ISSN Number: 2454-2024(Online)

International Journal of Technical Research & Science

INVESTIGATION OF DIFFERENT MACHINE LEARNING CLASSIFIERS FOR TIMELY FORECAST OF BREAST CANCER USING WISCONSIN DATASET

Mayuree Katara, Akash Saxena
E-Mail id: mayurikatara@gmail.com
Compucom Institute of Technology and Managment, Jaipur, Rajasthan, India

Abstract- The main aims, methods, and results of the study on breast cancer detection using various machine learning classifiers. It seems that the study focused on analyzing the performance of different classifiers such as Logistic Regression, KNN, SVM-LC, SBM-RBF, Gaussian Naïve Bayes, Decision Tree, and Random Forest Classifier on the Wisconsin dataset. The study aimed to measure the accuracy of these classifiers in detecting breast cancer at an early stage. The Wisconsin dataset is a well-known dataset frequently used for breast cancer research and contains relevant features for classification. According to the testing accuracy results you provided, each classifier achieved the following accuracy scores: Logistic Regression=0.9440, K Nearest Neighbor=0.9580, Support Vector Machine (Linear Classifier) =0.9650, Support Vector Machine (RBF Classifier) =0.9650, Gaussian Naïve Bayes=0.9230, Decision Tree=0.9510 and Random Forest Classifier=0.9650. Based on these accuracy outcomes, it can be concluded that the proposed machine learning models, particularly Support Vector Machines (both linear and RBF), as well as K Nearest Neighbor and Random Forest Classifier, performed well in classifying breast cancer using the Wisconsin dataset. Logistic Regression, Decision Tree, and Gaussian Naïve Bayes also achieved reasonably good accuracy scores. The study suggests that the proposed models have the potential to assist medical professionals in accurately classifying breast lesions, which can lead to early detection and better management of breast cancer.
Keywords: Breast cancer, Breast Cancer Wisconsin (BCW) diagnostic dataset, Foggy and random centroid, Logistic Regression, KNN, SBM, Gaussian Naïve Bayes, Decision Tree, Random Forest.

1. INTRODUCTION

The most prevalent cancer in female is breast cancer [1, 2]. Every year, about 2.3 million women are diagnosed with breast cancer, which is more than the number of new cases of lung cancer [3–6]. Cancerous lesions in the breast occur when abnormal cells in the breast combine with healthy tissue to form a lump [7]. Breast lesions are categorized as either malignant or benign based on the cancer stage, which is identified by breast imaging reporting and data system scores [8, 9]. Most doctors recommend surgery as the first suggestion for treating breast cancer and improving the survival rate [10, 11]. Breast cancer screening is typically performed using X-ray mammography, MRI, and ultrasound images [9, 10]. Treatment for breast cancer depends on the cancer stage. Higher pixel intensities in cancer tissues make them easier to be detected than in other breast regions. However, in dense breasts, the pixel intensity of non-cancerous parts is the same as that of cancerous parts; therefore, correct classification is crucial. Conventional computer-aided diagnosis (CAD) techniques employ handcrafted features for the classification of breast masses [8–10]. Recent technological advancements in deep learning (DL) methodologies, which make use of convolutional neural network (CNN) structures, have resulted in sophisticated models. These models are used for general computer vision tasks and require large datasets for training [12, 13] . Standard transfer learning (TL) methods used for extracting features and fine-tuning the weights based on pertained CNNs have reduced the requirement for large datasets and saved the time required for training from scratch [8, 10, 12]. Extensive research has been performed on classification algorithms; however, selecting the most effective algorithm for a particular dataset can be challenging due to concerns such as computational complexity, adherence to minimal local criteria, and overfitting problem when using traditional algorithms [12–14]. Ensemble learning improves the effectiveness of a classifier and overcomes the aforementioned issues. Ensemble classifiers consider the output from multiple classifiers and allow many classifiers to be used in combination to categorize new data to increase prediction accuracy [11, 12, 14]. Various ensemble methods, such as bagging, boosting, and stacking, have been developed.

The proposed system includes two main stages: preprocessing and TL using a pretrained model and an ensemble of pertained models. In preprocessing, images are resized and normalized, and imbalanced datasets are also handled by using preprocessing. Pertained CNN models are used to obtained robust ensemble results. We aimed to develop an effective CAD model for the classifying breast masses into positive and negative classes. The most important goal of this study is to develop an ensemble classifier by stacking of deep models that have been trained previously on large datasets and to enhance the model's overall performance. We used strategies such as balancing imbalanced data, TL, and early stopping.

1.1 Gap in Existing Research

Breast cancer is the utmost extensive category of cancer amongst women. The finding of breast cancer in its premature periods is quiet a substantial tricky thought out the global. To minimize extraordinary hazard of breast cancer, in last few years' radiologist are using variety of CAD systems to speed up the process of breast cancer classification and its localization accuracy. But the problem with CAD system is that most of preprocessing of MRI image or mammogram image is done using handcrafted features extraction, which can result in information loss and thus affects its accuracy. As from above literature review it has been observed that ordinary knowledgeable procedure commonly indications diverse recital on dissimilar datasets, it funds around of the time a particular algorithm shows strong classifier on some dataset but the predefined methods skilled on additional datasets exhausting the identical algorithm may contribute meager consequences.

1.2 Problem Statement's

To Designing an architecture that involves CNN-based feature extraction from the original dataset (Wisconsin Dataset) and performing fusion of multiple classifiers is a common approach to enhance the overall performance in machine learning tasks, including breast cancer classification.

1.3 Objective of Dissertation

The objective of a breast cancer study using machine learning is to leverage computational techniques to improve the accuracy, efficiency, and effectiveness of breast cancer detection, diagnosis, prognosis, and treatment.

2. LITERATURE REVIEW

2.1 Related Work

Rapid increase of the Breast Cancer patents [1] in India and all over the world and what methods/ ways we can use which can support in Primary detection and opinion of these sickness so that the lives of these cancer patients can be saved. In [1] various machine learning classifiers such as SVM, Biclustering, Ada boost Techniques, CNN were used to analysis the breast cancer statistics.

The main objectives in paper [2] is to investigate, study and suggest methods, techniques which can lead to accurate diagnostics of mammography. There are many techniques being used at the present for studying the result of mammography, but they all have some limitations towards their application. These limitations lead the researchers to study and find new, more precise detection methods. Through this research the researchers suggested the best technique for analyzing mammograms.

Main objective of [3] is to discovers a breast CAD technique established on feature fusion with Deep CNN. A CAD scheme established on mammograms allows primary breast cancer discovery, analysis, and action. Though, correctness of the current CAD schemes leftovers inacceptable. Primary discovery of lumps may successfully decrease the death ratio of breast cancer. In [4] Negative screening digital mammography from 115 women who had unilateral breast cancer at least one year later and 460 matched controls were evaluated retrospectively. Texture features were weighted by their position and underlying dense vs fatty tissue composition in different breast areas specified by an anatomically orientated polar grid. The goal of this research was to see if including breast architecture information can help to strengthen the links between mammographic parenchymal texture phenotypes and breast cancer risk.

In [5] To demonstrate that autonomous DL algorithms may quickly skilled to achieve great accurateness on a variety of mammography stages, and that they have a lot of potential for enhancing clinical tools to decrease false positive and false negative showing mammography outcomes. The study reveals that end-to-end deep learning models can be extremely accurate and potentially applicable across a variety of mammography platforms. As accessible training datasets and computer resources grow, deep learning approaches offer huge potential to increase the accuracy of breast cancer diagnosis on screening mammography. Our method could aid future development of improved CAD systems that could be used to help prioritizes the most worrisome instances for radiologist review, or as an automatic second reader following a first independent interpretation. Our end-to-end approach can be used to solve various medical imaging challenges with limited ROI annotations. In [6] mammography, researchers evaluated the viability of a data-driven imaging biomarker based on weakly supervised learning (DIB; an imaging biomarker created from large-scale medical image data using deep learning technology) (DIB-MG). A total of 29,107 digital mammograms from five institutions were included in the study (4,339 cancer cases and 24,768 normal cases).

In [7], Considerable upsurge in job load and density of histo-pathologic cancer diagnosis owing to the beginning of custom-made drugs. So, analytical conventions have to emphasis likewise on effectiveness and correctness. Deep Learning method was introduced in [7] to enhanced the effectiveness and correctness. From the findings it is clearly indicate that the images contain prostate and micro-metastases of breast cancer can find out automatically 30-40%

containing benign and normal. In [8], proposed Empirical Mode Decomposition for masses classification form mammogram images into benign or malignant using Bidimensional Empirical Mode Decomposition. Later BEMD was modified and new method was introduced called Modified Bidimensional Empirical Model Decomposition for classification of mammogram mass.

In [9], Automatic breast cancer multi-classification from histopathological images is a crucial aspect of computer-aided breast cancer diagnosis and prognosis. While the classification of binary classes, benign and malignant, has been studied extensively, multi-classification methods pose greater challenges. This is primarily due to the subtle differences between multiple classes caused by the vast variability in high-resolution image appearances, the high coherency of cancerous cells, and the extensive inhomogeneity of color distribution within histopathological images. The research paper titled "Multi-View Feature Fusion Based Four Views Model for Mammogram Classification Using Convolutional Neural Network for DDSM dataset" [10]focuses on developing a system for mammogram classification using a Convolutional Neural Network (CNN) and a multi-view feature fusion (MVFF) approach. The DDSM dataset, which contains mammogram images, is utilized for the experiments. One of the challenges encountered in the experiments is the issue of overfitting due to the limited and unbalanced number of images for each view in the dataset. To address this problem, the researchers employed data augmentation techniques. By artificially increasing the number of training samples through transformations like rotation, scaling, and flipping, data augmentation helped mitigate overfitting and improved the testing accuracy by a margin of 3% to 5%.

