Proceedings of the DataNinja sAIOnARA 2024 Conference 39-42 DOI: 10.11576/dataninja-1166

Trade-offs Between Privacy and Performance in Encrypted Datasets using Machine Learning Models

Sanaullah Sanaullah SANAULLAH@HSBI.DE
Bielefeld University of Applied Sciences and Arts, Bielefeld, Germany

Hasina Attaullah HASINA.ATTAULLAH@HSBI.DE
Bielefeld University of Applied Sciences and Arts, Bielefeld, Germany

Thorsten Jungeblut THORSTEN.JUNGEBLUT@HSBI.DE
Bielefeld University of Applied Sciences and Arts, Bielefeld, Germany

1. Introduction

In recent years, with the increasing importance of dataset privacy in machine learning (ML) applications, there has been an increased demand for secure and privacy-preserving solutions [1, 2]. Consequently, encryption techniques have become known as a critical tool for protecting data privacy in an era of massive data use, exchange, and analysis [3, 4]. Encryption protects data against illegal access and disclosure by changing it into unreadable ciphertext that can only be decrypted by authorized parties [5–7]. In the field of ML, where sensitive data is often utilized, in such a process the use of encryption techniques has significant potential for providing privacy-preserving model training and inference [7, 8].

Therefore, this article analyzes, investigates, and compares three widely used encryption techniques. Each encryption method offers unique advantages and trade-offs [9–11]. Thus, we evaluate the performance of Convolutional Neural Network (CNN) models trained on encrypted datasets using these encryption techniques to provide detailed information on the effectiveness, practical concerns, and applicability of various methods for real-world applications by completely analyzing them within the context of computer vision. We test the performance of CNN models trained on encrypted data with several encryption approaches using neural models based-architecture [12]. Parameters such as training time, memory usage, and classification accuracy are analyzed and compared between encryption methods. We also look into the effect of encryption on model interpretability and robustness against adversarial attacks. Furthermore, to support our study we demonstrate our approach by using practical implementation—to showcase the performance and efficiency of each encryption strategy in protecting data privacy while keeping model accuracy and testing in a real-time recognition application using an edge device such as NVIDIA Jetson. Through this comparative analysis, researchers and developers can achieve a more in-depth understanding of the importance and issues involved with the integration of encryption techniques into ML especially in computer vision application workflows.

2. Analysis Methodology

A CNN architecture is utilized for the classification task due to its effectiveness in handling image data [13–15]. The architecture comprises two convolutional layers followed by max-pooling layers that enable hierarchical feature extraction from the input images. Each convolutional layer is activated using the ReLU activation function to introduce nonlinearity. The resulting feature maps are downsampled using max-pooling layers to reduce spatial dimensions and extract dominant features. Subsequently, a flattened layer transforms the 2D feature maps into a 1D vector which helps the compatibility with densely connected layers. Two fully connected (dense) layers with ReLU activation functions further process the extracted features, promoting non-linear transformations and capturing intricate patterns in the data. Finally, a dense layer with a softmax activation function is employed for multi-class classification that generates probability distributions over the output classes. This CNN architecture uses the hierarchical nature of neural networks to effectively classify the handwritten digit images present in the MNIST dataset [16–18].

Table 1: Accuracy with Resource Consumption

Encryption Model	Test Accuracy	CPU Time (sec.)	Memory (KB)
Original Data	99.25%	85.50	306,140
XOR	96.70%	76.61	156
S.Cipher	11.35%	75.92	30,744
Homomorphic	98.02%	75.92	30744

2.1. Encryption Techniques

Three encryption techniques are utilized to protect the privacy of the MNIST dataset during model training and evaluation. First, XOR encryption involves applying the bitwise XOR operation between the image pixel values and a randomly chosen integer encryption key. This process effectively rearranges the pixel values based on the binary representation of both the image and the key. Second, substitution cipher encryption shifts pixel values by a fixed integer value (encryption key) using modulo addition. In this process, each pixel value in the image is shifted by the chosen encryption key, wrapping around if the resulting value exceeds the maximum pixel intensity. Last, homomorphic encryption enables computations on encrypted data without decryption, preserving data privacy during processing. An encryption homomorphic scheme is applied to the pixel values using a predefined encryption key for arithmetic operations on encrypted data while maintaining confidentiality.

3. Experimental Results

The experimental results showcase the performance and resource consumption of the CNN model trained on different encryption techniques applied to the MNIST dataset. Therefore, XOR and homomorphic encryption techniques showed reasonable performance with minimal computational overhead, and substitution cipher encryption significantly degraded the model's accuracy, indicating its limitations in preserving data integrity for image classification tasks, details of each model results can be seen in Table 1. These results highlight the trade-offs between data privacy and model performance. Furthermore, to support our study exploration, we analyze statistical properties. The statistical properties provide information into the distribution characteristics of pixel values in the encrypted datasets obtained using different encryption techniques. Therefore,

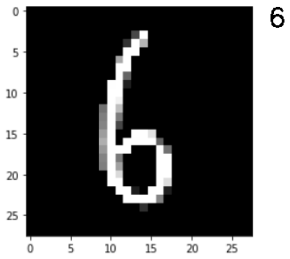

Figure 1: A handwritten digit on the left and on the right the model predicted number in real-time using XOR-encrypted dataset.

