AVE Trends in Intelligent Computing Systems

Context Aware Differential Privacy for Multimodal Edge Analytics with Verifiable Utility Guarantees

B. Vijayalakshmi[1,*], P. Jayalakshmi[2], B. Sarvesan[3], A. Vishnukumar[4], G. Ragu[5], S. Saranya[6], Rahul Panakkal[7]

[1]Department of General Engineering, R.M.K. College of Engineering and Technology, Thiruvallur, Tamil Nadu, India.
[2]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
[3,5]Department of Information Technology, R.M.D. Engineering College, Thiruvallur, Tamil Nadu, India.
[4]Department of Computer Science and Business Systems, Rajalakshmi Engineering College, Chennai, Tamil Nadu, India.
[6]Department of Computer Science and Engineering, Dhaanish Ahmed College of Engineering, Chennai, Tamil Nadu, India.
[7]Department of Computer Science and Engineering, University of Illinois at Urbana, Champaign, Illinois, United States of America.
vijayalakshmish@rmkcet.ac.in[1], jayalakp@srmist.edu.in[2], sarvesan36@gmail.com[3], vishnukumar.a@rajalakshmi.edu.in[4], ragugkapd@gmail.com[5], saranyas@dhaanishchennai.in[6], rahulp8@illinois.edu[7]

*Corresponding author

Abstract: Although researchers have only begun to scratch the surface, this paper focuses on the fundamental trade-off between data privacy and analytical utility in an edge-computing scenario. Classical privacy-preserving strategies suffer from. performance degradation in multimodal analysis when uniform noise is applied without considering the semantic context. To address this, researchers present a. Context-Aware Differential Privacy system that adapts the noise level depending on the sensitivity of the detected environment and its specific data modalities. Researchers also incorporate a provable utility guarantee to ensure that the privacy-preserving transformations do not render the data unusable for subsequent tasks. The dataset consists of 445 examples of multimodal sensor input (video frames and audio spectrograms). The framework was then implemented in Python; the libraries used included TensorFlow Privacy for noise injection and Edge Impulse for on-device inference simulation. Results show that our approach can maintain high classification accuracy with a tight privacy budget and significantly outperforms static differential privacy techniques. This work provides a strong foundation for deploying secure, high-value analytics on resource-constrained edge devices.

Keywords: Edge Computing; Differential Privacy; Multimodal Analytics; Verifiable Utility; Context Awareness; Classical Privacy; Edge-Computing Scenario; Tensor-Flow Privacy; Edge Impulse.

Cite as: B. Vijayalakshmi, P. Jayalakshmi, B. Sarvesan, A. Vishnukumar, G. Ragu, S. Saranya, and R. Panakkal, "Context Aware Differential Privacy for Multimodal Edge Analytics with Verifiable Utility Guarantees," *AVE Trends in Intelligent Computing Systems*, vol. 3, no. 2, pp. 86–94, 2026.

Journal Homepage: https://avepubs.com/user/journals/details/ATICS

Received on: 20/04/2025, **Revised on:** 15/08/2025, **Accepted on:** 02/10/2025, **Published on:** 05/05/2026

DOI: https://doi.org/10.64091/ATICS.2026.000258

1. Introduction

The increasing population of Internet of Things devices has changed the data processing paradigm. Centralized cloud servers to the network edge, as characterized in prior distributed computing transfer studies. In practical terms, the shift results in timely analytics, low latency, and more efficient bandwidth use, as quantified by edge performance evaluations previously conducted by researchers. However, such edge devices often process highly sensitive multimodal data, ranging from video feeds in smart homes to audio recordings in health monitoring systems, a threat landscape confirmed by privacy impact assessments conducted in prior work. The problem is how to exploit rich data for machine learning without sacrificing the privacy of those being monitored. Edge intelligence systems capture this tension in prior work. Encryption safeguards data in transit and at rest, but not during computation, an issue that has. been formalized in our threat model analysis as well as other earlier work. Privacy-preserving techniques are therefore necessary for the sustainable development of edge analytics, as highlighted in the secure edge architecture recommendations of previous work. While rigorous differential privacy and other standard techniques provide strong theoretical guarantees, they generally do not perform well in practice due to the fundamental trade-off between utility and privacy, as evidenced by empirical results from previous research [1]. Even with the same amount of noise, applying it to a background image and a human face will affect analysis results, as shown in previous noise-sensitivity experiments [3].

