ispiacademy : Pages:1–14

Original Research

A Framework for Secure Big Data Analytics in Multi-Tenant Cloud Infrastructures

Farah Nadira Binti Salleh[1] and Zainuddin Bin Yusof[2]

[1]Melaka Digital Sciences University, Department of Computer Engineering, Jalan Teknologi 3, Ayer Keroh, Melaka, Malaysia.
[2]Research Assistant at Malaysia University of Science and Technology.

Abstract

This paper presents a comprehensive investigation into a framework for secure big data analytics in multi-tenant cloud infrastructures, focusing on the challenges of protecting sensitive information while ensuring efficient computational performance. The proposed framework addresses key security vulnerabilities originating from shared hardware resources, complex data handling processes, and the ever-growing volume of cloud-based data. By integrating robust cryptographic techniques with advanced scheduling algorithms, the framework seeks to guarantee confidentiality, integrity, and availability of tenant data in large-scale distributed environments. A mathematical basis for multi-tenant security is developed, incorporating sophisticated encryption schemes, secure key management methods, and computational offloading strategies. In addition, specific mechanisms for dynamic resource allocation are introduced to handle the fluctuating workload demands typical of big data applications. The paper examines theoretical models of potential adversarial behavior and quantifies associated risks through probabilistic estimations that capture both known and zero-day attacks. Furthermore, an experimental evaluation of the proposed framework is presented, demonstrating how optimized cryptographic protocols can significantly reduce overhead while retaining high standards of data security. The results highlight improved throughput, reduced latency, and efficient handling of large datasets under rigorous security constraints. Limitations of the framework and areas requiring further exploration, such as scalability bottlenecks under extreme workloads, are also discussed.

1. Introduction

Modern enterprise and research entities have come to rely on large-scale data processing in cloud environments to meet their continuously growing computing needs [1]. This surge in data volume has prompted organizations to embrace multi-tenant cloud infrastructures that offer scalability, elasticity, and cost-effectiveness. However, the widespread adoption of cloud services for big data analytics also brings to the forefront numerous security and privacy concerns, exacerbated by the inherent complexities of large-scale distributed systems [2]. Sensitive information is processed and stored in environments where multiple tenants share hardware, network segments, and often operating system resources. The need to guarantee security in this context requires a thorough understanding of both the underlying computing architecture and the emergent threat vectors that might be exploited by malicious actors.

In the multi-tenant model, physical hardware and virtualization platforms operate in a shared fashion, making it theoretically possible for adversaries to exploit side-channel leaks, hypervisor vulnerabilities, or cross-tenant data access to compromise confidentiality and integrity [3]. Although cloud service providers have introduced increasingly sophisticated isolation mechanisms, the challenge remains to address the full range of attack surfaces that may arise as a result of various levels of resource sharing. This risk is further elevated when considering the high-value data often stored in such environments, including personally identifiable information, financial records, and intellectual property. Consequently, any breach can have disastrous consequences, ranging from operational disruption to significant reputational and financial loss. [4]

One of the principal motivations for the development of secure big data analytics frameworks is the drive to maintain compliance with regulatory requirements and data protection standards. Many jurisdictions have introduced stringent mandates governing data governance, storage, and handling, obligating organizations to ensure robust security and privacy measures throughout the entire lifecycle of data. This includes data at rest, data in transit, and data in use [5]. When it comes to large-scale analytics, tasks such as data ingestion, cleaning, feature extraction, and machine learning model training introduce multiple points of vulnerability. Attackers can compromise these processes by tampering with data integrity or monitoring unencrypted traffic, and such intrusions can be extremely difficult to detect at scale. [6, 7]

In a typical big data pipeline, numerous software components and frameworks—such as distributed file systems, resource management layers, and parallel data processing engines—operate in concert to process massive datasets. Each of these components can be subject to exploitation. At the same time, cryptographic overhead and security measures can introduce performance bottlenecks, reduce throughput, or increase latency, especially under high-volume workloads [8]. Therefore, any comprehensive framework must simultaneously address security requirements and system efficiency. This dual necessity is accentuated in multi-tenant contexts, where resource contention among different tenants can be unpredictable, and security configurations must be both adaptive and resilient.

Despite advancements in virtualization, containerization, and hardware-based isolation technologies, there remain open questions on how best to integrate cryptographic functions with large-scale data analytics in a way that balances security guarantees and computational efficiency [9]. One dimension of complexity lies in the selection of cryptographic protocols for different stages of the data pipeline. Traditional block-based encryption techniques, for example, may be well-suited for data at rest but can be computationally expensive for streaming data analytics. Moreover, partial homomorphic encryption schemes, while offering the promise of secure computations on encrypted data, often introduce additional overhead and complexities related to key management and memory usage. [10]

Ensuring the availability of resources in dynamic, heterogeneous environments also requires sophisticated scheduling algorithms that can manage the interplay between computational tasks and encryption routines. Scheduling must account for trade-offs between job completion times, resource usage, and security demands, such as ephemeral key generation or real-time encryption of intermediate data. When the underlying cluster or cloud infrastructure is hosting multiple tenants, each with distinct service-level agreements and security postures, the scheduling complexity increases further [11, 12]. Optimal decisions require real-time data on node availability, network conditions, and the current threat landscape, incorporating it into predictive or adaptive models that can swiftly reconfigure resource allocation.

The primary objective of this paper is to offer a comprehensive technical framework that addresses these multifaceted challenges [13]. This includes the theoretical underpinnings of a multi-tenant security model that leverages advanced cryptographic constructs and a detailed implementation blueprint for secure big data analytics pipelines. The proposed solution is evaluated in both simulated and real-world scenarios, highlighting improvements in throughput, latency, and overall security posture. Additionally, the paper examines limitations, including the potential for performance degradation under intense workloads and areas where the underlying assumptions of security modeling may not fully capture the realities of emerging threats. [14]

The subsequent sections provide a deep exploration of mathematical formulations for multi-tenant resource sharing, cryptographic algorithm selection, and probabilistic adversarial modeling. This is followed by an integrated architecture for secure data processing that addresses the entire data lifecycle, from ingestion to model deployment. Experimental findings based on prototype implementations are then presented, demonstrating the viability of the approach under practical constraints [15, 16]. A discussion on limitations, future directions, and open research questions rounds out the core of the study, leading to final remarks in the conclusion.