XOR-Encrypted Data - Mean: 0.4887 and Variance: 0.0088, demonstrate a mean value close to 0.5, indicating a balanced distribution of pixel values around the midpoint. The variance is relatively small, suggesting moderate dispersion of pixel values around the mean. However, Substitution-Encrypted Data - Mean: 0.1966 and Variance: 1.4825e-06, similarly Homomorphic-Encrypted Data - Mean: 0.1236 and Variance: 0.0882, shows a significantly lower mean value compared to XOR-encrypted, indicating a shift in the pixel value distribution. The variance suggested a minimal dispersion of pixel values around the mean.

3.1. Real-Time Implementation

In the real-time implementation, we demonstrate the system's capability to predict handwritten digits from uploaded images seamlessly and efficiently while preserving data privacy using encryption techniques. To provide a concrete example, Figure 1 showcases an uploaded image of a handwritten digit on the left and on the right the model predicted number in real-time testing. Therefore, through this demonstration, we illustrate the system's ability to preprocess incoming images, feed them into the XOR encrypted model, and seamlessly provide accurate predictions without compromising data privacy. Additionally, all test results are available on our GitHub channel.

Acknowledgments

This research was supported by the research training group "Dataninja" (Trustworthy AI for Seamless Problem Solving: Next Generation Intelligence Joins Robust Data Analysis) funded by the German fed-

eral state of North Rhine-Westphalia. This project is also supported by SAIL (grant no NW21-059B) and the "Transformation in Care and Technology" (funding program "Profiling 2020", and the grand number is PROFILNRW-2020-095), project is funded by the Ministry of Culture and Science of the State of North Rhine-Westphalia.

References

[1] Petr Velan, Milan Čermák, Pavel Čeleda, and Martin Drašar. A survey of methods for encrypted traffic classification and analysis. *International Journal of Network Management*, 25 (5):355–374, 2015.

[2] Sanaullah, Shamini Koravuna, Ulrich Rückert, and Thorsten Jungeblut. Snns model analyzing and visualizing experimentation using ravsim. In *International conference on engineering applications of neural networks*, pages 40–51. Springer, 2022.

[3] Rongxing Lu, Hui Zhu, Ximeng Liu, Joseph K Liu, and Jun Shao. Toward efficient and privacy-preserving computing in big data era. *IEEE Network*, 2014.

[4] Sanaullah, Hasan Baig, Jan Madsen, and Jeong-A Lee. A parallel approach to perform threshold value and propagation delay analyses of genetic logic circuit models. *ACS Synthetic Biology*, 9 (12):3422–3428, 2020.

[5] Karthik Nandakumar, Nalini Ratha, Sharath Pankanti, and Shai Halevi. Towards deep neural network training on encrypted data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 0–0, 2019.

[6] Sanaullah, S Koravuna, Ulrich Rückert, and Thorsten Jungeblut. Transforming event-based into spike-rate datasets for enhancing neuronal behavior simulation to bridging the gap for snns. In *IEEE Conference on Computer Vision (ICCV)*, 2023.

[7] Eva Papadogiannaki and Sotiris Ioannidis. A survey on encrypted network traffic analysis applications, techniques, and countermeasures. *ACM Computing Surveys (CSUR)*, 54(6):1–35, 2021.

[8] Shamini Koravuna, Ulrich Rückert, Thorsten Jungeblut, et al. Evaluation of spiking neural nets-based image classification using the runtime simulator ravsim. *International Journal of Neural Systems*, pages 2350044–2350044, 2023.

[9] Mohamed Elhoseny, Xiaohui Yuan, Hamdy K El-Minir, and Alaa Mohamed Riad. An energy efficient encryption method for secure dynamic wsn. *Security and Communication Networks*, 2016.

[10] Mohammed Abbas Fadhil Al-Husainy. A novel encryption method for image security. *International Journal of Security and Its Applications*, 6(1):1–8, 2012.

[11] Sana Ullah and Thorsten Jungeblut. Analysis of mr images for early and accurate detection of brain tumor using resource efficient simulator brain analysis. In *19th International Conference on Machine Learning and Data Mining MLDM*, 2023.

[12] Sanaullah, Shamini Koravuna, Ulrich Rückert, and Thorsten Jungeblut. Exploring spiking neural networks: a comprehensive analysis of mathematical models and applications. *Frontiers in Computational Neuroscience*, 17:1215824, 2023.

[13] Samir S Yadav and Shivajirao M Jadhav. Deep convolutional neural network based medical image classification for disease diagnosis. *Journal of Big data*, 2019.

[14] Neha Sharma, Vibhor Jain, and Anju Mishra. An analysis of convolutional neural networks for image classification. *Procedia computer science*.

[15] Sanaullah Sanaullah. A hybrid spiking-convolutional neural network approach for advancing machine learning models. In *Northern Lights Deep Learning Conference*, pages 220–227. PMLR, 2024.

[16] Savita Ahlawat, Amit Choudhary, Anand Nayyar, Saurabh Singh, and Byungun Yoon. Improved handwritten digit recognition using convolutional neural networks (cnn).

[17] Sanaullah, Shamini Koravuna, Ulrich Rückert, and Thorsten Jungeblut. Streamlined training of gcn for node classification with automatic loss function and optimizer selection. In *International Conference on Engineering Applications*

of Neural Networks, pages 191–202. Springer, 2023.

[18] Feiyang Chen, Nan Chen, Hanyang Mao, and Hanlin Hu. Assessing four neural networks on handwritten digit recognition dataset (mnist). *arXiv preprint arXiv:1811.08278*, 2018.