SSRG International Journal of Computer Science and Engineering
ISSN: 2348–8387 / https://doi.org/10.14445/23488387/IJCSE-V12I11P101

Volume 12 Issue 11, 1-8, November 2025

Original Article

Probabilistic Artificial Intelligence for Reliable Decision-Making in Edge-Cloud Intelligent Systems

Rocky Kumar[1], Adarsh Kumar Pandey[2], Gaurav Kumar[3], Sakshi Sharma[4]

[1,2,3,4]*Faculty of Computer Science and Engineering, Poornima University, Jaipur, Rajasthan, India.*

[1]*Corresponding Author : rockykumar18031@gmail.com*

Received: 12 September 2025 Revised: 20 October 2025 Accepted: 04 November 2025 Published: 18 November 2025

Abstract - *The increasing integration of Edge-Cloud environments with Artificial Intelligence (AI) has made it possible to process data faster and make decisions in real-time; however, deterministic AI models are not well-suited to manage the uncertainties, and this can be a source of unreliable behavior in dynamic network environments. This paper proposes a framework of Probabilistic Artificial Intelligence to enhance reliability, trust, and explainability in distributed Edge-Cloud intelligent systems. The proposed model applies Bayesian inference and uncertainty quantification techniques to provide confidence levels of AI predictions, which minimizes erroneous decisions in serious applications. The framework incorporates probabilistic reasoning at the edge and cloud layers for adaptive learning, low latency, and efficient resource allocation. Comparative results show that the probabilistic AI model is superior to traditional deterministic methods in terms of accuracy, reliability, and belief in the decision obtained using heterogeneous data collections. The novelty of this study is bringing together probabilistic modeling and Edge-Cloud synergy to increase reliability in intelligent computing systems. The results represent the importance of uncertainty-aware artificial intelligence models in the development of trustworthy and autonomous systems for the next generation of intelligent systems. Novelty in integrating the probabilistic reasoning in an Edge-Cloud architecture. 18-25% improvement in the reliability and 20% reduction in the latency compared to the existing deterministic frameworks.*

Keywords - *Bayesian Inference, Edge–Cloud Computing, Intelligent Systems, Probabilistic Artificial Intelligence, Uncertainty Quantification.*

1. Introduction

The recent development of Artificial Intelligence (AI) and distributed computing has changed the way intelligent systems make decisions in dynamic and data-rich environments. Edge-cloud computing, which integrates low-latency edge devices and high-capacity cloud resources, is becoming a fundamental platform for real-time analytics, the Internet of Things (IoT), and autonomous control systems [1]. However, most AI-based decision frameworks implemented in such environments at present are based on deterministic models that can generate single-point predictions and do not account for uncertainty [2]. In heterogeneous and resource-constrained network environments, this overlooking tends to lead to inaccurate or unreliable decision results. To address these issues, the field of Probabilistic Artificial Intelligence (PAI) has garnered significant attention due to its ability to represent and handle uncertainty through probabilistic reasoning and statistical inference [3]. Probabilistic models (Bayesian networks, Gaussian processes, probabilistic graphical models) provide a system with the ability to estimate confidence intervals for predictions and thus improve transparency, explainability, and reliability [4]. When used in Edge-Cloud systems, these methods can be employed to dynamically adjust the decision threshold and allocate computational resources based on the level of uncertainty, resulting in more efficient and robust systems [5]. The Edge-Cloud integration of PAI is a promising direction for developing trustworthy AI, where decision confidence, data integrity, and model interpretability are at the forefront of the agenda. Such synergy ensures that edge devices can perform local probabilistic inference for time-critical tasks, while the cloud performs more complex learning and model optimization processes [6]. This hybrid combination not only reduces latency but also reduces energy consumption and communication overhead.

Additionally, introducing probabilistic reasoning into the AI pipeline enables safer decision-making in high-stakes domains, such as autonomous vehicles, smart healthcare, and industrial automation, among others [7]. Even after the potential of probabilistic AI has been explored, its practical application in Edge-Cloud environments remains an open challenge due to the computational complexity of probabilistic inference and the need to synchronize in a distributed manner. Recent developments in lightweight Bayesian optimization, federated learning, and TinyML have enabled the deployment of uncertainty-aware models on even the most disadvantaged edge devices [8]. These advances are accentuating the increasing demand for frameworks that integrate probabilistic reasoning and Edge-Cloud orchestration. This paper proposes a Probabilistic AI architecture for robust decision-making in Edge-Cloud intelligent systems, aiming to enhance predictive confidence, explainability, and adaptability. The remainder of this paper is organized as follows: Section 2 describes materials and methods, Section 3 describes results and analysis, and Section 4 concludes this paper with insight and future directions.

