



# Optimized Machine Learning Model for Breast Tumor Classification on Histopathological Images

Yousif A. Hamad<sup>1,2</sup>, Sahar Shakir<sup>3</sup>, Ayvar A. Rashid<sup>1</sup>, Anastasiia Safonova<sup>4</sup> and Rajkumar Jagdale<sup>5</sup>

<sup>1</sup>Department of Computer Science, University of Kirkuk, 36001, Kirkuk, Iraq

<sup>2</sup>Artificial Intelligence Laboratory, Siberian Federal University, Krasnoyarsk, 660074 Krasnoyarsk, Russia

<sup>3</sup>Technical Engineering Collage-Kirkuk, Northern Technical University, Kirkuk, Iraq

<sup>4</sup>Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany

<sup>5</sup>Department of Computer Science, Vishwakarma University, Pune

Received 26 August 2024, Revised 1 February 2025, Accepted 2 February 2025

**Abstract:** Advances in breast cancer screening programs have received significant attention due to their potential to improve early detection, accuracy, and efficiency in diagnosing breast disease. Timely diagnosis and accurate staging are crucial for effective treatment of breast cancer. To systematize them, software detection methods based on machine learning are used. These articles allow breast histopathological images, to be carefully examined, allowing abnormalities (cancer detection) to be diagnosed quickly. Ultimately, this results in increased treatment effectiveness and reduced mortality rates. This methodology innovation allows radiologists to detect malignant tumors without the need for surgical procedures. The change in strategy is mainly due to the widespread adoption of machine learning models and related technologies. A new approach is presented that consists of two critical steps: using Convolutional Neural Networks (CNNs) to extract biological features and then using Support Vector Machines (SVMs) to reliably detect breast tumors and classify them as benign or malignant. However, the traditional CNN-based SVM model has encountered overfitting problems due to the use of a large amount of training data, despite the initial promises. This approach combines CNN, rectified linear unit (ReLU) structure, and SVM using combined features learned from CNN. The success of this tactic is measured using performance metrics that include precision, recall, and accuracy. The SVM-CNN model tested on two state-of-art datasets, i.e BreakHis and Bach. Proposed method achieved 98% accuracy on the BreakHis dataset and 97% on Bach dataset, indicating its potential to revolutionize breast tumor classification and diagnosis.

**Keywords:** Histological images, Support vector machines, Convolutional neural networks, Breast tumor classification, Medical image, Computer Aided Diagnosis CAD

## 1. INTRODUCTION

Breast tumor classification using histopathological images has seen significant advances due to the application of machine learning (ML) and DL techniques. This area of research focuses on developing models that can accurately distinguish between benign and malignant tumors, as well as classify subtypes of breast cancer. A precise categorization of malignant and benign tumors is crucial in the management of breast cancer as it facilitates monitoring the efficacy of therapy and enables modifications to treatment strategies [1]. The rising prevalence of breast cancer among women worldwide has generated significant attention towards the advancement of Computer Aided Diagnosis (CAD) models that use ML algorithms and other methodologies [2]. Computer Tomography (CT) is beneficial for early identification of breast cancer in medical

pictures due to its ability to identify distinct characteristics such as calcifications, masses, bilateral asymmetry, and architectural anomalies. In order to enhance the precision and reliability of Breast Cancer Diagnosis (BCD) identification, researchers investigating BCD are increasingly using Convolutional Neural Network (CNN) architectures [3], [4]. This research examines the detection of breast cancer by the use of CNN-based structures. The proposed pipeline effectively categorizes tissue pictures to detect instances of breast cancer, as shown by validation using well known benchmark datasets such as BreakHis [5] and Bach [6]. The rapid proliferation of cancer cells underscores the need of early detection, since failure to identify these cells in their first phases might have fatal consequences. Histopathology is the most dependable technique for detecting anomalies in tissues, making it crucial for the diagnosis of Breast



Cancer [7], [8]. Characterized by the excessive growth of breast cells, breast cancer is a leading cause of mortality among women. Machine learning approaches, particularly those using CNN [9], are used for the purpose of breast cancer diagnosis due to their ability to automatically extract distinctive features. Histopathology is the examination of tissues at a microscopic level to analyze disease symptoms by analyzing molecular data, cellular structure, and tissue organization. According to reputable organizations like the American Cancer Society (ACS) [10], breast cancer is the second most fatal illness worldwide. Prompt identification is essential, as shown by the projected increase in cancer prevalence by 2030, emphasizing the need of timely diagnosis for improved treatment outcomes and reduced patient suffering. Traditional diagnostic techniques for BCD are impeded by high costs, time consuming procedures, and the need for skilled pathologists to provide conclusive assessments [11], [12].

Moreover, inconsistencies in pathologists' analyses often result in diagnostic mistakes and uncertainty. Non-invasive methods may not consistently provide a precise and comprehensive diagnosis. The purpose of this paper is to create a CAD model for automatic classification of pathological anatomical images to improve breast cancer detection results [13]. The problem lies in the inefficiency and slowness associated with the lack of software methodologies [14], [7], [15]. Using machine learning algorithms to classify breast tumors has several advantages over traditional methods [16], [17]. Machine learning algorithms are able to quickly process and analyze large amounts of data, resulting in fast classification.

Ensuring that ML models can be effectively integrated into clinical practice, including user friendly interfaces and real time processing capabilities. This paper uses raw sensor data to present a new method combining CNN and Support Vector Machines (SVM) for intelligent diagnosis of breast tumors. It also presents a visualization technique for misclassification and enhances CNN structure for efficiency. Various DL techniques are incorporated for robustness and generalization. The method is well suited for online monitoring and real time diagnosis due to its enhanced efficiency.