In [11], Focuses on the objective classification of breast tissue images using Quantitative Transmission ultrasound tomography (QT), which is an emerging imaging technique with the potential to provide accurate three-layered image reconstruction of organic tissue. In the context of breast imaging, QT allows for non-ionizing and harmless imaging of the entire breast in vivo. The researchers proposed the primary demonstration of breast tissue image classification using QT imaging. They conducted a systematic analysis of the ability of QT image features to distinguish between different types of normal breast tissue. Three QT features were employed in Support Vector Machines (SVM) classifiers, and the classification of breast tissue into categories such as skin, fat, organs, ducts, or connective tissue achieved an overall accuracy of over 90%. In [12], paper titled "Discriminative Pattern Mining for Breast Cancer Histopathology Image Classification via Fully Convolutional Autoencoder" presents a practical and self-interpretable solution for invasive cancer diagnosis in breast histopathology images. The proposed method aims to identify contrast patterns between normal and malignant images using minimal annotation information in a weakly supervised manner. To achieve this, the researchers utilize a fully convolutional autoencoder. The autoencoder learns to encode the input histopathology images into a low-dimensional latent space and then reconstructs the images from this latent representation. By training the autoencoder on a large dataset of histopathology images, it learns to capture discriminative patterns specific to normal and malignant tissue. The approach is considered weakly supervised because it requires minimal annotation information. The model does not rely on pixel-level annotations or detailed manual segmentation of cancerous regions. Instead, it leverages the inherent contrast patterns between normal and malignant tissue to make predictions.

In [13] The adoption of Digital Breast Tomosynthesis (DBT) for breast cancer screening has been increasing, and there is a potential shift towards using synthesized Digital Mammography (DM) instead of traditional DM in combination with DBT. However, the interpretation process for DBT requires novel techniques to enhance its effectiveness. To address this, researchers collected data from DBT, digitized screen-film mammography, and digital mammography to create a dataset consisting of 4039 distinct regions of interest, including 1797 malignant and 2242 benign cases. The aim was to develop an intelligent visualization tool that could improve the efficiency of reading DBT volumes while enhancing diagnostic accuracy. In this paper [14], the context of locally advanced breast cancer, preoperative neoadjuvant chemotherapy (NAC) is commonly used as a systemic treatment approach. Recently, there has been interest in leveraging deep learning methods to predict early response to NAC in breast cancer patients in the United States. the researchers explored the potential of deep learning for early NAC response prediction using two transfer learning approaches. Transfer learning is a technique where a pre-trained neural network model, trained on a large dataset, is utilized as a starting point for a specific task. n the first approach, a Convolutional Neural Network (CNN) that was pre-trained on the ImageNet dataset, which is a large dataset with a wide range of images, was employed.

In this paper [15] highlights the potential of using deep learning techniques, specifically the densely deep supervision approach, for automated cancer detection in ABUS images. By leveraging the capabilities of convolutional neural networks and optimizing detection sensitivity and specificity, the proposed network offers an effective solution to aid in breast cancer inspection using ABUS. A 3D convolutional network is utilized in ABUS for automated cancer detection. The aim is to accelerate the review process while maintaining high detection sensitivity and minimizing false positives. The proposed approach introduces a densely deep supervision strategy that effectively utilizes multi-layer characteristics to significantly improve detection sensitivity. In [16], objective was to train deep CNNs to obtain a more discriminative representation of breast tissues, specifically in distinguishing malignant instances from benign ones. The introduction of metric learning layers in the CNN architecture, forming a parasitic relationship, improved the performance of the network for breast mass classification. In [17], focused on detecting and classifying lesions in mammography using deep learning techniques. The system was built based on the Faster R-CNN, which is a widely successful object detection framework. The goal was to develop a system that could automatically detect and classify malignant or benign lesions on mammograms without the need for human intervention. The proposed method demonstrated high classification performance on the INbreast database, achieving an Area Under the Curve (AUC) of 0.95. This indicates a strong ability to discriminate between malignant and benign lesions in mammograms. Moreover, in the Digital Mammography Dream Challenge, the approach secured the 2nd position with an AUC of 0.85. The system's performance was evaluated on the INbreast dataset using the receiver operating characteristics (ROC) metric, and it achieved an AUC of 0.95. The 95 percentiles were estimated through 10,000 bootstraps, indicating the robustness and reliability of the reported results.

In [18], findings indicate that utilizing transfer learning with DNN models, particularly GoogLeNet, can effectively contribute to the diagnosis of mammographic breast lesions. The high ACC and AUC values demonstrate the potential of deep learning techniques in improving breast cancer diagnosis using mammography. The analysis was conducted on a histologically verified database consisting of 406 lesions, with 230 of them being benign and 176 malignant. Several models were used in the study, including transferred DNNs such as GoogLeNet and AlexNet, as well as shallow CNNs like CNN2 and CNN3. These models were fully trained using medical instances and further enhanced with Support Vector Machine (SVM). The results showed that the performance of GoogLeNet was the highest among the models, with an accuracy (ACC) of 0.81 and an area under the curve (AUC) of 0.88. AlexNet also performed well with ACC=0.79 and AUC=0.83.

In recent years, breast cancer [19] has become unique prevalent reasons of death between women. The Globan project Report suggested that breast cancer is common in India. Though researchers have made progressive approach in understanding the fundamental biology of cancer leading to new preventive measures but still the existing system of various hybrid models require better accuracy and computational time to increase the performance of the models and help the radiographer for better diagnosis as early detection of such diseases helps the survivors and the patient to be cured and recover the worth of their lifetime, through lesser sufferings and risks. In study [20], an method that exploits DL methods with convolutional layers to remove the greatest valuable graphic structures for breast cancer cataloguing. As, it is mostly essential to analyses the removal of tissue called biopsy which may sound as carry but importantly it should be remain entirely pain free and low risk procedures. In [21], authors studies about only knowledge methods for breast cancer diagnosis beside with collaborative culture and law removal approaches. The proposed improved Random Forest based rule extraction method is being deliberated in terms of analysis correctness and interpretability.

In [22], Overall, the study demonstrates the effectiveness of the proposed DL approach with SVM for automated mammogram breast cancer detection. The combination of deep learning techniques, K-means clustering, and MSVM leads to improved accuracy and performance compared to traditional approaches. The validation results further support the robustness of the proposed method in breast cancer detection. novel method was introduced that incorporates Speed-Up Robust Features (SURF) selection based on pre-processing and intrinsic feature extraction using K-means clustering. The proposed approach was evaluated using the Mini-MIAS dataset, which consists of 322 images. At the classification level, a new layer is introduced, performing 70% training and 30% testing of the deep neural network and multi-class SVM (MSVM). Results demonstrate that the suggested automated DL technique, utilizing K-means clustering with MSVM, outperforms a decision tree model in terms of accuracy. The average accuracy (ACC) rates for normal, benign, and malignant cancers using the suggested method are 95%, 94%, and 98%, respectively. When compared to the Multi-Layer Perception (MLP) and J48+K-means clustering WEKA manual approaches, SVM shows a 3% increase in sensitivity and 2% increase in specificity. In [24], the study highlighted the challenges faced in developing an automatic breast cancer diagnosis system using CNNs based on the available BreakHis dataset. It further investigated the potential of alternative Deep Learning models like Deep Belief Networks to potentially enhance the accuracy of breast cancer diagnosis. In Paper [25], The mentioned work addresses the therapeutic limitations in the diagnosis and detection of biomarkers in early-stage breast cancer, which pose a barrier to effective treatment strategies. To overcome these limitations, the study proposes the use of deep learning-based Radiomics in breast cancer diagnosis using different imaging modalities, namely ultrasound, mammography, and MRI. The work highlights the developments in deep learning-based CAD systems for breast cancer across ultrasound, mammography, and MRI modalities. It explores the potential of deep learning frameworks to extract meaningful features and patterns from the imaging data, enabling more accurate and efficient diagnosis. In Paper [26], introduces a novel approach for breast cancer image classification using deep transfer learning and interactive cross-task ELM. The proposed method shows promising results in improving breast cancer diagnosis by leveraging high-level features and an interactive classification approach. The findings suggest that cancer cell overgrowth images can play a significant role in guiding anti-cancer treatment strategies. In work introduces a breast cancer image classification method that combines double deep transfer learning and interactive cross-task Extreme Learning Machine (ELM).

