Hybrid Privacy-preserving Histopathological Image Classification Using Fully Homomorphic Encryption

Dominik Moskalewicz[1[0009−0000−6200−2456]] and Bogusław Cyganek[1[0000−0001−5185−1145]]

Faculty of Computer Science, Electronics and Telecommunication, AGH University of Krakow, Al. Mickiewicza 30, 30-059 Kraków, Poland
{dmoskalewicz, cyganek}@agh.edu.pl

Abstract. Automated analysis of histopathological scans allows for cancer diagnosis, yet deploying deep learning models for automated classification brings the risk of misuse of patient data and privacy violations. Fully homomorphic encryption (FHE) is a cryptographic paradigm that allows for computation to be performed on encrypted data by untrusted parties, without decrypting the data. While FHE solves the problem of preserving privacy for remote computation on sensitive data, it is prohibitively computationally expensive when used with deep neural networks. To address this bottleneck, this study proposes a hybrid architecture that distributes computation between the client and the server. Our method utilizes a pretrained DINO ViT model for local image feature extraction on the client, followed by dimensionality reduction using the principal and neighborhood component analysis methods. These reduced features are then encrypted and classified remotely by a support vector machine (SVM), that can be executed in untrusted environments using FHE. We evaluated this approach on demonstrating that feature dimensions can be reduced by approximately 50% with less than one percentage point decrease in classification accuracy, while dramatically reducing encrypted model inference time.

Keywords: Histopathological image classification · ViT · DINO · PCA · NCA · Homomorphic encryption

1 Introduction

Histopathological image analysis is a key element in cancer diagnosis, determining tumor malignancy and grading disease progression. Despite its clinical importance, manual analysis of whole-slide images (WSI) is a labor-intensive and time-consuming process. Traditionally, this task is performed by specialist histopathologists, whose number and available time are limited, while the public demand for cancer diagnostics is constantly growing [16].

Machine learning can address these limitations - deep learning models trained on histopathological images can perform tissue type classification, assisting physicians with diagnosis, or performing diagnosis automatically when trained on

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_38

large quantities of images. Many studies show the efficacy of such models in datasets such as Kather CRC 2016 [14], Kather NCT CRC 100k [15], or BreaKHis [26]. Early works in this domain have relied mainly on convolutional neural network (CNN) model architectures, such as VGG, ResNet or DenseNet [16, 15, 2, 9, 29, 24, 27, 1]. With the emergence of transformer architecture, the state of the art in computer vision has shifted to models like vision transformers (ViT) [10, 28, 25, 17]. Unlike CNNs, transformer models can utilize self-attention to capture long-range dependencies and patterns in images [10]. Further on, self-supervised frameworks based on ViT, such as DINO (self-supervised vision transfomers) [3], have proven very effective in extracting semantic features from images that can be used to perform downstream tasks like image classification or image retrieval.

However, deploying these models risks exposing sensitive patient data. Fully homomorphic encryption (FHE) [11] provides a solution by enabling computations directly on encrypted data via schemes like CKKS [6] or TFHE [7, 8]. Although alternative privacy-preserving approaches exist, they present distinct trade-offs: federated learning primarily protects local data but remains vulnerable to model inversion and gradient leakage attacks [30], while secure multi-party computation requires massive latency-bound communication overhead between nodes [22]. Despite its security guarantees, FHE is multiple orders of magnitude slower than classical inference, making state of the art deep learning model architectures like CNNs and ViTs prohibitively expensive to execute [21, 20].

To overcome this issue, we propose a hybrid architecture that divides computation between the client and the server. We use a DINO ViT model to extract image features locally on the client. The encrypted features are classified by a support vector machine model remotely on the server. The encrypted classification result is sent back to the client, who decrypts it using the secret key. The main novelty of our proposed solution is that the extracted features have significantly lower dimensionality than the original images, at the same time conveying semantic information that allows classification models to have significantly fewer parameters [23].

To reduce the computational overhead of fully homomorphic encryption even further, we introduce feature dimensionality reduction algorithms on the client side; principal component analysis (PCA) [13] and neighborhood component analysis (NCA) [12]. Our choice of using dimensionality reduction for image features is based on our previous research [18, 4, 19, 5], which shows that even with high feature reduction, models such as ResFeats or DINO ViT perform well in image classification and content-based image retrieval tasks. Our experiments confirm these results, demonstrating that a reduction in the number of features does not necessarily cause a big decrease in classification accuracy. With some configurations, 50-80% feature reduction, which drastically reduces computational requirements of FHE model execution, is possible with less than one percentage point loss in accuracy compared to baseline results without feature dimensionality reduction.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_38

2 Methodology

The image classification system proposed in our study consists of three main components:

1. **Image feature extraction model** - a pretrained model with open-source weights, without domain-specific fine-tuning. It is used on the client side to transform images to the much lower dimensional feature space. These semantic features can contain a lot of relevant information in a relatively small amount of data compared to images, allowing simpler models to be trained while achieving high classification accuracy.
2. **Feature dimensionality reduction algorithm** - algorithms like PCA or NCA that further reduce the number of input parameters, while preserving as much information as possible, further reducing the amount of computational resources required by the classification model.
3. **Classification model** - a linear SVM that is trained on features extracted from input images and reduced with PCA/NCA. Model training is performed on cleartext data, then the model needs to be compiled to work in FHE. Classification is performed on the server-side using exclusively encrypted data, hence without revealing any sensitive information.