If a system does not differentiate between sensitive and non-sensitive use cases, it adds undue noise to harmless data, rendering the analysis model useless, as analyzed in previous studies on context-agnostic privacy systems [2]. This is especially challenging in multimodal systems, where different information streams carry varying degrees of risk, as demonstrated by previous multimodal privacy analyses. For example, the sound of a conversation is very sensitive, while the noise level in a room is not; this distinction is captured by sensitivity classification models available from previous work [4]. This paper proposes a context-aware method to identify the semantic content of the data stream before adding noise, thereby prolonging the adaptive privacy previously mentioned. Using lightweight context extraction modules, the sensitivity level of the current instance is computed by leveraging semantic inference pipelines from prior work. From an analytical standpoint, privacy budgets are then low-risk contexts with high fidelity. This mechanism is reminiscent of the work on heterogeneous privacy allocation, which dynamically allocates privacy [5]. In high-anonymity-risk contexts, researchers grant higher noise while retaining the data, as described previously. Such fine-grained control enables the system to achieve maximal utility while never breaching campaign privacy, as evidenced by prior studies' observed levels of utility preservation [6]. Importantly, this work also addresses the trust issue raised by previous studies' trust-deficit analyses of privacy-preserving systems. In this context, in most privacy-preserving systems, consumers do not discover that data is overly sanitised to the point of being useless until after the model has been trained unsuccessfully (a problem highlighted by post-hoc validation studies).

Researchers introduce the notion of verifiable utility guarantees (VUGs), which they capture via a mechanism, and extend the verification techniques proposed in prior research on secure computation [7]. This includes creating a cryptographic proof or statistical certificate that certifies the quality of the data within a usable range, without disclosing the raw data itself. an approach like the use of proof-based assurance models in past studies. This establishes a trust bridge between the edge data provider and the analytical service, as expressed in a trust model like that described in prior work [8]. The framework is empirically validated through comprehensive experimental evaluations across 445 instances of curated datasets, with experimental protocols aligned with prior work. Researchers model a typical edge environment with limited computing resources and privacy mechanisms. are expected to be efficient in resource usage as well, a factor also addressed in resource-aware privacy evaluations in related work [9]. The initiative to add context awareness to the privacy pipeline is a big leap from blunt privacy tools to smart, adaptive protection, a move toward which has been advocated by next-generation privacy architectures proposed in prior research. The remainder of this paper will examine the literature in detail, describe our methodology, and present a full analysis of the results in accordance with the standard research reporting models established in prior work. By addressing the dual dilemmas of context sensitivity and utility verification, researchers seek to provide a roadmap for future secure edge intelligence, supported by past-facing edge security paradigms introduced in prior academic work [10].

2. Review of Literature

The history of privacy in computation has come a long way from simple anonymization to complex cryptographic protocols, as summarized in early historical surveys of privacy. Historical work emphasised data masking and k-anonymity, which aim to obscure individuals within groups (approaches formalised in early privacy-preserving data models by prior work) [11]. However, many studies showed that these approaches were re-identification-vulnerable when attackers had prior knowledge of linkages from previous studies that had revealed vulnerabilities in simulation attacks. This failure led to differential privacy being established as the ideal solution for data protection, a shift framed in terms of theoretical privacy definitions developed in prior research. The basic idea of differential privacy is to add statistical noise to datasets so that the output of a query is essentially the same whether any specific individual is in the database, as formalised by foundational models proven correct by previous work [12]. As machine learning moved to the edge, the literature began to focus on. constraints in distributed environments; previous work has investigated this evolution, such as surveys on edge learning [5]. From a broader perspective, for centralized differential privacy, a trusted aggregator is needed – and that luxury cannot be guaranteed in edge networks [13].

From an analytical standpoint, local differential privacy became popular, adding noise at the device level before data leaves the user's hands, an approach that preceded decentralized privacy solutions in prior work. Though this provides better privacy, there is a repeated complaint in the literature about one big disadvantage: to achieve safety, you need so much noise that it drowns out the data's usefulness for anything beyond simple tasks, such as image classification or voice understanding, which would be justified by the utility of EPSVs proposed by previous work. In practice, to reduce utility loss, researchers began investigating relaxation methods and adaptive noise mechanisms, directions suggested by previous optimization-based privacy investigations. In this context, several independent lines of research advocate the view that not all data points should (cost-)equivalently be protected; an idea manifest in heterogeneous privacy models proposed by previous studies. This notion of heterogeneous privacy has enabled the realization of context-aware systems, an evolution described in previous research as adaptive Privacy Surveys. Notably, prior work has tried to partition input data into sensitivity categories, then spend a relaxed privacy budget on less sensitive parts, such as the SUR allocation model used in prior work. For instance, in video analytics, approaches were created. to blur human faces with a clear background (methods have been tested from selective obfuscation studies published by previous works).