ISSN Number: 2454-2024(Online)
International Journal of Technical Research & Science
The authors propose a model called D2TL ICELM, which involves fine-tuning a pre-trained ImageNet Break His model on breast cancer histopathology images and extracting high-level features from the fully connected layers of the transfer learning and double-step transfer learning models.
The paper aims to [27], address the classification of benign and malignant mammogram images using deep learning methods. The proposed DenseNet-II neural network model holds promise for improving the accuracy of breast cancer diagnosis by leveraging advanced techniques in digital image processing and artificial intelligence. In [28], By leveraging deep learning methods and image analysis techniques, the study aims to enhance the recognition and classification of breast cancer, particularly in its early stages. The utilization of advanced neural network models and image processing approaches holds promise for improving breast cancer diagnosis and contributing to more effective treatment strategies. In above mentioned article focuses on the application of deep learning methods for breast cancer diagnosis using image classification. Cancer has become a significant threat to people's lives worldwide, with a notable increase in the number of cancer patients over the past few decades. Breast cancer, in particular, has seen a rise in both urban and rural areas, and even younger women in the age group of 20-30 are now at risk. In article[29], aims to leverage the power of deep learning and image analysis to improve the detection and prediction of invasive disease in cases of DCIS, which can ultimately have significant clinical implications for treatment planning and patient outcomes. The above article focuses on utilizing deep learning-based algorithms for the analysis of breast Magnetic Resonance Imaging (MRI) in order to predict the presence of occult invasive disease in cases of ductal carcinoma in situ (DCIS) following a core needle biopsy (CNB). A core needle biopsy is a common diagnostic procedure used to diagnose breast cancer. However, due to the limited size, number, and location of the samples obtained during the biopsy, there is a possibility of under sampling, which can result in the missed detection of invasive disease. The objective of this study is to investigate whether deep learning algorithms applied to breast MR images can assist in predicting the occurrence of occult invasive disease in patients initially diagnosed with DCIS through a core needle biopsy.

Table-2.1 Year wise list of Published Papers

Year	No. of Publications
2016	7
2017	1
2018	7
2019	21
2020	16
2021	1

Year wise publication details which is used while conducting this study.

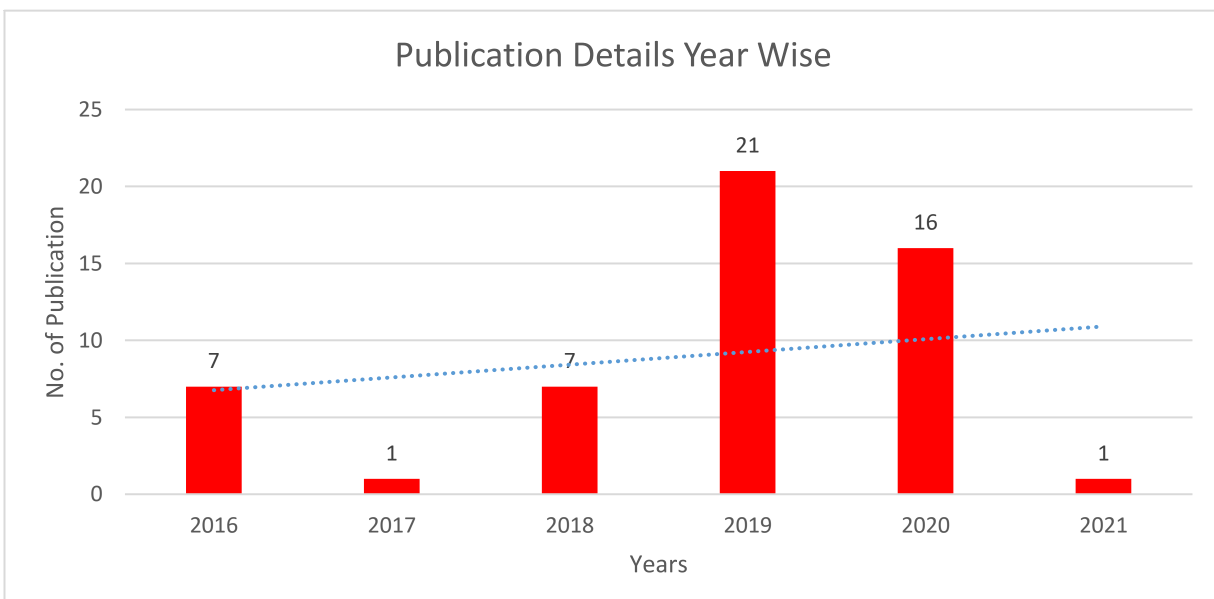

Fig. 2.1 Publication details year wise included in this Study

Table-2.2 No. of Article included in study Publisher Wise

Publisher	No. of Articles
Elsevier	29
IEEE	11
Springer	6
Others(BMC, Nature & Research Gate)	3

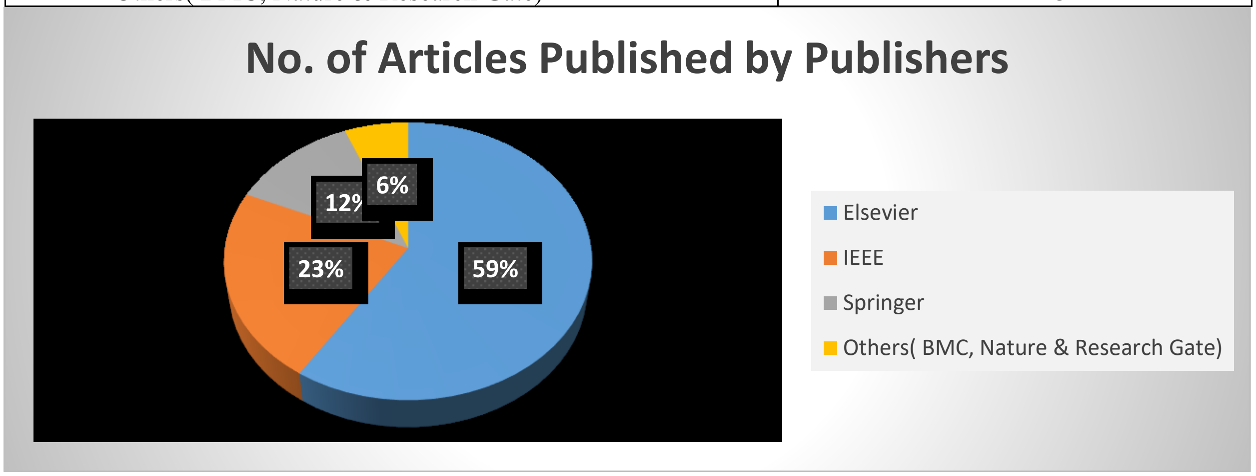

Fig. 2.2 List of Papers include in study publisher Wise

Table2.2 Comparative study of existing Algorithms

Publication Year	Publisher Name	Data Set Used	Data Set Source Name	Method Used in Paper	Performance measure in %
2020 [1]	Elsevier	8009 images over 683 Patients	MGCH& RI Visakhapatnam, India.	DNN	97%
2019 [2]	BMC	83 Research Articles were Reviewed	MIAS, Digital Database, BCDR, IRMA.	Survey	NA
2019 [3]	IEEE	400 with 200 Malignant & 200 benign mass	Pertained FCN-AlexNet	Mass Discovery	NA
2018 [4]	Springer Nature	460 images from 160 females	Selenia Dimensions (Hologic, Inc., Bedford, MA, USA) units.	LIBRA	AUC values from 0.62 to 0.78
2019 [5]	Springer Nature	FFDM images	INbreast database	Deep Learning, CAD software	96.1%
2018 [6]	Springer Nature	29,107 digital mammograms with 4339 diseased and 24,768 usual cases	Collected from Five different organizations.	An imaging Biomarker Derived with Deep Learning Technology	In case of Cancer cases-: 27.2–29.2%, Normal Case: 7.6–8.6%.
2016 [7]	Elsevier	Image Dataset (Microscopic & Histopathology); Digital whole Slide Images	Olympus VS120-S5 system CNN & Prostate Specific Antigen Testing	Deep Learning Method	99.9% sensitivity (ROC analysis
2019 [8]	Elsevier	MIAS Database; DDSM Database	SVM, Linear Discriminate Analysis .	Empirical Mode Decomposition Bimensional Intrinsic Mode Functions (BIMFs)	The classification accuracy is 88.8%, 96.2%

ISSN Number: 2454-2024(Online)

