Hardware Accelerated Privacy-Preserving Ensemble Learning for X-Ray Image Diagnostics

Joseph O'Neill[†,*], Lydia Bouzar-Benlabiod[‡], Nur Zincir-Heywood[†]
[†] Dalhousie University. Halifax, Nova Scotia
[‡] Acadia University. Wolfville, Nova Scotia

Abstract

The adoption of machine learning (ML) in highly regulated and sensitive domains such as healthcare is constrained by escalating concerns regarding data privacy and stringent legal and regulatory frameworks. Although Privacy-Preserving Machine Learning (PPML) techniques provide strong formal guaranties against data leakage, they frequently incur a non negligible reduction in predictive performance. This inherent privacy–accuracy trade-off constitutes a primary obstacle to the practical deployment of PPML systems. This research introduces a novel PPML framework that leverages hardware acceleration techniques in conjunction with ensemble learning to alleviate accuracy degradation and improve performance simultaneously. X-ray images are ideal for PPML diagnostics, as they provide clear, high contrast visualizations that promote fast detection of complex ailments. The resulting system constitutes a robust PPML architecture that attains state of the art performance, achieving an accuracy of 94% relative to existing single model baselines, while simultaneously reducing false negatives by an average of 62%.

Keywords: Privacy-Preserving Machine Learning, Ensemble Learning, Homomorphic Encryption, Hardware Acceleration, Quantization, Image Diagnostics

1. Introduction

The adoption of artificial intelligence (AI) across a wide range of sectors has been associated with measurable gains in productivity, operational efficiency, and processing capacity. Sectors including healthcare, defence, and finance are similarly poised to benefit from the integration of AI technologies. In particular, AI tools can support diagnostic workflows, therapeutic decision-making, administrative and documentation tasks, resource allocation, and large-scale data analytics.

Within the medical domain, multiple specific and well-defined areas of application for artificial intelligence have been identified. Convolutional Neural Networks (CNNs) [1] have been extensively employed for medical image analysis, including the interpretation of radiographic images such as X-rays [2][3], as well as neuro imaging modalities used in brain scans [4][5]. Predictive AI systems have been developed to assist clinical decision making in the diagnosis of various cancers [6][7], Alzheimer's disease [8], pulmonary nodules [9], and numerous other pathologies, often achieving high levels of diagnostic accuracy. However, the development and deployment of these AI systems depend critically on large volumes of high quality, domain-specific training data.

Protecting patient privacy is therefore a central consideration in the collection and utilization of healthcare data for AI development. In Canada, federal legislation governs the acquisition, storage, and sharing of sensitive health information, imposing stringent requirements to ensure data confidentiality and security. To comply with these regulatory frameworks, a variety of privacy-preserving techniques and software infrastructures have been introduced to secure and standardize the handling of medical data. Among these,

[*]joseph.oneill@dal.ca

fully homomorphic encryption (FHE) has emerged as a key cryptographic approach to enabling privacy-preserving analytics on encrypted clinical datasets. Homomorphic encryption schemes, such as fast fully homomorphic encryption over the torus (TFHE), have been integrated into privacy-preserving machine learning pipelines, enabling computations directly on encrypted data without decryption. Furthermore, these encryption schemes are designed to remain secure in the presence of quantum computing capabilities and are thus classified as post-quantum cryptography (PQC).

As with all cryptographic techniques, the principal limitation of FHE is its computational overhead. In the case of FHE, this overhead extends well beyond the costs of encryption and decryption. The substantial complexity of the encryption process also affects machine learning workloads, significantly increasing training and classification latency. Consequently, FHE performance optimization has become a major focus of current research. Core algorithmic components, such as bootstrapping [10], have undergone continuous refinement over the past several years, yielding substantial efficiency gains in FHE schemes, exemplified by constructions such as TFHE. Hardware based acceleration, particularly the adaptation and optimization of FHE computations for Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs) [11], constitutes a complementary strategy that shows promising improvements in the practical performance of FHE.

The primary goal of this research is a novel PPML framework that leverages hardware acceleration techniques in conjunction with ensemble learning to alleviate accuracy degradation while also improving performance. In light of this research goal, the following research questions will be explored. (i) What constitutes an optimal system architecture for implementing a hardware-accelerated, quantization-aware PPML pipeline, and how do GPUs and TPUs compare in this context with respect to computational performance and throughput gains? (ii) To what extent can advanced ensemble learning architectures, such as probability averaging or weighted voting, recover the lost classification accuracy compared to single-model PPML baselines? (iii) How can the predictions of these base learners be aggregated both securely and with the specific intent of lowering the rate of false negative predictions?

This paper is organized and structured as follows. Section 2 provides a synopsis on the state of the art related to FHE, quantization, and hardware acceleration. Section 3 outlines the methodology and high-level architecture of the proposed system. Section 4 gives a detailed account of the effectiveness of the proposed system. Finally, Section 5 concludes the paper and discusses a set of recommendations and proposes some future work.