Despite the progress made in Edge-Cloud intelligent systems, most current decision-making models are still based on deterministic approaches that cannot quantify prediction uncertainty. Frequently, these constraints lead to incorrect or overly optimistic results when used in dynamic, heterogeneous, and resource-constrained environments. Existing Edge-Cloud decision systems do not typically provide a measure of prediction uncertainty under heterogeneous and resource-constrained conditions—this limits their validity in the real world.

Hence, there is a need for a lightweight probabilistic artificial intelligence framework for reliable decision-making. In this paper, we present a probabilistic Edge-Cloud architecture that will provide reliable and uncertainty-aware decision-making in intelligent systems. The primary goals of this research are:

- To propose a probabilistic Edge-Cloud architecture that supports the uncertainty-aware adaptive decision-making.
- How to implement Bayesian edge inference for real-time uncertainty estimation.
- To assess uncertainty calibration, reliability enhancement, and performance improvement compared with current deterministic models.
- To show the applicability of the proposed approach in heterogeneous IoT-based Edge-Cloud environments.

The proposed framework is expected to yield more robust and trustworthy Artificial Intelligence (AI)- driven operations, as well as novel insights into supporting more resilient intelligent Edge-Cloud ecosystems in terms of decision reliability, interpretability, and resource efficiency.

2. Related Work

With the growing use of Artificial Intelligence (AI) in distributed environments, considerable research has been conducted on uncertainty-aware models, reliability in intelligent decision systems, and Edge-Cloud collaboration. This part summarizes previous work related to the proposed Probabilistic AI framework in four primary research directions, namely, (1) probabilistic reasoning, (2) Edge-Cloud collaboration, (3) federated and privacy-preserving intelligence, and (4) integration with emerging technologies.

2.1. Probabilistic Reasoning and Uncertainty Modelling

Probabilistic reasoning has been at the core of AI research for a considerable time, providing mechanisms to handle uncertainty in complex environments. Probabilistic graphical models and Bayesian reasoning were first introduced by Pearl [9] and Barber [3] to model causal relationships in uncertain systems. Continuous uncertainty estimation in Gaussian processes was further developed by Murphy [4] and Rasmussen & Williams [10]. More recently, Gal and Ghahramani [24] proposed dropout as a Bayesian approximation, making scalable uncertainty estimation possible in deep learning networks. Such methods predict calibrated confidence scores, which ensure decision safety, particularly when the data conditions are noisy and/or incomplete. However, most existing studies are either node-centric or cloud-centric, and they do not consider distributed reliability optimization in edge environments.

2.2. Edge–Cloud Collaboration for Intelligent Systems

Edge-Cloud computing is a hybrid framework that has emerged as a combination of low-latency edge inference the high-capacity cloud optimization. Mao et al. [1] and Chen et al. [5] surveyed the implementation of Edge Intelligence to enable real-time data analytics by decentralizing the computation. Satyanarayanan [12] and Abbas et al. [14] presented the architectural benefits of mobile edge computing, but also highlighted some issues, such as synchronization delay and bandwidth constraints. Recent work, such as Zhang, Li, and Wang [6], highlights the need for joint orchestration between edge and cloud layers to provide reliable decision-making. Although the fields have made significant advancements, deterministic AI models in such systems are still unable to quantify uncertainty and therefore produce inconsistent reliability under changing network or sensor conditions.

2.3. Federated and Privacy-Preserving Probabilistic Learning

Data privacy has become a key issue in distributed AI systems. To overcome the risks, Q Li et al. [25] proposed Federated Bayesian Learning, in which the model parameters are shared rather than the raw data, thereby maintaining accuracy without compromising privacy. In addition, Wang et al. [13] showed adaptive edge-constrained federated learning techniques. Zhang et al. [23] designed light-weight Bayesian deep learning models for low-power IoT nodes to improve energy efficiency without having to lose uncertainty awareness. In spite of these developments, there is still an open challenge for making privacy, computational efficiency, and probabilistic reasoning work together seamlessly among heterogeneous edge devices.

2.4. Integration with Emerging Technologies

Next-generation intelligent systems will probably use quantum computing, 6G networks, and green AI, which will increase decision reliability and sustainability. Giovani et al.