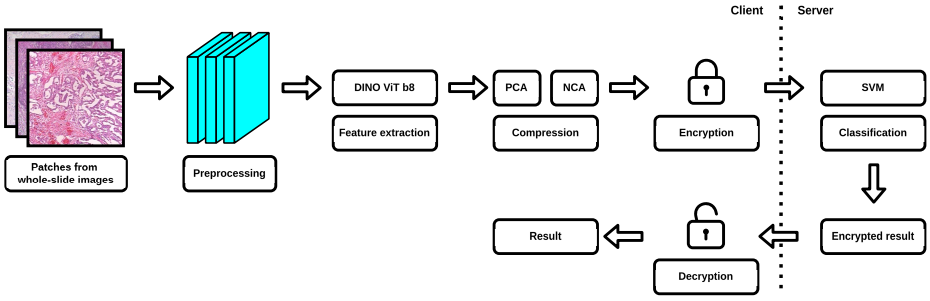


Fig. 1: Diagram showcasing the system architecture proposed in our study.

Figure 1 outlines the classification pipeline, with a dotted line separating the stages that are computed on the client from those computed on the server side, respectively. This architecture is highly suitable for scenarios where one party (the client, for example, a hospital) possesses sensitive data, while another party (the server, for example, a private company) holds a proprietary classification model. It ensures that the classification provider cannot access patient data, and the client cannot extract the underlying model, thereby protecting both patient privacy and the provider's intellectual property. The following steps are performed during classification:

1. **Obtaining patches** - A whole-slide image of a tissue scan is divided into smaller patches with dimensions matching requirements of the input layer of the feature extraction module.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_38

2. **Preprocessing** - Standard transformations for a DINO ViT model are applied to the patches, resizing them to 256 x 256 using bilinear interpolation, cropping out 224 x 224 squares from the center and normalizing the pixel values.
3. **Feature extraction** - Features are extracted from patches using a pre-trained model. In our experiments, we have used DINO ViT b8, without any fine-tuning.
4. **Feature dimensionality reduction** - Features are transformed into a lower dimensional space with the help of PCA or NCA.
5. **Encryption** - Transformed features are quantized to an integer tensor, encrypted by the client using the private key and then sent to the server.
6. **Classification** - Classification is computed on the encrypted features using a linear SVM on the server. The encrypted result is sent back to the client.
7. **Decryption** - The result is decrypted by the client, obtaining class probabilities which the client can use to determine the type of tissue present on the patch.

2.1 Datasets

We used three different datasets in our experiments: Kather NCT CRC 100K [15], Kather CRC 2016 [14] and BreaKHis [26]. All experiments were performed using 5-fold cross-validation, with the datasets divided into stratified folds with a uniform class distribution.

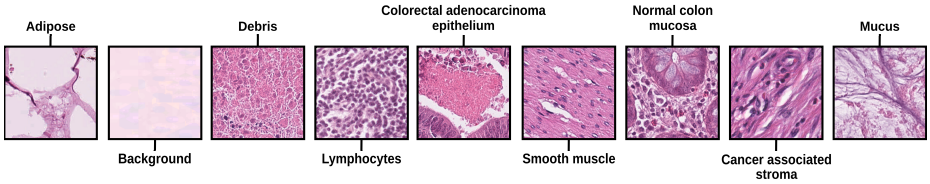


Fig. 2: Example patches from the Kather NCT CRC 100K dataset.

Kather NCT CRC 100K – Kather NCT CRC 100K [15] is a colorectal cancer dataset created by the National Cancer Centre (NCT) in Germany and Mannheim University Hospital. It consists of 100,000 224 x 224 images of H&E stained tissue sections. The images in the data set are split into nine roughly balanced classes. Figure 2 shows exemplary patches for each class.

Kather CRC 2016 – The Kather CRC 2016 [14] dataset contains 5000 150 x 150 pixel patches obtained from ten anonymized H&E stained CRC (colorectal cancer) whole-slide images. The samples are evenly distributed among eight classes, with 625 samples per class. Figure 3 shows sample patches for each class.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_38

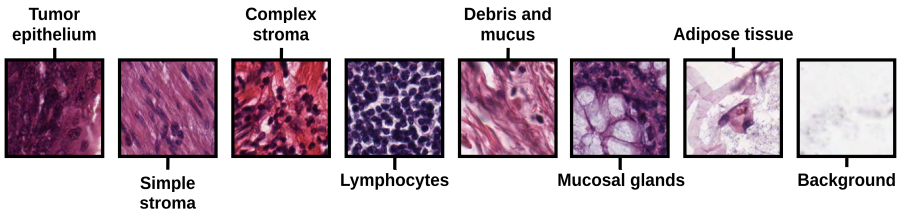


Fig. 3: Example patches from the Kather CRC 2016 dataset.