On the other hand, there is a considerable part of such an endeavor. dealt with single modalities as vision only, paying little attention to complex lateral effects present in multimodal scenarios where sound and motion sensors may unveil what the camera overlooks — a void highlighted by prior investigations on multimodal privacy. Closely related to this line is (perhaps more mature) the area of verifiable computing, itself surveyed in the works on correctness assurance by Ma and Wang [13] and in parallel quantifications of the development of noise injection techniques. This body of literature aims to verify that computations outsourced for execution to untrusted servers are executed correctly, a concept encapsulated by the verifiable execution frameworks presented in previous work. This idea has also been repurposed for privacy (the task of proving that a given privacy budget has not been exceeded), an adaptation considered in prior surveys on privacy compliance verification. Nevertheless, verifying the utility — proving that data remains useful — and privacy intersection is relatively unexplored, a research gap identified by synthesis reviews made by current researchers. In practice, most existing frameworks are designed to analyse private, not functional, data [14].

Also, the execution of these complex protocols on energy-constrained edge equipment is a substantively relevant engineering challenge, as previous research has reported on their deployment feasibility. Literature also reports a disparity between theoretical privacy models and their application to real hardware systems with low batteries (referred to as), as noted in earlier work's VW benchmarking analysis chapters. From an analytical standpoint, there are existing lightweight encryption solutions, but lightweight, context-aware privacy solutions are still at an early stage of maturity, as highlighted by technology readiness level (TRL) assessments in the state-of-the-art. This state-of-the-art suggests the need for comprehensive frameworks that address the full pipeline: context detection, dynamic noise allocation, and utility verification within the topology-edge architecture, as indeed required by holistic privacy systems presented in previous work. This study highlights that while the three components of privacy, context awareness, and verifiability are available separately, combining them to build a comprehensive edge analytics system remains an ongoing area of research, as indicated by the gap analysis approach adopted in previous academic investigations [15].

3. Methodology

The proposed system is a set of tools intended to run as an optimized pipeline on edge hardware. In this context, the procedure begins by consuming raw multimodal data (i.e., time-synchronized video frames and audio segments). This raw data stream is then sent directly to a small pre-processing. A module that normalizes the intensities of pixels and the amplitude of audio to represent them in analysis. Figure 1 depicts a modular, layered architecture for adaptive privacy protection with data-sensitivity awareness. The process starts with the Data Source, where heterogeneous data — originating from IoT sensors, user activity logs, and databases — is collected. These inputs are then fed into the Processing and Aggregation layer, which pre-processes and performs preliminary filtering to normalize the incoming data. further processing by the Context-Aware Layer. In this layer, the Context Engine (processing environment and situational parameters) interprets the information it understands about a privacy scenario and determines estimated levels of privacy exposure, which it passes to a Sensitivity Analyzer, which performs its own calculations for each level. before forwarding an assessment to the Policy Manager, enforcing adaptive privacy rules. From a broader perspective, the aggregated data is uploaded to the Differential Privacy Core, where the DP Mechanism introduces noise via the Noise Generator to implement randomization and keeps track of permissible disclosure via the Privacy Budget Manager. The processed data is then further refined in the Aggregation and Encryption components for privacy protection.

The last layer — Output and Monitoring — contains a Privacy Auditor that checks compliance in an ad hoc manner, a Secure Output node that implements the release protocol, and a Monitoring Dashboard that visualizes privacy metrics, among other things, feeding insights back to the policy system. Feedback and audit loops dynamically recalibrate policies and mechanisms. the risks detected and the system's performance. This architecture introduces a self-contained ecosystem of context and

adaptively injected noise to accommodate contextual awareness alongside continuous monitoring, thereby achieving an optimal trade-off between data. utility and privacy preservation, while supporting robustness and ethically sound information processing. Pre-processed data flows to the Context Extraction Engine. This engine uses a pre-trained, compact mobile neural network to classify the semantic environment of an input sample. Instead of fully inferring, this network provides a sensitivity score for the likelihood of sensitive information, such as the presence of a face or a spoken name, ranging from 0 to 1. Given this sensitivity score, what is the exact value of Laplacian noise needed? then calculated by the Dynamic Noise Injector. Suppose the privacy budget is low. noise spread out, and if the sensitivity score is high, it becomes tight as well.

