Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/JZFRM

A NOVEL METHOD FOR IMAGE BASED BREAST CANCER CLASSIFICATION

SOBIA YAQOOB
PhD Scholar, Department of Computer Science, Superior University Lahore Pakistan. Lecturer, Department of Computer Science, University of Okara, Okara Pakistan. Email: sobia.yaqoob@uo.edu.pk

ARSLAN AKRAM
PhD Scholar, Department of Computer Science, Superior University Lahore, Pakistan.
Email: aaadoula11@gmail.com

Dr. ARFAN JAFFAR
PhD, Professor Department of Computer Science and Information Technology, Superior University Lahore, Pakistan. Email: arfan.jaffar@superior.edu.pk

Dr. WASEEM IQBAL
PhD, Associate Professor Department of Software Engineering, Superior University Lahore, Pakistan.
Email: Waseem.iqbal@superior.edu.pk

AMNA NADEEM
Lecturer, Department of Computer Science, Lahore Garrison University Lahore, Pakistan.

Email: amnanadeem@lgu.edu.pk

SOHAIB SALEEM
PhD Scholar, School of Cyber Security, HuaZhong University of Science and Technology, China.
Email: sohaibchaudhery@gmail.com

Abstract

Breast common cancer kind of cancer that affects women globally. Around 30% of all new cases of cancer in women are anticipated to be breast cancer by 2022. This life-threatening disease is an incurable disease but controllable. Nevertheless, early diagnosis through routine inspection can boost recovery and survival chances. A computer-aided breast cancer diagnosis can show promising results to automatically classify breast histopathology images. The classification of benign and malignant patients using mammography pictures is suggested in this work utilizing a mix of curvelet transformations and support vector machines. Curvelet transformation fetches the optimal features for breast cancer classification. Images of Mammograms used in this paper for evaluation is provided by two databases i.e., Break His and BisQUE. The model's performance is measured in form of average and standard deviation of accuracy, False Negative Rate, True Positive Rate, and AUC of ROC. Results show the remarkable performance of the model by achieving accuracy of 91.0 %, and 85.9%, for Break His and BisQUE respectively.

Keywords: Breast Cancer Classification, Curvelet Transformation, Machine Learning, Support Vector Machine

1. INTRODUCTION

Breast cancer is among the most important public health problems. It is a frequent cause of death, right alongside heart attacks. By 2022, it is anticipated that breast cancer would account for nearly 30% of all newly diagnosed malignancies in women.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/JZFRM

287,850 new cases of invasive breast cancer are expected to be identified in women in the United States this year, compared to 51,400 new instances of non-invasive breast cancer. According to estimates, 43,250 Americans will lose their lives to breast cancer this year. Moreover, statistics tells that approximately 3.8 million women had breast cancer as of January 2022. Research tells that around 1 in 833 men may suffer from this disease. Invasive breast cancer in men is expected to increase by 2,710 new cases in the United States by 2022. [1].

Mutations or aberrant alterations in the genes that govern cell development and health are the causes of cancer. It might develop into a malignant/cancerous or benign/non-cancerous breast cyst. Histopathology is a discipline of pathology that entails taking a small sample of tissue from a contaminated breast and examining it under a microscope. There are many approaches to diagnosing cancer, including magnetic resonance imaging and fine-needle aspiration cytology. However, histopathology is a standardized procedure for the diagnosis of breast cancer and guarantees its result accuracy when compared to other cancer diagnostic approaches. Investigation through histopathology is a complex and time-consuming road to cancer diagnostic information. However, the strong professional experience of pathologists is crucial for the physical diagnosis of breast cancer which is time consuming, costly and misdiagnosis may also occur [2]. By introducing a faster and more accurate diagnostic approach, CAD “computer-assisted diagnostic” may help to reduce the use of specialists. CAD models working with “Machine learning algorithms” have shown favorable results in examining breast tissues using histology [3] [18]. Authentic dataset or medical records in form of textual or image data is fed into CAD models to get trained. Data is said to be imbalanced if one sample type has a higher no. of instances than the other. In the case of unfair distribution of data records, classification results are weighted toward the class having more instances which leads to misdiagnosis. Several algorithms, such as J48, K-NN, and Naive Bayes, which are some of the extant classifiers in machine learning, are used for classification and prediction. In this work, the categorization outcomes of breast cancer illnesses are created utilising the support vector machine "SVM" approach. Accuracy, precision, and sensitivity are the most common metrics used to evaluate the model's performance [4] [5].

This study presents an ML-based method for classifying the breast cancer images as malignant and benign. A curvelet transformation is used to extract features of interest. Then the support vector machine is utilized to save time and enhance overall precision. This research work provides a low-cost high-precision solution for a variety of time modalities.

The author’s promises to the research are discussed as follows:

- The proposed strategy has successfully produced positive comparable results.
- One of the distinguishing features of this study is that it covers a broad range of disease informational records. The generalizability of the study is demonstrated by the results obtained for three data sets with distinct features. To put it another

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/JZFRM

way, the proposed method is not dependent on the data set. Therefore, a different set of data can also be passed to the model with comparable accuracy.

- Traditional learning procedures can provide as many successful results as deep learning methods, according to this study.
- Attempt to make a decision-support system that can assist medical professionals in making decisions by allowing accurate detection of breast cancer malignancy using Machine Learning and image processing methods.