2. Background

A wide range of privacy-preserving techniques has been developed over the years. These approaches span from perturbative data-transformation methods, such as K-anonymity and L-diversity, to various forms of cryptographic mechanisms. Figure 1 presents a taxonomy of the principal privacy-preserving techniques currently available. Among the techniques depicted in Figure 1, homomorphic encryption has recently been demonstrated to be particularly effective in the context of privacy-preserving machine learning [12].

2.1. Homomorphic Encryption

The conceptual foundations underlying fully homomorphic encryption originate in algebra, where the term "homomorphism" was first introduced. These homomorphic principles have been transferred to the field of cryptography, giving rise to encryption schemes whose transformations from plaintext to ciphertext preserve the underlying algebraic structure. As a consequence of this structural preservation, homomorphic encryption enables secure computation directly on encrypted data.

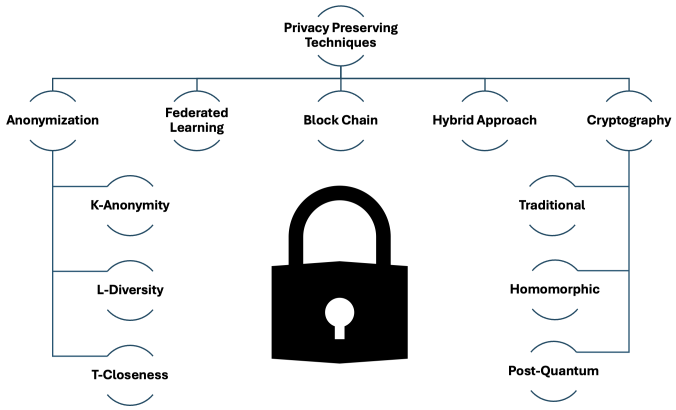


Figure 1. Privacy-Preserving ML Techniques

Among contemporary FHE schemes, TFHE has gained particular traction due to its performance, accessibility, and suitability for privacy-preserving machine learning applications. TFHE is one of the most recent FHE constructions and introduces a novel bootstrapping mechanism that significantly enhances computational efficiency [10]. In the context of FHE, bootstrapping is a key technique used to refresh ciphertexts by homomorphically evaluating the decryption circuit on a noisy ciphertext, thereby reducing its noise level and restoring its capability to support further homomorphic operations. TFHE also contains advanced hardware acceleration techniques that can be enabled. One such technique allows PPML image classifiers to use Compute Unified Device Architecture (CUDA) cores on an available GPU for accelerated processing.

2.2. **Hardware Acceleration**

Graphics processing units, such as NVIDIA's A100, have become widely adopted in AI and machine learning because of their CUDA cores and their capacity to perform large numbers of operations in parallel [13]. Parallel matrix computations are central to many PPML image classification algorithms; consequently, GPUs are frequently employed to accelerate such computationally intensive AI tasks. This is particularly relevant for CNNs, whose feature maps are generated through extensive matrix and tensor operations. Other companies, such as Google, have pursued an alternative hardware strategy by developing TPUs, for example the V6e1, that are explicitly designed to optimize matrix multiplications, which dominate the computational cost in deep neural networks.

To fully exploit these hardware accelerators, the proposed system must be appropriately configured and initialized using dedicated software libraries. To leverage the CUDA cores available on GPUs, CONCRETE-ML [14] exposes additional hyperparameters that can be tuned for each PPML image classifier. When the PPML classifiers are instantiated, a special "device" parameter is used to specify the CUDA cores for GPU useage. Beyond configuring the ensemble itself, a specific, GPU-enabled version of the CONCRETE-ML library must be selected within the execution environment. GPU optimization depends on a set of backend routines that are explicitly tuned for CUDA. Because of this, a CUDA-accelerated build of the library is required to achieve the intended performance.

In contrast, Google's TPUs do not necessitate additional hyperparameters or substantial modifications to the runtime configuration. These TPUs can execute code that is originally written for CPU based computation with minimal or no changes, thereby improving the flexibility and interoperability of the overall system. Furthermore, TPUs have been developed with explicit support for particular Python machine learning frameworks. TensorFlow

is a widely used ML library, and Google has optimized its TPUs for tight integration with TensorFlow [15]. In this work, the implementation has been adapted to exploit TensorFlow's data loader functionality for preparing and managing datasets during the stages of encryption, secure analysis, and decryption.

2.3. **Quantization**

Quantization is a technique widely employed in implementations of FHE to optimize PPML models with respect to the trade-off between computational performance and predictive accuracy. Increasing the quantization bit value in an FHE computation generally leads to longer execution times, while typically improving model accuracy [16]. If the quantization bit value is too small, discarded bits cause quantization error and reduce accuracy. A final quantization value of **8** bits was determined to produce the best accuracy while keeping the system performance high.

For linear models such as SVM and LR, quantization is applied in a post-training manner. PPML models are first trained in floating point to find high-quality parameters, which are then quantized for encrypted inference. For CNNs, the quantization strategy differs and CNNs are trained using quantization-aware training (QAT).