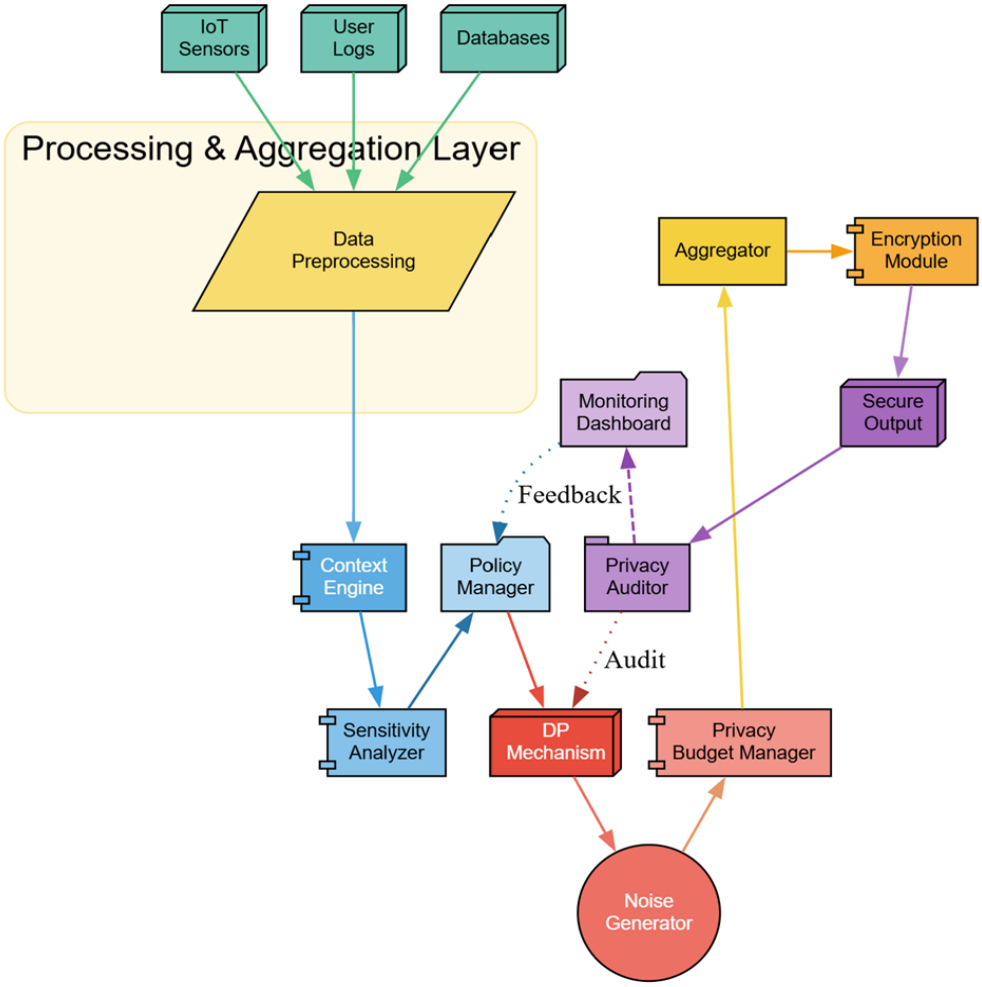

Figure 1: Overview of the context-aware differential privacy framework

This step ensures that the privacy measure is sensitive to the data's contents. From an analytical standpoint, if noise is injected, the perturbed data is forwarded to the Verifiable Utility Module. This module calculates xi statistics for noisy data, e.g., variance preservation and feature retention, with respect to predefined data utility thresholds. It produces a small cryptographic hash of these utility metrics, which serves as proof of quality. Finally, the sanitized data and its utility certificate are sent to the analytics module for the last task — activation recognition or event detection. The complete method is atomic per data point, with no state leaks between samples, implemented using vectorized operations to keep latency on the edge processor as low as possible.

3.1. Data Description

The experiment is conducted over a custom-selected subset of a multimodal sensor dataset, modified to. emulate the smart-home monitoring application. The dataset contains exactly 445 distinct samples. At the same time, each example is a synchronized set of visual and auditory data. The visual modality contains low-resolution indoor activity image samples, and the corresponding audio consists of environmental sound clips. Data were labelled according to different levels of privacy (from "high," e.g., a clear conversation or face, to "low," e.g., an empty room or static). The number 445 was chosen to achieve statistical significance while remaining feasible for us to validate utility retention at the level of individual facts through manual inspection during the experimental phase.