The following are the four sections of this work. Section 2 relates the associated existing work while Section 3 contains the highlights of the dataset used in this evaluation and recommended framework. Section 4 presents and analyses the findings of the evaluation, and comparisons with previous studies. The final section 5 summarizes the work and provides guidelines for future research.

2. LITERATURE REVIEW

The second-leading cause of fatalities worldwide, particularly among women, is breast cancer. As a result, academic research has drastically grown. Recent studies have used CAD frameworks extensively, especially in this field for rapid and precise diagnosis. Deep learning and machine learning are showing up more and more in instructional design. This section describes the research that has already been done.

In [6], s. Sandhu khan et al. proposed a model based on expected features of the nucleus of cells. This study compares the accuracy of two distinct algorithms, “K means nearest neighbor” (KNN) and “Support Vector Machine” (SVM). Using image processing, the nuclei of the cells are then determined by examining a digital image of a “Fine Needle Aspirate” (FNN) of breast tissue. The trained model is then used to determine whether the tumour is classified as benign or malignant by using lowest number of feature values. KNN outperforms with a 97.4893% accuracy rating compared to the SVM model's 97.273%. A World Cup Optimization technique is presented in [7] and is used as the optimization method over thresholding. In contrast to Kapur's approach, the proposed technique considers as potential solutions random examples from the solution space within the picture histogram. The "Mammography Image Analysis Society" (MIAS) database was sampled and selected using the specified methodology, which involved systematic sampling. The given model's accuracy outperforms ICA and PSO-based algorithms.

The dataset is subjected to a dimensionality reduction technique in order to categorize the breast cancer classifications [8]. In this study, combined elastogram and B-mode pictures were categorized using the supervised learning technique SVM. Model validation is carried out using K-fold cross-validation for generalizability. The effectiveness, confusion matrix, and logistic loss of the employed method are evaluated. Classification accuracy for SVM using the "radial basis function" (RBF) kernel is 94.12%. Researchers in [9] provide a brand-new feature extraction method based on

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/JZFRM

"local binary pattern" characteristics. This method produces twenty-five sliding windows from each image. Each window's features are recorded and utilized to create a SVM “Support Vector Machine” classifier. Based on the most common windows classes, each image is classified as benign or malignant using the SVM classifier. From a full histology image, the approach can be utilized to locate malignant tissues. The proposed approach has a 91.12 percent accuracy, a 94.01 percent specificity, and an 85.22 percent sensitivity. [10] Proposes a hybrid machine learning methodology to handle the problem of class imbalance. Breast cancer histopathology images were obtained using the BreakHis and BisQue datasets. The pre-trained ResNet50 was given the non-overlapping patches created from the photos. Patches were processed using ResNet50 to extract their features, which were then sent to the kernelized weighted extreme learning machine (KWELM) for classification. The suggested method did a fantastic job of classifying both the majority class of malignant pictures and the minority class of benign images. However, the suggested model did not consider stain normalization or feature selection when determining quality features.

In [11], the researchers used a new collection of histopathology images of canine mammary tumors to test their hypothesis. Using the "CMTHis" and "BreakHis," datasets, a VGGNet-16-based framework evaluated the fused framework’s performance with several classifiers. This study also looks at the effects of stain normalization, data augmentation, and magnification on the performance of the suggested technique. The suggested structure achieved the accuracy of 97 percent on CMT dataset and 93 percent of human breast cancer for binary classification with support vector machine. The s2sp classifier, presented in [12], is a one-of-a-kind approach for identifying tumor based on 22 features. To work with the MIAS data base, they developed a back propagation neural network. A classifier based on genetic algorithms. The approach supports the use of a swarm intelligence optimization classifier. Value can be used to encode the classification feature. This method delivers the best categorization by improving sensitivity and receiver operating parameters. The mammography function was used to detect breast cancer. To categorize distinct stages of cancer, the DWT and SWT categorization systems are used. The features for tumor detection were extracted from 58 biopsy monographs. Stepwise linear classification functions are utilized for diagnosis. A fuzzy-based SVM classifier performs well on ultrasound breast images. A system known as a Bayesian network classifier is used to classify data. The Relief and Bayesian networks have joined to build a rapid and easy-to-use cancer-diagnosis solution. Multiple methods with SNR feature selection and other features were utilized to discriminate between malignant and benign breast tumors.

In [13], Author presents an improved method for removing speckle noise for enhanced image enhancement. In this work, a hybrid architecture that combines a cross-task extreme learning machine and a double deep transfer learning is shown. Deep transfer learning and double-step deep transfer learning are used in the initial phase to extract high-level properties. An interactive cross-task extreme learning machine uses the high-level feature sets as regularization terms to further improve classification performance.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/JZFRM

The suggested method was tested on 134 histological breast cancer pictures and proved to be incredibly accurate (96.67 percent, 96.96 percent, and 98.18 percent).

In [14], researchers are attempting to uncover the limitations of digital mammography in detecting breast cancer. To classify the benign and malignant tumors, they used image analysis and machine learning classifiers. They also examined Supervised Machine Learning classifiers like K Nearest Neighbor, Random Forest, Gaussian Naive Bayes, and Decision Tree by researching the biomarkers which are involved in the risks associated with breast cancer. The goal is to present a comprehensive picture of breast cancer prediction utilizing Machine Learning and image and data analytics, which will aid in the prevention of future misdiagnoses.