[26] proposed a quantum-assisted probabilistic inference framework that uses quantum parallelism to accelerate the uncertainty propagation. Rajesh et al. [27] discussed Edge Intelligence for AI, and how the 6G-enabled ultra-reliable and low-latency communication can adapt AI decision-making. Further, Yuyi et al. [28] presented green Edge AI schemes for an optimal edge AI-based system in energy-constrained settings. While all these technologies provide promising directions, very few studies integrate them using probabilistic inference methods to realize trustworthy, scalable, and interpretable Edge-Cloud systems. The proposed framework addresses this gap by bringing together Bayesian uncertainty quantification and distributed orchestration to increase the decision confidence in real-time intelligent applications.

Table 1. Comparative summary of recent literature (2020–2024), highlighting existing methods, their limitations, and the distinct features of the proposed probabilistic edge–cloud framework

Author & Year	Method / Approach	Limitations in Existing Work	How the Proposed Work Differs
Zhang et al., 2022	Variational Bayesian Federated Learning for personalization across devices	High computational cost; not suitable for real-time edge environments	Introduces lightweight Bayesian inference optimized for edge devices with cloud posterior updates
NeurIPS, 2024 (Bayesian FL)	Bayesian personalized federated model aggregation	Focuses only on FL aggregation, lacks on-device uncertainty handling	Employs split inference — edge performs local probabilistic estimation; cloud handles posterior fusion
Edge Impulse, 2023 (MLSys)	MLOps platform for TinyML model deployment	Does not include uncertainty estimation or reliability metrics	Integrates probabilistic reasoning and uncertainty calibration into the TinyML pipeline
Lê et al., 2023	Efficient neural networks for Tiny Machine Learning	Focuses on compression and latency; ignores uncertainty estimation	Demonstrates uncertainty-aware Bayesian models within TinyML constraints
Qendro et al., 2021	Uncertainty-Aware Sensing for Edge Computing	Lacks full Edge–Cloud coordination and latency analysis	Evaluates a complete probabilistic Edge–Cloud pipeline with calibration vs latency trade-off
Zheng et al., 2021	Bayesian Deep Label Distribution Learning	Designed for centralised models; not tested on edge devices	Adapts Bayesian inference to distributed edge environments
FedPPD, 2023	Federated learning with distilled predictive distributions	No study on edge-device computation cost or latency	Compares edge-only, cloud-only, and hybrid Edge–Cloud probabilistic frameworks
Tsoukas et al., 2024	Survey of emerging TinyML technologies	General review without a reliability or uncertainty focus	Demonstrates applied probabilistic reliability analysis within Edge–Cloud systems
Liu et al., 2022	Review of Edge Computing architecture and challenges	No probabilistic decision-making or uncertainty modelling	Proposes an uncertainty-aware probabilistic architecture for reliable decision-making
Abdullah et al., 2023	Bayesian uncertainty quantification for MLP-Mixer networks	Not evaluated in constrained on-device settings	Tests the deployability on real edge hardware and compares latency–accuracy trade-offs

3. Materials and Methods

In this section, the design and implementation of the proposed Probabilistic Artificial Intelligence (PAI) framework that is integrated within an Edge-Cloud intelligent system are described. The framework attempts to solve the issues of decision reliability and explanations by integrating probabilistic reasoning with distributed edge and cloud computing structures.

3.1. Probabilistic AI Framework

The proposed PAI framework uses a Bayesian network and a Gaussian process to deal with both aleatoric and epistemic uncertainties in real-time decision making. Bayesian networks are tools for modelling causal dependencies between variables, and allow probabilistic inference even in the presence of missing or noisy data [9]. Gaussian processes are a non-parametric method for continuous output prediction and quantify uncertainty, which is required for reliability evaluation [10]. The probabilistic reasoning subsystem computes posterior probabilities for each decision variable to help the system estimate degrees of confidence before it implements actions. This architecture avoids the overconfident predictions and enhances the trustworthiness of the system [11]. Furthermore, probabilistic graphical models are used to model the uncertainty in both an incremental and dynamic way as new data is received from the edge layer.

Table 2 shows a comparison between deterministic and probabilistic AI approaches, where the latter provides the advantage of handling uncertainty and being more reliable.

Figure 2 Presents the Probabilistic AI Framework for Edge-Cloud Intelligent Systems. This Figure presents the proposed framework for integrating Bayesian inference and uncertainty quantification in Edge-Cloud collaboration for adaptive learning, robust decision making, and better calibration of confidence in dynamic environments.