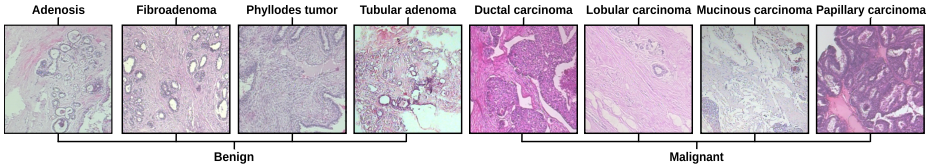


Fig. 4: Example patches from the BreaKHis dataset at 40 times magnification.

BreaKHis – BreaKHis [26] is a dataset of H&E stained tissue images taken from breast cancer patients. It is split into four subsets with different magnification levels: 40x, 100x, 200x, and 400x magnification. The dataset can be evaluated in a binary scenario with two classes (benign and malignant) or a multiclass scenario with eight classes. The dataset contains 7909 images in total, split between magnifications, with the subset for each magnification level being around 2000 images. Figure 4 shows exemplary patches.

2.2 Implementation Details

Our system was implemented in Python3, using PyTorch for feature extraction, scikit-learn used for splitting the dataset, grid search, PCA and NCA implementations and measuring model performance. We have used Concrete-ML for training and running encrypted model inference.

FHE Implementation Concrete-ML is a machine learning library built on the FHE compiler library Concrete. Concrete implements the TFHE encryption scheme, allowing for compiling high level code to circuits that can be evaluated using fully homomorphic encryption. Since Concrete can only perform computation on booleans and integers, these models are trained to work in integer space using quantization, with model weights being integers between 2 and 16 bits of precision. After training, these models are compiled to TFHE-compatible circuits for encrypted inference. Concrete-ML models can operate in three distinct modes after compilation:

1. **Disable** - disable encryption. We used this mode to compute the time required to classify samples without FHE.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_38

2. **Simulate** - simulates the encrypted circuit without using encryption. Used for calculating model performance metrics such as classification accuracy.
3. **Execute** - execute the model using FHE. Used for measuring encrypted model inference time.

Image Feature Extraction – For image feature extraction, we have utilized the DINO ViT b8 model, using open-source model weights, without applying any fine-tuning.

Feature Dimensionality Reduction – For feature dimensionality reduction we have used the PCA and NeighborhoodComponentsAnalysis classes from scikit-learn. Additionally, the NeighborhoodComponentsAnalysis class implements the linear discriminant analysis method for initializing NCA, which can be used when the number of components is lower than the number of classes in the dataset. In our experiments, linear discriminant analysis is only applied for the multiclass datasets (with 8 or 9 classes) when the number of components is 5.

Classification – We used the LinearSVC class from the Concrete-ML library for classification, which is a linear SVM classification algorithm implemented to work in FHE.

3 Results and Discussion

To ensure robust results, all experiments utilized 5-fold cross-validation, with reported values representing the fold averages. Plaintext and encrypted SVM inference times were measured on an AMD Ryzen 9 9950X CPU. In our study three experiments were performed: (i) examining the classification performance of our architecture on different datasets, (ii) measuring the impact of weights bit-width and (iii) influence of feature dimensionality reduction on time and classification accuracy of encrypted SVM inference. The inference time was measured by calculating the time required to classify 100 samples. Results of these experiments are reported in the subsequent sections.

3.1 Impact of Quantization on Model Performance

We measured the impact of model weight bit-width precision on model classification accuracy and encrypted inference time. We tested the performance of the model for bit-width values of 2 to 16 in the Kather NCT CRC 100k dataset. Figure 5 shows two plots, one showing classification accuracy and the second the time needed to classify 100 encrypted samples, respectively. Our results show that 6 to 8 bits of precision are enough to achieve close to peek classification accuracy. Based on these results, we have opted to use 8 bits of integer precision for further experiments, which allows for a good balance between model performance and required computational resources.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_38

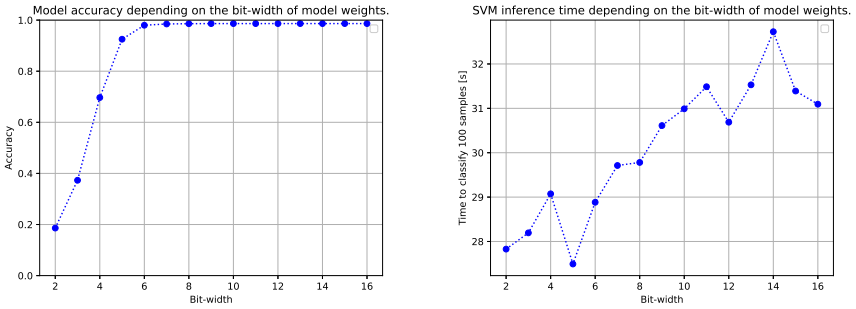


Fig. 5: Encrypted model performance depending on the bit-width of model weights.