International Journal of Technical Research & Science

2017 [9]	Elsevier	BreaKHis; Augmentation of breast cancer histopathological images.	Mammogram Images - Magnetic Reasonance, Mammogram Images - Computed Tomography (CT) and Histopathological Images.	CSDCNN Architecture with Input Layer, Convolutional layer; Pooling Layer.	Performance, average 93.2% accuracy
2019 [10]	IEEE Access	(CBIS-DDSM), Image Datasets,	CBIS-DDSM: VGG16, VGG19, InceptionV3, ResNet50 MVFF, CBIS- and MIAS	Mammographic Images, Convolutional Neural Network (CNN)	NA
2016 [11]	Research Gate	Image Datasets, CAD , QT images	Image Datasets,	PZT array, BI-RADS, sequential floating forward selection (SFFS), SVM	On four tissue 50-fold cross-validation done for p-value <0.05.
2019 [12]	IEEE Access	Experimental, Epidemiological, CNN, Data Augmentation, Transfer learning.	SSIM and MSE, SVM, 2-D T-SNE,	Rectied Linear (Relu), MSE, H&E stained and image patch	Statistical significance at the 5% significance level.
2019 [13]	IEEE	SFM,DM,DBT	Single-stage CNN	Multi Stage Transfer Learning Approach	0.005
2021 [14]	IEEE	BREAST MASS DATASET	Kaggle	Siamese CNN model	0.847
2020 [15]	IEEE	ABUS	PyPI	Automated breast ultrasound	96%
2018 [16]	Elsevier	CNN's, metric learning layers	NA	Parasitic metric learning net	NA
2018 [17]	Nature	DDSM dataset breast dataset	CAD Method Created on Faster-R-CNN	CAD Method Created on Faster-R-CNN	90%
2019 [18]	Springer	ImageNet	DNNs, GoogLeNet,	DNNs, GoogLeNet,	0.81, 0.79
2020 [19]	Elsevier	Image Datasets	MIAS &DDSM	LWT PCA + Classifier MFO-ELM	100% (For Male) (BenignVs. Malignant) =100%
2019 [20]	Elsevier	Image Datasets (Bioimaging& BreaKHisdataset), Convutionalmethod.	Breast Cancer Database (BCDB), CBI, PDB and PubMed.	Gradient Boosting algorithms,DCNN	For Normal97.8%; Benign is 100%, Insitu 88.9%, Invasive 98.9%, Non-carcinoma 98.9%.
2019 [21]	Elsevier	Image Datasets	Epidemiological,CNN, Data Augmentation, Transfer learning	SVM, K-NN, ANN,	WDBC= 5.8227,WOBC= 8.7108, SEER=.2.1699
2019 [22]	Elsevier	Mini- MIAS	Centre Processing Unit	automated DL approach	93.8%
2019 [23]	Elsevier	MAMMOSET	Department of Radiology at University of Vienna	MARRow (Medical Active leaRning and Retrieval)	87.3%
2019 [24]	Elsevier	Break-HIS	CAD systems	Hematoxylin and Eosin (H&E)	70%
2020 [30]	Elsevier	MIAS, DDSM.	MIAS,DDSM.	K- fold stratified, SCV method.	97.49% 92.61%

ISSN Number: 2454-2024(Online)

International Journal of Technical Research & Science

2020 [26]	Elsevier	D2TL-ICELM	acquisition and categorizationof cell images	D2TL, ICELM	94.55% and 96.36%
2020 [25]	Elsevier	Differentdata sets with different methods	Compilationof many	Compilationof many	Compilationof many
2019 [27]	Elsevier	2042	First Hospital of Shanxi Medical University	10- fold cross validation method	94.55%
2019 [28]	Elsevier	143	Case pathologies	Pixel gray scale or gradient	
2019 [29]	Elsevier	131	-	10- fold cross validation method	95%
2020 [31]	IEEE	IRMA, RetinaNet	Kaggle	Gabor filters method	99.4%
2018 [32]	IEEE	RCC, RMLO, LCC, LMLO	GitHub	whole image classification of both 2D mammogram and 3D tomosynthesis images	NA
2019 [33]	Elsevier (Neurocomputing)	CBIS-DDSM BCDR INbreast MIAS	BCDR – MA Guevara & Co-authors, BCDR consortium INbreast - Breast research group, INESC Porto Portugal	Multi label image classification	NA
2020 [34]	Elsevier (Computer Methods & Programmes in Biomedicine)	DDSM INbreast	DDSM – University of South Florida	Data balancing & Augmentation, Detection of breast lesion, Classification of breast lesion	NA
2020 [35]	Elsevier	Women mammograms from 11 hospitals under Spanish Breast cancer screening network	CC and MLO views for each women Analysis of Mammograms by 2 experienced radiologists using DMscan.	Histogram normalization	NA
2018 [36]	Elsevier	breast cancer patients and control subjects	MIAS	CNNI-BCC method	89.47%, 90.50%, 0.901±0.0314 and 90.71%
2020 [37]	Elsevier	*Identify Variables *Demographic History *Screening History *Current Health *Pathologic Variables *Laboratory Result *Excisional Biospy *Radiologic History Data *First Assesment	mammogram classification system using Matlab software.	*Unsupervised Anisotropic-Feature Transformation Method*	96-100%
2019 [38]	Elsevier	1,912 training dataset and 311 test dataset	FFDM	DL Technique	Annotated samples<20%

ISSN Number: 2454-2024(Online)

International Journal of Technical Research & Science

2018 [39]	Elsevier	breast cancer patients and control subjects	MIAS	CNNI-BCC method	89.47%, 90.50%, 0.901±0.0314 and 90.71%
2016 [40]	Elsevier	Modified SVM and ELM	Mini-MIAS; DDSM; IRMA; Image feature extraction – Morphological Spectrum; Zernike Movements; Neural Networks.	Segmentation Method	69.26% and 68.95%, respectively.
2016 [41]	Elsevier	MIAS Database; Mammography Image; ACS; BancoWeb Database	Artificial Neural Network.	Rough-set approach.	MIAS DATA BASE – ACS: 96%; MIAS – ROUGH SET REDUCTION PROCESS: 92%
2016 [42]	Springer	Micro-calcification features	Scientific Reports (*Sci Rep*)	Participation Population, Imaging and Analysis, Deep Learning Model	78%
2019 [43]	Springer	Representative histological	npj Breast Cancer (*npj Breast Cancer*)	Study population, Breast Biopsy Specimens, Assessment of breast desnsity, Analytical population, Statistical analysis	13.8% were aggressive carcinoma
2020 [44]	Springer	Distribution of slice thickness	Scientific Reports (*Sci Rep*)	Study Participants, multiparametric methods, Inter-sequence image registration, statistical analysis	-
2020 [45]	IEEE	BreaKHis Database	Laboratório Visão Robótica e Imagem	DMAE	200X Magnification
2020 [46]	IEEE	DDSM	DDSM	DL Methods	96.84%
2019 [47]	ELSEVIER	Wisconsin breast cancer, DDSM, MIAS, SEER, BI-RADS,NCBI, REIS		ensemble classification	47%

2.2 Conclusion of the Literature Review

Provide a comprehensive summary and analysis of existing research studies and findings related to breast cancer. These papers aim to synthesize the available evidence, identify trends, gaps, and areas for further research, and provide insights into the current understanding of breast cancer. The outcomes of a breast cancer literature review will depend on the research question, the methodology used, and the current state of knowledge in the field. The aim of a literature review is to provide a valuable synthesis of existing evidence and contribute to the advancement of knowledge in breast cancer research.

ISSN Number: 2454-2024(Online)

International Journal of Technical Research & Science

3. PROPOSED WORK

3.1 Introduction

In this paper we analysis various ML procedures know as Logistic Regression, KNN, SVM (Linear Classifier), SBM (RBF Classifier), Gaussian Naïve Bayer, Decision Tree and Random Forest Classifiers for Wisconsin Dataset to detects breast cancer, based on data. Breast Cancer (BC) is a public cancer aimed at women around the globe and early detection of BC can greatly improve prognosis and survival chances by promoting clinical treatment to patients.

3.2 Technology Use

3.2.1 Implementation Setup Tools

- Deep learning frameworks
- Python, OpenCV, Kears, Sklearn
- GPU processor to train deep network.

3.3 Datasets

From literature Review it has been identified that there are different dataset available for Breast cancer prediction such as Wisconsin Dataset which is available at url https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data.

3.4 Methodology

We analyse Wisconsin Dataset using various machine leering classifiers for the grouping of mammogram images. When working with image datasets, pre-processing is an essential step to prepare the images for further analysis or machine learning tasks. Pre-processing techniques help enhance the quality of the images, reduce noise, standardize the format, and extract relevant features for better analysis.
Here are some common pre-processing steps performed on images for Wisconsin Dataset.

3.4.1 Resizing

Images in a dataset may have different resolutions or sizes. Resizing them to a consistent size is often necessary to ensure uniformity and compatibility with the chosen analysis or model. This step can involve increasing or decreasing the image dimensions while maintaining the aspect ratio.

3.4.2 Normalization

Normalizing pixel values is crucial to ensure that the images have consistent intensity ranges. This step typically involves scaling the pixel values to a predefined range, such as [0, 1] or [-1, 1]. Normalization prevents some pixels from dominating others and helps in achieving faster convergence during training.

3.4.3 Grayscale Conversion

In certain cases, converting color images to grayscale might be beneficial, especially when color information is not essential for the task at hand. Grayscale images reduce the dimensionality of the data and can simplify subsequent processing steps.

3.4.4 Noise Reduction

Image datasets can contain various types of noise, such as random noise or artefacts introduced during acquisition or storage. Applying denoising techniques, such as blurring filters or advanced algorithms like median filtering or wavelet denoising, can help reduce the impact of noise and improve image quality.

3.4.5 Contrast Enhancement

Adjusting the contrast of images can improve their visual quality and make the features more distinguishable. Techniques like histogram equalization, contrast stretching, or adaptive contrast enhancement can be used to enhance the image's dynamic range and improve its overall appearance.

3.4.6 Cropping and Region of Interest (ROI) Extraction

In some cases, images may contain irrelevant or redundant areas. Cropping the images to focus on the regions of interest can help remove unnecessary information and reduce computational requirements. Additionally, extracting specific regions or objects from images (ROI extraction) can be performed to isolate and analyse specific features.

3.4.7 Augmentation

Data augmentation is commonly used to artificially increase the diversity and size of the dataset. Techniques like rotation, flipping, scaling, translation, or introducing random noise can help improve the model's generalization ability by exposing it to a broader range of variations.

- Different base learners are employed by using pertained models.