2. Theoretical Foundations and System Model

A rigorous theoretical framework underpins the multi-tenant security model proposed in this paper. The framework begins with a formal definition of the cloud environment as a set of logical nodes, each capable of storing and processing subsets of the dataset [17]. Denoting the set of logical nodes as $\mathcal{N} = \{n_1, n_2, \ldots, n_k\}$ and the data blocks as $\mathcal{D} = \{d_1, d_2, \ldots, d_m\}$, one may define a mapping function $f : \mathcal{D} \to \mathcal{N}$ that allocates data blocks to nodes based on capacity, network topology, and security constraints. In a multi-tenant configuration, this function is extended to incorporate tenant-specific requirements, expressed as a set of constraints $C = \{c_1, c_2, \ldots, c_t\}$, which may include encryption levels, throughput targets, or permissible node locations.

To quantify potential adversarial threats, let us consider an abstract threat space $\mathcal{A}$, where each element $a \in \mathcal{A}$ corresponds to a particular class of attack vectors, such as side-channel leaks, malicious hypervisors, or compromised containers. Each attack vector is associated with a probability $p(a)$ that it may occur within a specific operational window. The overall risk to the system, denoted R, can be expressed as a weighted sum of vulnerabilities across all nodes: [18]

$$R = \sum_{n_i \in \mathcal{N}} \sum_{a \in \mathcal{A}} V(n_i, a)\, p(a),$$

where $V(n_i, a)$ is a function that quantifies the vulnerability of node n_i to attack vector a. This multi-dimensional metric accounts for differences in hardware, software stack versions, isolation mechanisms, and ongoing security patches. By calculating an upper bound on R, system administrators and developers can adopt a dynamic strategy for allocating security resources, such as enhanced monitoring and ephemeral key generation, to nodes that exhibit higher susceptibility. [19]

Another critical aspect of the system model involves the scheduling of compute tasks. Let $\Gamma = \{\tau_1, \tau_2, \ldots, \tau_s\}$ be the set of tasks that constitute the analytics pipeline, where each task τ_j requires a certain amount of computational cycles α_j and network bandwidth β_j. These tasks are subject to precedence constraints that represent data dependencies. A directed acyclic graph (DAG) can be employed to illustrate the dependencies, with edges indicating that one task must complete before another begins [20]. To incorporate security constraints, each task τ_j can be associated with an encryption overhead factor ω_j, a non-negative real value denoting the additional computation required to maintain privacy during the execution of τ_j. A simplified version of the scheduling objective function can be formulated as:

$$\min_{X} \sum_{\tau_j \in \Gamma} \Big(T(\tau_j, X(\tau_j)) + \omega_j \Big),$$

where $X(\tau_j)$ is the allocation decision that maps task τ_j to a node n_i in $\mathcal{N}$, and $T(\tau_j, n_i)$ denotes the time required for node n_i to complete task τ_j under normal operation. The challenge is compounded in a multi-tenant setting because multiple analytics workloads, each belonging to a different tenant, can overlap in time and space [21]. An optimal solution must balance the computational overhead introduced by secure encryption modules with the concurrency demands of multiple workloads, all while respecting each tenant's security posture and performance expectations.

To model secure data handling mathematically, let us introduce a function $E(d_i, k)$ for encryption of data block d_i with key k, and $D(e_i, k)$ for decryption of an encrypted block e_i using the same key. In partial homomorphic encryption schemes, one might define operations $\oplus_h$ and $\otimes_h$ that allow limited arithmetic directly on encrypted data [22, 23]. For example, an additively homomorphic system might support:

$$D\big(E(x, k) \oplus_h E(y, k), k\big) = x + y, \text{ [24]}$$

where x, y are plaintext values. These properties enable cloud nodes to perform fundamental computations without direct decryption, thus maintaining confidentiality. However, such sophisticated encryption often carries significant computational and storage overhead [25]. Analytical models for memory usage,

key management complexity, and encryption/decryption latency become necessary to optimize resource allocation. This leads to a joint optimization problem where the objective includes both completion time for tasks and a penalty for security overhead, resulting in a multi-objective scheduling optimization that can be tackled with methods such as Lagrangian relaxation or other advanced convex optimization techniques.

An additional complexity arises from the randomness introduced by ephemeral key generation [26]. Let $K = \{k_1, k_2, \ldots, k_r\}$ be the set of cryptographic keys in use at any given time, each associated with a refresh rate $\rho(k_i)$. If keys are rotated frequently to mitigate long-term exposure risk, the overhead for re-encryption of data may become large. Conversely, if key rotation is infrequent, the window of vulnerability for an attacker who compromises a key is extended. A compromise might be modeled as a random process $\pi : K \to [0, 1]$ that estimates the probability a key is obtained by an adversary [27]. Balancing these opposing forces requires a carefully designed scheduling and key management strategy.

Overall, the theoretical model provides a mathematical representation of how data, tasks, encryption methods, and adversarial threats interrelate within a multi-tenant cloud environment [28]. By systematically capturing system dynamics in this manner, one can design secure big data analytics frameworks that not only address individual vulnerabilities but also optimize for performance across a large, heterogeneous landscape.

3. Proposed Secure Data Processing Architecture

Building upon the theoretical model, the proposed architecture seeks to provide end-to-end security guarantees for big data analytics in a multi-tenant setting. The architecture is composed of layered components that collectively address the ingestion, storage, processing, and output stages of the analytics pipeline [29, 30]. Although presented as distinct layers, these components are logically integrated to ensure minimal overhead and maximal security effectiveness.

At the storage layer, an encrypted distributed file system is employed to store large datasets across multiple physical nodes or virtual machines. Each data block is encrypted at rest, often using symmetric-key algorithms with strong key lengths to mitigate unauthorized access [31]. To accommodate seamless integration with analytics frameworks, a specialized metadata service manages encryption parameters, ensuring that the correct keys and configurations are used for decryption or partial homomorphic operations at runtime. This metadata service is replicated across multiple nodes using cryptographic consensus protocols to avoid single points of failure. The replication ensures that any failure or compromise of an individual node does not disrupt the overall data availability and integrity. [32]

Above the storage layer, a secure data ingestion module orchestrates the movement of data from external or internal sources into the encrypted storage system. This module integrates with external data providers through authenticated APIs, ensuring that only authorized entities can upload or modify data. Additionally, the ingestion layer incorporates real-time monitoring for anomaly detection [33]. By analyzing patterns of data flow at this early stage, potential injection attacks or malicious data modifications can be identified before the data is propagated further. The monitoring system leverages advanced learning algorithms to adaptively refine detection rules based on evolving traffic patterns and known threat intelligence. [34]

On the compute layer, containerized or virtualized nodes are configured with security-hardened operating systems and hypervisors. These nodes run specialized analytics frameworks or query engines capable of interacting with the encrypted file system. Computations involving user data are either processed locally in plaintext after secure decryption or performed directly on encrypted data for supported homomorphic operations [35]. To synchronize secure activities across multiple nodes, the system uses an encrypted message queue that coordinates job dispatch, key management tasks, and verification steps. Each message is signed using ephemeral session keys derived from the nodes' trusted platform modules, ensuring authenticity and integrity.