3.2 Classification Performance

In order to validate the performance of the architecture proposed in this study, we evaluated classification accuracy across the Kather CRC 2016, Kather CRC NCT 100k and BreaKHis datasets, respectively. We established a baseline by using the full 768 dimension feature vector obtained from DINO ViT b8 without dimensionality reduction. For multiclass classification tasks we obtain 98.59% accuracy on Kather NCT CRC 100k, 95.02% accuracy for Kather CRC 2016, 90.32%, 86.25%, 85.34%, and 84.56% accuracy on BreaKHis with 40x, 100x, 200x, and 400x magnification respectively. For binary classification on BreaKHis we obtained 96.84%, 95.19%, 97.06% and 96.04% classification accuracy respectively.

Impact of Dimensionality Reduction on Multiclass Classification Performance – Table 1 demonstrates classification accuracy results obtained by reducing the dimensionality of features using PCA before training on multiclass datasets. We observed a very strong resilience to dimensionality reduction in the model trained on the Kather NCT CRC 100k dataset, with 97.89% accuracy, less than one percentage point drop from the baseline, while using only 120 components. Surprisingly, we obtained slightly higher accuracy results, while using 550, 500, 450, 400, 350, and 310 components than the baseline. This suggests that it is possible to maintain accuracy while reducing the feature amount to less than half of the original number. Results obtained on the Kather CRC 2016 dataset are slightly worse, we observed around one percentage drop in accuracy while using 310 components, again less than 50% less than the original number of features. Feature dimensionality reduction did not work as well on the BreaKHis datasets. Reducing the number of features to 700 decreased the accuracy by more than one percentage point for all magnification levels. There is a consitent downward trend, with a five percentage point drop in accuracy from baseline results with around 50% feature reduction. These results suggest that dimensionality reduction performs better on datasets with strong baseline

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_38

Table 1: Multiclass classification accuracy comparison across different datasets using PCA depending on the number of components used. Underlined text shows values with around 1 percentage percentage point accuracy drop compared to results without dimensionality reduction.

Components	Kather CRC NCT 100k	Kather CRC 2016	BreaKHis 40x	BreaKHis 100x	BreaKHis 200x	BreaKHis 400x
768[1]	98.59	**95.02**	**90.32**	**86.25**	**85.34**	**84.56**
700	98.19	94.58	88.27	83.99	82.51	81.04
650	98.17	94.44	88.42	83.85	82.26	80.65
600	98.33	94.36	88.07	83.42	81.66	80.93
550	98.65	94.26	87.96	83.22	81.76	80.38
500	**98.77**	94.42	88.17	82.98	80.97	80.65
450	**98.77**	94.40	88.22	83.22	81.22	79.83
400	98.72	94.36	88.27	82.12	79.73	79.89
350	98.67	94.28	87.66	81.20	79.28	78.18
310	98.62	94.26	86.26	79.76	78.24	75.60
270	98.55	93.88	85.31	76.88	77.34	75.16
230	98.42	93.84	83.00	82.22	79.53	75.49
200	98.27	93.48	85.16	81.01	78.58	75.87
160	98.15	92.92	84.91	77.94	77.24	75.16
120	97.89	92.42	81.55	79.91	80.12	74.34
80	97.33	91.10	83.50	80.24	78.54	75.54
60	96.83	90.32	82.50	78.90	78.34	75.05
50	96.62	90.06	82.35	77.46	76.40	73.57
40	96.11	89.58	81.80	75.78	74.16	70.38
30	95.56	88.42	78.89	72.27	69.54	67.14
20	93.99	87.46	73.58	68.81	66.12	62.47
10	89.65	83.36	63.90	59.82	62.34	57.85
5	81.43	64.46	52.48	51.22	56.73	53.46

[1] Results using features without dimensionality reduction included for comparison.

Table 2: Multiclass classification accuracy comparison across different datasets using NCA depending on the number of components used. Highest accuracy values are emboldened.

Components	Kather CRC NCT 100k	Kather CRC 2016	BreaKHis 40x	BreaKHis 100x	BreaKHis 200x	BreaKHis 400x
768[1]	98.59	**95.02**	90.32	86.25	**85.34**	**84.56**
700	98.44	94.90	**90.33**	86.26	85.25	83.74
650	98.34	94.84	90.28	**86.54**	85.20	83.46
600	98.46	94.86	90.03	85.97	85.25	83.79
550	98.69	94.72	90.33	86.06	84.50	83.41
500	98.78	94.84	89.97	85.54	84.10	83.30
450	**98.78**	94.64	89.77	84.77	83.51	82.80
400	**98.75**	94.60	89.62	84.48	82.31	82.14
350	98.71	94.58	88.97	83.13	81.17	80.44
310	98.63	94.58	88.17	81.50	79.88	79.12
270	98.54	94.48	87.37	78.71	79.68	76.37
230	98.44	94.32	83.96	83.28	82.46	77.31
200	98.27	94.16	86.37	81.79	80.28	78.02
160	98.12	94.06	85.56	77.32	76.60	75.22
120	97.90	93.48	80.60	80.06	79.78	72.20
80	97.40	92.52	82.81	78.71	77.80	75.05
60	96.98	91.82	82.46	76.74	77.00	75.77
50	96.83	91.20	81.15	77.37	76.31	74.07
40	96.63	90.68	81.35	76.07	74.22	70.33
30	96.38	89.22	79.40	72.85	69.70	67.25
20	96.09	88.02	74.59	69.01	66.92	63.52
10	95.03	85.68	65.86	61.08	62.59	58.52
5	96.16	87.54	81.70	77.17	75.86	73.96