ISSN Number: 2454-2024(Online)

IJTRS International Journal of Technical Research & Science

- Ensemble models are constructed, and meta-learners are employed.
- The performances of individual base learners and proposed models are evaluated.

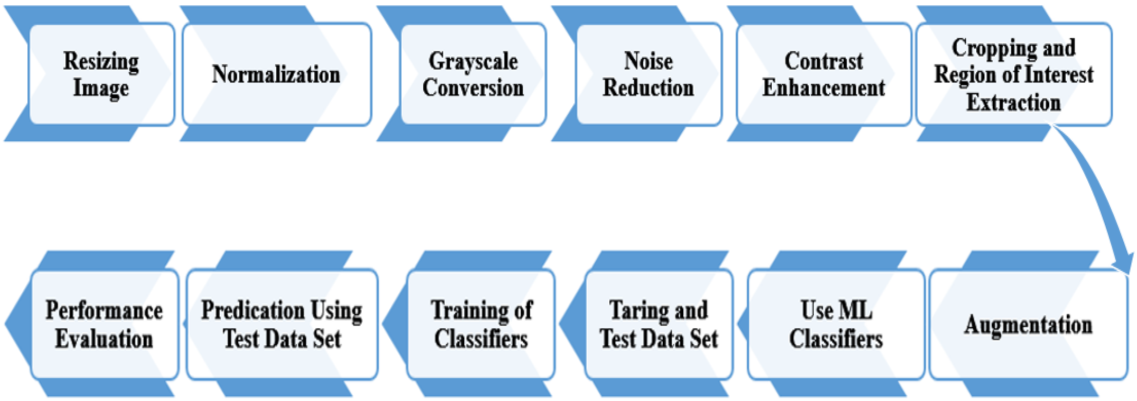

Fig. 3.1 Flow for analyzing machine learning Classifiers

3.5 Methodology

The proposed method comprises the following components:

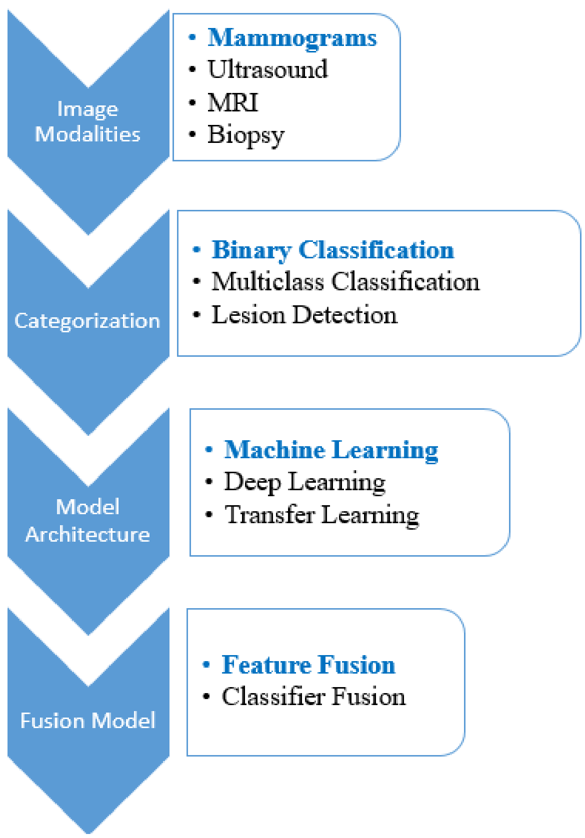

Fig. 3.2 Methods used for Breast Cancer Diagnosis

4. ANALYSIS OF PROPOSED WORK

4.1 Comparative Study Based on Training Accuracy and Testing Accuracy

Table 4.1 shows the training accuracy and testing accuracy for Logistic Regression, KNN, SVM (Linear Classifier), SVM (RBF Classifier), Gaussian Naïve Bayer, Decision Tree Classifier and Random Forest Classifier. From the simulation result it is clear that Random Forest classifier perform better for training and testing accuracy at 0.9953 and 0.965 for Wisconsin Dataset.

ISSN Number: 2454-2024(Online)

International Journal of Technical Research & Science

Table-4.1 Training Accuracy Vs. Testing Accuracy

S. No.	Method	Training Accuracy	Testing Accuracy
1	Logistic Regression	0.9906	0.944
2	K Nearest Neighbor	0.9765	0.958
3	Support Vector Machine (Linear Classifier)	0.9882	0.965
4	Support Vector Machine (RBF Classifier)	0.9835	0.965
5	Gaussain Naïve Bayes	0.9507	0.923
6	Decision Tree Classifier	1	0.951
7	Random Forest Classifier	0.9953	0.965

From Fig. 4.1 it clearly shows that the Decision Tree Classifier has very high training accuracy as compared to other machine learning classifiers.

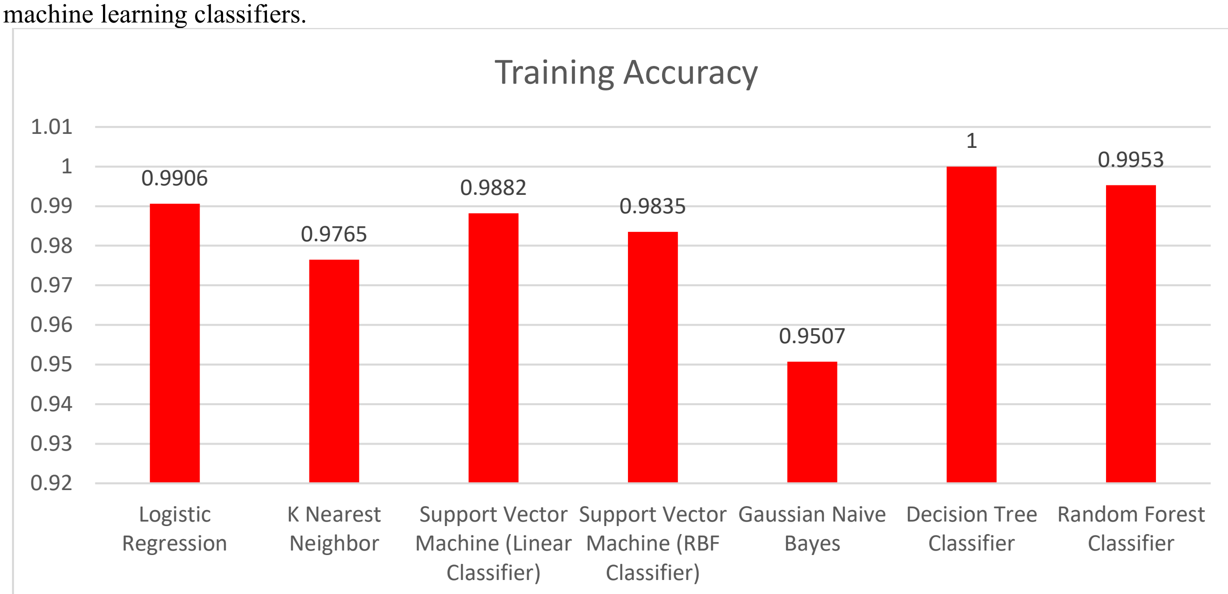

Fig. 4.1 Performance Evaluation for Various Machine Learning classifier for Training Accuracy

But Random Forest, Support Vector Machine (Linear Classifier) and Support Vector Machine (RBF Classifier) perform better in terms of testing accuracy as per the fig. 4.2.

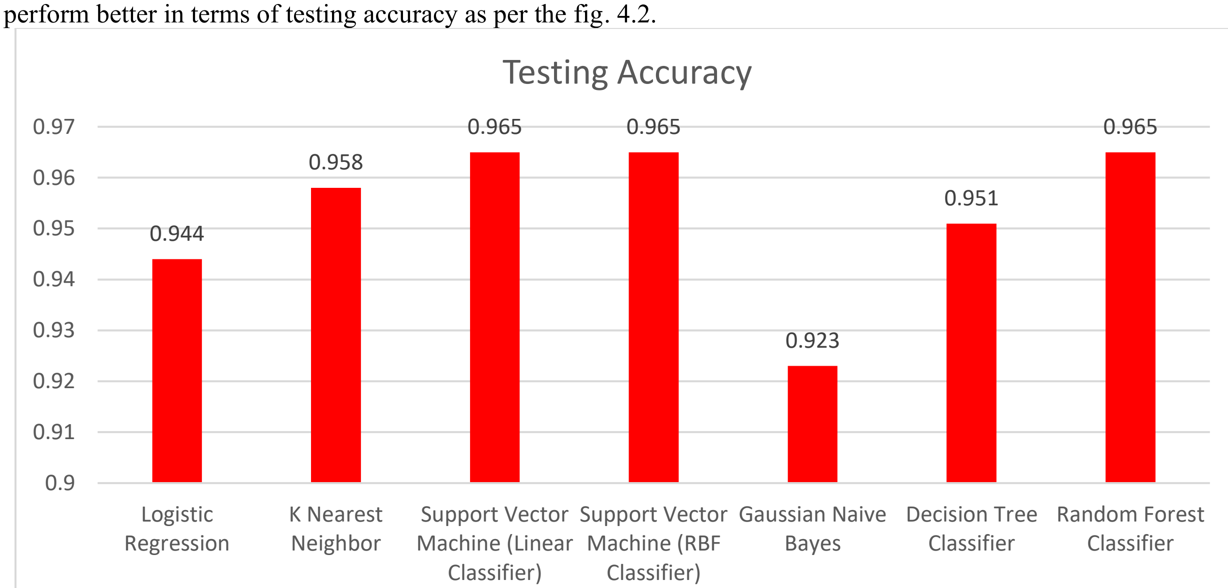

Fig. 4.2 Performance Evaluation for Various Machine Learning classifier for Testing Accuracy

ISSN Number: 2454-2024(Online)
IJTRS International Journal of Technical Research & Science

Precision Comparison for Diagnosis Type [B=0 and B=1]

Table 4.2 and fig. 4.3 shows the comparative analysis of precision for Diagnosis Type [B=0] using various machine learning classifiers.