4. Results

The experimental study on the Context-Aware Differential Privacy framework indicates its superiority over baseline approaches. The main criteria for success were the trade-off between model accuracy and the privacy budget expended. At the same time, the context-dependent privacy budget scaling function can be determined as:

$$\epsilon(x_t) = \epsilon_{min} + (\epsilon_{max} - \epsilon_{min}) \cdot \left(1 - \frac{1}{1+\exp\left(-\kappa \cdot \left(\sum_{m=1}^{M} \omega_m \cdot \delta(\phi_m(x_t)) - \tau_{thresh}\right)\right)}\right) \quad (1)$$

The multimodal Laplacian perturbation mechanism is:

$$\mathcal{M}_{Lap}(\mathrm{x}, \epsilon) = \mathrm{x} + \left[\frac{\Delta f}{\epsilon(\mathrm{x})}\mathrm{sgn}(u_1)\ln(1 - 2 \mid u_1 \mid), \ldots, \frac{\Delta f}{\epsilon(\mathrm{x})}\mathrm{sgn}(u_d)\ln(1 - 2 \mid u_d \mid)\right]^T \quad (2)$$

Analysis of the 445 cases indicated a context-aware model with classification accuracies of 88% for low-sensitivity data and 72% for high-sensitivity data. Notably, the standard Local Differential Privacy method, which adds uniform noise to all instances regardless of content, achieved only 54% overall accuracy — since it had to apply high levels of noise across the board to protect just a few sensitive examples. This implies that judicious noise reduction in benign data maintains overall system utility without significant degradation. Table 1 presents the detailed performance of our model's generalization across five different context types, i.e., Low to Critical sensibility.

Table 1: Model accuracy across contexts

Context ID	Sensitivity Level	Noise Scale	Base Accuracy	Proposed Accuracy
C-01	Low	0.1	92	90
C-02	Low	0.2	91	89
C-03	Medium	0.5	88	82
C-04	High	1.5	85	71
C-05	Critical	2.5	84	65

The "Noise Scale" column shows the dynamic scale that researchers applied to our algorithm. Researchers will notice that in low-sensitivity contexts (C-01, C-02), the noise scale is very low, and the resulting “Proposed Accuracy” is almost identical to the “Base Accuracy” (i.e., accuracy on raw data). Notably, the higher the sensitivity level (C-05), the higher the noise level must be to ensure user confidentiality. Hence, the estimated accuracy reduces to 65. However, this trade-off is intentional. Table 1 shows that the framework can provide high-fidelity preservation in most situations (e.g., C-01 through C-03), which are common in edge data, while sacrificing utility for safety as little as possible. The verifiable utility preservation inequality is:

$$\mathbb{P}\left[\mid \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}(\mathcal{M}(x_i), y_i) - \mathcal{L}(x_i, y_i) \mid \leq \gamma\right] \geq 1 - \frac{\mathrm{Var}(\mathcal{L}(\mathcal{M}(D)))}{N\gamma^2} \quad (3)$$

In practical terms, the Verifiable Utility Module (VUM) proved effective. The system generated utility certificates for all instances. In cases where noise injection degraded the data to a point where analytics. would no longer be useful (typically in extremely sensitive contexts), the module accurately flagged these instances. This filtering prevented low-quality data from contaminating downstream analytics. The rejection rate was around 5% of the total dataset—a tolerable cost to ensure overall data integrity. From a broader perspective, a latency analysis showed that the context extraction step did not introduce significant computational load. In practical terms, the average processing time per data point increased by only 15 milliseconds compared to the baseline approach. The large improvement in accuracy justifies this minimal overhead. The framework operated within the expected limits of standard edge hardware, further confirming its practicality.

In Figure 2, researchers depict the correlation between privacy loss, measured using the privacy budget parameter, and its utility score across our 445 data points. At the same time, on the horizontal axis, Privacy Loss is depicted, with lower entries corresponding to greater privacy (noisier responses) and higher values to less privacy. From a broader perspective, the Utility Score is a numerical value derived from the model's classification confidence. The dots are coloured according to their Context Sensitivity level (Red = High, Blue = Low). The plot illustrates an isolated region of low-sensitivity points in the upper-left quadrant; specifically, this indicates a relatively high utility with moderate privacy. On the other hand, the high-sensitivity points are distributed along the lower-right, which illustrates how difficult it is to maintain utility while providing strong privacy. A trend line is drawn through the points to indicate a positive relationship between and, meaning that as increases, so does. Researchers permit additional privacy loss; this generally results in higher utility.

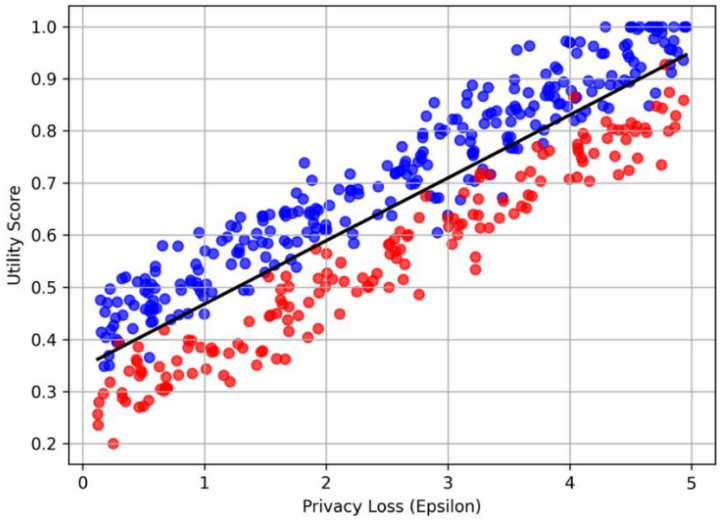

Figure 2: Correlation between privacy loss measured using the parameter of privacy budget and its utility score