[1] Results using features without dimensionality reduction included for comparison.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_38

Table 3: Binary classification accuracy comparison across different datasets using PCA depending on the number of components used. Underlined text shows values with around 1 percentage percentage point accuracy drop compared to results without dimensionality reduction.

Components	BreaKHis 40x	BreaKHis 100x	BreaKHis 200x	BreaKHis 400x
768[1]	96.84	**95.19**	97.06	**96.04**
700	96.94	95.04	**97.26**	95.98
650	97.04	94.80	96.96	95.71
600	**97.19**	94.95	96.52	95.32
550	96.84	94.85	96.91	95.27
500	**97.19**	94.61	96.57	95.05
450	96.79	94.76	96.76	95.16
400	96.69	94.47	96.37	95.60
350	96.74	93.99	96.32	94.61
310	96.64	93.36	95.92	94.12
270	96.64	92.69	95.52	94.45
230	94.98	92.45	95.57	94.06
200	95.58	92.40	94.93	93.51
160	94.63	92.16	94.23	92.85
120	94.38	93.17	93.59	93.02
80	95.23	94.13	94.13	92.96
60	94.28	93.07	93.19	92.52
50	94.38	93.07	92.74	92.03
40	94.08	93.03	92.39	91.70
30	92.38	92.35	91.85	90.49
20	91.02	91.20	89.76	88.07
10	86.61	87.36	88.47	85.82
5	83.35	79.96	86.68	85.32

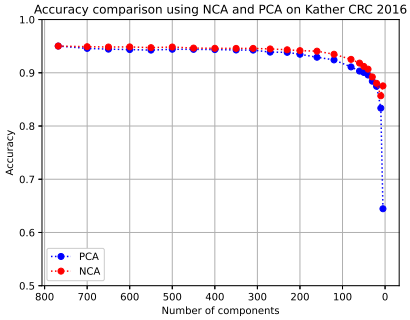
[1] Results using features without dimensionality reduction included for comparison.

Table 4: Binary classification accuracy comparison across different datasets using NCA depending on the number of components used. Highest accuracy values are emboldened.

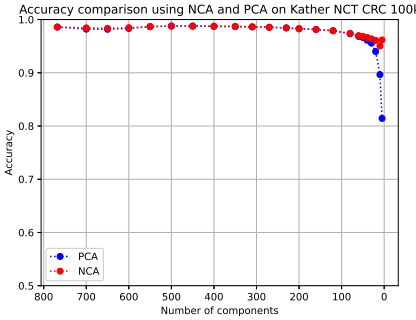
Components	BreaKHis 40x	BreaKHis 100x	BreaKHis 200x	BreaKHis 400x
768[1]	96.84	95.19	97.06	96.04
700	**97.19**	95.10	97.12	**96.32**
650	97.09	94.90	97.02	95.77
600	96.89	95.09	97.02	95.82
550	96.79	**95.24**	96.97	96.10
500	96.84	94.86	**97.22**	95.71
450	96.79	94.91	96.92	95.66
400	96.84	94.71	96.87	95.49
350	96.84	94.33	96.57	95.05
310	96.59	94.09	96.52	94.62
270	96.14	93.56	96.07	94.56
230	95.58	93.37	96.22	94.45
200	95.63	92.89	95.88	93.68
160	95.28	92.31	94.19	93.08
120	94.63	93.22	94.04	92.64
80	94.93	94.09	94.04	92.97
60	94.23	92.94	93.34	92.69
50	94.08	93.03	92.90	92.25
40	93.73	92.55	92.50	91.81
30	92.43	92.31	91.85	90.27
20	91.17	91.45	89.86	88.35
10	89.02	88.18	89.32	86.37
5	89.07	86.40	89.67	86.15

[1] Results using features without dimensionality reduction included for comparison.

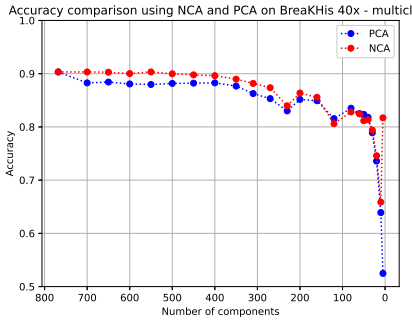
ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_38



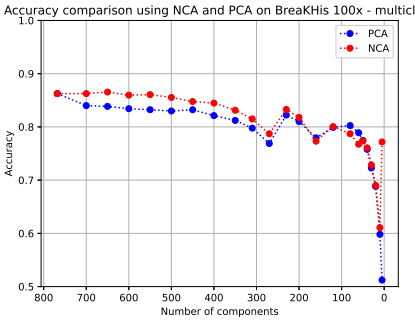
(a) Kather CRC 2016.