A notable aspect of PPML hardware acceleration is that it is frequently designed around integer arithmetic for activations and intermediate representations. Quantization schemes can be made particularly hardware-efficient by selecting integer scaling factors that are powers of two. When the effective quantization scale is a power of two, divisions in PPML computations can be replaced by bitwise right shift operations. Since TFHE efficiently supports right bit shifts, this quantization configuration both matches common hardware accelerators and exploits optimized homomorphic operations.

2.4. **Related Work**

Privacy-preserving machine learning has undergone substantial growth as a research area over the past several years. Numerous PPML based medical image diagnostic systems have been proposed; however, these systems predominantly rely on a single classifier to generate privacy-preserving predictions [17]. Moreover, prior studies employ earlier generations of fully homomorphic encryption schemes, which are less efficient and less robust than more recent implementations. With advances in homomorphic encryption, increasingly complex classification methods are now being investigated within the PPML domain. The exploration of hardware acceleration mechanisms for PPML represents another promising and increasingly important direction in privacy-preserving machine learning research [18].

The work in [19] and [20] both demonstrate that PPML models can be trained directly on encrypted images and subsequently used to generate predictions on encrypted inputs. Some drawbacks of training on encrypted images include increased training times and decreased prediction accuracies. These studies also focus on medical image diagnostics, employing singular CNN classifiers to make encrypted predictions. In [21], the authors present a class of CNN architectures that exploit memory mechanisms to perform predictions over encrypted images. The research outlined in [22] shows that both the PPML model and the images can be encrypted using FHE and still retain accuracy.

The research presented in [23] introduces a PPML framework based on the TFHE fully homomorphic encryption scheme, with a specific emphasis on medical image diagnostics. In that work, the authors implement a novel ensemble learning approach and develop a system referred to as "EDM-Diagnostic." One of the proposed ensemble methods is explicitly designed to reduce the incidence of false negative classifications. The implications of false negatives for clinical outcomes and system level performance are examined in detail in [24]. Building on the foundations established in [23], the primary objective of the present research

is to extend this framework by incorporating additional, more advanced ensemble techniques [25] and by enhancing performance through hardware acceleration.

To the best of our knowledge, these works constitute the most recent and relevant literature on PPML for medical image diagnostics. Collectively, existing approaches in this area have focused primarily on the use of a single image classifier typically a deep neural network to perform image evaluation. Alternative image classification models can achieve comparable predictive performance under FHE while offering substantially reduced computational costs. Building on this observation, we propose an ensemble-based PPML framework that combines multiple image classifiers, advancing previous work done in [23]. Advanced ensemble techniques such as probability averaging, hard voting, and weight voting are compared to determine the best performance.

3. Methodology

The architectural diagram shown in Figure 2 has been developed with two principal objectives in mind. First, the experimental framework is intended to extend and enhance the work presented in [23]. Potential improvements in overall system performance will be investigated through hardware acceleration methodologies, such as model quantization. Additionally, software level optimizations will be explored to further increase system efficiency and to more effectively exploit hardware accelerators, including GPU CUDA cores. Furthermore, TPU cores available through Google Cloud services constitute an additional class of hardware accelerator, specifically engineered to support machine learning and privacy-preserving machine learning workloads.

Second, the high-level architecture is designed to introduce a novel ensemble learning paradigm within the context of privacy-preserving machine learning. The proposed ensemble comprises a convolutional neural network, logistic regression (LR), and support vector machines (SVM), which represent the principal machine learning models deployed in the current SOTA literature. The outputs of these models are subsequently integrated using a meta-learner, which is employed to compute ensemble-based performance statistics, including improvements in predictive accuracy and reductions in false negative rates. Two new novel PPML ensemble techniques are implemented and demonstrated in this work.

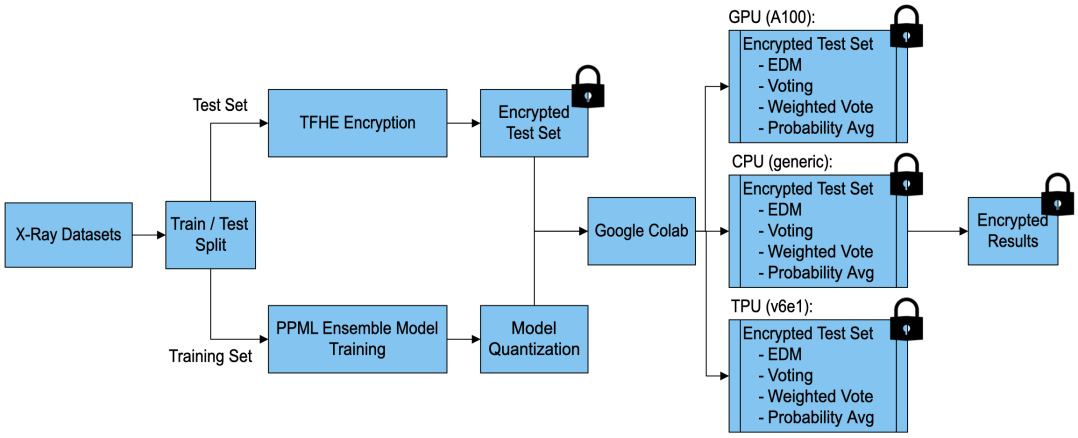


Figure 2. Proposed High Level System Architecture

The system begins by analyzing the X-ray datasets. All images undergo a preprocessing pipeline and are partitioned into disjoint training and test subsets. The test subsets are subjected to encryption, whereas the training subsets are used to train the individual PPML