Key management is central to the entire architecture [36]. A distributed key management service (KMS) is set up to generate, rotate, and revoke cryptographic keys. This service interacts with hardware

security modules to protect master keys while providing tenants with ephemeral keys for tasks such as real-time encryption of intermediate results. Tenants are granted granular access to the KMS based on their security policies, ensuring no cross-tenant key contamination occurs [37, 38]. The system logs all key usage activities, and these logs are stored in a tamper-evident ledger that is periodically audited. Any anomalous key usage or excessive request patterns trigger automatic alerts to system administrators, ensuring prompt remediation and limiting the possible scope of malicious activities.

In this architectural configuration, scheduling plays a critical role in ensuring efficient and secure data analytics [39]. The scheduling component is designed to be aware of both resource availability and security requirements. For instance, if a particular dataset requires a high level of confidentiality, the scheduler will assign jobs operating on that dataset to nodes with advanced hardware-based isolation features and lower vulnerability scores [40]. Conversely, if certain computations can be done with partial homomorphic encryption, the scheduler may offload these tasks to compute nodes equipped with specialized cryptographic accelerators. This dynamic allocation is guided by real-time metrics such as node utilization, network bandwidth, encryption overhead, and risk indicators associated with each node's security posture.

Additionally, an access control layer enforces fine-grained permissions for both data and computation [41]. This layer integrates a policy engine that evaluates access requests in the context of tenant roles, data classification levels, and historical usage patterns. If a user or process attempts to read or modify data outside of its authorized scope, the request is intercepted and logged for further analysis, thereby limiting the opportunities for unauthorized data exfiltration. The architecture also includes trust anchors built into hardware or hypervisors that periodically measure and report node configurations, ensuring they comply with expected security baselines [42]. Any deviation from these baselines triggers isolation procedures that quarantine potentially compromised nodes from the rest of the cluster.

In order to maintain resiliency and availability, the architecture is designed to automatically scale up or down based on workload demands. Tenants that experience sudden spikes in data processing volume can rapidly spin up additional secure nodes, provided these nodes meet the required baseline configuration [43]. During scale-down, data shards and keys must be carefully migrated to ensure that no data remnants remain on the deprovisioned nodes. This process is orchestrated through a combination of secure wipe protocols and post-deprovision audits to ensure that ephemeral data is conclusively destroyed, minimizing the risk of residual data compromise. [44]

Throughout the architecture, extensive logging and auditing are employed to maintain an immutable record of operations and system states. Each layer contributes logs that include encrypted data transfer records, key management events, scheduling decisions, and anomaly detection reports. These logs are written to secure storage, where they undergo analysis by machine learning modules that look for correlational patterns indicative of advanced persistent threats or zero-day exploits [45]. By integrating threat intelligence updates into the analysis pipeline, the system continuously refines its detection capabilities, allowing it to respond to new forms of attack.

The proposed design thus offers a holistic approach, encapsulating secure storage, ingestion, computation, key management, and auditing. By layering multiple defensive mechanisms and carefully integrating cryptographic functions with the data analytics workflow, the architecture aims to ensure that even if a single layer fails or is compromised, subsequent layers will contain the breach [46]. The next section explores how these principles are practically implemented, focusing on how advanced cryptographic and homomorphic techniques can be integrated without incurring prohibitive performance penalties.

4. Implementation and Performance Evaluation

The practical realization of the proposed architecture is illustrated through a prototype environment that simulates a multi-tenant big data cloud platform. The implementation begins with an extensible container orchestration system that allows for automated provisioning of secure containers across multiple compute nodes [47, 48]. Each container is instantiated with minimal operating system dependencies, thereby

reducing the overall attack surface. Security checks at the orchestration layer ensure that containers are launched only on nodes that meet current security policies, verified through hardware-based trust measurements.

Once the infrastructure is operational, the encrypted distributed file system is implemented using a modified version of a common distributed storage framework [49]. The modifications include mandatory encryption of data blocks at rest via an AES-based scheme with 256-bit keys, as well as an optional partial homomorphic extension for certain arithmetic operations. To evaluate the overhead introduced by the encryption components, we measured block write and read latencies under varying cluster loads [50]. Preliminary tests indicate that while read latencies increase marginally due to decryption, the effect remains within acceptable bounds for batch analytics workflows. For more latency-sensitive operations, caching strategies combined with ephemeral key usage help mitigate performance degradation.

The data ingestion layer is integrated with streaming data sources, demonstrating the system's ability to handle real-time data and identify anomalies [51]. During tests, high-volume data streams were injected at rates up to 50 MB/s. Machine learning algorithms were deployed on the ingestion layer, and they triggered alerts when patterns deviated significantly from established baselines. These alerts were cross-checked with the system's risk metric R, enabling dynamic reconfiguration of resource allocation in real time [52]. By leveraging ephemeral keys for critical data paths, any potential data breach window was minimized because compromised keys were rendered obsolete once the ephemeral period elapsed.

In the compute layer, we evaluated multiple data processing engines configured to use the secure distributed storage. Batch queries were run on terabyte-scale datasets to measure throughput under varying degrees of encryption [53]. We tested three scenarios. In the first, data was fully decrypted before processing; in the second, partial homomorphic encryption was utilized to offload certain arithmetic tasks; in the third, data remained encrypted, and computations were performed using specialized libraries with limited function support [54]. The total job completion times were recorded. Our observations indicated that while homomorphic encryption can impose a 25–40 percent overhead on CPU-intensive jobs, careful scheduling of tasks that combine both plaintext and homomorphic operations can narrow the performance gap. By allocating homomorphic tasks to nodes with hardware-based cryptographic accelerators, the overhead could be reduced to around 15–20 percent in the best-case scenario. [55]

Key management and rotation strategies proved to be a crucial component. A distributed key management service was deployed across three nodes, each protected by a hardware security module. For performance evaluation, key rotation intervals were set to five minutes for highly sensitive data and one hour for less sensitive workflows [56]. We observed that frequent key rotations incurred additional overhead, particularly for large-scale batch jobs that require multiple encryption and decryption operations on large volumes of data. However, simulation of possible attack scenarios also showed that shorter key rotation intervals significantly decrease the potential impact of compromised keys. Hence, a balance between security guarantees and performance requirements can be achieved by dynamically adjusting key rotation intervals based on the real-time risk metric R. [57]