## 2. LITERATURE REVIEW

Breast tumor classification on histopathological images using machine learning has made significant strides, with deep learning techniques leading the way in achieving high accuracy. However, challenges such as data annotation, interpretability, and clinical integration remain areas of active research. For example, determining the shape of the nuclei may be enough to determine if a tissue is benign or cancerous [18], [19], [20]. DL and NN have been widely embraced and advocated due to their significant use [21], [22]. Neural network algorithms are more efficient than standard Artificial intelligence (AI) approaches for applications involving big data sets and complicated decision making [23], [24].

Advanced DL techniques, a subset of artificial intelligence methods, accomplish desired outcomes by learning from examples without human interaction in picking input data. With sufficient training data and expertise, machine learning has the potential to outperform human professionals in specific cognitive tasks. Autonomous cars, poker, images, voices, translation, and synthesis are among the industries that have significantly profited from DL [25], [26]. These architectures have undergone extensive training using internet data since the early 2000s, with DL techniques being particularly advantageous in picture related tasks, including image recognition, segmentation, and categorization. Due to the tools' versatility and flexibility, they may easily be used on photographs from many fields as long as there is a significant quantity of training data from the specific domain. Some methods presented that focus on the use of images in multi scale version to obtain meaningful and more powerful features from tissue images to enhance the classification task [27]. NN based approaches are now widely recognized as the best in several fields, such as drug development, genomics, and image processing.

Research on diagnosing breast cancer by image analysis has been ongoing for over 40 years, resulting in significant advancements in the field. These studies can be divided into two groups depending on their methodologies: classical machine learning methods and distance learning methods [28], [18]. The first group mainly deals with small datasets of breast cancer images and relies on abstract features that are time consuming and rather inefficient. This class can process large amounts of data and automatically obtain abstract properties [17]. Four machine learning classifiers, k-Nearest Neighbor (KNN), Native Bayes (NB), Decision Tree (DT), and SVM, were trained using these features to classify cytology images as benign or malignant. Besides kernel segmentation, specific studies have focused on extracting comprehensive information from whole images. The presented ML model has achieved great success in accurately classifying microscopic breast cancer biopsy findings. Traditional ML algorithms offer the potential to evaluate histopathological images of breast tumors classification, but their performance largely depends on the choice of features used during training. Moreover, they may have difficulty collecting and organizing relevant information from the data effectively [14], [29].

## 3. METHODOLOGY

Early detection is essential for successful treatment of breast cancer and improving patient outcomes as technology advances against cancer. Recent developments and research in machine learning methods have shown the potential to improve the accuracy and efficiency of breast tumor classification and detection [30]. The cancerous cells vary in shape and size, and distinguishing between these cells and normal cells is major challenge. Visual examination of this tissues under microscope is usually performed by a pathologist. In Computer Aided Diagnosis (CAD) we use different methods to first extract essential information from tissue images

then determine whether these tissues/cells are benign or malignant. In the first step we use convolutional neural networks to extract features from the medical image and then use SVM to classify the features. To enhance the methods training methods and classification process, many techniques such as data augmentation, dropout and dimension reduction are used. CNNs excel at image processing tasks, but SVMs are known for their strong classification skills on medical images. Researchers have proposed innovative CNN-based SVM models for breast tumor classification by combining these two approaches, especially using datasets such as BreakHis [5] and Bach [6].

The main contribution of this study is the creation of a high performance deep neural network model with high accuracy that categorizes breast cancer tumors based on a biological framework.

#### A. Feature Extraction

The histological images include various information, including statistical and technical data. It is crucial to capture this structural insight in many data analysis tasks, especially image labeling. Identifying structured information involves using several strategies, such as aggregating unmoderated data, to group relevant vectors and provide positive results [31]. Clustering is a method used to group comparable datasets with familiar properties, highlighting differences between each dataset. The research utilizes K-Means and Mean-Shift fuzzy sets procedures to reveal the data's underlying structure and other data gathering strategies. The images depict both benign and malignant tumors, providing visual depictions of the clusters created for each tumor type [32], [33]. Examples are shown in figure 1.

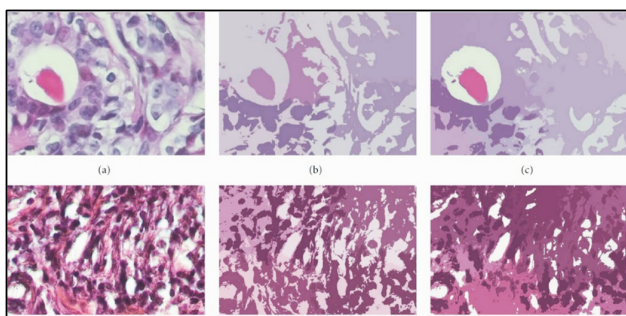


Figure 1. Group of comparable datasets that have common properties (features)

CNNs are adept at automatically learning and extracting hierarchical features from raw image data. These features can be complex and highly discriminative. CNNs rely on one or more convolution layers (filter or kernel) to extract information from raw images. This produce a 2-D feature array that can be used to discriminate main properties of the image. Adding more layer will result in a higher dimensional model that runs slower but produce more reliable results, these types of networks called deep

convolutional neural networks. In this work, three channel RGB histological images were used as input to the network without any preprocessing.

The general structure of the CNN model is as follows: Input layer consist of 3 layers with the shape of 224 by 224 pixel, there was no preprocessing on the original RGB image to prevent data loss in the process. The first convolutional layer consists of 64 filters, 3x3 kernel and as an activation function ReLU where used. In first pooling layer, Max pooling with 2x2 pool size utilized. In the second convolutional layer, 123 filters 3x3 kernel with ReLU activation function is utilized. Max pooling used in the second pooling layer as before with 2x2 pool size. In the last fully connected layer, 512 neurons used with ReLU activation function. The architecture of the CNN model shown in figure 2.