In [15], two datasets, a collection from Helsinki University Central Hospital and the FinProg collection, were utilized to train and test a machine learning based system for predicting patient outcomes. The samples are categorized based on their digital risk scores, which range from low to high. In univariate survival analysis, the hazard ratio for the DRS classification was 2.10. (95 percent CI 1.33–3.32, p=0.001). In a multivariate Cox model, The DRS classification has been shown to be a reliable predictor of breast cancer survival, with a ratio of 2.04 (95 percent CI 1.20–3.44, p=0.007). Human expert predictions based on comparable TMA samples had a C-index of 0.58 (95 percent confidence interval: 0.53–0.63), whereas DRS grouping had a C-index of 0.60 (95 percent confidence interval: 0.55–0.65). It's been proposed that predictive signals in tumor tissue images can be learned without any prior domain knowledge. For women, breast cancer has become a severe health issue. The gold standard in the clinic is pathological image-based diagnosis. This study [16] proposes a strategy for automatically detecting breast cancer by combining images of disease with hybrid features. A three-output convolutional neural network is employed to identify packed and chromatin-sparse nuclei, resulting in better segmentation results. Because the hematoxylin (H) and eosin (E) channels have a weak correlation, textural characteristics are extracted independently for the two channels, resulting in more representative findings. From diverse angles, morphological, texture and spatial structural features are retrieved and combined. A support vector machine classifier with increased generalization based on the relief technique for feature selection is used to classify an image as benign or malignant. For the University of California, Santa Barbara database (UCSB), the method's classification accuracy is 96.7 percent, and the area under the curve (AUC) is 0. 983.To identify patients as benign or malignant based on mammography images, the authors of this study used a mix of Gray-level co-occurrence matrix and Support Vector Machine [17]. The ideal Gray-level co-occurrence matrix angle for breast cancer categorization instances was discovered using mammography data in this work. The mammography pictures utilized in this study came from the “Curated Breast Imaging Subset of Digital Database Screening Mammography” (CBIS-DDSM) dataset. The accuracy was 63.03 percent, and the specificity was 89.01 percent, according to the trial data.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/JZFRM

3. METHODOLOGY

The proposed framework to classify breast cancer images is shown in Figure 1 and the overall process is described in next subsections.

3.1. Preprocessing

Images are read from directory then block division is implemented using non-overlapping block division. Each image is divided into blocks of size 256 x 256. After this process an image is converted into multiple images of size as shown in Figure 1.

3.2. Feature Extraction

Each block of image is passed to curvelet transform to get texture features of an image those will be used for differentiation and learning for benign and malignant images.

3.2.1. Fast Discrete Curvelet Transform

Since it takes several terms to precisely reconstruct a discontinuity, the Fourier transform is not recommended for many image processing applications, according to the literature. As a response to the Fourier transform conundrum, the wavelet transform has gained popularity. Researchers are interested in it because of its localization and multi-scale characteristics. This transform works well in 2D, i.e., it can handle point singularities, but because of limited directional selectivity and isotropic scaling, it cannot accurately describe higher dimensional singularities like lines, curves, and so on. The ability of the curvelet transform to overcome the difficulties presented by different traditional wavelet transformations has led to its increased adoption. Curvelet is recognized as a multi-scale tool of geometric analysis that efficiently accomplishes curve singularities in an image. Curvelet transforms have better directional selectivity, multiresolution, localization, and anisotropy as its key characteristics. Furthermore, due to its parabolic scaling capabilities, it allows for a sparse representation of objects with curve singularities that is practically ideal. To create curve singularities, a picture is first split into smaller pieces, and all of the resulting sub-images are then subjected to the ridgelet transform. The Curvelet transform was the name of this block-based transformation. However, the first-generation curvelets' use in a variety of applications has been hampered by the ridgelet' imprecise geometry. Then, in [5,] a second generation curvelet transform was offered as a solution to the difficulties with first generation curvelets. In place of wavelets and its derivatives for capturing more directional data, discrete curvelet transform is chosen to employ because the tumor shape has numerous curves and lines in brain MRI imaging. The essential mathematical basics of the continuous and discrete curvelet transforms are given below. For a given signal, the curvelet transform can be defined as an inner product as follows:

$C(j, l, k) = \langle f, \varphi j,l,k \rangle$Where, the curvelet basis function is $\varphi j, l, k$ and parameters are j, k, and l, which are scale, position, and direction are respectively.The continuous curvelet transform is performed in two dimensions (R2) and can be expressed in the frequency domain using the variables x as the spatial and frequency domain variables.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/JZFRM

In the frequency domain, the polar coordinates are r and. W(r) and V (t), commonly known as radial and angular windows, with r ($\frac{1}{2}$, 2) and t [1, 1], respectively, are two smooth, non-negative, real-valued windows. W and V will always meet the following conditions.

$$\sum_{j=-\infty}^{\infty} W^2\left(2^j r\right) = 1, \qquad r \in \left(\frac{3}{4}, \frac{3}{2}\right)$$

$$\sum_{l=-\infty}^{\infty} V^2(t-1) = 1, \qquad t \in \left(\frac{-1}{2}, \frac{1}{2}\right)$$