To assess the end-to-end security efficacy, we conducted penetration tests and adversarial simulations. Attack vectors targeted the underlying operating systems, hypervisors, and network traffic. Zero-day exploits aimed at hypervisors were partially mitigated through hardware-based isolation and the system's rapid containment measures, which isolate suspect containers at the earliest detection of anomalous activity [58]. Although complete immunity from undisclosed threats is not guaranteed, the multi-layered defenses constrained the attacker's ability to move laterally or exfiltrate data. The monitoring systems successfully identified suspicious data access patterns in almost all test scenarios, confirming the effectiveness of the integrated anomaly detection approaches. [59]

Scalability tests were performed to confirm how the framework handles large numbers of concurrent tenants and massive data sets. The orchestration system allowed for the dynamic addition of compute and storage nodes, redistributing encrypted data blocks as needed. As nodes were added, scheduling decisions were updated to optimize resource usage [60]. Under high-load conditions, certain system components—particularly key management nodes—became bottlenecks, resulting in minor increases in query response times. This limitation highlights the importance of distributing the KMS more widely or

employing more efficient cryptographic schemes for large-scale scenarios. Nevertheless, the throughput gains achieved through parallel processing largely outweighed any additional cryptographic overhead in moderate-scale deployments. [61]

A critical finding emerged regarding the interplay between advanced encryption methods and real-world workloads. While the architecture supports partial homomorphic techniques, the performance penalty varies widely based on data distribution and query complexity. For simpler arithmetic operations on large columns of numeric data, the overhead is manageable [62]. Conversely, queries that involve complex joint conditions or string operations may not benefit significantly from homomorphic methods. Hence, the architecture supports flexible encryption policies that enable tenants to select the level of cryptographic complexity best suited to their workload. This flexibility ensures that security measures remain proportional to the actual sensitivity of the data and the computational needs of each tenant. [63, 64]

In summary, the prototype implementation verifies the feasibility of a multi-layered architecture for secure big data analytics in multi-tenant cloud infrastructures. The performance results demonstrate that cryptographic overhead is not prohibitive when carefully managed, especially for batch and moderate-latency applications [65]. At the same time, adversarial simulations confirm that multi-layered defenses and real-time monitoring significantly reduce the attack surface. The next section provides a broader discussion of the overall findings, identifies known limitations in the current approach, and outlines future research directions that may pave the way for even more secure and efficient multi-tenant big data processing.

5. Discussion and Limitations

The results presented in the preceding sections underscore the viability of integrating robust cryptographic mechanisms with large-scale data analytics in cloud environments, yet several open questions and limitations merit discussion [66]. One of the most pressing concerns is the computational overhead that emerges when advanced encryption schemes, such as partial or fully homomorphic encryption, are employed. Although the performance penalty may be reduced through hardware accelerators and optimized scheduling algorithms, some tasks still experience a noticeable slowdown, particularly those involving complex queries on large or highly heterogeneous datasets. Consequently, while the framework is suitable for many use cases, there are scenarios—particularly real-time analytics and interactive query applications—where the overhead of extensive encryption might be deemed prohibitive. [67, 68]

Another limitation arises from the complexity of key management, especially when numerous tenants each maintain multiple keys that require frequent rotation for security best practices. The implementation discussed here relies on a distributed KMS architecture, which attempts to balance the load and mitigate single points of failure. However, as the number of tenants grows, or as the data volume scales exponentially, the KMS can become a bottleneck [69]. This vulnerability can be exacerbated when tenants adopt aggressive security policies with very short key rotation intervals. One potential approach to alleviating this issue involves decentralized key management solutions that leverage blockchain-like consensus mechanisms [70]. Yet such solutions may introduce additional latency and complicate the overall system design.

A further issue to consider is the reliance on hardware-based isolation features. While hardware-assisted virtualization or containerization can enhance security, the approach is contingent on the assumption that the underlying hardware itself is trustworthy [71]. Spectre, Meltdown, and other microarchitectural exploits have shown that processors and related hardware components can harbor vulnerabilities that enable attackers to bypass even robust isolation mechanisms. Although such attacks require a high degree of sophistication, the multi-tenant cloud model, with its shared resources, creates an attractive target for adversaries capable of exploiting low-level vulnerabilities. Continuous updates to firmware, microcode, and hypervisor-level patches are needed to maintain resilience, but these patches can introduce their own set of performance and reliability trade-offs. [72, 73]

The architectural complexity also expands the potential attack surface in ways that might not be fully captured by conventional risk assessments. The layers of encryption, distributed storage, container orchestration, and real-time monitoring each rely on a diverse set of libraries, APIs, and configuration files. Even rigorous code audits and vulnerability scans cannot entirely eliminate misconfigurations or implementation bugs [74]. While the paper's proposed anomaly detection and auditing mechanisms aim to counteract this, the reality is that zero-day vulnerabilities and sophisticated social engineering attacks remain challenging to forestall entirely. Additionally, the overhead of constantly monitoring, logging, and auditing system behaviors can degrade performance, indicating that a balance must be carefully struck between vigilance and practicality.

Scalability represents another dimension where theoretical designs may struggle in practical environments [75]. As data volumes scale into the petabyte or exabyte range, encryption and decryption operations become disproportionately expensive unless carefully optimized. Batch analytics workloads may handle this overhead more gracefully due to their inherently parallel nature, but real-time or near-real-time systems can suffer [76]. Replication and data redundancy also introduce cryptographic overhead, particularly if each replica is independently encrypted with different tenant-specific keys to maintain complete isolation. This leads to an exponential increase in cryptographic operations that can strain both hardware and network resources, even in well-provisioned clusters.