models. Following training, each of the PPML models are quantized to ensure fast execution. The encrypted test images are then independently evaluated by each PPML model. Subsequently, the prediction outputs from the SVM, LR, and CNN models are aggregated by the proposed meta-learner, which produces the ensemble predictions. The meta-learner combines the three sets of predictions using four different techniques. The EDM technique that is proposed in [23] is calculated along with a voting technique, weighted voting technique, and probability averaging technique. This final step of encrypted image prediction and ensemble combination is replicated using different forms of hardware. CPUs, TPUs and GPUs are directly compared in this manner.

At all stages of this process, the prediction results remain fully encrypted, as indicated by the lock icons in Figure 2. Both the PPML models and the meta-learner operate directly on ciphertexts, generating predictions that are embedded within the encrypted domain. This property is a key feature of fully homomorphic encryption, ensuring that data confidentiality is preserved even after model inference. The final outputs can only be accessed by applying the private key associated with the initial encryption process. Consequently, the originator of the data must perform decryption in order to inspect the prediction results.

3.1. **Datasets Employed**

Table 1 summarizes the key characteristics of each dataset, including the total number of images and their respective sizes. In total, three publicly available Kaggle datasets [26] [27] [28] are used to evaluate the proposed system. Each of the selected data sets contains X-ray images exclusively. X-ray images are inherently well suited for use with FHE as they are typically represented in grayscale. This property is advantageous in an PPML context, where image resolution and data volume substantially influence encryption time, decryption time, and inference latency in encrypted data. Consequently, many state of the art approaches convert images to grayscale during the preprocessing stage. The datasets considered in this work encompass tasks ranging from binary to multiclass classification.

Name	**Image Type**	**Year**	**Image Count**	**Classification**
COVID-19 [26]	Lung X-Ray	2022	33,920	MultiClass
Pediatric Pneumonia [27]	Lung X-Ray	2025	5,856	Binary
BreakHIS [28]	Cancer Histogram	2021	9,109	Binary

Table 1. Datasets Used To Evaluate Our System

3.2. **Ensemble Techniques**

As described in Figure 2, the proposed meta-learner combines the encrypted results generated from the SVM, LR, and CNN [23] classifiers. The proposed system implements four different techniques to combine the results. The following ensemble techniques are employed by the meta-learner:

- **EDM**: The meta-learner looks to see if any single classifier detected an ailment. If so, this takes precedence and is reflected in the final result.
- **Soft Voting**: The results of all three classifiers are averaged to produce the final result. Given the three classifiers, the voting technique requires that at least two classifiers detect and agree on an ailment.
- **Weighted Voting**: A weight factor is applied to each of the ensemble classifiers. These weights are combined with the classifiers vote to produce a final value. The SVM classifier has a weight of **0.**4, LR has a weight of **0.25**, and CNN has a weight of **0.35**. These weights are assigned on the basis of the performance of the classifier across the tested datasets.

- **Probability Averaging**: This technique accounts for how certain each individual PPML classifier is in terms of a probability and combines that with the certainty of the others. This advanced technique is the only one that uses these extra data when generating its final prediction.

Exactly how EDM and soft voting techniques are implemented is explained in [23]. Algorithm 1 details how the probability averaging technique is implemented.

Algorithm 1 Probability Averaging Technique

```
for (i = 0 to image_results.length) do
    zero_prob = (svm-probability[i][0] + lr-probability[i][0] + cnn-probability[i][0])/3
    one_prob = (svm-probability[i][1] + lr-probability[i][1] + cnn-probability[i][1])/3

    if (one_prob ≥ zero_prob) then
        ensemble-results[i] = 1 ← Ailment Detected!
    else
        ensemble-results[i] = 0 ← No Ailment
    end if
end for
return ensemble-results[]
```

3.3. Hyperparameter Tuning

Each hyperparameter is adjusted with two primary objectives in mind. First, we seek to maximize the predictive accuracy of the system. Second, we aim to balance the first objective against the overall computational performance. This trade-off is particularly critical when configuring the CNNs, as these classifiers constitute the vast majority of the total ensemble runtime. Across our three X-ray datasets, CNNs account on average for **98.22%** of the total system time. In one instance, CNN prediction alone represented **99.1%** of the total system time.