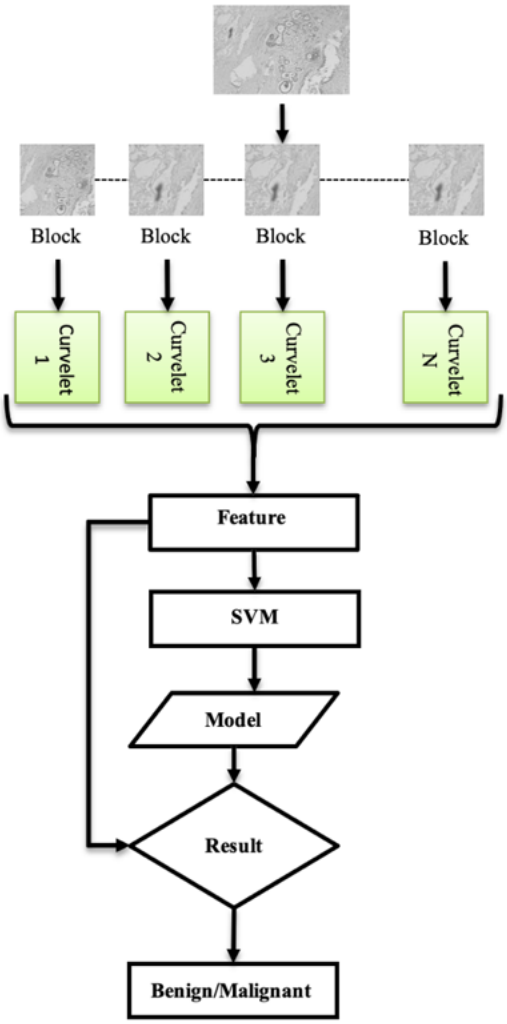

Figure 1: Proposed Framework

In the Fourier domain, the frequency window Uj is then given by for each j ≥ j0.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/JZFRM

$$U_j(r,\theta) = 2^{\frac{-3j}{4}} W\left(2^{-j}r\right) V\left(\frac{2^{\left\lfloor \frac{j}{2} \right\rfloor}\theta}{2\pi}\right)$$

Were, $\frac{j}{2}$ represents the integer portion of j. As a result, the support of Uj is a polar wedge that is inferredby the sponsorship of W and V, which are applied with scale-dependent window widths in each direction. To generate real valued curvelets, consider the symmetric variant of Uj s, i.e. Uj (r + θ) + Uj (r, $\theta + \pi$) and the waveform $\varphi j(x)$ using its Fourier transform$\varphi j(\omega) = Uj(\omega)$, where Uj (ω1, ω2) is the polar coordinate system window. Keeping that all curvelets at size 2-j are formed by rotations and translations of φj. it can be considered the mother curvelet. Curvelets can be generated at scale 2-j, orientationθl, and position xj,l using the function x = (x1, x2) by

$\varphi_{j,l,k}(x) = \varphi_j\left(R_{\theta_l}\left(x - x_k^{(j,l)}\right)\right)$Where, $k = (k_1, k_2) \in Z^2$ specifies the order in which the translation parameters are applied, $\theta_l = 2\pi . 2^{-\left\lfloor \frac{j}{2} \right\rfloor} . l, l = 0, 1, \ldots$ such that $0 \leq \theta_l < 2\pi$, and $x_k^{(j,l)} = R_{\theta_l}^{-1}\left(k_1 . 2^{-j}, k_2 . 2^{\frac{-j}{2}}\right)$. R_θ and R_θ^{-1} denotes the rotation by θ radians and its inverse respectively, and are defined as

$$R_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, R_\theta^{-1} = R_\theta^T = R_{-\theta}$$

The inner product of an element f and a curvelet is then the curvelet coefficients $\varphi_{j,l,k}$ i.e.,

$$C(j,l,k) = \langle f, \varphi_{j,l,k} \rangle = \int_{R^2} f(x)\overline{\varphi_{j,l,k}}\, dx$$

We can express the inner product as an integral over the frequency plane by expressing it as an integral over the frequency plane.

$$C_{j,l,k} = \frac{1}{(2\pi)^2} \int \hat{f}(\omega)\ \overline{\hat{\varphi}_{j,l,k}(\omega)}\, d\omega = \frac{1}{(2\pi)^2} \int \hat{f}(\omega)\, U_j\left(R_{\theta_l}\omega\right) e^{i\langle x_k^{(j,l)}, \omega \rangle} d\omega$$

The linear digital curvelet transform of an input Cartesian array $f(t_1, t_2); 0 \leq t_1, t_2 < n$ is defined by a set of coefficients as

$$C^D(j,l,k) = \sum_{0 \leq t_1, t_2 < n} f[t_1, t_2]\ \overline{\varphi_{j,l,k}^D[t_1, t_2]}$$

3.3. Classification

Our focus of work is feature extraction than design of the best classifier, we depend on the other existing works for the selection of the classification algorithm. The classification block's goal is to classify chest histopathological pictures based on the residual encoder block's extracted features. As the backbone of our classification

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/JZFRM

system, we use the Support Vector Machine, which has a powerful feature representation ability. With its excellent resilience and classification ability, the SVM has shown to be a powerful tool in machine learning and data mining. In practice, the most important task is to detect cancer cases with great sensitivity and efficiency. As a result, the labels were divided into two categories: cancer infection cases and others. We focused on more practical work, such as identifying confirmed cases so that appropriate measures could be taken as soon as possible to prevent the epidemic from spreading. The pre-processed image was passed to feature extraction using Curvelet transformation. These features were being used for classification process using non-linear SVM kernel.

4. RESULTS AND ANALYSIS

Proposed method was tested and authenticated using different training and testing environments. Details of these settings are as under in this section.