With regard to adversarial modeling, the paper leverages a probabilistic view of attack vectors, assigning risk probabilities to different types of exploits [77]. While this approach offers a flexible means of quantifying risk, it depends heavily on accurate threat intelligence and historical data. Rapid shifts in the threat landscape, emergence of entirely new vulnerabilities, or sudden changes in attacker tactics can invalidate the underlying assumptions in these models. Future research needs to explore adaptive frameworks that can rapidly recalibrate probabilities and reassign resources in near real-time to maintain a relevant security posture. [78]

In addition, the reliance on partial homomorphic encryption for certain computations highlights a limitation related to function support. Many practical analytics tasks, including complex joins, string operations, or certain machine learning algorithms, may not be compatible with partial homomorphic schemes. Fully homomorphic encryption, while theoretically appealing, still suffers from extreme overhead [79]. Hybrid models that selectively apply homomorphic methods to the most sensitive subsets of the dataset or the most critical computations offer a promising avenue but require intricate orchestration and careful partitioning of data. Achieving an optimal partitioning strategy remains an open research problem, as does the task of designing dynamic scheduling algorithms that can handle the interplay between encrypted and unencrypted computations. [80]

Finally, it is crucial to recognize that security is not a static property but an evolving process. The framework described here, despite its layered defenses and mathematical rigor, cannot provide an absolute guarantee of security. It instead forms a robust baseline from which iterative improvements can be made [81]. In practice, organizations using this framework would still need dedicated security teams to monitor system health, analyze intrusion attempts, and keep abreast of the latest patches, vulnerabilities, and adversarial tactics. The cost of maintaining this level of operational security may be non-trivial, especially for smaller organizations with limited resources.

In summary, while the proposed framework demonstrates a strong potential for securing big data analytics in multi-tenant clouds, it is not devoid of practical constraints and areas necessitating further innovation [82]. Addressing the limitations in key management, handling hardware-level vulnerabilities, reducing cryptographic overhead, and strengthening adaptive adversarial modeling constitutes the frontier of research in this domain. The final section offers concluding remarks and suggests directions for future investigation, particularly in bridging the gap between cryptographic theory and scalable real-world implementations.

6. Conclusion

This paper has presented a comprehensive framework for secure big data analytics in multi-tenant cloud infrastructures, offering a multi-layered architecture grounded in robust theoretical models and practical implementation strategies [83]. By integrating mathematical formulations for multi-tenant resource sharing with partial homomorphic encryption, distributed key management, and dynamic scheduling, the approach addresses the dual challenges of data protection and high-throughput processing. Through experimental evaluations, the framework has demonstrated its ability to maintain stringent security guarantees, such as encryption at rest, in transit, and partially in use, while imposing an acceptable level of performance overhead on batch-oriented workloads. Simulated adversarial tests suggest that the layered defenses and real-time monitoring can significantly reduce the risk of data breaches, particularly for organizations operating under strict regulatory mandates. [84]

Despite these advancements, several areas remain open for further exploration. The cost of high-level encryption methods, especially those that enable computations on encrypted data, can still be substantial for large-scale, real-time analytics [85]. Key management strategies, while crucial to ensuring isolation among tenants, may introduce system bottlenecks and require innovative distributed solutions. The reliance on hardware-based isolation also highlights an inherent vulnerability if microarchitectural flaws are discovered, underscoring the importance of rapid patching and continuous system validation. Additionally, the probabilistic risk model employed here, although offering a flexible approach to threat quantification, depends on timely and accurate threat intelligence [86]. Its efficacy may be diminished by emerging vulnerabilities or novel attack vectors that were not considered during the model's calibration phase.

Future research can focus on developing hybrid cryptographic schemes that intelligently blend homomorphic operations with selective decryption, thus maximizing security for the most critical data while minimizing overhead for less sensitive tasks. Advances in decentralized or blockchain-based key management could also help distribute the load, preventing performance degradation as the system scales to thousands of tenants and petabytes of data. Techniques in anomaly detection and machine learning-based threat modeling are continually evolving, and their integration with the core scheduling and encryption layers may offer more adaptive and agile responses to sophisticated adversaries. Another promising direction involves fine-tuning scheduling algorithms that incorporate real-time risk metrics, ephemeral key usage patterns, and hardware acceleration capabilities, ensuring that resources are optimally allocated to meet both performance and security objectives.

Ultimately, the results discussed here validate the feasibility of deploying secure, large-scale analytics frameworks in multi-tenant clouds without completely sacrificing efficiency [87, 88]. They also highlight the real-world considerations and trade-offs that system architects, administrators, and security professionals must balance. By continuing to refine cryptographic performance, key management strategies, hardware-level defenses, and adaptive security policies, the community can advance toward a future where multi-tenant cloud environments deliver not only the scale and cost benefits long associated with cloud computing but also the robust security and privacy protections that modern data-driven enterprises demand.

References

[1] B. Yi, X.-W. Wang, M. Huang, and Q. He, "A qos based reliable routing mechanism for service customization," *Journal of Computer Science and Technology*, vol. 37, pp. 1492–1508, 11 2022.

[2] E. Adi, A. Anwar, Z. A. Baig, and S. Zeadally, "Machine learning and data analytics for the iot," *Neural Computing and Applications*, vol. 32, pp. 16205–16233, 5 2020.

[3] A. Bhargava, D. Bhargava, P. N. Kumar, G. S. Sajja, and S. Ray, "Industrial iot and ai implementation in vehicular logistics and supply chain management for vehicle mediated transportation systems," *International Journal of System Assurance Engineering and Management*, vol. 13, pp. 673–680, 1 2022.

[4] Y.-Y. Teing, A. Dehghantanha, and K.-K. R. Choo, "Greening cloud-enabled big data storage forensics: Syncany as a case study," *IEEE Transactions on Sustainable Computing*, vol. 4, pp. 204–216, 4 2019.

[5] K.-P. Lee and S. Song, "Developing insights from the collective voice of target users in twitter.," *Journal of big data*, vol. 9, pp. 75–, 6 2022.

[6] A. M. Alzarooni, S. A. Khan, A. Gunasekaran, and M. S. Mubarik, "Enablers for digital supply chain transformation in the service industry," *Annals of Operations Research*, 11 2022.

[7] M. Muniswamaiah, T. Agerwala, and C. Tappert, "Big data in cloud computing review and opportunities," *arXiv preprint arXiv:1912.10821*, 2019.

[8] Y. Abduallah and J. T. L. Wang, "New algorithms for inferring gene regulatory networks from time-series expression data on apache spark," *International Journal of Big Data Intelligence*, vol. 6, no. 3/4, pp. 153–162, 2019.

[9] M. Shabaninezhad and G. Ramakrishna, "Theoretical investigation of plasmonic properties of quantum-sized silver nanoparticles," *Plasmonics*, vol. 15, pp. 783–795, 12 2019.

[10] O. Mypati, A. Mukherjee, D. Mishra, S. K. Pal, P. P. Chakrabarti, and A. Pal, "A critical review on applications of artificial intelligence in manufacturing," *Artificial Intelligence Review*, vol. 56, pp. 661–768, 7 2023.