The prediction time of a CNN is strongly influenced by its architectural depth, i.e. the number of layers. To manage this trade-off between computational cost and predictive performance, we adopt a six layer CNN within our PPML ensemble learner. Multiple empirical evaluations were conducted to identify an appropriate balance between runtime efficiency and classification accuracy. The configuration of our PPML classifiers is described as follows.

$$\textbf{SVM: } C = 1,\ dual = false,\ penalty = L1,\ n\_bits = 8$$
$$\textbf{LR: } C = 1,\ dual = false,\ penalty = None,\ n\_bits = 8$$
$$\textbf{CNN: } Convolutional\ = 5,\ ReLu\ = 5,\ Linear\ = 1,\ n\_bits = 8$$

Both the SVM and LR classifiers share the same set of hyperparameters. We vary the regularization parameter C, which controls the strength of the regularization applied to the classifiers. In addition, we disable the dual optimization procedure via the corresponding dual configuration parameter and adjust the penalty terms that govern the regularization scheme. Finally, each of these PPML classifiers is configured with a quantization bit size of 8. This value was determined to provide the best balance of accuracy and system performance.

The convolutional neural network comprises six layers in total, five of which are two-dimensional convolutional layers. The final layer is a fully connected linear layer that maps the learned feature representation to a vector whose dimensionality equals the number of classes in the dataset. The CNN is also configured using a quantization bit size of 8. Similarly to SVM and LR, this value was determined to provide the best balance of accuracy and

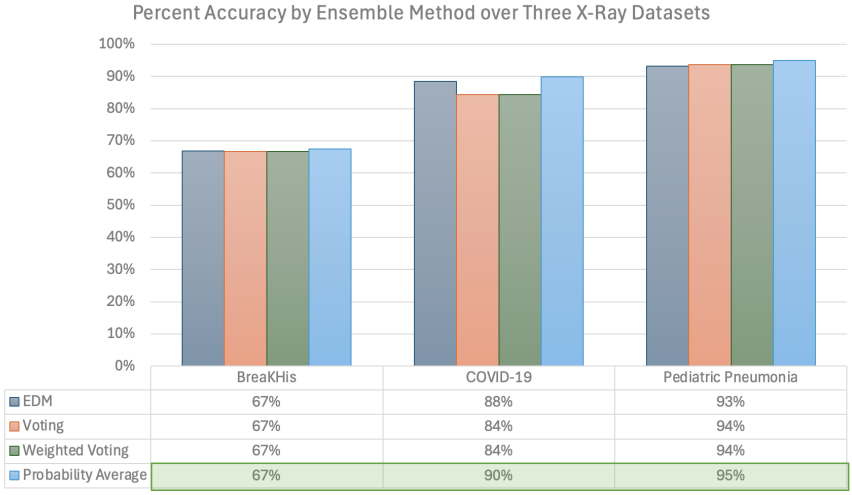
system performance for CNN. This value is also a power of two, which aligns with the hardware acceleration associated with CONCRETE-ML.

4. Evaluation and Results

In this study, key performance indicators include the comparative evaluation of CPU, GPU, and TPU hardware acceleration performance, ensemble model accuracy and ensemble false negative rates (FNR). Each of the hardware accelerators will be evaluated by comparing how many encrypted images per hour they can process. This value is calculated by first determining how many images per hour are processed for each individual dataset, and then an average of the three values is used. The PPML ensemble techniques are compared based on their performance on each individual dataset, and again the values are averaged.

4.1. Ensemble Learning Results

The three X-ray datasets were evaluated using the ensemble. In each case, the meta-learner was tuned to evaluate the output of four different ensemble techniques. Figure 3 details the output of the privacy-preserving ensemble against each of the datasets. This chart outlines the individual accuracies of the EDM approach, voting, weighted voting, and probability averaging techniques. Of the techniques shown, probability averaging achieved the highest accuracy. This is expected as it is the most advanced of the ensemble techniques. While only accuracy values are shown, precision and F-1 values were also captured as part of this research but are not included due to space considerations. Precision and F-1 scores match that of accuracy.



	BreaKHis	COVID-19	Pediatric Pneumonia
EDM	67%	88%	93%
Voting	67%	84%	94%
Weighted Voting	67%	84%	94%
Probability Average	67%	90%	95%

Figure 3. Ensemble Techniques Accuracy Results

As shown in Figure 4, the EDM approach achieves a substantially lower rate of false negatives. A false negative represents someone who has an ailment, but the classifier did not find it. To reduce false negatives, we allow the meta-learner to detect if any one classifier detects an ailment. If any classifier detects the illness, we flag it as detected. This sets a precedent over the majority vote. Although PPML should not be relied upon exclusively, highlighting ailments such as this would be a valuable decision support system for a physician who can further investigate. Figure 4 compares the FNR for each of the ensemble techniques on encrypted datasets.