4.1. Experimental Environment

To create a simulated environment to verify method discussed earlier we have used MATLAB R2021a with Intel Core I7 based processor having 8 GB DDR4 ram.

4.2. Data Processing

Different publicly published datasets used for this experiment. Images were from different datasets, so they are of different dimensions and types. Details of datasets used to evaluate proposed method is described as following two sections. Two types of datasets are being used for this experimentation. BreakHis include images at different magnification levels to better describe cancer traces. Table 1 shows details of datasets used for this experiment.

Table 1: Dataset Details

Database	Magnification	Benign	Malignant
Break His	40	625	1370
	100	644	1437
	200	623	1390
	400	588	1232
BisQUE		32	26

4.3. Evaluation Policy

Nonlinear SVM classifier is employed for classification. We employ a grid-search technique to identify the SVM kernel's ideal parameter values. We employ 10-fold cross validation to evaluate the classifier's performance. The most common measurements used in detection are the True Positive Rate (TPR), False Negative Rate (FNR), and Area Under Curve (AUC) of the Receiver Operating Characteristic Curve (ROC). We considered these measurements to be the best way to compare our work to other

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/JZFRM

methods already in use. Accuracy is expressed as a ratio of samples that were correctly classified to all samples that were classified.

$$\mathrm{ACC} = \frac{100\,(\mathrm{TP} + \mathrm{TN})}{\mathrm{TP} + \mathrm{TN} + \mathrm{FN} + \mathrm{FP}}$$

True positive rate deals with the percent ratio of correctly classified samples and is calculated using

$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$False positive rate deals with the percent ratio of incorrectly classified samples and is calculated using

$$\mathrm{FPR} = \frac{\mathrm{TP}}{\mathrm{TN} + \mathrm{FN}}$$

Here ‘TP’, ‘TN’, ‘FP’ and ‘FN’ are respectively amount of true positive values, true negatives values, false positives values and false negative values described in previous section. As it is noticed that dataset available for this experiment is unbalanced so, we need to add more evaluation parameter. So, precision is to be calculated using following formula.

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

4.4. Training Results of Proposed Method

Multiple experiments were conducted to validate the proposed solution of this classification problem. The results of the proposed model are presented with respect to following aspects:

1. Differentiatian of the histopathology images into benign and malignant images.
2. The evaluation of the proposed model’s performance on the publicly available datasets. First experiment was done on BreakHis dataset by applying different magnification levels of 40, 100, 200, and 400 to trace the cancer more accurately. Second experiment was done with BisQUE dataset. And the third experiment was executed by combining BreakHis and BisQUE datasets.
3. Comparison of the results with other state-of-the-art methods

Proposed method outperformed on different available datasets. Results of training process according to evaluation policy mentioned in previous section are presented in Table 2. In the first experiment, magnification level of 40x, 100x, 200x and 400x were applied to BreakHis images and accuracies of 91.0 %, 87.7 % accuracy 88.8% and 88.7% respectively for each level of magnification. Keeping magnification rate at 40x, Area under curve plotted in Figure 2 (a) shows the classification accuracy for Benign class as 95% and Figure 2 (b) shows the classification accuracy for Malignant class as 95%. Area under curve plotted in Figure 3, 4, 5 shows the classification accuracy for

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/JZFRM

Benign class and Malignant class at 100x, 200x and 400x respectively. The results exhibited that the proposed model achieved admirable identification rates using magnification at level 40 with best overall accuracy results i.e., 91%. The second experiment was executed with BisQUE dataset. Area under curve plotted in Figure 6 (a) shows the classification accuracy for Benign class as 93% and Figure 6 (b) shows the classification accuracy for malignant class as 93%. Table 2 also shows the values of True positive Rate, False Negative Rate and Positive Predicted values for both classes for each dataset. Moreover, excellent results are achieved when both datasets were combined with 88.8% accuracy.

Table 2: Training Results

Dataset	Magnification	ACC	TPR	FNR	PPV
BreakHis	40	91.0	85.8	14.2	85.5
			93.4	6.6	93.5
	100	87.7	80.1	19.9	80.1
			91.1	8.9	91.1
	200	88.8	80.9	19.1	82.5
			92.3	7.7	91.5
	400	88.7	83.7	16.3	81.7
			91.1	8.9	92.1
BisQUE		85.9	86.8	13.2	89.2
			84.6	15.4	81.5
Combined		88.8	81.8	18.2	82.3
			92.0	8.0	91.7

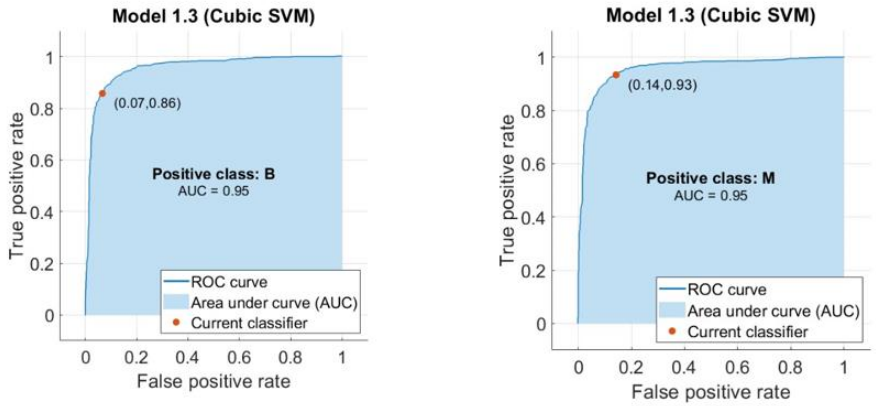

Figure 2: Area under Curve (BreakHis 40x) Dataset using SVM (Left) Benign (Right) Malignant