[11] M. Horton, S. Dwaraknath, and K. A. Persson, "Promises and perils of computational materials databases," *Nature computational science*, vol. 1, pp. 3–5, 1 2021.

[12] M. Abouelyazid and C. Xiang, "Architectures for ai integration in next-generation cloud infrastructure, development, security, and management," *International Journal of Information and Cybersecurity*, vol. 3, no. 1, pp. 1–19, 2019.

[13] K. Stereńczak, M. Lisańczuk, and Y. Erfanifard, "Delineation of homogeneous forest patches using combination of field measurements and lidar point clouds as a reliable reference for evaluation of low resolution global satellite data," *Forest Ecosystems*, vol. 5, pp. 1–12, 3 2018.

[14] G. Sonnert, "Sociological research and modernity: the rise and fall of the survey subject," *International Journal of Politics, Culture, and Society*, vol. 32, pp. 259–277, 9 2018.

[15] A. Khan and P. Tandon, "Realizing the end-of-life considerations in the design of food packaging," *Journal of Packaging Technology and Research*, vol. 2, pp. 251–263, 10 2018.

[16] R. Avula *et al.*, "Data-driven decision-making in healthcare through advanced data mining techniques: A survey on applications and limitations," *International Journal of Applied Machine Learning and Computational Intelligence*, vol. 12, no. 4, pp. 64–85, 2022.

[17] A. Singh, S. Garg, K. Kaur, S. Batra, N. Kumar, and K.-K. R. Choo, "Fuzzy-folded bloom filter-as-a-service for big data storage in the cloud," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2338–2348, 2019.

[18] W. Dekens, L. Andreoli, J. de Vries, E. Mereghetti, and F. Oosterhof, "A low-energy perspective on the minimal left-right symmetric model," *Journal of High Energy Physics*, vol. 2021, pp. 1–74, 11 2021.

[19] Y. Lin and J. Wessel, "The continuing evolution of precision health in type 2 diabetes: Achievements and challenges.," *Current diabetes reports*, vol. 19, pp. 1–10, 2 2019.

[20] I. A. Ahmed, null Shahfahad, D. Dutta, M. R. I. Baig, S. S. Roy, and A. Rahman, "Implications of changes in temperature and precipitation on the discharge of brahmaputra river in the urban watershed of guwahati, india," *Environmental monitoring and assessment*, vol. 193, pp. 518–, 7 2021.

[21] J. Zhang and R. X. Gao, "Deep learning-driven data curation and model interpretation for smart manufacturing," *Chinese Journal of Mechanical Engineering*, vol. 34, pp. 1–21, 7 2021.

[22] E. M. Onyema, V. Gude, A. Bhatt, A. Aggarwal, S. Kumar, M. E. Benson-Emenike, and L. O. Nwobodo, "Smart job scheduling model for cloud computing network application," *SN Computer Science*, vol. 5, 11 2023.

[23] M. Muniswamaiah, T. Agerwala, and C. C. Tappert, "Federated query processing for big data in data science," in *2019 IEEE International Conference on Big Data (Big Data)*, pp. 6145–6147, IEEE, 2019.

[24] A. Elizondo-Noriega, N. Tiruvengadam, and D. Güemes-Castorena, "An economic feasibility study using a system-dynamics-based archetype of rfid implementation in a manufacturing firm," *International Journal on Interactive Design and Manufacturing (IJIDeM)*, vol. 15, pp. 187–210, 8 2021.

[25] F. Wang and G. Luo, "Guest editorial: special issue on data management and analytics for healthcare," *Distributed and Parallel Databases*, vol. 37, pp. 233–234, 5 2019.

[26] J. Dong and C. Rudin, "Exploring the cloud of variable importance for the set of all good models," *Nature Machine Intelligence*, vol. 2, pp. 810–824, 12 2020.

[27] T. Dee, R. A. Scheel, N. Montelibano, and A. Tyagi, "User-silicon entangled mobile identity authentication," *Journal of Hardware and Systems Security*, vol. 4, pp. 208–229, 7 2020.

[28] S. K. Rizvi, M. A. Azad, and M. M. Fraz, "Spectrum of advancements and developments in multidisciplinary domains for generative adversarial networks (gans)," *Archives of computational methods in engineering : state of the art reviews*, vol. 28, pp. 1–19, 4 2021.

[29] M. García-Valls, A. Dubey, and V. Botti, "Introducing the new paradigm of social dispersed computing: Applications, technologies and challenges," *Journal of Systems Architecture*, vol. 91, pp. 83–102, 2018.

[30] A. K. Saxena and A. Vafin, "Machine learning and big data analytics for fraud detection systems in the united states fintech industry," *Emerging Trends in Machine Intelligence and Big Data*, vol. 11, no. 12, pp. 1–11, 2019.

[31] O. Ruiz-Alvarez, V. P. Singh, J. Enciso-Medina, C. L. Munster, R. A. Kaiser, R. E. Ontiveros-Capurata, L. Diaz-Garcia, and C. A. C. dos Santos, "Spatio-temporal trends in monthly pan evaporation in aguascalientes, mexico," *Theoretical and Applied Climatology*, vol. 136, pp. 775–789, 5 2018.

[32] Y. Chen, Y. Lu, L. Bulysheva, and M. Y. Kataev, "Applications of blockchain in industry 4.0: a review," *Information Systems Frontiers*, vol. 26, pp. 1715–1729, 2 2022.

[33] M. Zhang, G. M. Abrahão, and S. E. Thompson, "Sensitivity of soybean planting date to wet season onset in mato grosso, brazil, and implications under climate change," *Climatic Change*, vol. 168, pp. 1–28, 10 2021.

[34] A. Haghighat, V. Ravichandra-Mouli, P. Chakraborty, Y. Esfandiari, S. Arabi, and A. Sharma, "Applications of deep learning in intelligent transportation systems," *Journal of Big Data Analytics in Transportation*, vol. 2, pp. 115–145, 8 2020.

[35] J.-W. Lee, "Quantum entanglement of dark matter," *Journal of the Korean Physical Society*, vol. 73, pp. 1596–1602, 11 2018.

[36] A. A. T. Bafrooee, E. Moniri, H. A. Panahi, M. Miralinaghi, and A. H. Hasani, "Ethylenediamine functionalized magnetic graphene oxide (fe3o4@go-eda) as an efficient adsorbent in arsenic(iii) decontamination from aqueous solution," *Research on Chemical Intermediates*, vol. 47, pp. 1397–1428, 1 2021.

[37] K. Hariss, H. N. Noura, and A. E. Samhat, "An efficient fully homomorphic symmetric encryption algorithm," *Multimedia Tools and Applications*, vol. 79, pp. 12139–12164, 1 2020.