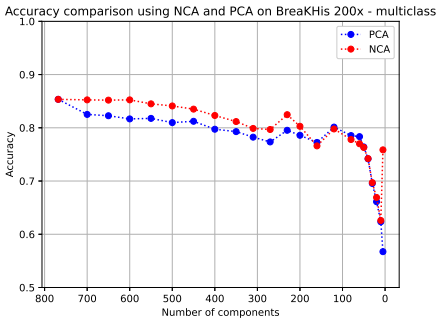
(b) Kather NCT CRC 100k.



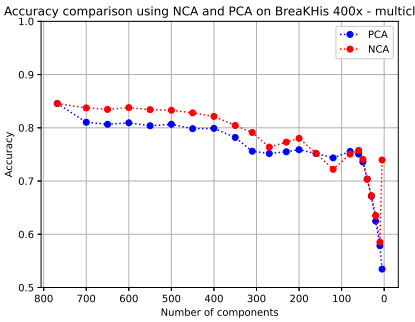
(c) BreaKHis 40x multiclass.



(d) BreaKHis 100x multiclas.



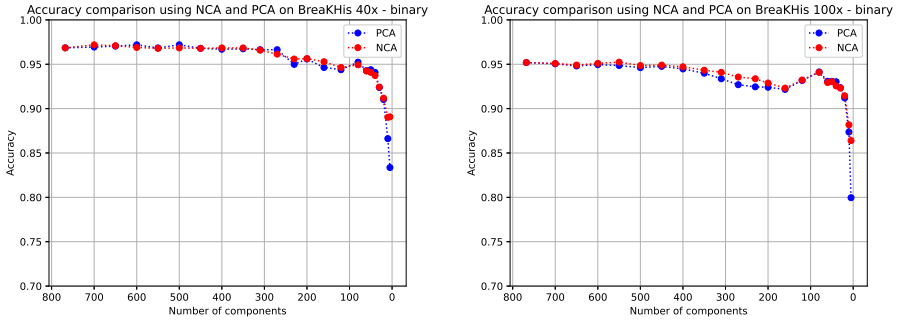
(e) BreaKHis 200x multiclass.



(f) BreaKHis 400x multiclass.

Fig. 6: Multiclass classification accuracy comparison, using PCA and NCA for feature dimensionality reduction.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_38



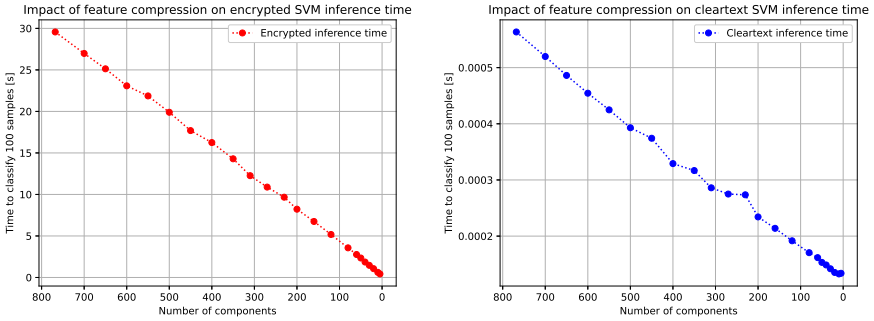
(a) BreaKHis 40x binary.

(b) BreaKHis 100x binary.

(c) BreaKHis 200x binary.

(d) BreaKHis 400x binary.

Fig. 7: Binary(benign or malignant) classification accuracy comparison, using PCA and NCA for feature dimensionality reduction.



(a) Encrypted inference time plot.

(b) Cleartext inference time plot.

Fig. 8: Encrypted and cleartext SVM inference time measurements.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_38

results. It is important to note that the Kather NCT CRC 100k dataset is significantly larger and maintains a stable class distribution compared to BreaKHis. These factors could have a significant impact on both model training and fitting of dimensionality reduction algorithms such as PCA or NCA to the data.

Table 2 details the classification accuracy results obtained while using NCA for feature dimensionality reduction. Unlike PCA, which is unsupervised, NCA utilizes class labels to preserve discriminative information, resulting in improved performance. Figure 6 illustrates this advantage, with NCA outperforming PCA for most data points on all multiclass datasets. This is especially evident when examining BreaKHis results; for the dataset with 40x magnification, we reduced the number of features to 400, while only losing 0.7 of a percentage point in classification accuracy. Similarly for 100x, 200x and 400x magnification datasets, accuracy was preserved within one percentage point difference from the baseline with 500, 500 and 550 components, respectively. NCA managed to preserve accuracy on Kather CRC 2016 while using only 160 components, compared to 310 using PCA. On Kather NCT CRC 100k, NCA maintains stability even with a very low number of feature dimensions, keeping 97.9% accuracy using only 120 components and never dropping below 95% regardless of the component number. Notably, while reducing the number of features down to 5 we were able to use linear discriminant analysis (LDA) to initialize NCA, which increased classification accuracy significantly. This configuration achieved higher results than NCA without LDA with 10-40 components depending on the dataset used. Histopathological image datasets usually have a low number of classes (9 or fewer in the case of datasets used in our study), which limits the usability of NCA with LDA in practice. We speculate that on datasets with a higher number of classes, feature reduction using NCA initialized with LDA might show even more significant improvements over other methods.