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/JZFRM

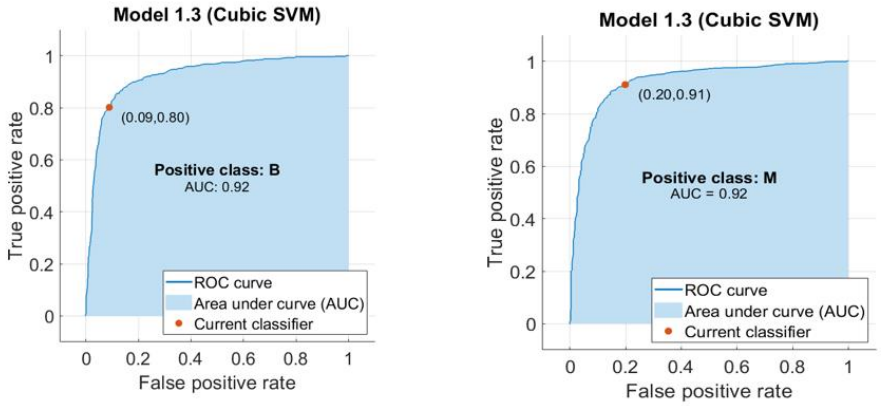

Figure 3: Area under Curve (BreakHis 100x) Dataset using SVM (Left) Benign (Right) Malignant

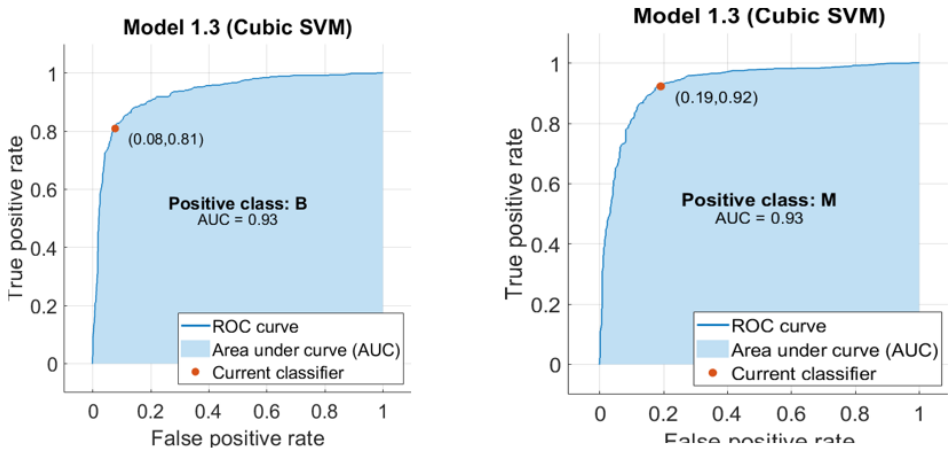

Figure 4: Area under Curve (BreakHis 200x) Dataset using SVM (a) Benign (b) Malignant

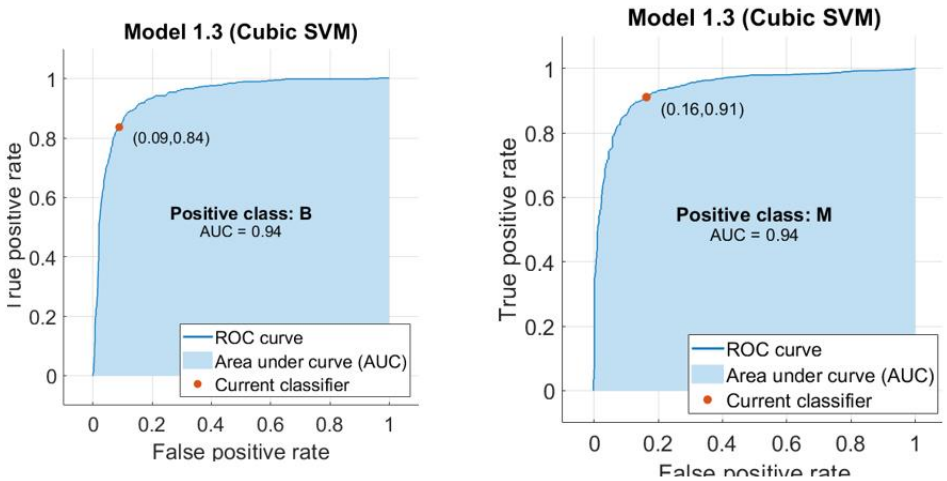

Figure 5: Area under Curve (BreakHis 400x) Dataset using SVM (a) Benign (b) Malignant