[38] M. K. Kansara, "Overcoming technical challenges in large-scale it migrations: A literature-based analysis and practical solutions," *JNRID*, vol. 1, no. 3, 2023.

[39] G. Nigro, F. Pegoraro, and F. Valentini, "Plasma physics and astrophysics: retrospects, state-of-the art, and prospects," *Rendiconti Lincei. Scienze Fisiche e Naturali*, vol. 32, pp. 25–44, 11 2020.

[40] P. Ploton, F. Mortier, M. Réjou-Méchain, N. Barbier, N. Picard, V. Rossi, C. F. Dormann, G. Cornu, G. Viennois, N. Bayol, A. Lyapustin, S. Gourlet-Fleury, and R. Pélissier, "Spatial validation reveals poor predictive performance of large-scale ecological mapping models.," *Nature communications*, vol. 11, pp. 1–11, 9 2020.

[41] V. Chunduri, A. Kumar, A. Joshi, S. R. Jena, A. Jumaev, and S. More, "Optimizing energy and latency trade-offs in mobile ultra-dense iot networks within futuristic smart vertical networks," *International Journal of Data Science and Analytics*, 12 2023.

[42] S. G. Langer, G. Shih, P. Nagy, and B. A. Landman, "Collaborative and reproducible research: Goals, challenges, and strategies.," *Journal of digital imaging*, vol. 31, pp. 275–282, 2 2018.

[43] F. Azadi, P.-S. Ashofteh, A. Shokri, and H. A. Loáiciga, "Development of the fa-knn hybrid algorithm and its application to reservoir operation," *Theoretical and Applied Climatology*, vol. 155, pp. 1261–1280, 10 2023.

[44] T. Taami, S. Azizi, and R. Yarinezhad, "Unequal sized cells based on cross shapes for data collection in green internet of things (iot) networks," *Wireless Networks*, vol. 29, pp. 2143–2160, 2 2023.

[45] H. Lammer, M. Scherf, H. Kurokawa, Y. Ueno, C. Burger, T. I. Maindl, C. P. Johnstone, M. Leizinger, M. Benedikt, L. Fossati, K. G. Kislyakova, B. Marty, G. Avice, B. Fegley, and P. Odert, "Loss and fractionation of noble gas isotopes and moderately volatile elements from planetary embryos and early venus, earth and mars," *Space Science Reviews*, vol. 216, pp. 1–50, 6 2020.

[46] M. Abukhaled, N. Guessoum, and N. Alsaeed, "Mathematical modeling of light curves of rhessi and agile terrestrial gamma-ray flashes," *Astrophysics and Space Science*, vol. 364, pp. 1–16, 8 2019.

[47] E. M. Dovom, A. Azmoodeh, A. Dehghantanha, D. E. Newton, R. M. Parizi, and H. Karimipour, "Fuzzy pattern tree for edge malware detection and categorization in iot," *Journal of Systems Architecture*, vol. 97, pp. 1–7, 2019.

[48] R. Avula, "Applications of bayesian statistics in healthcare for improving predictive modeling, decision-making, and adaptive personalized medicine," *International Journal of Applied Health Care Analytics*, vol. 7, no. 11, pp. 29–43, 2022.

[49] H. Saboonchi, D. Blanchette, and K. Hayes, "Advancements in radiographic evaluation through the migration into nde 4.0.," *Journal of nondestructive evaluation*, vol. 40, pp. 17–17, 1 2021.

[50] V. J. Sterken, A. J. Westphal, N. Altobelli, D. Malaspina, and F. Postberg, "Interstellar dust in the solar system," *Space Science Reviews*, vol. 215, 10 2019.

[51] M. Liebrenz, A. Gamma, A. Buadze, R. Schleifer, S. Baggio, B. J. Schwartz, A. R. Schneeberger, and A. Uchtenhagen, "Fifteen years of heroin-assisted treatment in a swiss prison—a retrospective cohort study," *Harm reduction journal*, vol. 17, pp. 67–67, 10 2020.

[52] F. S. Vahidy, S. L. Jones, M. E. Tano, J. C. Nicolas, O. Khan, J. Meeks, A. Pan, T. Menser, F. Sasangohar, G. Naufal, D. H. Sostman, K. Nasir, and B. A. Kash, "Rapid response to drive covid-19 research in a learning health care system: Rationale and design of the houston methodist covid-19 surveillance and outcomes registry (curator).," *JMIR medical informatics*, vol. 9, pp. e26773–, 2 2021.

[53] I. Sandu, A. van Niekerk, T. G. Shepherd, S. Vosper, A. Zadra, J. T. Bacmeister, A. Beljaars, A. Brown, A. Dörnbrack, N. A. McFarlane, F. Pithan, and G. Svensson, "Impacts of orography on large-scale atmospheric circulation," *npj Climate and Atmospheric Science*, vol. 2, pp. 1–8, 5 2019.

[54] M. Noura, M. Atiquzzaman, and M. Gaedke, "Interoperability in internet of things: Taxonomies and open challenges," *Mobile Networks and Applications*, vol. 24, pp. 796–809, 7 2018.

[55] M. Grigorieva, M. Titov, A. A. Alekseev, A. Artamonov, A. Klimentov, T. A. Korchuganova, I. Milman, T. Galkin, and V. Pilyugin, "Evaluation of the level-of-detail generator for visual analysis of the atlas computing metadata," *Lobachevskii Journal of Mathematics*, vol. 40, pp. 1788–1798, 11 2019.

[56] R. Krishna and V. V. Elisseev, "User-centric genomics infrastructure: trends and technologies.," *Genome*, vol. 64, pp. 467–475, 11 2020.

[57] S. Luo, "Addressing military ai risks in u.s.–china crisis management mechanisms," *China International Strategy Review*, vol. 4, pp. 233–247, 9 2022.

[58] M. Ienca, P. Haselager, and E. J. Emanuel, "Brain leaks and consumer neurotechnology.," *Nature biotechnology*, vol. 36, pp. 805–810, 10 2018.

[59] J. M. Tien, "Convergence to real-time decision making," *Frontiers of Engineering Management*, vol. 7, pp. 204–222, 6 2019.

[60] S.-H. Hsu, H.-T. Hung, Y.-Q. Lin, and C.-M. Chang, "Defect inspection of indoor components in buildings using deep learning object detection and augmented reality," *Earthquake Engineering and Engineering Vibration*, vol. 22, pp. 41–54, 1 2023.