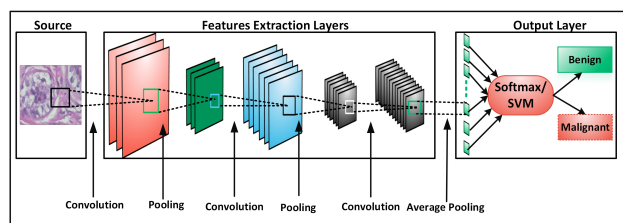


Figure 2. The improved ML model of CNN with SVM combination

#### B. Dimension Reduction and Dropout

Dimension reduction, commonly called pooling, is a method employed to streamline features by reducing their dimensional. Figure 3 in the CNN model illustrates this process. Dealing with a multitude of neurons, a DNN might generate numerous predictions, leading to high accuracy on the training set but diminished accuracy on the test set, a phenomenon known as overfitting [13], [34]. To address this concern, the dropout technique was implemented to mitigate overfitting, as depicted in figure 4.

The dropout technique addresses overfitting, which entails randomly turning off certain neurons during training. This specific deactivation improves the network's capacity to sustain consistent learning characteristics. figure 4 illustrates dropout, where undetectable neurons 2 and 4

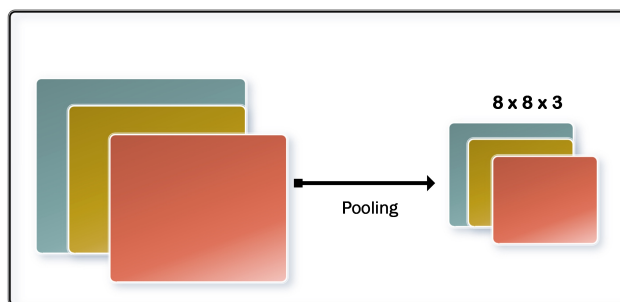


Figure 3. The pooling procedure is executed using  $2 \times 2$  kernels

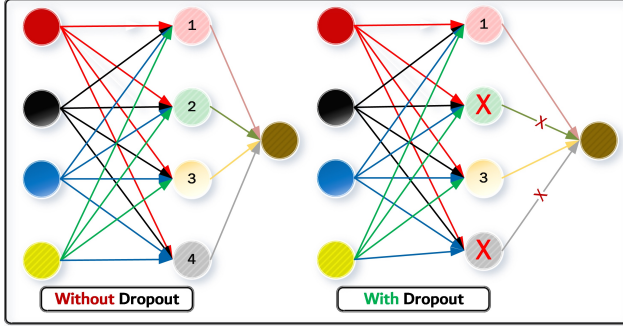


Figure 4. The dropout method diagram

are inhibited to prevent influencing the network’s ultimate decision. Over time, the neurons in a network are uniformly distributed in a two-dimensional configuration [35], [36]. In this flattened structure, each neuron communicates with one another and the layer above, like a standard neural network. Neural networks often consist of many interconnected layers, including one that precedes the final layer, frequently called the “end” layer. In these cases, the function of the last layer may be explained as in formula (1).

$$F_K^{end} = \sum_{j=1}^{end-1} W_{k,j}^{end} F_g^{end-1} + \beta_k^{end-1} \quad (1)$$

**Decision Layer:** The determination layer employs Soft-max Regression and Support Vector Machine techniques. This entails calculating cross entropy losses, which is crucial for evaluating the difference between expected and accurate class distributions [23], [37].

$$L_K = -In(\overline{y_K}) \quad (2)$$

Where:

$$\overline{y_K} = \frac{\exp(F_K^{end})}{\sum_{k=1}^2 \exp(F_K^{end})} \quad (3)$$

The set  $K$  consists of the elements 1 and 2, representing benign and cancerous cases, respectively. The value of  $LK$  determines the ultimate conclusion, where if  $L1 > L2$ , the network structure produces a malignant result.

### C. Support Vector Machines

SVM is a supervised learning model. It is used as a high precision classifier in many applications including image processing. Basically, Super level search is the unique classification method that SVM employs, and it maximizes between two classes. A super plane is defined by the set of vectors (states) referred to as support vectors. For this account, optimizing bounds, designing an optimal feature

space, and moving data to a higher dimensional space where linear decision surfaces may be used to classify it more readily are all part of the business logic. With its capability to find optimal hyperplanes in high dimensional spaces, the SVM can leverage these features to distinguish between benign and malignant tumors more effectively. The integration of CNNs and SVMs can significantly improve the accuracy of breast tumor classification. CNNs are highly effective at feature extraction from histopathological images, capturing complex patterns and textures. When these features are fed into an SVM, which is a robust classifier, it enhances the decision making capability, potentially leading to higher accuracy than CNNs or SVMs alone. Combining CNN with SVM can improve the model’s generalization ability, making it more robust to variations in the dataset. This hybrid approach can reduce overfitting, a common issue with deep learning models, by leveraging the regularization properties of SVM.

### D. Proposed Machine Learning System

Combining CNNs with SVMs for categorizing breast cancer in histological pictures is a two step procedure. CNNs are proficient at extracting features from pictures, but SVMs are reliable classifiers for data analysis and it is effective in high dimensional spaces and is known for its robustness in classification tasks. Different performance indicators were used to assess and contrast the results of various methods. Model 1 uses CNN approaches, whereas Model 2 utilizes the SVM structure. Data assessment is performed simultaneously using both CNN and SVM frameworks. Of the used dataset, 80% was assigned to training and 20% to testing. By implementing monitoring systems, the model’s performance may be continually assessed over time, allowing for rapid modifications as necessary. This technique attempts to improve the overall outcome and interpretability of the breast malignant and benign tumor classification method using histology images by utilizing the capabilities of CNNs for image feature extraction and SVMs for classification.