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/JZFRM

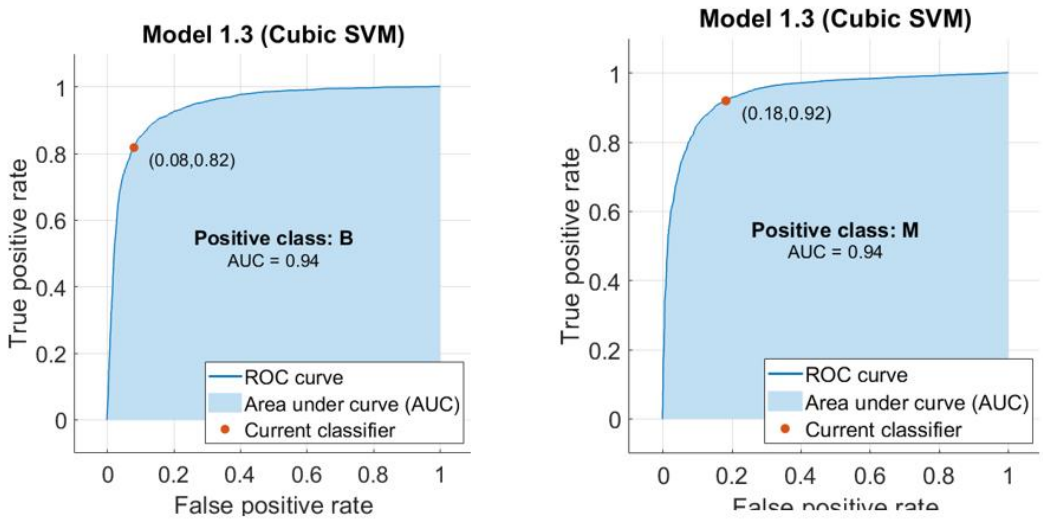

Figure 6: Area under Curve (BreakHis) Dataset using SVM (a) Benign (b) Malignant

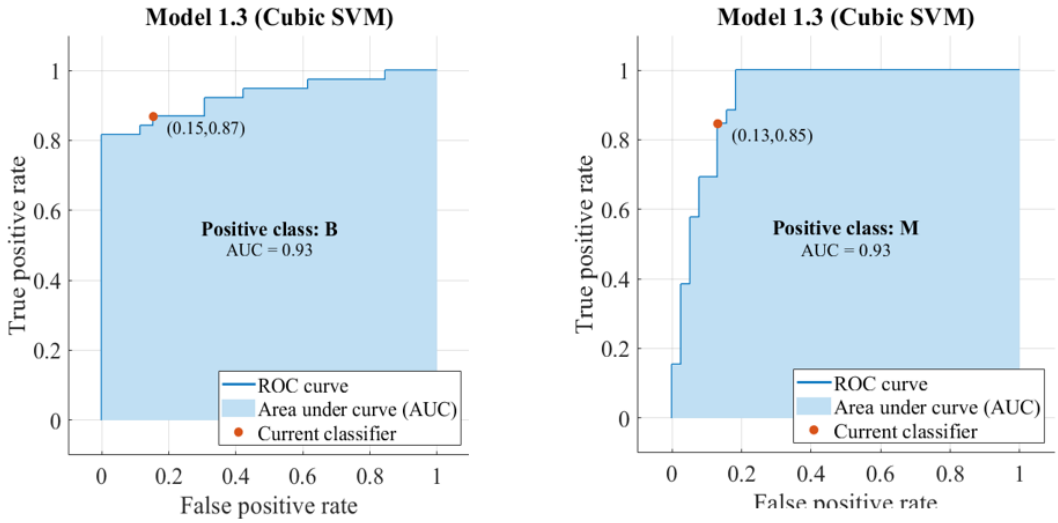

Figure 7: Area under Curve (BisQUE) Dataset using SVM (a) Benign (b) Malignant

COMPARATIVE ANALYSIS WITH STATE OF THE ART METHODS

Some of recent machine learning methods were selected for comparative evaluation of proposed method. Table 3 shows efficiency of proposed method against classical machine learning and deep learning methods using BreakHis dataset.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/JZFRM

Table 3: Comparative Analysis

Ref	Method	ACC	TPR	AUC
[10]	Resnet + ELM	88.35 %	--	--
[11]	VGGNet-16+SVM/Random Forest	86.40 %	--	--
Proposed	Curvelet + SVM	88.80 %	92.	0.94

CONCLUSION

One of the most life-threatening diseases is breast cancer which causes large number of death cases among women worldwide. A new method based on mammography processing and a newly provided tool were employed in this study due to the importance of early and timely detection of this disease. In this research, a mammography image is used to categorize malignant and benign patients using a combination of curvelet modification and SVM. The best characteristics for breast cancer classification are obtained via the curvelet transformation. BreakHis and BisQUE are two databases that give images of mammograms that were used in this research for evaluation. The accuracy is measured using the average and standard deviation of accuracy, the False Negative Rate, the True Positive Rate, and the (AUC) of the model (ROC). The model's excellent performance is demonstrated by accuracy levels of 88.8% and 85.9% for BreakHis and BisQUE, respectively. In future, a well-balanced dataset will be used to produce best results. Moreover, convolutional neural network-based model will be demonstrated to achieve better performance using that newly designed dataset.