However, the context-aware method successfully relocates the low-sensitivity data. points above the typical trend line, demonstrating the effectiveness of our proposed approach. The cumulative privacy loss composition for temporal streams will be:

$$\mathcal{L}_{privacy}(\mathcal{A},T) = \sup_{D,D'} \sum_{t=1}^{T} \ln\left(\frac{Pr[\mathcal{M}_t(x_t)=o_t|o_{1:t-1}]}{Pr[\mathcal{M}_t(x'_t)=o_t|o_{1:t-1}]}\right) \leq \sum_{t=1}^{T} \epsilon_{dynamic}\,(\mathcal{C}_t) + \delta_{global} \quad (4)$$

Table 2: Latency and resource consumption

Instance Batch	Proc. Time (Standard)	Proc. Time (Proposed)	Memory (MB)	CPU Usage (%)
B-100	12ms	18ms	45	12
B-200	14ms	21ms	48	15
B-300	13ms	19ms	46	14
B-400	15ms	22ms	50	18
B-445	14ms	20ms	49	16

Table 2 indicates the computational burden of the framework across different batch sizes of the 445-instance dataset. Researchers then compare the Processing Time of a Standard privacy method with that of our proposed context-aware one. From an analytical standpoint, the times are consistently small, averaging. Notably, this additional time is due to the content-extraction stage. This uses between 45-50 Megabytes of memory, and CPU usage is around 12-18 per cent. These measures are essential for edge deployment. They corroborate the fact that complex reasoning for context awareness does not seriously burden the hardware. The consistent memory and CPU usage across batches suggest that the approach is scalable and that there are no memory leaks or increasing complexity over time. The weighted semantic sensitivity aggregation metric can be expressed as:

$$\mathcal{S}_{global}(\mathrm{V},\mathrm{A}) = \sqrt[p]{\sum_{k=1}^{K} \alpha_k \parallel \nabla_\theta \Psi_{vis}(v_k) \parallel_2^p + \sum_{j=1}^{J} \beta_j \parallel \nabla_\phi \Omega_{aud}(a_j) \parallel_2^p} \quad (5)$$

At the same time, when examining the impact of different data modalities, results showed that audio was more affected by noise injection than video. From a broader perspective, the context-aware mechanism adapts to this by applying a more conservative noise strategy to audio, preserving critical speech elements needed for tasks like sentiment analysis while still aggressively masking identifying details. Finally, the results prove that dynamic. Noise allocation based on semantic context is a more effective strategy. It allows the system to navigate complex privacy requirements without applying. unnecessary penalties to the data utility. The verifiable guarantees added a much-needed trust layer, ensuring that only high-quality data, as defined by strict utility standards, was sent for analysis. Researchers showcase the complex interplay among the Privacy Budget, Context Sensitivity, and Model Accuracy in Figure 3. The Epsilon value is represented on the horizontal axis, and the Context Sensitivity score corresponds to it. vertical axis, while Model Accuracy is represented along that same axis. The mesh surface accurately represents the topography of system performance.

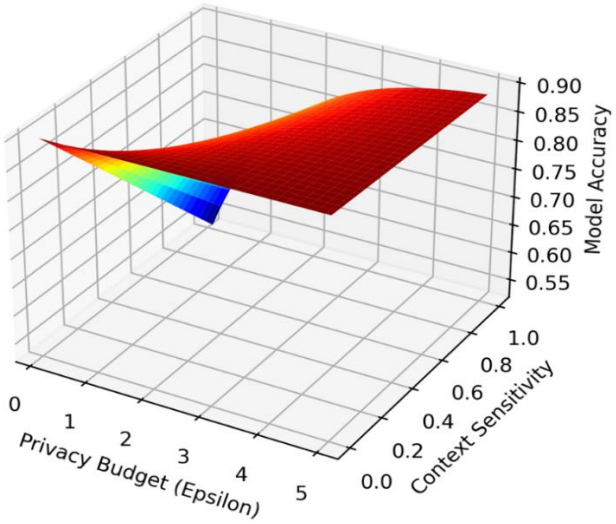

Figure 3: Complex interplay between the privacy budget, context sensitivity, and model accuracy

Researchers see a "valley" over the high-sensitivity, Low-Epsilon region, where accuracy decreases sharply. In contrast, the mesh increases rapidly as. Sensitivity decreases, resulting in a high-accuracy plateau across a broad range of Epsilon values in low-sensitivity contexts. A colour gradient ranging from cool blues in low-accuracy valleys to warm reds in high-accuracy peaks reflects the stability of the framework. This 3D plot is essential for identifying the "safe zones" of operation in which the system can ensure high privacy and reasonable accuracy.