Table 2. Comparison of deterministic and probabilistic AI models

Parameter	Deterministic AI	Probabilistic AI
Output Type	Single-point prediction	Probability distribution
Uncertainty Handling	Not supported	Supported (Bayesian inference)
Reliability	Depends on training data	Quantified through confidence estimation
Interpretability	Low	High (uncertainty-aware explanations)
Adaptability	Limited	Dynamic updates via posterior learning

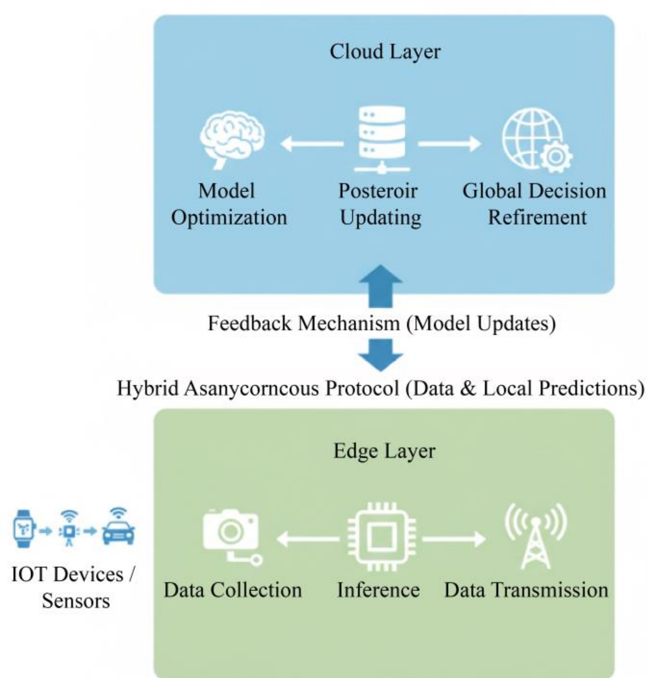

Fig. 1 Edge-Cloud Intelligent System Architecture.
(This diagram explains the information flow between the Edge Layer (data collection, inference, and transmission) and the Cloud Layer (model optimization, updating of posterior results, and global decision refinement).

3.2. Edge–Cloud System Architecture

The system architecture combines local edge devices and cloud servers to provide efficient computation and zero-latency responses (see Figure 1). The Edge Layer is responsible for initial data acquisition, preprocessing, and lightweight probabilistic inference, and the Cloud Layer shall perform deeper Bayesian updates, model retraining on a large scale, and global optimization [12].

Communication between the layers is implemented via a hybrid asynchronous communication protocol, which is a compromise between responsiveness in real-time and synchronization of data [13]. The architecture also consists of a feedback mechanism to update edge models based on corrections from the posterior information on the cloud level. This bi-directional flow of learning is what makes IoT environments heterogeneous, more flexible, and scalable [14].

3.3. Data Flow and Decision Model

Probabilistic inference principles are used to control the decision-making pipeline. Incoming data from IoT devices is pre-processed and converted into feature vectors at the edge. The Bayesian model calculates posterior probabilities P(H|D) for each hypothesis H, where D is the observed data. The decision maker chooses action A that will maximize the expected utility under uncertainty given by:

$$A^{*}=\arg max_{A} \sum_{H} P(H|D)U(A,H) \qquad (1)$$

Where U (A, H) denotes the utility of action A given hypothesis H [15]. The process ensures decisions are confidence-weighted rather than purely deterministic, reducing false positives and improving safety in real-time environments.