Table-4.2 Comparative Analysis of precision for Diagnosis Type[B=0]

Method	**Precision**	**Diagnosis Type [B=0]**
Logistic Regression	0.96	0
K Nearest Neighbor	0.95	0
Support Vector Machine (Linear Classifier)	0.98	0
Support Vector Machine (RBF Classifier)	0.97	0
Gaussian Naïve Bayes	0.93	0
Decision Tree Classifier	0.99	0
Random Forest Classifier	0.98	0

Simulation result show that Decision Tree Classifier perform better for the Diagnosis type[B=0].

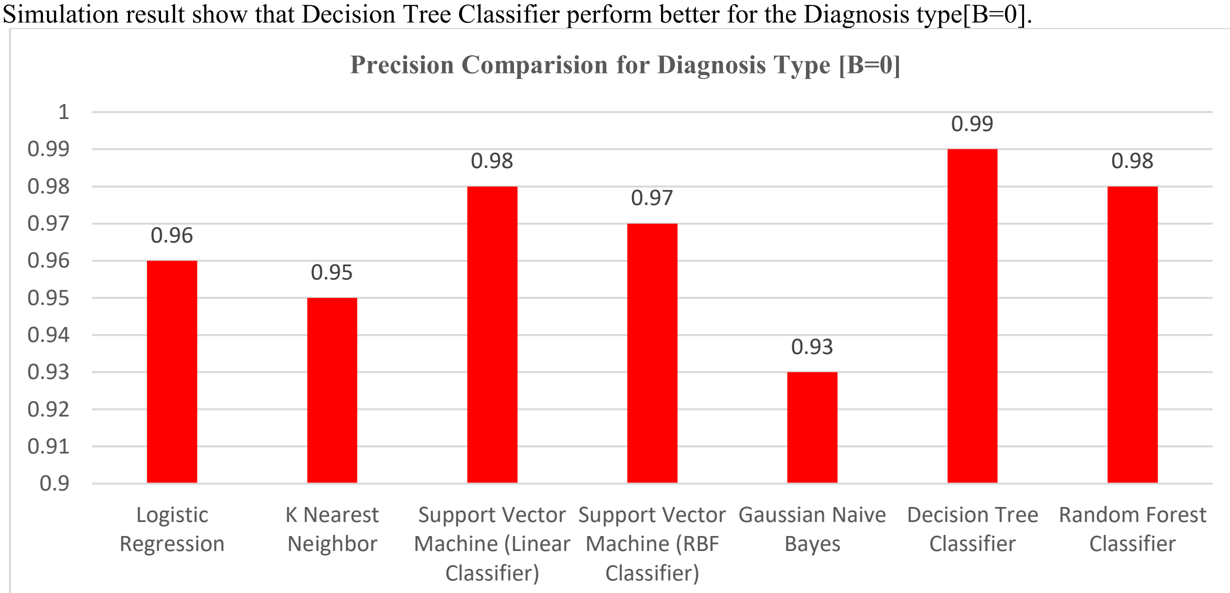

Fig. 4.3 Simulation Results Analysis of precision for Diagnosis Type[B=0]

Table 4.3 and Fig. 4.4 shows the comparative analysis of precision for Diagnosis Type [B=1] using various machine learning classifiers.

Table-4.3 Comparative analysis of precision for Diagnosis Type [B=1]

Method	**Precision**	**Diagnosis Type [B=1]**
Logistic Regression	0.92	1
K Nearest Neighbor	0.98	1
Support Vector Machine (Linear Classifier)	0.94	1
Support Vector Machine (RBF Classifier)	0.96	1
Gaussian Naïve Bayes	0.9	1
Decision Tree Classifier	0.9	1
Random Forest Classifier	0.94	1

Simulation result show that Support Vector Machine (RBF Classifier) perform better for the Diagnosis type[B=0].

ISSN Number: 2454-2024(Online)

International Journal of Technical Research & Science

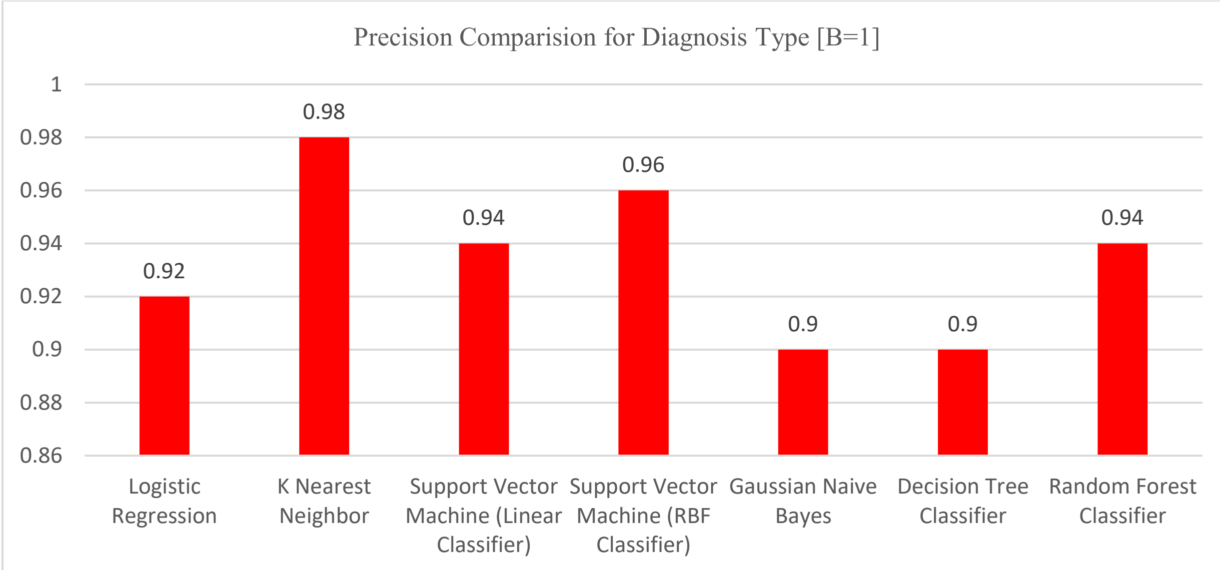

Fig. 4.4 Precision Comparison of Diagnosis Type [B=1]

Recall Comparison for Diagnosis Type [B=0 and B=1]

Table 4.4 and figure 4.5 shows the comparative analysis of recall for Diagnosis Type [B=0] using various machine learning classifiers.

Simulation result show that Support Vector Machine (RBF Classifier) perform better for the Diagnosis type[B=0].

Table-4.4 Comparative analysis of recall for Diagnosis Type [B=0]

Method	**Precision**	**Diagnosis Type [B=0]**
Logistic Regression	0.96	0
K Nearest Neighbor	0.99	0
Support Vector Machine (Linear Classifier)	0.97	0
Support Vector Machine (RBF Classifier)	0.98	0
Gaussian Naïve Bayes	0.94	0
Decision Tree Classifier	0.93	0
Random Forest Classifier	0.97	0

Below graphs sows the performance analysis for Recall with Diagnosis Type[B=0]

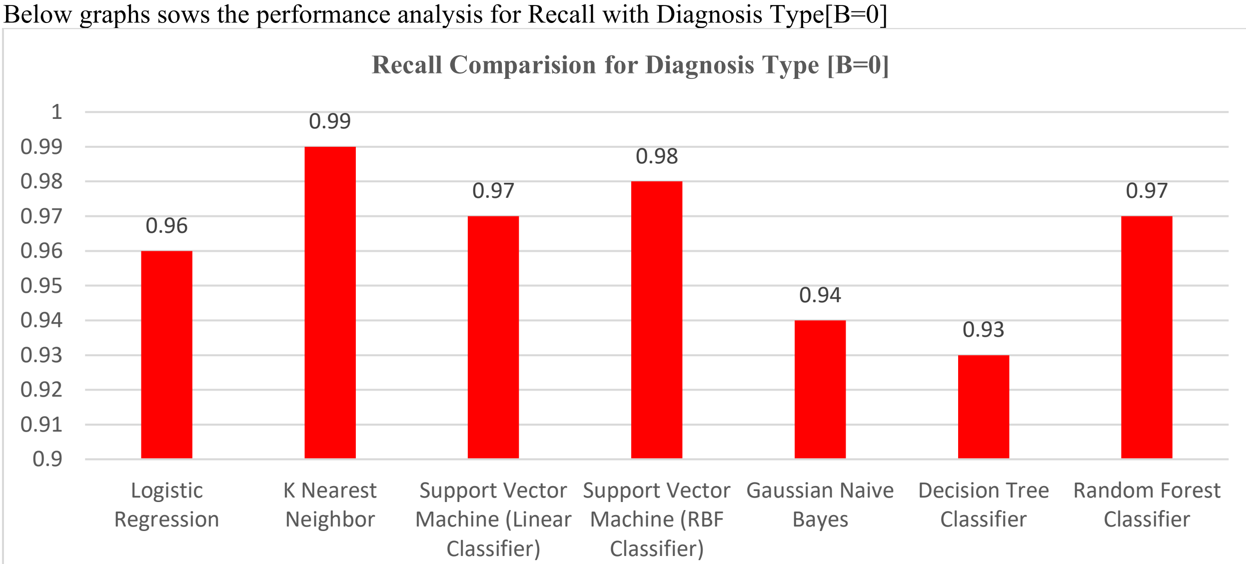

Fig. 4.5 Recall Comparison of Diagnosis Type [B=0]

ISSN Number: 2454-2024(Online)

International Journal of Technical Research & Science
Table 4.5 and fig. 4.6 shows the comparative analysis of recall for Diagnosis Type [B=0] using various machine learning classifiers.