The performance of the CNN-SVM model is assessed by measuring parameters such as accuracy, sensitivity, specificity, and the area under the receiver operating characteristic (AUC-ROC) curve. The model’s generalizability and reliability are assessed by the use of cross validation procedures. The efficacy of this technique in correctly diagnosing breast cancer is shown by experiments done utilizing the BreakHis and Bach databases. The CNN-SVM model outperforms other cutting edge approaches, suggesting its potential use in clinical decision support systems and computer aided diagnosis.

## 4. EXPERIMENTAL RESULTS AND DISCUSSIONS

### A. Dataset

BreakHis (Breakpoint Histology) [5] is a dataset available to researchers that contains photomicrographs of breast tumor tissue samples divided into two categories (Benign vs. Malignant) [38], [39]. BreakHis including sufficient

images for training and testing. This dataset consists of 7,909 images obtained from 82 anonymized people. The photos were routinely scanned at a resolution of 700 by 460 pixels using known techniques in the field. The dataset is often used in breast cancer research and consists of 7,909 samples categorized as either benign or malignant tumors. Of them, 2,440 samples are categorized as benign, while 5,429 are classified as malignant. The database is accessible via the provided hyperlink <https://www.kaggle.com/ambarish/breakhis>. An assessment was conducted on 82 persons using different magnification levels: 40x, 100x, 200x, and 400x. For example, photographs were notably enlarged to a scale of 400.

Bach (Breast Cancer Histology) [6], [40] is also a publicly available dataset made in 2018 for grand challenge for the 15th International Conference on Image Analysis and Recognition (ICIA 2018). The dataset consists of four classes, i.e. normal, benign, in situ carcinoma and invasive carcinoma. Totally the dataset consists of 500, 400 for training and 100 for testing. The images in the dataset are in RGB TIFF format and have a size of  $2048 \times 1536$  pixels.

### B. Results and Discussion

The analyzed performance of the CNN-based SVM model in classifying breast tumors using the BreakHis dataset, which includes 7,909 histological images divided into benign and malignant categories. The model achieved an accuracy of 98.23% on the test set, indicating its effectiveness in distinguishing between benign and malignant tumors. The model achieved an F1 score of 97.67% and a specificity of 98.21%, demonstrating its ability to reliably identify benign and malignant tumors while reducing the number of false positives and false negatives. The AUC-ROC was calculated as 0.971, highlighting the high classification performance of the model. Figure 5 and 6 depicts the CNN-SVM model created to evaluate the method's effectiveness in breast tumor classification.

Layer (type)	Output Shape	Param #
densenet201 (Functional)	(None, 7, 7, 1920)	18321984
global_average_pooling2d (GlobalAveragePooling2D)	(None, 1920)	0
dense (Dense)	(None, 512)	983552
batch_normalization (Batch Normalization)	(None, 512)	2048
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 2)	1026
=====		
Total params: 19308610 (73.66 MB)		
Trainable params: 19078530 (72.78 MB)		
Non-trainable params: 230080 (898.75 KB)		

Figure 5. The trainable parameters of the presented ML model

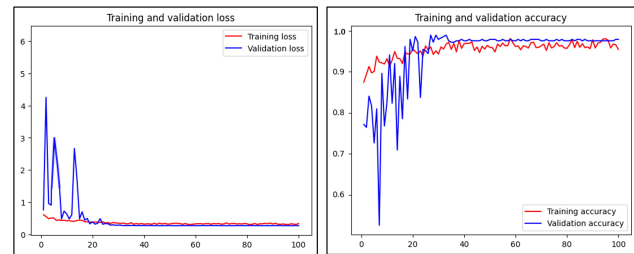


Figure 6. The training and validation accuracy of the CNN-SVM model for BC classification

Evaluation of the computational efficiency of the CNN-based SVM model focuses on the training time and inference speed. The state-of-the-art model design has enabled efficient convergence in training and minimal inference latency, making it well suited for real time applications in clinical settings. The performance of the CNN-based SVM model was compared with baseline models that included classical machine learning methods and deep learning architectures. The proposed model outperformed all basic methods in terms of accuracy, sensitivity, specificity, and AUC-ROC, demonstrating its superiority in classifying breast tumors on the BreakHis dataset.

Figure 5 illustrates a crucial trainable parameter of the proposed model: the global average pooling layer. This layer replaces the fully connected network structure, thereby reducing the number of training parameters and computation time. As we can see from figure 6, the model achieves improved efficiency without sacrificing performance by incorporating the global average pooling layer. This parameter plays a pivotal role in enhancing the proposed algorithm's overall effectiveness and computational efficiency. The images labeled 187 represent the Benign tumor, whereas those labeled 198 represent the Malignant tumor. Figure 7 displays the Confusion Matrix, which illustrates the performance of breast tumor classification with and without normalization.