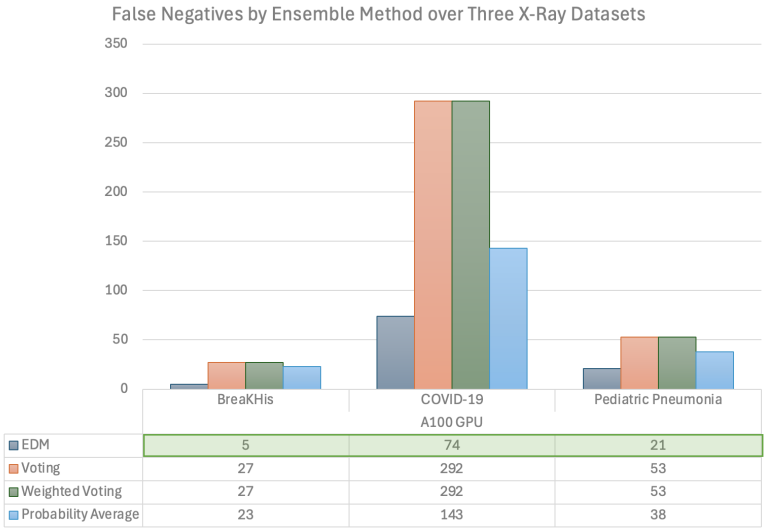


Figure 4. Ensemble Techniques False Negative Results

The proposed approach demonstrates two different sets of key performance metrics. First, by testing the four different ensemble techniques, we can show that the probability averaging technique produces the highest accuracy of all the ensemble techniques. These results are also higher than for any single PPML classifier used. This result is consistent across the three X-ray datasets tested. This result matches the original prediction, as this technique is the most advanced of the four tested. Probability averaging uses both the ensemble prediction, and its confidence in that prediction to make a final decision. It is the only technique that utilizes this additional information.

The second key performance metric evaluated was how EDM reduced the false negative prediction rate. In this case, the EDM technique reduced the FNR by an average of **62%** across all three datasets. Even in the worst case scenario the EDM approach still reduced false negative predictions by **51%**. This is a significant reduction in false negatives, which are the cause of widespread problems in the medical industry. The EDM approach also retains high overall accuracy and precision compared to the other four PPML classifiers, although it does not achieve top accuracy. In the best case scenario, the EDM reduced overall false negatives by **79%** while also achieving a high accuracy at **67%**.

4.2. **System Performance Results**

When comparing TFHE across CPUs, GPUs, and TPUs, the primary focus was on performance increases in the area of encrypted prediction times. The overall performance of HE has been its primary drawback, so any improvements in this area would be a great improvement to PPML. The TFHE encryption scheme boasts improved bootstrapping techniques, which should reduce prediction times while retaining accuracy. After testing across all three hardware platforms, the prediction times on encrypted data were reduced significantly by the TPUs. The TPUs were able to process **25,000 images per hour** compared to the GPUs **11,500 images per hour**, and the CPUs **7,500 images per hour**. This represents a significant increase in system performance of nearly **300%** compared to the CPU. The precision of each of the ensemble techniques was tracked and the prediction accuracy remained consistent across all hardware tested. Figure 5 provides details of this data.

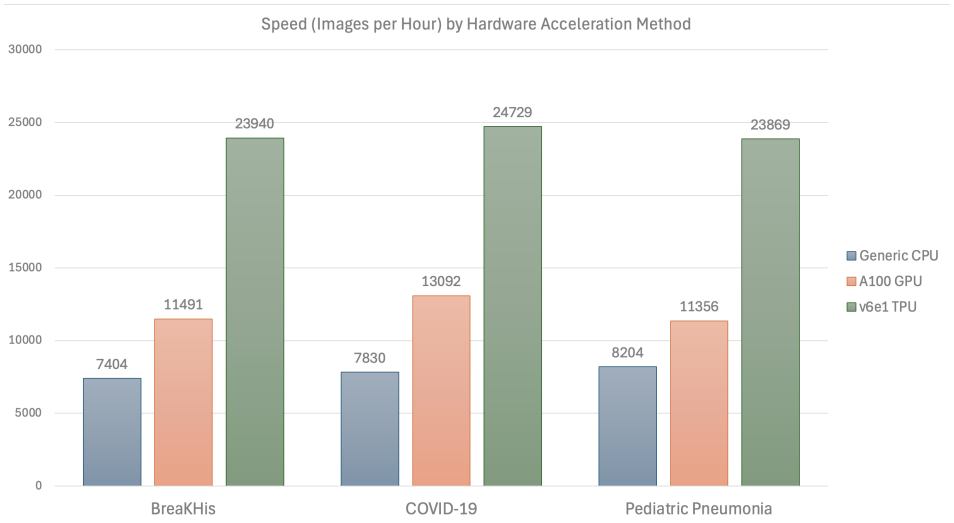


Figure 5. Hardware Performance Results

5. **Conclusion and Future Work**

The original research goal was a novel PPML framework that leverages hardware acceleration techniques in conjunction with ensemble learning to alleviate accuracy degradation while also improving performance. The proposed PPML ensemble learning system examines and empirically addresses the three research questions introduced in Section I. By evaluating the impact of PPML ensemble learning techniques, we demonstrate that the secure aggregation of multiple base learners can be leveraged to reduce the false negative rate in privacy-preserving predictions, while concurrently enhancing overall predictive accuracy. Furthermore, the evaluation of the proposed system underscores the performance advantages of the TFHE fully homomorphic encryption scheme when optimized for hardware accelerators such as TPUs. Collectively, these findings emphasize the importance of adopting more robust and advanced machine learning methodologies in privacy-sensitive environments.