Table-4.5 Recall Comparison for Diagnosis Type[M=1]

Method	**Recall**	**Diagnosis Type [M=0]**
Logistic Regression	0.92	1
K Nearest Neighbor	0.91	1
Support Vector Machine (Linear Classifier)	0.96	1
Support Vector Machine (RBF Classifier)	0.94	1
Gaussian Naïve Bayes	0.89	1
Decision Tree Classifier	0.98	1
Random Forest Classifier	0.96	1

Simulation result show that Support Vector Machine (RBF Classifier) perform better for the Diagnosis type[B=1].

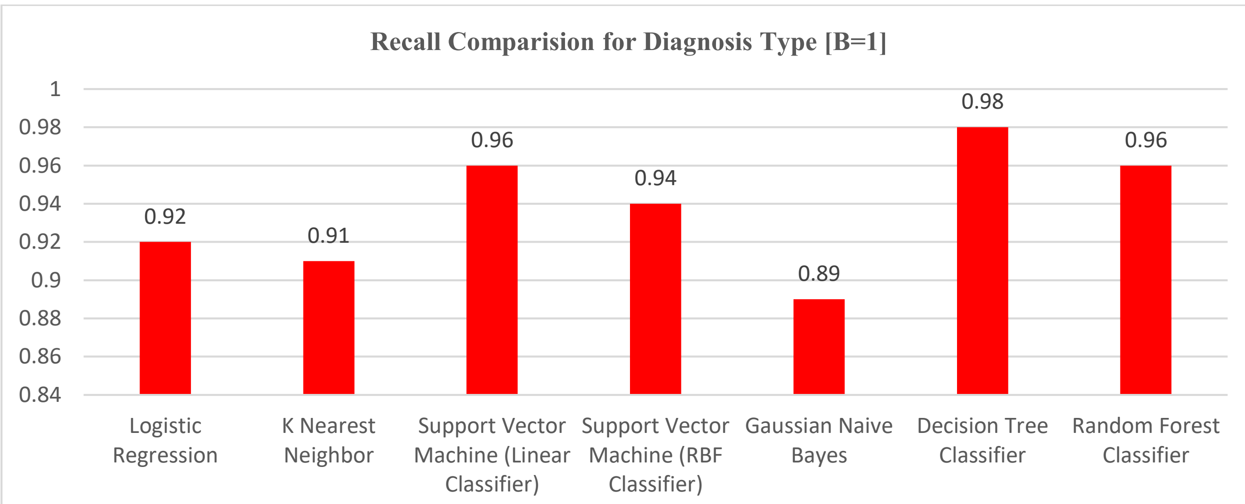

Fig. 4.6 Recall Analysis for Diagnosis Type [B=1]

F1-Score Comparison for Diagnosis Type [B=0 and B=1]
Table 4.6 and fig. 4.7 shows the comparative analysis of F1- Score for Diagnosis Type [B=0] using various machine learning classifiers.

Table-4.6 Comparative analysis of F1- Score for Diagnosis Type [B=0]

Method	**F-1 Score**	**Diagnosis Type [B=0, M=1]**
Logistic Regression	0.96	0
K Nearest Neighbor	0.97	0
Support Vector Machine (Linear Classifier)	0.97	0
Support Vector Machine (RBF Classifier)	0.97	0
Gaussian Naïve Bayes	0.94	0
Decision Tree Classifier	0.96	0
Random Forest Classifier	0.97	0

Simulation result show that Support Vector Machine (RBF Classifier) perform better for the Diagnosis type[B=0].

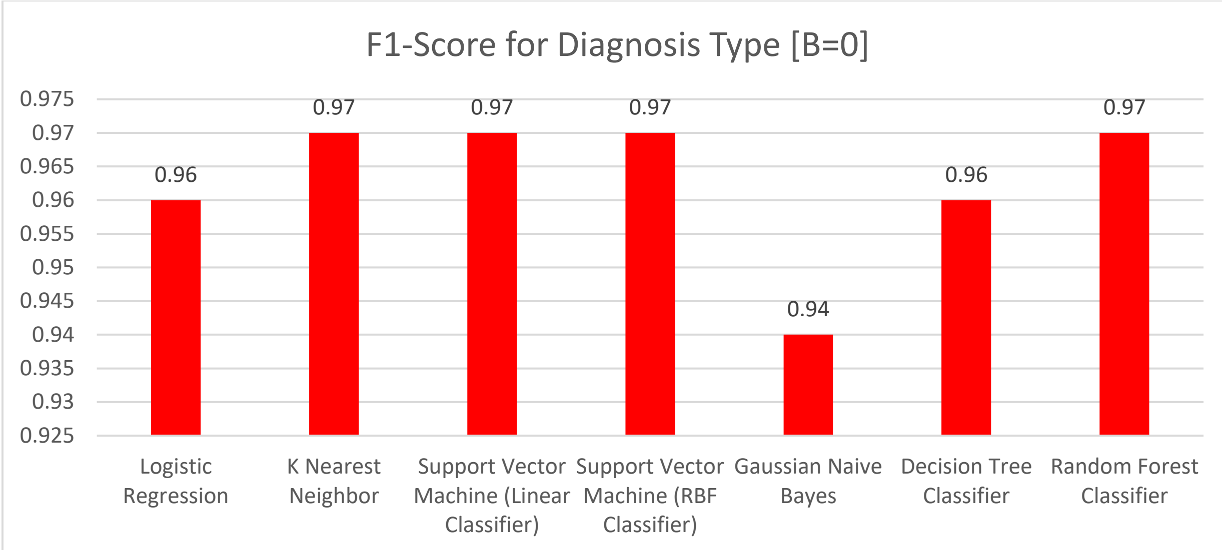

Fig. 4.7 Analysis of comparative study of F1- Score for Diagnosis Type [B=0]

Table 4.7 and fig. 4.8 shows the comparative analysis of F1- Score for Diagnosis Type [B=1] using various machine learning classifiers.

Table-4.7Comparative analysis of F1- Score for Diagnosis Type [B=1]

Method	**F-1 Score**	**Diagnosis Type [B=0, M=1]**
Logistic Regression	0.92	1
K Nearest Neighbor	0.94	1
Support Vector Machine (Linear Classifier)	0.95	1
Support Vector Machine (RBF Classifier)	0.95	1
Gaussian Naïve Bayes	0.9	1
Decision Tree Classifier	0.94	1
Random Forest Classifier	0.95	1

Simulation result show that Support Vector Machine (RBF Classifier) perform better for the Diagnosis type[B=1].

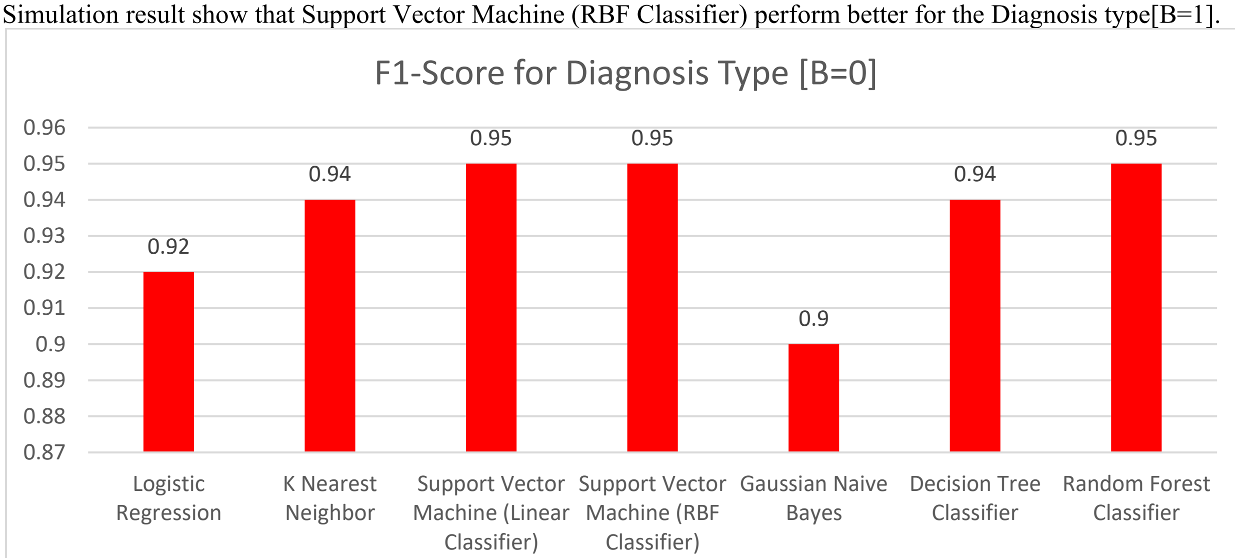

Fig. 4.8 Analysis of F1- Score for Diagnosis Type [B=1]