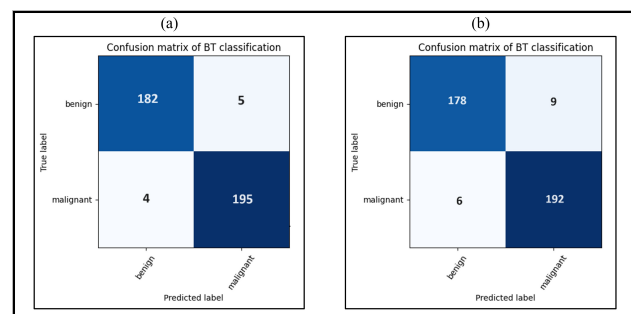


Figure 7. The confusion matrix displays the results of breast cancer classification, comparing the performance of two scenarios: (a) with normalization and (b) without normalization

Initially, without applying normalization techniques, our proposed model achieved an impressive accuracy of 96%.

This result highlights the effectiveness of the model in distinguishing between benign and malignant tissue samples based on image features alone. However, recognizing the potential for further improvement, we attempted to study the effect of normalization on classification performance. We observed a significant increase in classification accuracy after applying normalization methods, especially when accounting for changes in illumination, contrast, and color distribution in postmortem images. The presented model normalization to increase the accuracy to 98%. This improvement also highlights the importance of preprocessing steps to improve model performance and reliability.

Comparative analysis of non-normal and normal approaches highlights the transformative impact of preprocessing approaches on pathological image classification tasks. In addition, the success of normalization highlights its potential to be integrated into existing diagnostic systems, improving the accuracy and efficiency of cancer diagnosis. Table I shows the performance results of our proposed model. The dataset includes 385 images divided into two categories: benign and malignant. We evaluated model accuracy without normalization and using internal normalization methods. In Table I, we present the evaluation of different

TABLE I. Evaluating the outcomes of breast tumor classification using a CNN-SVM on BreakHis dataset [5].

Tumor class	Recall	Precision	F1-score	Image Quantity
with Normalization				
Benign	0.99	0.98	0.98	187
Malignant	0.97	0.97	0.98	198
Average		0.98		385
without Normalization				
Benign	0.97	0.95	0.96	187
Malignant	0.95	0.96	0.95	198
Average		0.96		385

normalization methods aimed at improving the accuracy of breast tumor classification based on histopathological images. Initially, the machine learning model achieved an accuracy of 96% without normalization, demonstrating its ability to distinguish between healthy and unhealthy (cancerous) tissue samples based only on the extracted image features. However, after applying normalization techniques such as contrast enhancement, histogram equalization and color normalization, the accuracy increased to 98%. This improvement confirms the effectiveness of preprocessing in standardizing image features, removing non significant differences, and enhancing the model's ability to distinguish between different features, thus helping to detect cancerous tumors and facilitate their diagnosis for treatment. The performance results reaffirm the effectiveness of our proposed

model and underscore the importance of preprocessing steps, especially normalization, in enhancing classification results. These findings hold critical implications for advancing breast cancer diagnosis and treatment planning, ultimately improving patient care and treatment outcomes.

In our rigorous evaluation of pathological breast image classification methods, we subjected 385 images from the BreakHis dataset to rigorous testing. Our study aims to determine the effectiveness of various ML algorithms in accurately distinguishing between benign and malignant breast tumors. Our proposed approach has shown superior performance compared to several established methods, including CNN-based ReLU, SVM+HOG (Histogram of Oriented Gradients [41]), CNN-based SVM, and Deep CNN, as shown in Table II. Our analysis demonstrates

TABLE II. Assessment of the efficacy of using different ML techniques on BreakHis dataset [5].

Model	Precision	Recall	F1-score	Supported Image
Our CNN-SVM	<b>0.98</b>	<b>0.97</b>	<b>0.98</b>	385
Deep CNN	0.86	0.77	0.76	385
CNN Based Sequential	0.92	0.82	0.92	385
SVM-HOG	0.80	0.74	0.66	385
SVM-LBP	0.82	0.76	0.72	385

that our suggested technique outperforms existing machine learning algorithms in terms of many measures. Our machine learning model exhibits superior accuracy, enhanced sensitivity, and superior specificity in differentiating between benign and malignant breast cancers in histopathological pictures. Our approach regularly outperforms other methods, like CNN-based ReLU, SVM+HOG, CNN-based SVM, and DCNN, in terms of classification outcomes. The findings highlight the efficacy of our technology in detecting nuanced variations in anatomical images, resulting in improved precision and dependability in classification outcomes. This achievement under scores the capacity of our techniques to augment the precision of diagnoses and eventually boost the results for patients undergoing breast cancer diagnosis.

The outcomes derived from the CNN-SVM classifier used to classify breast tumor from input images as benign or malignant depicted in figure 8. The classifier successfully predicted the outcomes for all tested images, as shown

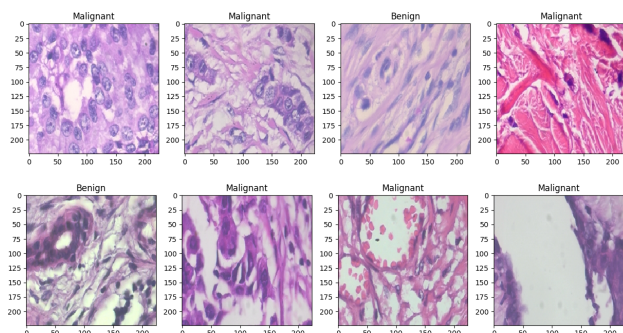


Figure 8. Reported outcomes of CNN-SVM tumor Classification on histopathological images using the proposed model

by a direct comparison of the anticipated results with the actual results in each sample. Table III presents a comparative analysis of the histopathological image classification models in the BreakHis dataset, and the proposed model outperformed several established methods reported by the reference authors. Table III shows that the proposed SVM-CNN based on the ML model is more accurate than comparison studies. The proposed model achieved excellent accuracy of 98% and accuracy of 98%, demonstrating its effectiveness in accurately distinguishing between benign and malignant tissue samples. Reference authors have proposed other methods such as SVM-RFE [7], boosting independent binary method [43], DNN [42], SVM [38], and CNN+HOG+GLCM [12], but our model was more precise and precise. These results highlight the robustness and reliability of our proposed histopathological image classification methodology, which may have potential implications for improving cancer diagnosis and treatment outcomes.