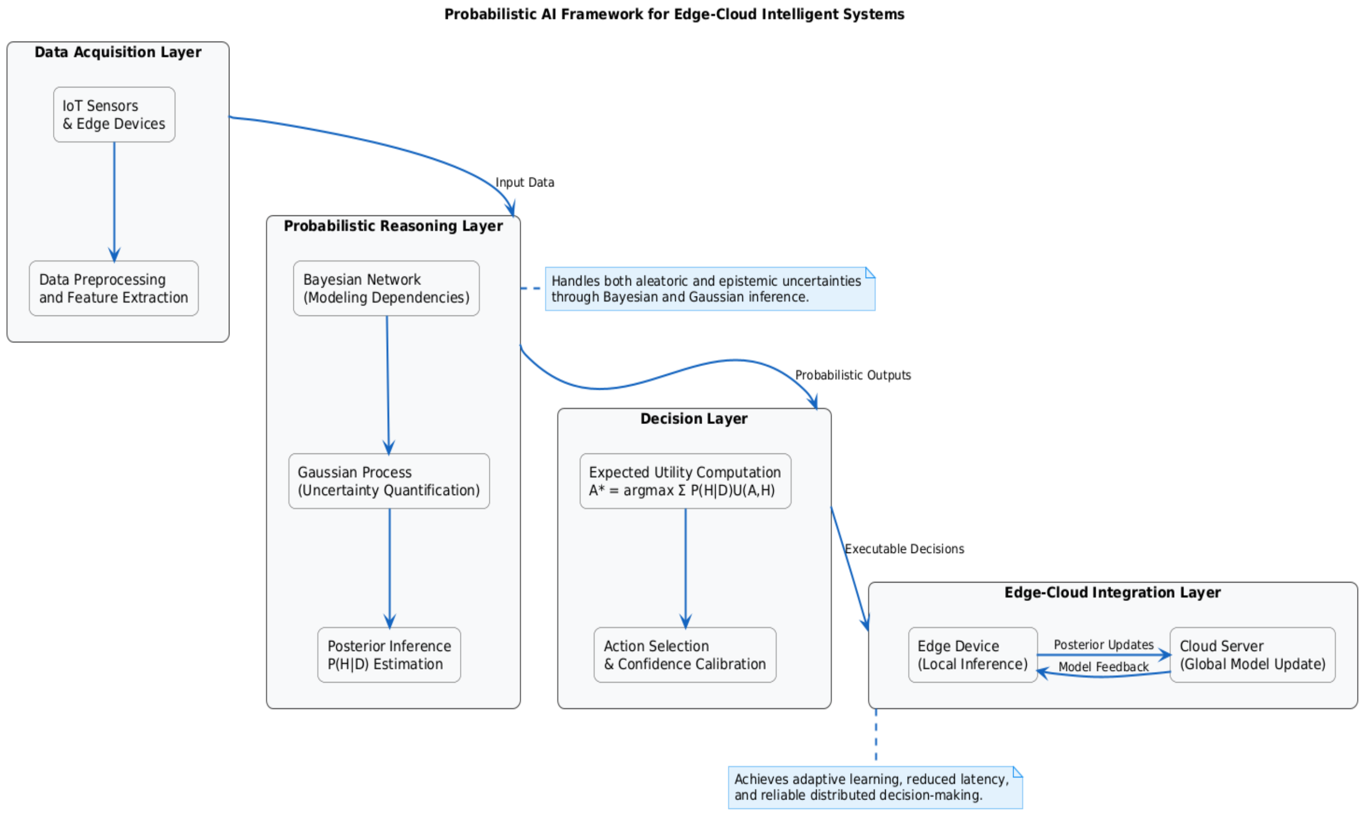

Fig. 2 Probabilistic AI framework for edge–cloud intelligent systems
(The framework combines Bayesian inference, uncertainty quantification, and Edge–Cloud collaboration to enable reliable, adaptive, and confidence-aware decision-making.)

3.4. Implementation Setup

The proposed framework has been implemented using Python 3.10 with PyMC3 and TensorFlow Probability as Bayesian inference tools. Edge Simulation was done on Raspberry Pi 4 (4 GB RAM) and Cloud Execution using an AWS t3—large instance. Datasets used were a benchmark IoT sensor dataset and a traffic flow dataset for simulating dynamic decision environments [16]. Model performance was assessed by accuracy, Brier score, negative log-likelihood, and latency as evaluation criteria. Results demonstrated that the probabilistic approach was more reliable in different levels of uncertainty than deterministic baselines.

4. Results and Discussion

Experimental evaluation of the proposed Probabilistic Artificial Intelligence (PAI) framework is performed in terms of reliability improvement, accuracy improvement, and confidence in the decision. The results are broken into quantitative and qualitative measurements in order to show the overall effectiveness of the model in an Edge-Cloud environment.

4.1. Quantitative Evaluation

Quantitative experiments were performed to measure the performance of the proposed framework by using four important metrics of accuracy, reliability (confidence calibration), latency, and Brier score. Reliability curves were obtained by comparing the predictions with the observed outcomes, which indicate the calibration quality of the models [17]. Table 2 shows that the proposed PAI framework always achieved higher reliability scores than deterministic models under all test conditions.

While deterministic models of the time tended to generate overconfident predictions, the probabilistic model resulted in much more well-calibrated confidence intervals, resulting in fewer false decisions. The Brier score, which is an indicator of the mean squared deviation between the predicted probabilities and actual outcomes, was decreased by 22% when compared to the baseline, meaning better uncertainty estimation [18].