[61] G. Henriques and J. H. Michalski, "Defining behavior and its relationship to the science of psychology.," *Integrative psychological & behavioral science*, vol. 54, pp. 328–353, 11 2019.

[62] S. J. Purkis, A. C. R. Gleason, C. R. Purkis, A. C. Dempsey, P. Renaud, M. Faisal, S. Saul, and J. M. Kerr, "High-resolution habitat and bathymetry maps for 65,000 sq. km of earth's remotest coral reefs," *Coral Reefs*, vol. 38, pp. 467–488, 4 2019.

[63] P. C. K. Hung and M. B. Blake, "Reflecting on two decades of services computing," *IEEE Internet Computing*, vol. 22, no. 5, pp. 3–7, 2018.

[64] M. Abouelyazid, "Forecasting resource usage in cloud environments using temporal convolutional networks," *Applied Research in Artificial Intelligence and Cloud Computing*, vol. 5, no. 1, pp. 179–194, 2022.

[65] Y. Zuo, "Non-asymptotic analysis and inference for an outlyingness induced winsorized mean," *Statistical Papers*, vol. 64, pp. 1465–1481, 9 2022.

[66] T. Bouley, P. Sørensen, and T.-T. Yu, "Constraints on ultralight scalar dark matter with quadratic couplings," *Journal of High Energy Physics*, vol. 2023, 3 2023.

[67] J. Thomas and P. Mantri, "Axiomatic cloud computing architectural design," *MATEC Web of Conferences*, vol. 301, pp. 24–00024, 12 2019.

[68] M. Muniswamaiah, T. Agerwala, and C. C. Tappert, "Approximate query processing for big data in heterogeneous databases," in *2020 IEEE international conference on big data (big data)*, pp. 5765–5767, IEEE, 2020.

[69] A. Weeger, X. Wang, H. Gewald, M. S. Raisinghani, O. P. Sanchez, G. Grant, and S. Pittayachawan, "Determinants of intention to participate in corporate byod-programs: The case of digital natives," *Information Systems Frontiers*, vol. 22, pp. 203–219, 4 2018.

[70] T. Bates, C. Cobo, O. Marino, and S. Wheeler, "Can artificial intelligence transform higher education," *International Journal of Educational Technology in Higher Education*, vol. 17, pp. 1–12, 6 2020.

[71] C. Bignami, E. Valerio, E. Carminati, C. Doglioni, P. Tizzani, and R. Lanari, "Volume unbalance on the 2016 amatrice - norcia (central italy) seismic sequence and insights on normal fault earthquake mechanism.," *Scientific reports*, vol. 9, pp. 4250–4250, 3 2019.

[72] B. Ball and A. Koliousis, "Training philosopher engineers for better ai," *AI & SOCIETY*, vol. 38, pp. 861–868, 7 2022.

[73] R. Avula, "Assessing the impact of data quality on predictive analytics in healthcare: Strategies, tools, and techniques for ensuring accuracy, completeness, and timeliness in electronic health records," *Sage Science Review of Applied Machine Learning*, vol. 4, no. 2, pp. 31–47, 2021.

[74] J. M. Tien, "Toward the fourth industrial revolution on real-time customization," *Journal of Systems Science and Systems Engineering*, vol. 29, pp. 127–142, 4 2020.

[75] V. Navale, D. von Kaeppler, and M. McAuliffe, "An overview of biomedical platforms for managing research data," *Journal of Data, Information and Management*, vol. 3, pp. 21–27, 1 2021.

[76] A. Kusiak, "Extreme science and engineering.," *Journal of intelligent manufacturing*, vol. 31, pp. 1607–1610, 8 2020.

[77] A. C. Yoshikuni, R. Dwivedi, R. G. D. de Lima, C. Parisi, and J. C. T. Oyadomari, "Role of emerging technologies in accounting information systems for achieving strategic flexibility through decision-making performance: An exploratory study based on north american and south american firms," *Global Journal of Flexible Systems Management*, vol. 24, pp. 199–218, 1 2023.

[78] M. Ayachi, H. Nacer, and H. Slimani, "Cooperative game approach to form overlapping cloud federation based on inter-cloud architecture," *Cluster Computing*, vol. 24, pp. 1551–1577, 3 2021.

[79] R. Q. Cao, D. G. Schniederjans, and V. C. Gu, "Stakeholder sentiment in service supply chains: big data meets agenda-setting theory," *Service Business*, vol. 15, pp. 151–175, 2 2021.

[80] M. Aktaş, E. Akbas, and A. E. Fatmaoui, "Persistence homology of networks: methods and applications," *Applied Network Science*, vol. 4, pp. 1–28, 8 2019.

[81] M. H. de Menendez, C. A. E. Díaz, and R. Morales-Menendez, "Engineering education for smart 4.0 technology: a review," *International Journal on Interactive Design and Manufacturing (IJIDeM)*, vol. 14, pp. 789–803, 7 2020.

[82] N. N. Ho-Dac, M. Kumar, and R. J. Slotegraaf, "Using product development information to spur the adoption of continuous improvement products," *Journal of the Academy of Marketing Science*, vol. 48, pp. 1156–1173, 6 2020.

[83] R. A. Hasan, H. Irshaid, F. Alhomaidat, S. Lee, and J.-S. Oh, "Transportation mode detection by using smartphones and smartwatches with machine learning," *KSCE Journal of Civil Engineering*, vol. 26, no. 8, pp. 3578–3589, 2022.

[84] A. Farzam, P. Moradi, S. Mohammadi, Z. Padar, and A. A. Siegel, "Opinion manipulation on farsi twitter.," *Scientific reports*, vol. 13, pp. 333–, 1 2023.

[85] R. Yang, "Privacy and surveillance concerns in machine learning fall prediction models: implications for geriatric care and the internet of medical things," *AI & SOCIETY*, vol. 39, pp. 1969–1973, 4 2023.

[86] A. Hsu and R. Rauber, "Diverse climate actors show limited coordination in a large-scale text analysis of strategy documents," *Communications Earth & Environment*, vol. 2, pp. 1–12, 2 2021.

[87] K. Budhrani, Y. Ji, and J. H. Lim, "Unpacking conceptual elements of smart learning in the korean scholarly discourse," *Smart Learning Environments*, vol. 5, pp. 1–26, 10 2018.

[88] D. Ge, Y. Pan, Z.-J. Shen, D. Wu, R. Yuan, and C. Zhang, "Retail supply chain management: a review of theories and practices," *Journal of Data, Information and Management*, vol. 1, pp. 45–64, 7 2019.