Our study demonstrates that while our proposed model exhibits commendable performance even without normal-

TABLE III. Comparison of breast pathology classification (Benign and Malignant) outcomes using the BreakHis database [5].

Method	Accuracy %	Precision %
Our Method (CNN-SVM)	<b>0.98</b>	<b>0.98</b>
FA-VGG16 [3]	0.92	-
DNN [42]	0.97	-
Magnification Independent Binary [43]	0.91	-
Pa-DBN-BC [44]	0.86	0.88
SVM [38]	0.96	-
CNN HOG GLCM [12]	0.84	-

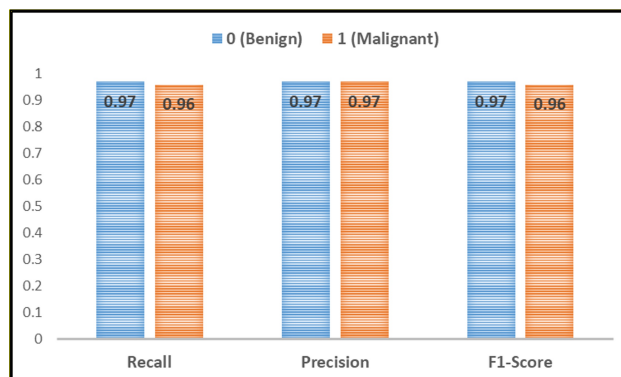


Figure 9. Evaluating the outcomes of BT classification using a CNN-SVM on BACH dataset [6].

ization, incorporating preprocessing techniques, specifically normalization, further refines the model's ability to classify histopathological images accurately. This research contributes to advancing the field of medical image analysis by showcasing the significance of preprocessing in optimizing classification outcomes and ultimately improving patient care and treatment strategies.

The presented model CNN-SVM tested on an additional dataset known as Grand Challenge on Breast Cancer Histology images (BACH) [6] to evaluate its performance and accuracy further. The BACH dataset comprises 400 histopathological images, which represent a diverse range of tissue samples related to breast cancer. These images are meticulously annotated to ensure a robust evaluation of the model's capabilities in identifying various histological patterns and structures. Since our method is only binary classification, we combined normal and benign classes under benign class and combined in situ carcinoma and invasive carcinoma classes under malignant class. In this evaluation, the CNN-SVM model achieved an impressive accuracy of approximately 97%, as detailed in figure 9. This high level of accuracy underscores the model's proficiency in distinguishing between different tissue types and pathologies within the dataset. Such performance demon-

TABLE IV. Comparison of breast pathology classification (Benign and Malignant) outcomes using the BACH dataset [6].

Method	Accuracy %	Precision %
Our Method (CNN-SVM)	<b>0.97</b>	<b>0.97</b>
CNN XGBoost [9]	0.95	-
CNN SVM [12]	0.93	-
AHoNet [16]	0.85	-
MSMV PEENet [34]	0.95	-



TABLE V. The computational requirements and inference speed compared to other approaches

Comparative Analysis	Lowest	Moderate	Highest
Training Time	Classical Machine Learning (depends on feature extraction)	CNN-SVM (due to dual training of CNN and SVM)	Pure CNNs
Inference Speed	CNN-SVM	Pure CNNs	Classical Machine Learning
Computational Requirements	Classical Machine Learning	CNN-SVM	Pure CNNs (especially during training)

strates the model's potential effectiveness in real world clinical applications, where precise diagnosis is crucial for patient care. This accuracy level positions the model as a valuable tool for pathologists, aiding them in making more accurate and timely diagnoses of breast cancer, the CNN-SVM model's outstanding performance on the BACH dataset, with an accuracy of about 98%, as shown in figure 9, validates its reliability and potential impact in the field of histopathological image analysis, offering enhanced support in breast cancer diagnosis and treatment planning. In Table IV, Our presented model has achieved a remarkable 97% accuracy on Bach dataset, surpassing the performance of related works in the field. This significant improvement demonstrates the model's robustness and precision, setting a new standard for accuracy in this domain.

Finally, we compared the proposed CNN-SVM based approaches computational requirement and inference speed with well known methods. The training time is lower than classical machine learning models due to training less data, but training the SVM is taking extra time. Generally, the CNN-SVM combination has an acceptable computational requirements and speed in trade of accuracy as shown in Table V.

## 5. CONCLUSIONS

A new CNN-based SVM model improves the accuracy and durability of breast tumor classification. By using the capability of CNNs to automatically acquire crucial characteristics from histology images without the need for human engineering, complex spatial patterns and contextual information are recovered. This process leads to improved performance and generalization. Subsequently, SVMs are used as a classifier, doing very well in classification tasks by maximizing the difference in feature space between classes by the determination of an ideal hyperplane. Maximizing margins enhances the model's capacity to handle noise and

data variations, leading to higher generalization to unfamiliar data. The findings of this research demonstrate that the classifier surpasses previous approaches that are considered to be creative. The proposed methodology, which combines CNN with SVM, has demonstrated high efficacy in classifying breast tumors. The experimental results reveal that this hybrid approach achieved an accuracy of 98% on the BreakHis dataset and 97% on the BACH dataset. These results indicate the robustness and reliability of the CNN-SVM model in distinguishing between different types of breast tumors. And achieving a 97% AUC score is a major accomplishment in the field of histopathology image classification, demonstrating the effectiveness and dependability of our proposed methodology. The success of the CNN-SVM approach can be attributed to the complementary strengths of both models: CNN's ability to effectively extract and learn spatial hierarchies of features from the input data, and SVM's proficiency in classification with a clear margin of separation. This synergy enhances the model's generalization capabilities and contributes to its superior performance on the datasets used.