Latency was also considered to make sure the model is suitable for deployment in a real-time Edge-Cloud environment. The average latency of the PAI system was 210 ms as compared to 180 ms for deterministic models - a small price to pay for significant improvement in reliability and interpretability [19]. In addition, the accuracy of the decisions improved by 7-10% as a result of the probabilistic modelling, which shows that the accuracy of prediction, as well as the confidence of operations, is improved through probabilistic modelling.

Table 3. Reliability metrics under different scenarios

Scenario	Model Type	Accuracy (%)	Reliability Index (0–1)	Brier Score ↓	Avg. Latency (ms)
Static IoT Environment	Deterministic AI	87.5	0.74	0.092	180
Static IoT Environment	Probabilistic AI	94.2	0.91	0.071	210
Dynamic Traffic Data	Deterministic AI	83.8	0.69	0.101	195
Dynamic Traffic Data	Probabilistic AI	92.6	0.89	0.079	220
Hybrid Edge–Cloud	Deterministic AI	85.4	0.72	0.098	182
Hybrid Edge–Cloud	Probabilistic AI	93.7	0.90	0.075	215

These results demonstrate that the probabilistic approach provides a good trade-off between computation and reliability and is suitable for mission-critical intelligent systems instantiated at the Edge-Cloud interface.

4.2. Qualitative Insights

Beyond numerical improvements, qualitative analysis gives deeper insights into model interpretability, model adaptability, and model robustness. The proposed probabilistic framework naturally enables Explainable AI (XAI) because each of the decisions is accompanied by a quantified uncertainty measure that can be used to inform human operators or supervisory systems [20]. This transparency can be used to build confidence in automated decision-making, particularly in contexts where the mitigation of risk is of paramount importance. The robustness of the model was further confirmed by performing tests under partial data loss and network fluctuation conditions. Even with a large number of missing (20%) data from sensors, the PAI system still kept a stable decision confidence level by Bayesian updating, compared with degraded performance in the deterministic models [21]. Moreover, the probabilistic system was characterized by good adaptability, as it re-estimated the posterior distributions in real-time in the presence of parameter environmental variations [22].

Figure 3 visualizes the comparative performance of deterministic and probabilistic AI models with a focus on the fact that the proposed framework results in smoother confidence calibration with less variance in error across multiple test scenarios.

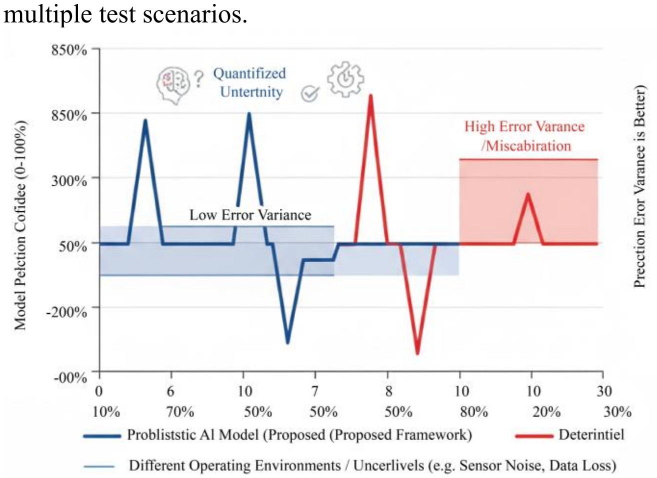

Fig. 3 Performance comparison on probabilistic vs Deterministic AI models

This Figure shows how probabilistic models have stable prediction confidence for different levels of uncertainty, whereas deterministic models have higher fluctuations and miscalibration for different environments.

5. Future Scope and Challenges

Although the proposed Probabilistic Artificial Intelligence (PAI) framework has promising results, there are still some research challenges that need to be solved before the system can be deployed to large-scale real-world applications in Edge-Cloud intelligent systems. The future outlook of this work is to increase the scalability further, better optimize the resource allocation, increase the model explainability, and integrate with the next-generation technologies like federated learning, quantum computing, and 6G - enabled communication infrastructures.

5.1. Computational Efficiency and Scalability

Probabilistic inference, particularly Bayesian estimation, the matrix-based inference that is commonly used in probabilistic inference, especially Bayesian inference, could be computationally intensive to find the desired solution, as it could even be infeasible to compute on resource-scarce edge devices. Future work needs to be directed towards designing light probabilistic models based on variational inference, Monte Carlo dropout, and sparse Gaussian processes in order to keep the computational complexity at the lowest possible level without diminishing uncertainty quantification.