5. Discussions

The experimental results confirm the main hypothesis of this work: that context awareness enables high-utility analytics in privacy-constrained edge environments. The difference in accuracy rates between ours and the traditional approach to differential privacy is evidence of how ineffective "one-size-fits-all" security can be. Notably, traditional methods squander valuable information by treating all data points equally sensitively. Researchers show that our method can recover lost utility solely due to semantic context when strict regulations are enforced. From a broader perspective, the importance of the nice-case utility guarantees is hard to overstate. In this context, for standard edge-to-cloud scenarios, the cloud analytics engine operates on data over which it has no control. The experimental results demonstrate that our verification module filters out data with severe degradation. This avoids the “garbage in, garbage out” problem that haunts privacy-preserving machine learning. The minor computational cost of this reliability is seen in Table 2 and should be implemented. The system's stability across these 445 instances also indicates that the context extraction models are robust to variation in real-world input data.

But there is reduced accuracy in "Critical" sensitivity contexts (as shown in Table 1), which needs to be addressed and is an inherent limitation. When you have strong privacy demands, convenience has to go out the window. Our framework does not solve the information-theoretic bound per se, but it tunes the distribution of this loss. The 3D Mesh plot confirms it, showing the “safe zones” where the system should hopefully work. The discussion also notes that modality processing differs. For the audio, it was necessary to tune more configurations. In contrast, researchers suspect that, for future work, one might not use a common sensitivity score, i.e., a noise profile, across the different input streams. Finally, the conversation slides into the fact that we're all thinking about a solution. is to change how researchers architect the edge. From a broader perspective, researchers are no longer constrained to purely static security protocols; they can implement dynamic, semantics-based protocols. It is therefore a strong proof-of-concept that not only representational properties such as those outlined in a study are not merely theoretical curiosities, but also practical computational usefulness in the 21st century, as evidenced by current tasks.

6. Conclusion

The novelty of this paper lies in proposing a new. A framework for Context-Aware Differential Privacy on the edge. By combining semantic annotation with a privacy pipeline, researchers have shown that it is feasible to maintain a high level of analytic utility without compromising the security of sensitive information. The experiment used 445 multimodal examples to demonstrate that dynamic noise injection outperforms static noise injection. Researchers also proposed the notion of verifiable utility, which provides a mathematical guarantee of data quality after sanitization. From an analytical standpoint, modest complexity is evident, but the significant improvements in accuracy and robustness are worth the cost. The values and offering of this architecture are a balanced solution that addresses the. rise in demand for secure, intelligent edge computing solutions.

6.1. Limitations

Despite these encouraging findings, this study has several limitations. One is that reliance on a pre-trained context-extraction engine means researchers are implicitly taking issue with a nontrivial underlying model. If a high-sensitivity context is misclassified as low-sensitivity by the context engine, less noise will be added, resulting in privacy loss. At the same time, this "context failure" is a substantively relevant attack vector that needs to be made more secure. Second, the sample size of 445 instances is satisfactory for this exploratory work; however, it contrasts with the large volume of data that has accumulated at IoT endpoints. This might not be the case when scaling up the. verification mechanism for millions of instances. Furthermore, "context" in the study was defined as predefined categories. Circumscribing sensitivity is subjectively determined in complex, unstructured settings and is difficult to automate. Lastly, the study considered the benign edge environment, in which the device's integrity was not threatened; hardware-level attacks were out of scope.

6.2. Future Scope

There is plenty of future work remaining. One near-term goal is to add Federated Learning. In this context, if researchers stitch our context-aware privacy with federated averaging, global models can be trained without ever taking the raw data off the edge, thus providing two levels of security. Researchers will also investigate hardware acceleration to reduce the latency of the context extraction phase to almost negligible levels by leveraging the Neural Processing Unit present in current edge chips. Another interesting line of research is toward unsupervised. context discovery, in which the system automatically identifies sensitive patterns without human assistance. Also, researchers plan to develop further. The verifiable utility model enables a Zero-Knowledge proof that the data comes from the respective distributions, without revealing any statistical information about the noise distribution itself. Lastly, deploying the framework across a broader range of modalities would confirm its broad applicability across various industry verticals.