The work presented here demonstrates the design, implementation, test and evaluation of the a PPML medical image diagnostic system for X-ray images. Novel contributions include designing and implementing two new advanced ensemble techniques, as well as increasing system performance through hardware acceleration using GPUs, TPUs, and FHE. The evaluation of the system included a direct comparison of several advanced PPML ensemble techniques. To test the proposed system, three medical image datasets were analyzed using our novel PPML ensemble learner. The results shown by our ensemble learner demonstrate that false negatives can be reduced by an average of **62%** and as high as **79%**, while improving accuracy and precision over any singular PPML base learner. Our research further demonstrates that the system can analyze up to **25,000** encrypted images per hour while maintaining this level of accuracy and precision.

The proposed PPML ensemble learning system further demonstrates that the optimized TFHE encryption scheme for TPUs increases the number of encrypted predictions that the system can execute per hour by **300%** compared to CPUs. At the same time, the accuracy and precision of the prediction increased slightly when tested on the exact same datasets. GPUs have also been shown to increase the number of encrypted predictions per hour by **51%** compared to CPUs. These increases represent a significant improvement in the performance of TFHE over other comparable FHE techniques.

Future work includes improvements to the bootstrapping technique to decrease the computational cost of encrypted prediction times while also reducing any noise introduced into the ciphertext. The concept of PPML could also be expanded beyond image classification. The concepts demonstrated through the proposed system could be leveraged and utilized in many other privacy sensitive domains.

6. Acknowledgments

This research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through the NSERC CREATE program in Cybersecurity for Emerging Technologies, and NSERC Discovery grants.

References

[1] O. L. Usman, R. C. Muniyandi, K. Omar, and M. Mohamad. "Privacy-Preserving Classification Method for Neural-Biomarkers using Homomorphic Residue Number System CNN: HoRNS-CNN". In: *2022 International Conference on Business Analytics for Technology and Security (ICBATS)*. 2022, pp. 1–8. DOI: 10.1109/ICBATS54253.2022.9759007.

[2] H. S. Alghamdi, G. Amoudi, S. Elhag, K. Saeedi, and J. Nasser. "Deep Learning Approaches for Detecting COVID-19 From Chest X-Ray Images: A Survey". In: *IEEE Access* 9 (2021), pp. 20235–20254. DOI: 10.1109/ACCESS.2021.3054484.

[3] E. Çallı, E. Sogancioglu, B. van Ginneken, K. G. van Leeuwen, and K. Murphy. "Deep learning for chest X-ray analysis: A survey". In: *Medical Image Analysis* 72 (2021), p. 102125. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2021.102125. URL: https://www.sciencedirect.com/science/article/pii/S1361841521001717.

[4] H. Mohsen, E.-S. A. El-Dahshan, and A.-B. M. Salem. "A machine learning technique for MRI brain images". In: *2012 8th International Conference on Informatics and Systems (INFOS)*. 2012, BIO–161–BIO–165.

[5] A. M. Hasan, H. A. Jalab, F. Meziane, H. Kahtan, and A. S. Al-Ahmad. "Combining Deep and Handcrafted Image Features for MRI Brain Scan Classification". In: *IEEE Access* 7 (2019), pp. 79959–79967. DOI: 10.1109/ACCESS.2019.2922691.

[6] W. Yue, Z. Wang, H. Chen, A. Payne, and X. Liu. "Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis". In: *Designs* 2.2 (2018). ISSN: 2411-9660. DOI: 10.3390/designs2020013. URL: https://www.mdpi.com/2411-9660/2/2/13.

[7] E. Alabdulkreem, M. K. Saeed, S. S. Alotaibi, R. Allafi, A. Mohamed, and M. A. Hamza. "Bone Cancer Detection and Classification Using Owl Search Algorithm With Deep Learning on X-Ray Images". In: *IEEE Access* 11 (2023), pp. 109095–109103. DOI: 10.1109/ACCESS.2023.3319293.

[8] M. Tanveer, B. Richhariya, R. U. Khan, A. H. Rashid, P. Khanna, M. Prasad, and C. T. Lin. "Machine Learning Techniques for the Diagnosis of Alzheimer's Disease: A Review". In: *ACM Trans. Multimedia Comput. Commun. Appl.* 16.1s (Apr. 2020). ISSN: 1551-6857. DOI: 10.1145/3344998. URL: https://doi.org/10.1145/3344998.

[9] Y. Gu, J. Chi, J. Liu, L. Yang, B. Zhang, D. Yu, Y. Zhao, and X. Lu. "A survey of computer-aided diagnosis of lung nodules from CT scans using deep learning". In: *Computers in Biology and Medicine* 137 (2021), p. 104806. ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2021.104806. URL: https://www.sciencedirect.com/science/article/pii/S0010482521006004.

