efforts. Overall, the working principle of the framework emphasizes proactive risk management, continuous monitoring, and a multi-faceted approach to data protection, ensuring the confidentiality, integrity, and availability of sensitive information within Big Data environments.

Table 4. Data Encryption AlgorithmsPerformance

Algorithm	Key Length	Encryption Time (ms)	Decryption Time (ms)	Throughput (Mbps)
AES-128	128 bits	2.5	2.2	1000
AES-256	256 bits	3.2	2.9	800
RSA-2048	2048 bits	10.5	9.8	400
ECC (secp256k1)	256 bits	4	3.5	1200

Table 4 presents a comparative analysis of the performance of different data encryption algorithms, including AES-128, AES-256, RSA-2048, and ECC (secp256k1), across various metrics such as key length, encryption time, decryption time, and throughput (Mbps). AES-128 and AES-256, both symmetric encryption algorithms, demonstrate efficient encryption and decryption processes, with AES-128 showcasing slightly faster performance. RSA-2048, an asymmetric encryption algorithm, exhibits longer encryption and decryption times due to its larger key size. ECC (secp256k1), another asymmetric encryption algorithm, offers faster encryption and decryption times compared to RSA-2048, along with significantly higher throughput, making it particularly suitable for scenarios requiring high-speed data encryption and decryption operations while maintaining robust security.

V.Conclusion and Future Scope

In conclusion, the Cybersecurity Enabled Improved Big Data Privacy Management Measures (IBPMM) framework offers a structured and comprehensive approach to tackling the multifaceted challenges of preserving privacy in the ever-expanding realm of Big Data, especially within sensitive sectors such as healthcare. By seamlessly integrating advanced cybersecurity measures with privacy management strategies and maintaining stringent adherence to regulatory requirements, IBPMM provides organizations with a robust solution to safeguarding sensitive information while leveraging the transformative insights derived from vast datasets. Through meticulously designed methodologies for data classification, risk assessment, and continuous monitoring, the framework ensures that privacy risks are systematically identified, evaluated, and mitigated, thereby fostering compliance and nurturing the trust of stakeholders. Furthermore, the incorporation of cutting-edge privacy-preserving technologies and proactive employee training initiatives underscores the unwavering commitment to upholding privacy principles and fortifying organizational resilience against potential threats. As data continues to burgeon in volume and complexity, IBPMM lays a sturdy foundation for future advancements in privacy management and cybersecurity, promising continued innovation in safeguarding information imbued with privacy concerns. Looking forward, the IBPMM framework presents a plethora of opportunities for further development and refinement to meet the evolving demands of the digital landscape. Future research endeavors could concentrate on enhancing the scalability and interoperability of the framework to accommodate the ever-expanding data ecosystems and emerging technologies seamlessly. In essence, IBPMM lays a sturdy foundation for future research and development endeavors aimed at advancing privacy management and cybersecurity paradigms in the intricate tapestry of Big Data environments, thereby fortifying organizations' capabilities to navigate the complexities of the digital age confidently.

REFERENCES

[1] Dr. M.John Basha1, at al., (2023), "Privacy-Preserving Data Mining and Analytics in Big Data", E3S Web of Conferences 399, 04033, ICONNECT-2023, DOI: 10.1051/e3sconf/202339904033

[2] P. K. et al "An Secure and Low Energy Consumption based Intelligent Street Light Managing System using LoRa Network," ICECA, Coimbatore, India, 2022, pp. 638-645, doi: 10.1109/ICECA55336.2022.10009408.

[3] Poltavtseva Maria, et al., (2023), "Data protection in heterogeneous big data systems", JCVHT, 19. 1-8, DOI: 10.1007/s11416-023-00472-3

[4] G. A et al "Efficient Internet of Things Enabled Smart Healthcare Monitoring System Using RFID Security Scheme" Intelligent Technologies for Sensors, 1st Edition, 2023, Apple Academic Press, ISBN: 9781003314851

[5] M. Abhineswari and R. Priyadarshini, "Analyzing Large-Scale Twitter Real Time Streaming Data with Manifold Machine Learning Algorithms in Apache SPARK," (ICDSAAI), Chennai, India, 2023, pp. 1-9, doi: 10.1109/ICDSAAI59313.2023.10452549.

[6] Jie Chen, et al., (2022), "Holistic big data integrated artificial intelligent modeling to improve privacy and security in data management of smart cities", Microprocessors and Microsystems, Volume 81, 103722, ISSN 0141-9331, DOI: 10.1016/j.micpro.2020.103722

[7] Mehdi Seydali, et al., (2024), "Streaming traffic classification: a hybrid deep learning and big data approach", Cluster Comput (2024), DOI: 10.1007/s10586-023-04234-0

[8] Maysaa Khalil, et al., (2023), "Privacy-Preserving federated learning: An application for big data load forecast in buildings", Computers & Security, Volume 131, 103211, ISSN 0167-4048, DOI: 10.1016/j.cose.2023.103211

[9] YanjunZuo, et al., (2023), "Big data and big risk: a four-factor framework for big data security and privacy", IJBIS, 42:2, 224-242, Volume 42, Issue 2, ISSN: 1746-0972, eISSN: 1746-0980, DOI: 10.1504/IJBIS.2023.128648

[10] IraklisVarlamis, et al., (2023), "Using big data and federated learning for generating energy efficiency recommendations", Int J Data Sci Anal 16, 353–369, DOI: 10.1007/s41060-022-00331-2

[11] Sreemoyee Biswas, et al., (2022), "Enhancing correlated big data privacy using differential privacy and machine learning", J Big Data 10, 30, DOI: 10.1186/s40537-023-00705-8

[12] R. Mühlhoff, et al., (2023), "Predictive privacy: Collective data protection in the context of artificial intelligence and big data", Big Data & Society, 10(1), DOI: 10.1177/20539517231166886

[13] Yongkai Fan, et al., (2023), "Privacy-preserving deep learning on big data in cloud", China Communications, vol. 20, no. 11, pp. 176-186, DOI: 10.23919/JCC.ea.2020-0684.202302

[14] Binjie Hua, et al., (2023), "Big data security and privacy protection model based on image encryption algorithm", Soft Comput, DOI: 10.1007/s00500-023-08548-4

[15] Muhammad Babar, et al., (2023), "An Improved Big Data Analytics Architecture for Intruder Classification Using Machine Learning", Security and Communication Networks, vol. 2023, Article ID 1216192, 7 pages, DOI: 10.1155/2023/1216192