Research Paper Summary
Medical Privacy Preservation Using Zero Knowledge Machine Learning

Team Member: Richard Huynh
Instructor: Dr. Pooria Yaghini

Abstract

Privacy is a major concern in machine learning, especially in healthcare, where regulations like HIPAA and GDPR govern data usage. Traditional models require large datasets, increasing the risk of exposure. Zero-Knowledge Machine Learning (ZKML) offers a solution by enabling private inferences while maintaining compliance. This research explores EZKL, a Python library for generating zero-knowledge proofs, to convert a logistic regression model into a z-SNARK circuit for disease risk prediction without revealing sensitive data. While ZKML enhances inference privacy, challenges remain in training methods and computational trade-offs, highlighting the need for further improvements before real-world deployment.

Introduction

Medical data is highly sensitive, containing details on patient history, diagnoses, and treatments, and its exposure can lead to severe consequences, including insurance discrimination and confidentiality breaches. Regulations like HIPAA and GDPR enforce strict privacy safeguards, especially as medical records transition from paper to electronic formats. While machine learning has revolutionized various industries, its application in healthcare raises concerns about data privacy, as traditional models require access to large volumes of sensitive patient information, making them vulnerable to attacks like membership inference and model inversion. Privacy-preserving machine learning (PPML) is essential to mitigate these risks. Zero-Knowledge Machine Learning (ZKML) offers a promising solution by enabling model predictions to be verified without exposing input data or model parameters. This paper explores transforming a machine learning model into a Zero-Knowledge circuit, allowing secure proof generation and verification on-chain via Ethereum Virtual Machine (EVM) smart contracts, ensuring privacy and compliance in medical AI applications.

Background

Machine Learning

- A subset of AI and is composed of an input layer, hidden layers, and an output layer

- During training, data goes through hidden layer where it is transformed in order to extract features
    - Activation function increases model complexity to identify more complex relations
    - Loss function evaluates how well the model's predictions are compared to the actual value to update model parameters

Zero Knowledge Proofs

- 3 Key Properties:
    - Completeness: If the statement is true, an honest prover can convince an honest verifier
    - Soundness: If the statement is false, no cheating power can convince an honest verifier
    - Zero-Knowledgeness: If the statement is true, the verifier learns nothing other than the fact that the statement is true
- Primitives:
    - Prover wants to prove they know a specific piece of information
    - Verifier needs to be convinced that the prover knows the secret, but without learning the secret itself

Zero Knowledge Machine Learning

- General Steps: Data Encryption, Model Training, Proof Generation, Proof Verification

Methodology

1. Preprocess dataset
    a. Ensure all data types are usable by transforming then into scalar values for binary classification
2. Training Model using a Logistic Regression model
    a. fit the data with 2500 iterations for convergence
3. Building EZKL Circuit from ONNX model
    a. EZKL creates a circuit with the necessary settings and keys
4. Generate Proof using input data along with the proof key
    a. Output is an inference and proof hash
5. Verify Proof using WASM or EVM Smart Contract

Results

The mean absolute error of ~1% and mean squared error of ~0.01% indicates the EZKL circuit closely represents the original model. During inferences, encrypted data took longer to generate a proof due to the additional overhead. However, verification time is significantly faster than proof generation, which prevents additional overhead for verifications on-chain. The output given by the ZK circuit only reveals the model inference and the proof has, indicating that despite being unable to train on encrypted data, the model still does not reveal any information

on its training data, the input data, or model parameters. This captures the privacy preservation feature of ZKML, ensuring that medical diagnostics can be made without revealing sensitive patient data.

Limitations

ZKML faces significant computational challenges due to the fundamental mismatch between machine learning models and zero-knowledge (ZK) circuits. Machine learning models rely on floating-point weights for precision adjustments, whereas ZK circuits use fixed-point arithmetic, requiring approximations that introduce performance bottlenecks. These limitations make it difficult to deploy ZKML for real-world applications, particularly in privacy-sensitive fields like healthcare, where real-time processing is crucial. Additionally, EZKL, the ZKML library used in this study, is still in early development and not yet production-ready. Its compatibility is currently limited to models trained with PyTorch, preventing the construction of ZK circuits for encrypted training data. This restricts ZKML's ability to ensure end-to-end privacy, as it can only guarantee privacy during inference, leaving training data vulnerable—an issue especially critical in medical applications where data sensitivity is paramount.

Conclusion

As ZKML continues to evolve, it holds significant potential for enabling secure and privacy-preserving machine learning in sensitive fields like healthcare. This study demonstrates the feasibility of using ZKML at the inference level, ensuring that no sensitive patient data is exposed during predictions. By leveraging EZKL, machine learning models can be transformed into zero-knowledge circuits capable of generating verifiable proofs, enhancing trust and security. However, the computational overhead and current limitations in encrypted training highlight the challenges that must be addressed before widespread adoption. Future advancements in proof efficiency, hardware acceleration, and compatibility with encrypted training will be crucial for making ZKML a viable solution for real-world medical applications.

Future Work

- *Working* On-chain Verification
- Implement homomorphic encryption for model training
- Increase the scaling of the model to increase the complexity alongside a larger dataset

References

1. Ankit, "Heart attack risk amp; prediction dataset in india," 2025. [Online]. Available: https://www.kaggle.com/dsv/10853291.
2. S. Bharath Babu and K. R. Jothi, "A secure framework for privacy-preserving analytics in healthcare records using zero-knowledge proofs and blockchain in multi-tenant cloud environments," IEEE Access, vol. 13, pp. 8439–8455, 2025.

3. EZKL, “Benchmarking zkml frameworks.” [Online]. Available: https://blog.ezkl.xyz/post/benchmarks/.
4. EZKL, “The ezkl system.” [Online]. Available: https://docs.ezkl.xyz/.
5. I. Keshta and A. Odeh, “Security and privacy of electronic health records: Concerns and challenges,” Egyptian Informatics Journal, vol. 22, no. 2, pp. 177–183, 2021. [Online]. Available: https: //www.sciencedirect.com/science/article/pii/S1110866520301365.
6. T. Liu, X. Xie, and Y. Zhang, “zkcnn: Zero knowledge proofs for convolutional neural network predictions and accuracy,” in Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security, ser. CCS ’21. New York, NY, USA: Association for Computing Machinery, 2021, p. 2968–2985. [Online]. Available: https://doi.org/10.1145/3460120.3485379.
7. Q. Team, “Zero-knowledge machine learning: A beginner’s guide.” [Online]. Available: https://www.quillaudits.com/blog/ai-agents/zero-knowledge-machine-learning-zkml.
8. Z. Xing, Z. Zhang, Z. Zhang, Z. Li, M. Li, J. Liu, Z. Zhang, Y. Zhao, Q. Sun, L. Zhu, and G. Russello, “Zero-knowledge proof-based verifiable decentralized machine learning in communication network: A comprehensive survey,” 2023.
9. S. Zapechnikov, “Privacy-preserving machine learning as a tool for secure personalized information services,” Procedia Computer Science, vol. 169, pp. 393–399, 2020, postproceedings of the 10th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2019 (Tenth Annual Meeting of the BICA Society), held August 15-19, 2019 in Seattle, Washington, USA. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050920303598.