Impact of Dimensionality Reduction on Binary Classification Performance – Table 3 and Table 4 present the results for the binary classification task (benign vs. malignant) on the BreaKHis dataset. Baseline results on binary BreaKHis are significantly higher compared to multiclass BreaKHis, 96.84%, 95.19%, 97.06%, and 96.04% for 40x, 100x, 200x, and 400x magnification datasets, respectively. Similarly to multiclass experiments, we observed that PCA allows for a reduction in feature dimensions while maintaining high classification accuracy in the binary setting. As shown in Table 3, we maintained accuracy within one percentage point of the baseline while using as few as 200 components for the 40x dataset. The models exhibit strong resilience, with accuracy remaining above 88% even when reduced to 20 dimensions.

Table 4 shows that NCA generally outperforms PCA. For example, on the 200x magnification data set, NCA maintained accuracy within one percentage point of the baseline using only 230 components. In particular, at very low dimensions (5 components), NCA still achieved accuracy between 86% and 89% across all magnifications, significantly outperforming PCA, which dropped to between 79% and 86% at the same dimensionality.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_38

In general, the binary classification task proved significantly more resilient to feature reduction than multiclass classification. While multiclass accuracy on BreaKHis dropped sharply at lower dimensions, the binary models retain high performance even with very few components. This suggests that the decision boundary between benign and malignant tissue is far more distinct and linearly separable in the lower-dimensional space than the boundaries between specific subtypes. Figure 7 compares the accuracy of PCA and NCA across the binary BreaKHis subsets.

3.3 Inference Time Benchmark

To evaluate the computational overhead introduced by Homomorphic Encryption and the efficiency gains provided by our proposed dimensionality reduction methods, we measured the time required to classify a batch of 100 samples. Figure 8 illustrates the inference time for both unencrypted (cleartext) and encrypted execution as a function of the number of input features. Looking at the baseline results with 768 features, encrypted model inference is more than 50,000 times slower than cleartext execution. For processing large quantities of data or real-time classification, this is a significant overhead.

Although cleartext execution is negligible, taking fractions of a millisecond regardless of feature count, the encrypted inference time exhibits a distinct linear correlation with the number of input dimensions. The baseline model, which uses the full 768 feature vector of the DINO ViT model, takes approximately 29.85 seconds to classify 100 encrypted samples.

The application of PCA or NCA significantly reduces this computational burden. By reducing the feature space to 350 dimensions, inference time drops to 13.77 seconds, less than 50% of the baseline processing time, while maintaining high classification accuracy across most datasets. This downward trend continues linearly; at 160 dimensions, the execution time is reduced to 6.64 seconds.

The most substantial performance gains are observed at the lower end of the dimensionality spectrum. Using only 5 components results in an inference time of just 0.50 seconds for 100 samples. This represents a more than 60-fold reduction in processing time compared to the baseline.

4 Conclusions

In this paper, we proposed a hybrid privacy-preserving architecture for histopathological image classification. By combining client-side image feature extraction using DINO ViT with server-side fully homomorphically encrypted SVM classification on the extracted features, we successfully balanced the computational requirements of FHE with accuracy requirements of histopathological image recognition in privacy-constrained environments.

Our experiments demonstrate that feature dimensionality reduction allows for optimizing FHE inference, significantly reducing model execution time. We showed that techniques such as PCA and NCA can reduce the feature space by

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_38

approximately 50% with less than a one percentage point drop in accuracy. This reduction translates directly to a linear decrease in encrypted inference time, processing a batch of 100 samples in under 14 seconds, a significant improvement over fully encrypted deep neural networks which often require hours of processing per image.

These results confirm that, while encrypted deep learning remains computationally prohibitive, our hybrid approach utilizing compact semantic features offers a practical path for secure remote medical diagnostics. Future work will focus on extending these ideas, by evaluating possible improvements to this architecture, such as the usage of different encrypted classification models.

Acknowledgments. This research was funded in whole by the National Science Centre, Poland, grant number: UMO-2024/55/B/ST6/01681.

Disclosure of Interests. The authors declare no conflicts of interest.