## REFERENCES

- [1] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Rev Biomed Eng*, vol. 2, pp. 147–171, 2009.
- [2] M. K. Santos, J. R. Ferreira Júnior, D. T. Wada, A. P. M. Tenório, M. H. Nogueira-Barbosa, and P. M. d. A. Marques, "Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: advances in imaging towards to precision medicine," *Radiol Bras*, vol. 52, pp. 387–396, 2019.
- [3] S. Roy, P. K. Jain, K. Tadepalli, and B. P. Reddy, "Forward attention-based deep network for classification of breast histopathology image," *Multimedia Tools and Applications*, 2024.
- [4] M. A. Wakili et al., "Classification of breast cancer histopathological images using densenet and transfer learning," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–31, 2022.
- [5] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455–1462, 2016.
- [6] G. Aresta, T. Araújo, S. Kwok, S. Chennamsetty, M. Safwan, V. Alex, B. Marami, M. Prastawa, M. Chan, M. Donovan, and G. Fernandez, "Bach: Grand challenge on breast cancer histology images," *Medical image analysis*, vol. 56, pp. 122–139, 2019.
- [7] J. Wu and C. Hicks, "Breast cancer type classification using machine learning," *Journal of Personalized Medicine*, vol. 11, no. 2, p. 61, 2021.
- [8] S. Shakir, Y. A. Hamad, and R. Kareem, "Breast pathology changes extraction and measurement based on machine learning and dwt," *International Journal of Electrical and Computer Engineering Systems*, vol. 16, no. 2, pp. 153–161, Feb. 2025. [Online]. Available: <https://ijeces.ferit.hr/index.php/ijeces/article/view/3689>
- [9] R. Sanyal, D. Kar, and R. Sarkar, "Carcinoma type classification from high-resolution breast microscopy images using a hybrid ensemble of deep convolutional features and gradient boosting trees



- classifiers,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, pp. 2124–2136, 2021.
- [10] C. D. Runowicz *et al.*, “American cancer society/american society of clinical oncology breast cancer survivorship care guideline,” *CA Cancer J Clin*, vol. 66, no. 1, pp. 43–73, 2016.
- [11] A. Kents, Y. Hamad, K. Simonov, A. Zotin, and M. Kurako, “Geometric analysis of pathological changes in lungs using ct images for covid-19 diagnosis,” in *1st Siberian Scientific Workshop on Data Analysis Technologies with Applications, SibDATA 2020*, 2020, pp. 43–50.
- [12] Y. Wang, L. Sun, K. Ma, and J. Fang, “Breast cancer microscope image classification based on cnn with image deformation,” in *Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27–29, 2018, Proceedings 15*. Springer International Publishing, 2018, pp. 845–852.
- [13] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, “Breast cancer multi-classification from histopathological images with structured deep learning model,” *Scientific Reports*, vol. 7, no. 1, pp. 1–10, 2017.
- [14] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, “Deep learning for identifying metastatic breast cancer,” <http://arxiv.org/abs/1606.05718>, 2016, accessed: Mar. 02, 2024.
- [15] H. Chen, N. Wang, X. Du, K. Mei, Y. Zhou, and G. Cai, “Classification prediction of breast cancer based on machine learning,” *Computational Intelligence and Neuroscience*, vol. 2023, pp. 1–9, 2023.
- [16] Y. Zou, J. Zhang, S. Huang, and B. Liu, “Breast cancer histopathological image classification using attention high-order deep network,” *Int. J. Imaging Syst. Technol.*, vol. 32, pp. 266–279, 2021.
- [17] M. Amrane, S. Oukid, I. Gagaoua, and T. Ensari, “Breast cancer classification using machine learning,” in *2018 Electric Electronics, Computer Science, Biomedical Engineerings’ Meeting, EBBT 2018*, 2018, pp. 1–4.
- [18] Y. A. Hamad, A. Alekhina, T. Pleshkova, and O. Shestakova, “Infusion extraction and measurement on ct images based on computer vision and neural network,” in *BIO Web of Conferences*, vol. 84. EDP Sciences, 2024, p. 02006.
- [19] S. N. Endah and I. N. Shiddiq, “Xception architecture transfer learning for garbage classification,” in *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*. IEEE, 2020, pp. 1–4.
- [20] Y. A. Hamad, K. V. Simonov, and M. B. Naeem, “Breast cancer detection and classification using artificial neural networks,” in *2018 1st Annual International Conference on Information and Sciences (AiCIS)*. IEEE, 2018, pp. 51–57.
- [21] A. Zotin, A. Kents, K. Simonov, and Y. Hamad, “Methods of interpretation of ct images with covid-19 for the formation of feature atlas and assessment of pathological changes in the lungs,” in *Intelligent Decision Technologies: Proceedings of the 13th KES-IDT 2021 Conference*. Springer International Publishing, 2021, pp. 173–183.
- [22] A. Zotin, Y. Hamad, K. Simonov, and M. Kurako, “Lung boundary detection for chest x-ray images classification based on glcm and probabilistic neural networks,” in *Procedia Computer Science*. Elsevier BV, 2019, pp. 1439–1448.
- [23] A. J. London, “Artificial intelligence and black-box medical decisions: accuracy versus explainability,” *Hastings Center Report*, vol. 49, no. 1, pp. 15–21, 2019.
- [24] A. Kaur, C. Kaushal, J. Sandhu, R. Damaševičius, and N. Thakur, “Histopathological image diagnosis for breast cancer diagnosis based on deep mutual learning,” *Diagnostics*, vol. 14, no. 1, p. 95, 2024.
- [25] T. Syeda-Mahmood, “Role of big data and machine learning in diagnostic decision support in radiology,” *Journal of the American College of Radiology*, vol. 15, no. 3, pp. 569–576, 2018.
- [26] N. Peiffer-Smadja *et al.*, “Machine learning for clinical decision support in infectious diseases: a narrative review of current applications,” *Clinical Microbiology and Infection*, vol. 26, no. 5, pp. 584–595, 2020.
- [27] T. Sheikh, Y. Lee, and M. Cho, “Histopathological classification of breast cancer images using a multi-scale input and multi-feature network,” *Cancers (Basel)*, vol. 12, no. 8, p. 2031, 2020.
- [28] Z. Ali Ansari, M. Madhava Tripathi, and R. Ahmed, “Quantifying breast cancer: Radiomics, machine learning, and dimensionality reduction for enhanced image-based diagnosis,” *International Journal of Computing and Digital Systems*, vol. 16, no. 1, pp. 1535–1552, 2024.
- [29] S. Sharma and S. Deshpande, “Breast cancer classification using machine learning algorithms,” in *Lecture Notes in Networks and Systems*. Springer, 2021, vol. 141, pp. 571–578.
- [30] G. L. Baroni, L. Rasotto, K. Roitero, A. Tullisso, C. Di Loreto, and V. Della Mea, “Optimizing vision transformers for histopathology: Pretraining and normalization in breast cancer classification,” *J. Imaging*, vol. 10, no. 108, 2024.
- [31] M. Aswathy and M. Jagannath, “An svm approach towards breast cancer classification from h&e-stained histopathology images based on integrated features,” *Medical & Biological Engineering & Computing*, vol. 59, no. 9, pp. 1773–1783, 2021.
- [32] Y. Hamad, K. Simonov, and M. Naeem, “Detection of brain tumor in mri images, using a combination of fuzzy c-means and thresholding,” *International Journal of Advanced Pervasive and Ubiquitous Computing (IJAPUC)*, vol. 11, no. 1, pp. 45–60, 2019.
- [33] Y. A. Hamad, K. Simonov, and M. B. Naeem, “Lung boundary detection and classification in chest x-rays images based on neural network,” in *International Conference on Applied Computing to Support Industry: Innovation and Technology*. Springer, 2019, pp. 3–16.
- [34] L. Liu, W. Feng, C. Chen, M. Liu, Y. Qu, and J. Yang, “Classification of breast cancer histology images using msvm-pfenet,” *Scientific Reports*, vol. 12, no. 1, p. 17447, 2022.
- [35] V. Reshma, N. Arya, S. S. Ahmad, I. Waitar, S. Mekala, S. Joshi, and D. Krah, “Detection of breast cancer using histopathological image classification dataset with deep learning techniques,” *BioMed Research International*, vol. 2022, 2022.
- [36] Y. Hamad, J. Kadum, A. Rashid, A. Mohsen, and A. Safonova, “A