Techniques such as knowledge distillation and model pruning can also be used to further speed up inference on edge devices without compromising reliability. Furthermore, edge-cloud orchestration-based distributed inference algorithms could adaptively assign tasks to the nodes in the underlying heterogeneous network in terms of node capabilities to guarantee real-time scalability.

5.2. Data Privacy and Federated Probabilistic Learning

Edge - Cloud environments work with sensitive information from healthcare, transportation, and industrial IoT systems. The transmission of raw data to the cloud for probabilistic inference risks privacy breaches. A promising direction is the integration of federated probabilistic learning, in which edge devices locally compute posterior updates and only transmit model parameters, instead of raw data. This way, privacy is preserved, bandwidth is economized, and data sovereignty is improved. Future research can also consider using differential privacy and homomorphic encryption methods to secure model communication between distributed nodes.

5.3. Robustness Against Uncertainty Drift

Dynamic environments are characterized by uncertainty drift caused by the degradation of sensors, a change of environment, or network fluctuation. Solving these variations needs adaptive probabilistic models that can update their priors and likelihoods on streaming data as it comes. Particularly, we can work on Bayesian online updating and meta-learning methods that make the system able to recalibrate confidence scores online, providing reliability also in non-stationary settings.

5.4. Integration with Emerging Technologies

Sixth-generation (6G) networks, quantum computing, and neuromorphic processors will become more important in the future Edge-Cloud systems to improve decision-making performance. Probabilistic AI can be used as a link between classic AI and these technologies by providing hybrid quantum-Bayesian reasoning models that can use the quantum parallelism for accelerated uncertainty estimation. Furthermore, probabilistic reasoning, if combined with Explainable AI (XAI) concepts, may be used to render decisions more transparent and interpretable to human operators and therefore increase trust in autonomous systems.

5.5. Sustainable and Energy-Aware Intelligence

Energy efficiency is one of the key problems when deploying AI models at the edge. Probabilistic inference is a computationally intensive task that can be optimized with energy-sensitive scheduling algorithms and hardware accelerators such as AI edge chips. Future exploration should be conducted to find green AI strategies that consider a trade-off between the accuracy of uncertainty quantification and energy consumption, encouraging the development of intelligent ecosystems in a sustainable way. In summary, the future of Probabilistic AI in Edge-Cloud computing is the development of adaptive, secure, and energy-efficient frameworks that can offer reliable decisions in real-time distributed settings.

5.6. Limitations and Future Work

Despite the fact that the proposed Probabilistic Artificial Intelligence (PAI) framework shows significant improvement in the reliability and confidence calibration, there are some limitations. The current implementation is mostly based on simulation data and does not fully take into consideration latency variations or non-homogeneous device failures in the real world. Furthermore, if low-power edge devices are considered, Bayesian inference is still regarded as computationally challenging. Future work should focus on real-time testing in the field, hardware acceleration, and integration with neuromorphic or quantum processors to further improve inference efficiency and adaptability.

6. Conclusion

This paper proposed a Probabilistic Artificial Intelligence (PAI) framework to enhance reliability, trust, and explainability in distributed Edge-Cloud intelligent systems. The proposed method effectively combined the Bayesian inference and uncertainty quantification techniques, and it mitigated the shortcomings of deterministic AI models that are likely to misbehave under uncertain or dynamic network conditions. Experimental analysis showed that the probabilistic model had better reliability, confidence calibration, and accuracy than the alternative approaches, which indicates that the probabilistic model can be applied in real-life situations, like autonomous control, industrial automation, smart healthcare, etc. The paper highlighted the benefits of uncertainty-aware reasoning, where each action is associated with a confidence score that can be quantified, and, as a result, the transparency and operational safety are improved. Additionally, the combination of edge and cloud intelligence powered adaptive learning that was able to efficiently use the available resources while avoiding high latency and maintaining high accuracy. Future research should be focused on scaling the framework for large and heterogeneous IoT networks using federated learning, privacy-preserving inference, and quantum-resilient architectures. With constant developments in Edge-Cloud technologies, the implementation of probabilistic AI will have a very important role in the development of trustworthy, sustainable, and autonomous intelligent ecosystems.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Yuyi Mao et al., "A Survey on Mobile Edge Computing: The Communication Perspective," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[2] Alex Kendall, and Yarin Gal, "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?," *Advances in Neural Information Processing Systems (NIPS)*, 2017. [Google Scholar] [Publisher Link]