[10] R. Wang, B. Wei, Z. Li, X. Lu, and K. Wang. "TFHE Bootstrapping: Faster, Smaller and Time-Space Trade-Offs". In: *Information Security and Privacy*. Ed. by T. Zhu and Y. Li. Singapore: Springer Nature Singapore, 2024, pp. 196–216. ISBN: 978-981-97-5025-2.

[11] S. Tan, B. Knott, Y. Tian, and D. J. Wu. "CryptGPU: Fast privacy-preserving machine learning on the GPU". In: *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE. 2021, pp. 1021–1038.

[12] R. Podschwadt, D. Takabi, P. Hu, M. H. Rafiei, and Z. Cai. "A Survey of Deep Learning Architectures for Privacy-Preserving Machine Learning With Fully Homomorphic Encryption". In: *IEEE Access* 10 (2022), pp. 117477–117500. DOI: 10.1109/ACCESS.2022.3219049.

[13] K. Ho, H. Zhao, A. Jog, and S. Mohanty. "Improving gpu throughput through parallel execution using tensor cores and cuda cores". In: *2022 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. IEEE. 2022, pp. 223–228.

[14] Zama. *Concrete ML: a Privacy-Preserving Machine Learning Library using Fully Homomorphic Encryption for Data Scientists*. https://github.com/zama-ai/concrete-ml. 2022.

[15] M. Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

[16] M. Shen, F. Liang, R. Gong, Y. Li, C. Li, C. Lin, F. Yu, J. Yan, and W. Ouyang. "Once quantization-aware training: High performance extremely low-bit architecture search". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 5340–5349.

[17] A. Vizitiu, L. C. Nita, A. Puiu, C. Sucio, and L. M. Itu. "Applying Deep Neural Networks over Homomorphic Encrypted Medical Data". In: *Computational and Mathematical Methods in Medicine* (2020). DOI: https://doi.org/10.1155/2020/3910250.

[18] R. Hagag, H. Nassar, J. Henkel, and M. A. A. El Ghany. "Hardware-Accelerated Mode-Switching Polymorphic Encryption for Privacy Preserving Machine Learning". In: *2025 14th International Conference on Modern Circuits and Systems Technologies (MOCAST)*. 2025, pp. 1–5. DOI: 10.1109/MOCAST65744.2025.11083924.

[19] Y. Bai, H. Zhaoa, X. Shi, and L. Chen. "Towards practical and privacy-preserving CNN inference service for cloud-based medical imaging analysis: A homomorphic encryption-based approach". In: *Computer Methods and Programs in Biomedicine*. Elsevier, 2025. DOI: https://doi.org/10.1016/j.cmpb.2025.108599.

[20] W. Boulila, A. Ammar, B. Benjdira, and A. Koubaa. "Securing the Classification of COVID-19 in Chest X-ray Images: A Privacy-Preserving Deep Learning Approach". In: *International Conference of Smart Systems and Emerging Technologies*. DOI: 10.1109/SMARTTECH54121.2022.00055.

[21] Z. Yue, S. Ding, L. Zhao, Y. Zhang, Z. Cao, M Tanveer, A. Jolfaei, and X. Zheng. "Privacy-preserving Time-series Medical Images Analysis Using a Hybrid Deep Learning Framework". In: *ACM Transactions onInternet Technology*. ACM, 2021. DOI: https://doi.org/10.1145/3383779.

[22] J. O'Neill. "A Fully Secure Approach to Privacy-Preserving Machine Learningfor Satellite Image Classification". In: *Proceedings of the Canadian Conference on Artificial Intelligence* (2024). https://caiac.pubpub.org/pub/h1004qc6.

[23] J. O'Neill, L. Bouzar-Benlabiod, and N. Zincir-Heywood. "Privacy-Preserving Ensemble Learning for Medical Image Diagnostics using Post-Quantum Cryptography". In: *Proceedings of the IEEE International Systems Conference (SYSCON)*. IEEE. Apr. 2026.

[24] T Burt, K. Button, H Thom, R. Noveck, and M. Munafò. "The Burden of the "False-Negatives" in Clinical Development: Analyses of Current and Alternative Scenarios and Corrective Measures". In: *Clinical and translational science* 10 (2017), pp. 470–479. DOI: doi:10.1111/cts.12478.

[25] K. Raza. "Chapter 8 - Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule". In: *U-Healthcare Monitoring Systems*. Ed. by N. Dey, A. S. Ashour, S. J. Fong, and S. Borra. Advances in Ubiquitous Sensing Applications for Healthcare. Academic Press, 2019, pp. 179–196. DOI: https://doi.org/10.1016/B978-0-12-815370-3.00008-6. URL: https://www.sciencedirect.com/science/article/pii/B9780128153703000086.

[26] T. RAHMAN. *COVID-19 Radiography Dataset*. 2022. URL: https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database.

[27] LARXEL. *Pediatric Pneumonia Chest X-ray*. 2025. URL: https://www.kaggle.com/datasets/andrewmvd/pediatric-pneumonia-chest-xray.

[28] Bukun. *BreakHIs*. 2021. URL: https://www.kaggle.com/datasets/ambarish/breakhis.