References

1. “BREAST CANCER FACTS AND STATISTICS.” HTTPS://WWW.BREASTCANCER.ORG/FACTS-STATISTICS (ACCESSED APR. 23, 2022).
2. R. L. SIEGEL, K. D. MILLER, AND A. JEMAL, “CANCER STATISTICS, 2019,” CA. CANCER J. CLIN., VOL. 69, NO. 1, PP. 7–34, 2019.
3. A. R. VAKA, B. SONI, AND S. REDDY, “BREAST CANCER DETECTION BY LEVERAGING MACHINE LEARNING,” ICT EXPRESS, VOL. 6, NO. 4, PP. 320–324, 2020.
4. J. WU AND C. HICKS, “BREAST CANCER TYPE CLASSIFICATION USING MACHINE LEARNING,” J. PERS. MED., VOL. 11, NO. 2, P. 61, 2021.
5. E. H. HOUSSEIN, M. M. EMAM, A. A. ALI, AND P. N. SUGANTHAN, “DEEP AND MACHINE LEARNING TECHNIQUES FOR MEDICAL IMAGING-BASED BREAST CANCER: A COMPREHENSIVE REVIEW,” EXPERT SYST. APPL., VOL. 167, P. 114161, 2021.
6. S. SADHUKHAN, N. UPADHYAY, AND P. CHAKRABORTY, “BREAST CANCER DIAGNOSIS USING IMAGE PROCESSING AND MACHINE LEARNING,” IN EMERGING TECHNOLOGY IN MODELLING AND GRAPHICS, SPRINGER, 2020, PP. 113–127.
7. N. RAZMJOOY, V. V. ESTRELA, AND H. J. LOSCHI, “ENTROPY-BASED BREAST CANCER DETECTION IN DIGITAL MAMMOGRAMS USING WORLD CUP OPTIMIZATION ALGORITHM,” INT. J. SWARM INTELL. RES. IJSIR, VOL. 11, NO. 3, PP. 1–18, 2020.
8. M. ADEL, A. KOTB, O. FARAG, M. S. DARWEESH, AND H. MOSTAFA, “BREAST CANCER

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/JZFRM

DIAGNOSIS USING IMAGE PROCESSING AND MACHINE LEARNING FOR ELASTOGRAPHY IMAGES," IN 2019 8TH INTERNATIONAL CONFERENCE ON MODERN CIRCUITS AND SYSTEMS TECHNOLOGIES (MOCAST), 2019, PP. 1–4.

9. A. ALQUDAH AND A. M. ALQUDAH, "SLIDING WINDOW BASED SUPPORT VECTOR MACHINE SYSTEM FOR CLASSIFICATION OF BREAST CANCER USING HISTOPATHOLOGICAL MICROSCOPIC IMAGES," IETE J. RES., PP. 1–9, 2019.
10. S. SAXENA, S. SHUKLA, AND M. GYANCHANDANI, "BREAST CANCER HISTOPATHOLOGY IMAGE CLASSIFICATION USING KERNELIZED WEIGHTED EXTREME LEARNING MACHINE," INT. J. IMAGING SYST. TECHNOL., VOL. 31, NO. 1, PP. 168–179, 2021.
11. A. KUMAR ET AL., "DEEP FEATURE LEARNING FOR HISTOPATHOLOGICAL IMAGE CLASSIFICATION OF CANINE MAMMARY TUMORS AND HUMAN BREAST CANCER," INF. SCI., VOL. 508, PP. 405–421, 2020.
12. D. V. SOUNDARI, R. PADMAPRIYA, C. THIRUMARISELVI, N. NANTHINI, AND K. PRIYADHARSINI, "DETECTION OF BREAST CANCER USING MACHINE LEARNING SUPPORT VECTOR MACHINE ALGORITHM," J. COMPUT. THEOR. NANOSCI., VOL. 16, NO. 2, PP. 441–444, 2019.
13. P. WANG ET AL., "CROSS-TASK EXTREME LEARNING MACHINE FOR BREAST CANCER IMAGE CLASSIFICATION WITH DEEP CONVOLUTIONAL FEATURES," BIOMED. SIGNAL PROCESS. CONTROL, VOL. 57, P. 101789, 2020.
14. R. RAY, A. A. ABDULLAH, D. K. MALLICK, AND S. R. DASH, "CLASSIFICATION OF BENIGN AND MALIGNANT BREAST CANCER USING SUPERVISED MACHINE LEARNING ALGORITHMS BASED ON IMAGE AND NUMERIC DATASETS," IN JOURNAL OF PHYSICS: CONFERENCE SERIES, 2019, VOL. 1372, NO. 1, P. 012062.
15. R. TURKKI ET AL., "BREAST CANCER OUTCOME PREDICTION WITH TUMOUR TISSUE IMAGES AND MACHINE LEARNING," BREAST CANCER RES. TREAT., VOL. 177, NO. 1, PP. 41–52, 2019.
16. C. YU, H. CHEN, Y. LI, Y. PENG, J. LI, AND F. YANG, "BREAST CANCER CLASSIFICATION IN PATHOLOGICAL IMAGES BASED ON HYBRID FEATURES," MULTIMED. TOOLS APPL., VOL. 78, NO. 15, PP. 21325–21345, 2019.
17. S. J. A. SAROSA, F. UTAMININGRUM, AND F. A. BACHTIAR, "MAMMOGRAM BREAST CANCER CLASSIFICATION USING GRAY-LEVEL CO- OCCURRENCE MATRIX AND SUPPORT VECTOR MACHINE," IN 2018 INTERNATIONAL CONFERENCE ON SUSTAINABLE INFORMATION ENGINEERING AND TECHNOLOGY (SIET), 2018, PP. 54–59.
18. Akram, Arslan, Saba Ramzan, Akhtar Rasool, Arfan Jaffar, Usama Furqan, and Wahab Javed. "Image splicing detection using discriminative robust local binary pattern and support vector machine." World Journal of Engineering (2022).