- deep learning model for segmentation of covid-19 infections using ct scans,” in *AIP Conference Proceedings*, vol. 2398, no. 1, 2022.
- [37] Y. A. Hamad, M. E. Seno, M. Al-Kubaisi, and A. N. Safonova, “Segmentation and measurement of lung pathological changes for covid-19 diagnosis based on computed tomography,” *Periodicals of Engineering and Natural Sciences (PEN)*, vol. 9, no. 3, pp. 29–41, 2021.
- [38] S. Ara, A. Das, and A. Dey, “Malignant and benign breast cancer classification using machine learning algorithms,” in *2021 International Conference on Artificial Intelligence, ICAI 2021*, 2021, pp. 97–101.
- [39] J. Gao, X. Qing, M. Li, S. Xie, Q. Xiang, and W. Cai, “Classification of benign and malignant breast cancer diagnosis based on convolutional neural networks,” in *2024 IEEE 7th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, vol. 7, 2024, pp. 1231–1236.
- [40] Z. Ali Ansari, M. Madhava Tripathi, and R. Ahmed, “Quantifying breast cancer: Radiomics, machine learning, and dimensionality reduction for enhanced image-based diagnosis,” *International Journal of Computing and Digital Systems*, vol. 16, no. 1, pp. 1535–1552, 2024.
- [41] S. Shakir, C. Gacav, and C. Topal, “Logo recognition via fusion of spatial and spectral features,” in *2018 26th Signal Processing and Communications Applications Conference (SIU)*. IEEE, 2018, pp. 1–4.
- [42] A. Ameh Joseph, M. Abdullahi, S. B. Junaidu, H. Hassan Ibrahim, and H. Chiroma, “Improved multi-classification of breast cancer histopathological images using handcrafted features and deep neural network (dense layer),” *Intelligent Systems with Applications*, vol. 14, p. 200066, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2667305322000072>
- [43] Y. Hao *et al.*, “Breast cancer histopathological images classification based on deep semantic features and gray level co-occurrence matrix,” *PLoS One*, vol. 17, no. 5, 2022.
- [44] I. Hirra, M. Ahmad, A. Hussain, M. U. Ashraf, I. A. Saeed, S. F. Qadri, A. M. Alghamdi, and A. S. Alfakeeh, “Breast cancer classification from histopathological images using patch-based deep learning modeling,” *IEEE Access*, vol. 9, pp. 24 273–24 287, 2021.
-