References

1. Alom, M.Z., et al.: Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network. Journal of Digital Imaging **32**(4), 605–617 (Aug 2019)
2. Bychkov, D., et al.: Deep learning based tissue analysis predicts outcome in colorectal cancer. Scientific Reports **8**(1), 3395 (Feb 2018)
3. Caron, M., et al.: Emerging properties in self-supervised vision transformers (2021)
4. Chandranegara, D.R., et al.: Compact dino-vit: Feature reduction for visual transformer. Electronics **13**(23) (2024)
5. Chandranegara, D.R., et al.: A study on highly efficient compact transformer features for histopathological image recognition. In: Artificial Intelligence and Soft Computing: 24th International Conference, ICAISC 2025, Zakopane, Poland, June 22–26, 2025, Proceedings, Part II. p. 38–49. Springer-Verlag, Berlin, Heidelberg (2025)
6. Cheon, J.H., et al.: Homomorphic encryption for arithmetic of approximate numbers. In: Takagi, T., Peyrin, T. (eds.) Advances in Cryptology – ASIACRYPT 2017. pp. 409–437. Springer International Publishing, Cham (2017)
7. Chillotti, I., et al.: Faster fully homomorphic encryption: Bootstrapping in less than 0.1 seconds. In: Cheon, J.H., Takagi, T. (eds.) Advances in Cryptology – ASIACRYPT 2016. pp. 3–33. Springer Berlin Heidelberg, Berlin, Heidelberg (2016)
8. Chillotti, I., et al.: Tfhe: Fast fully homomorphic encryption over the torus. Journal of Cryptology **33**(1), 34–91 (Jan 2020)
9. Coudray, N., et al.: Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nature Medicine **24**(10), 1559–1567 (Oct 2018)
10. Dosovitskiy, A., et al.: An image is worth 16x16 words: Transformers for image recognition at scale (2021)
11. Gentry, C.: Fully homomorphic encryption using ideal lattices. In: Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing. p. 169–178. STOC '09, Association for Computing Machinery, New York, NY, USA (2009)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_38

12. Goldberger, J., et al.: Neighbourhood components analysis. In: Proceedings of the 18th International Conference on Neural Information Processing Systems. p. 513–520. NIPS'04, MIT Press, Cambridge, MA, USA (2004)
13. Hotelling, H.: Relations Between Two Sets of Variates, pp. 162–190. Springer New York, New York, NY (1992)
14. Kather, J.N., et al.: Multi-class texture analysis in colorectal cancer histology. Scientific Reports **6**(1), 27988 (Jun 2016)
15. Kather, J.N., et al.: Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study **16**, 1–22 (01 2019)
16. Koziarski, M., et al.: Diagset: a dataset for prostate cancer histopathological image classification. Scientific Reports **14**(1), 6780 (Mar 2024)
17. Kumar, A., et al.: Vision transformer based effective model for early detection and classification of lung cancer. SN Computer Science **5**(7), 839 (Aug 2024)
18. Łażewski, S., Cyganek, B.: Highly compressed image representation for classification and content retrieval. Integr. Comput.-Aided Eng. **31**(3), 267–284 (Jan 2024)
19. Łażewski, S., Cyganek, B.: Vision transformer representations for efficient content-based image retrieval. In: Rutkowski, L., Scherer, R., Korytkowski, M., Pedrycz, W., Tadeusiewicz, R., Zurada, J.M. (eds.) Artificial Intelligence and Soft Computing. pp. 144–157. Springer Nature Switzerland, Cham (2026)
20. Lee, E., et al.: Low-complexity deep convolutional neural networks on fully homomorphic encryption using multiplexed parallel convolutions. Cryptology ePrint Archive, Paper 2021/1688 (2021)
21. Lee, J.W., et al.: Privacy-preserving machine learning with fully homomorphic encryption for deep neural network. IEEE Access **10**, 30039–30054 (2022)
22. Mohassel, P., Zhang, Y.: Secureml: A system for scalable privacy-preserving machine learning. In: 2017 IEEE Symposium on Security and Privacy (SP). pp. 19–38 (2017)
23. Ozkan, M., et al.: A histopathology aware dino model with attention based representation enhancement. Scientific Reports **15**(1), 45083 (Dec 2025)
24. Paladini, E., et al.: Two ensemble-cnn approaches for colorectal cancer tissue type classification. Journal of Imaging **7**(3) (2021)
25. Sabry, M., et al.: A vision transformer approach for breast cancer classification in histopathology. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). pp. 1–4 (2024)
26. Spanhol, F.A., et al.: A dataset for breast cancer histopathological image classification. IEEE Transactions on Biomedical Engineering **63**(7), 1455–1462 (2016)
27. Spanhol, F.A., et al.: Breast cancer histopathological image classification using convolutional neural networks. In: 2016 International Joint Conference on Neural Networks (IJCNN). pp. 2560–2567 (2016)
28. Sriwastawa, A., et al.: Vision transformer and its variants for image classification in digital breast cancer histopathology: a comparative study. Multimedia Tools and Applications **83**(13), 39731–39753 (Apr 2024)
29. Sun, K., et al.: Automatic classification of histopathology images across multiple cancers based on heterogeneous transfer learning. Diagnostics **13**(7) (2023)
30. Zhu, L., et al.: Deep leakage from gradients (2019)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_38