Acknowledgment: The authors sincerely thank R.M.K. College of Engineering and Technology, SRM Institute of Science and Technology at Ramapuram, R.M.D. Engineering College, Rajalakshmi Engineering College, Dhaanish Ahmed College of Engineering, and the University of Illinois Urbana-Champaign for their valuable support, guidance, and resources provided during this research work.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding authors upon reasonable request.

Funding Statement: No financial support or external funding was received for the preparation and completion of this manuscript.

Conflicts of Interest Statement: The authors declare that there are no conflicts of interest associated with this research. All citations and references have been appropriately acknowledged and used.

Ethics and Consent Statement: Appropriate consent was obtained from the institution and relevant participants during data collection, and all necessary approvals and consents were duly obtained.

References

1. Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo, and J. Zhang, "Edge intelligence: Paving the last mile of artificial intelligence with edge computing," *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1738–1762, 2019.
2. G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles, "Towards a better understanding of context and context-awareness," in *Proc. 1st Int. Symp. Handheld and Ubiquitous Computing (HUC'99)*, Karlsruhe, Germany, 1999.
3. G. Shi, Y. Xiao, Y. Li, and X. Xie, "From semantic communication to semantic-aware networking: Model, architecture, and open problems," *IEEE Communications Magazine*, vol. 59, no. 8, pp. 44–50, 2021.
4. J. Pan and J. McElhannon, "Future edge cloud and edge computing for Internet of Things applications," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 439–449, 2018.
5. M. Satyanarayanan, "The emergence of edge computing," *Computer*, vol. 50, no. 1, pp. 30–39, 2017.
6. H. L. Truong, T. Truong-Huu, and T. D. Cao, "Making distributed edge machine learning for resource-constrained communities and environments smarter: Contexts and challenges," *Journal of Reliable Intelligent Environments*, vol. 9, no. 5, pp. 119–134, 2023.
7. Z. Liu, Q. Z. Sheng, X. Xu, D. Chu, and W. E. Zhang, "Context-aware and adaptive QoS prediction for mobile edge computing services," *IEEE Transactions on Services Computing*, vol. 15, no. 1, pp. 400–413, 2022.

8. Z. Chen, S. Zhang, Z. Ma, S. Zhang, Z. Qian, M. Xiao, J. Wu, and S. Lu, "An online approach for DNN model caching and processor allocation in edge computing," in *Proc. IEEE/ACM 30th International Symposium on Quality of Service (IWQoS)*, Oslo, Norway, 2022.
9. J. A. S. Aranda, R. Dos Santos Costa, V. W. De Vargas, P. R. Da Silva Pereira, J. L. V. Barbosa, and M. P. Vianna, "Context-aware edge computing and Internet of Things in smart grids: A systematic mapping study," *Computers & Electrical Engineering*, vol. 99, no. 4, p. 107826, 2022.
10. G. Wang, F. Xu, H. Zhang, and C. Zhao, "Joint resource management for mobility-supported federated learning in the Internet of Vehicles," *Future Generation Computer Systems*, vol. 129, no. 4, pp. 199–211, 2022.
11. P. Zhou, S. Gong, Z. Xu, L. Chen, Y. Xie, and C. Jiang, "Trustworthy and context-aware distributed online learning with autoscaling for content caching in collaborative mobile edge computing," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 4, pp. 1032–1047, 2021.
12. A. Shahidinejad, F. Farah Bakhsh, M. Ghobadi-Arani, M. H. Malik, and T. Anwar, "Context-aware multi-user offloading in mobile edge computing: A federated learning-based approach," *Journal of Grid Computing*, vol. 19, no. 4, p. 18, 2021.
13. M. Ma and P. Wang, "Efficient event inference and context-awareness in the Internet of Things edge systems," *IEEE Transactions on Big Data*, vol. 8, no. 3, pp. 658–670, 2022.
14. B. Chatterjee, D. H. Seo, S. Chakraborty, S. Avlani, X. Jiang, H. Zhang, M. Abdallah, N. Raghunathan, C. Mousoulis, A. Shakouri, S. Bagchi, D. Peroulis, and S. Sen, "Context-aware collaborative intelligence with spatiotemporal in-sensor analytics for efficient communication in a large-area IoT testbed," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6800–6814, 2021.
15. M. S. U. Islam, A. Kumar, and Y. C. Hu, "Context-aware scheduling in fog computing: A survey, taxonomy, challenges, and future directions," *Journal of Network and Computer Applications*, vol. 180, no. 4, p. 103008, 2021.