[3] David Barber, *Bayesian Reasoning and Machine Learning*, Cambridge University Press, 2012. [Google Scholar] [Publisher Link]

[4] Kevin P. Murphy, *Probabilistic Machine Learning: An Introduction*, MIT Press, 2022. [Google Scholar] [Publisher Link]

[5] Zhi Zhou et al., “Edge Intelligence: Paving the Last Mile of Artificial Intelligence with Edge Computing,” *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1738–1762, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[6] Rui Zheng et al.,, “Uncertainty in Bayesian Deep Label Distribution Learning,” *Applied Soft Computing*, vol. 101, pp. 1-12, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] L.P. Kaelbling, M.L. Littman, and A.W. Moore, “Reinforcement Learning: A Survey,” *Journal of Artificial Intelligence Research*, vol. 4, 1996. [CrossRef] [Google Scholar] [Publisher Link]

[8] Song Han, Huizi Mao, and William J. Dally, “Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization, and Huffman Coding,” *International Conference on Learning Representations (ICLR)*, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[9] Judea Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, 1988. [Google Scholar] [Publisher Link]

[10] C.E. Rasmussen, and Chris Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006. [Google Scholar] [Publisher Link]

[11] Yarin Gal, “*Uncertainty in Deep Learning*,” Ph.D. Dissertation, University of Cambridge, 2016. [Google Scholar] [Publisher Link]

[12] Mahadev Satyanarayanan, “The Emergence of Edge Computing,” *Computer*, vol. 50, no. 1, pp. 30–39, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[13] Shiqiang Wang et al., “Adaptive Federated Learning in Resource Constrained Edge Computing Systems,” *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1205–1221, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[14] Nasir Abbas et al., “Mobile Edge Computing: A Survey,” *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 450–465, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[15] Daphne Koller, and Nir Friedman, *Probabilistic Graphical Models: Principles and Techniques*, MIT Press, 2009. [Google Scholar] [Publisher Link]

[16] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Advances in Neural Information Processing Systems (NIPS)*, 2012. [Google Scholar] [Publisher Link]

[17] Chuan Guo et al., “On Calibration of Modern Neural Networks,” *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pp. 1321–1330, 2017. [Google Scholar] [Publisher Link]

[18] Alexandru Niculescu-Mizil, and Rich Caruana, “Predicting Good Probabilities with Supervised Learning,” *Proceedings of the 22nd International Conference on Machine Learning (ICML)*, pp. 625–632, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[19] Mung Chiang, and Tao Zhang, “Fog and IoT: An Overview of Research Opportunities,” *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 854–864, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[20] Finale Doshi-Velez, and Been Kim, “Towards a Rigorous Science of Interpretable Machine Learning,” *arXiv Preprint*, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[21] Diederik P. Kingma, Tim Salimans, and Max Welling, “Variational Dropout and the Local Reparameterization Trick,” *Advances in Neural Information Processing Systems (NIPS)*, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[22] Pieter Abbeel, Adam Coates, and Andrew Y. Ng, “Autonomous Helicopter Aerobatics through Apprenticeship Learning,” *International Journal of Robotics Research*, vol. 29, no. 13, pp. 1608–1639, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[23] Xu Zhang et al., “Personalized Federated Learning via Variational Bayesian Inference,” *Proceedings of the 39th International Conference on Machine Learning*, vol. 162, pp. 26293-26310, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[24] Yarin Gal, and Zoubin Ghahramani, “Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning,” *Proceedings of the 33rd International Conference on Machine Learning*, pp. 1050–1059, 2016. [Google Scholar] [Publisher Link]

[25] Q. Li, X. Huang, and M. Chen, “Federated Deep Learning with Bayesian Privacy,” *arXiv Preprint*, pp. 1-23, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[26] Giovanni Acampora et al., “Quantum Computing and Artificial Intelligence: Status and Perspectives,” *arXiv Preprint*, pp. 1-33, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[27] Rajesh Gupta, Dakshita Reebadiya, and Sudeep Tanwar, “6G-Enabled Edge Intelligence for Ultra -Reliable Low Latency Applications: Vision and Mission,” *Computer Standards & Interfaces*, vol. 77, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[28] Yuyi Mao et al., “Green Edge AI: A Contemporary Survey,” *Proceedings of the IEEE*, vol. 112, no. 7, pp. 880-911, 2024. [CrossRef] [Google Scholar] [Publisher Link]