5. CONCLUSION AND FUTURE WORK

Discussion

We have carried out the comparative analysis of Machine Learning Classifiers so that Brest cancer can be detected and predicted by using Wisconsin Dataset. In terms of accuracy, sensitivity and specificity. The ensemble model is used as a reliable tool for breast cancer diagnosis and can assist radiologists in making accurate decisions. Early diagnosis of breast cancer is crucial for successful treatment and saving lives. Technological advancements in medical imaging have improved the accuracy of breast cancer diagnosis. Wisconsin Dataset is widely used dataset for experimental studies in breast cancer diagnosis. The imbalanced nature of the dataset can be addressed using oversampling techniques like SMOTE. Pre-trained deep learning models like logistic regression, K Nearest Neighbor, Support Vector Machine (Linear Classifier), Support Vector Machine (RBF Classifier), Gaussian Naïve Bayers, Decision Tree Classifier and Random Forest Classifier.

- Machine Learning Classifier have demonstrated remarkable performance is various computer vision tasks such as image classification, object detection and semantic segmentation. However, it requires a significant amount of computational power and time to train, especially for large datasets. In addition, combing features and classifiers can also help improve the results. Overall, Machine Learning classifiers are powerful tools for computer vision and transfer learning and feature/ classifier combination techniques can further enhance their performance while reducing the computational cost and training time.
- SVM (Support Vector Machines) is a popular algorithm for classification tasks, especially in case where the data is separable by a hyperplane in a high dimensional space. SVMs work well when there is a clear separation between classes in the data, and when the number of features is relatively small compared to the number of data points.
- Combining Machine Learning (ML) and Deep Learning (DL) approaches with optimization techniques can be highly effective for pattern identification and parameter section.
- One of the major challenges in machine learning is the unavailability of data with complete annotations, which can hinder the trading of accurate models. To address this challenge, researchers have developed various methods usch as transfer learning, semi-supervised learning and data augmentation.
- Overall, appropriate pre-processing techniques such as cropping, filtering, histogram equalization, and banalization can help to improve the quality of the data and enhance the accuracy and reliability of machine learning and computer vision models.
- Data augmentation is a useful technique that can help to improve the accuracy and reliability of machine learning and computer vision models, particularly when dealing with limited data size, data imbalance, and overfitting. The choice of augmentation techniques and parameters depends on the specific problem and the characteristic of the data being analyzed.
- The hybridization of human crafted features with automated features and the combination of classifiers can be effective approaches to improving the performance of machine learning and computer vision models, particularly in complex and challenging task. However, the choice of approach depends on the specific problem and the characteristics of the data being analyzed.

Conclusion

Overall, the thesis highlights the importance of technology in breast cancer diagnosis and the potential of Machine Learning and Deep Learning techniques to improve the accuracy of diagnosis. The thesis provides a comprehensive comparative analysis of the different machine learning techniques used in various studies and their results, highlighting the advantages, gaps, and challenges of each approach. The analysis of dataset also provides insights into the effectiveness of these techniques in different contexts. The thesis concludes that the combination of Machine Learning and Deep Learning approaches with optimization techniques can be useful for pattern identifications and parameter selection.

The conclusions are summarized given below:

- The Wisconsin Data dataset was used for the experimental study.
- Because above Wisconsin Dataset was imbalanced, we employed SMOTE to balance the dataset.
- Stacking were used where two or more base models are trained on the same dataset to create predictions, and the output of these models is used as input to a meta model that learns how to combine the base models.
- Base learning were implemented using pre-trained logistic regression, K Nearest Neighbor, Support Vector Machine (Linear Classifier), Support Vector Machine (RBF Classifier), Gaussian Naïve Bayers, Decision Tree Classifier and Random Forest Classifier. Among them, Support Vector Machine (RBF Classifier) shows the best performance on the dataset, got 0.97 Precision, Recall 0.98 and F1-Score0.97.

ISSN Number: 2454-2024(Online)

International Journal of Technical Research & Science

- The performance of each Machine Learning classifier is shown from table 4.1 to 4.7, which show the training accuracy, test accuracy, precision, recall, F1-score and Support.

Study Limitations and future directions

Limitations and potential sources of bias are important considerations in any research study, and it is commendable that these have been acknowledged in the analysis and discussion of the study. The exclusion of non-English language publication may limit the generalizability and diversity of the findings, particularly in fields where non-English publications are common or where language specific nuances may impact the result. Additionally, the use of only computational parameters for the analysis may overlook important contextual and environmental factors that can influence the performance of machine learning and computer vision models. For example, the quality and availability of the data, the expertise of the researcher, and the computational resources used can all impact the results.

Furthermore, the differences in pre-processing techniques, feature extraction techniques, algorithms, and training / testing data sizes can make it challenging to make direct comparisons between the results of different studies. This highlights the importance of carefully considering the specific context and characteristics of the data when interpreting and applying the result of machine learning and computer vision.

- Conducting more studies on multiclass classification is an important are of research that can lead to improved accuracy and reliability in disease diagnosis and other application of machine learning and computer vision.
- Dataset overfitting and misbalancing issue need to be handled.
- The fusion of deep learnings models can be a powerful approach to improve the performance of ML. Deep learning models are more capable of learning complex patterns and relationships in data.
- There are several challenges in breast cancer analysis, including the need of accurate and consistent detection and segmentation of lesions, the extraction of relevant features from the images, and the identification of the most relevant features for diagnosis and prognosis. A comprehensive, automated, and hybrid framework for breast cancer analysis could address these challenges by incorporating advanced ML Algorithms in medical imaging and cancer diagnosis.

REFERENCES

[1] A. R. Vaka, B. Soni, and S. R. K., “Breast cancer detection by leveraging Machine Learning,” ICT Express, vol. 6, no. 4, pp. 320–324, 2020, doi: 10.1016/j.icte.2020.04.009.

[2] D. Abdelhafiz, C. Yang, R. Ammar, and S. Nabavi, “Deep convolutional neural networks for mammography: Advances, challenges and applications,” BMC Bioinformatics, vol. 20, no. Suppl 11, 2019, doi: 10.1186/s12859-019-2823-4.

[3] Z. Wang et al., “Breast Cancer Detection Using Extreme Learning Machine Based on Feature Fusion with CNN Deep Features,” IEEE Access, vol. 7, pp. 105146–105158, 2019, doi: 10.1109/ACCESS.2019.2892795.

[4] A. Gastounioti, M. K. Hsieh, E. Cohen, L. Pantalone, E. F. Conant, and D. Kontos, “Incorporating Breast Anatomy in Computational Phenotyping of Mammographic Parenchymal Patterns for Breast Cancer Risk Estimation,” Sci. Rep., vol. 8, no. 1, pp. 1–10, 2018, doi: 10.1038/s41598-018-35929-9.

[5] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, “Deep Learning to Improve Breast Cancer Detection on Screening Mammography,” Sci. Rep., vol. 9, no. 1, pp. 1–12, 2019, doi: 10.1038/s41598-019-48995-4.

[6] E. K. Kim et al., “Applying Data-driven Imaging Biomarker in Mammography for Breast Cancer Screening: Preliminary Study,” Sci. Rep., vol. 8, no. 1, pp. 1–8, 2018, doi: 10.1038/s41598-018-21215-1.

[7] G. Litjens et al., “Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis,” Sci. Rep., vol. 6, no. April, pp. 1–11, 2016, doi: 10.1038/srep26286.

[8] V. Nagarajan, E. C. Britto, and S. M. Veeraputhiran, “Feature extraction based on empirical mode decomposition for automatic mass classification of mammogram images,” Med. Nov. Technol. Devices, vol. 1, no. June, p. 100004, 2019, doi: 10.1016/j.medntd.2019.100004.

[9] 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,” Sci. Rep., vol. 7, no. 1, pp. 1–10, 2017, doi: 10.1038/s41598-017-04075-z.

[10] H. Nasir Khan, A. R. Shahid, B. Raza, A. H. Dar, and H. Alquhayz, “Multi-View Feature Fusion Based Four Views Model for Mammogram Classification Using Convolutional Neural Network,” IEEE Access, vol. 7, pp. 165724–165733, 2019, doi: 10.1109/ACCESS.2019.2953318.

[11] B. Malik, J. Klock, J. Wiskin, and M. Lenox, “Objective breast tissue image classification using Quantitative Transmission ultrasound tomography,” Sci. Rep., vol. 6, no. December, 2016, doi: 10.1038/srep38857.

[12] X. Li, M. Radulovic, K. Kanjer, and K. N. Plataniotis, “Discriminative Pattern Mining for Breast Cancer Histopathology Image Classification via Fully Convolutional Autoencoder,” IEEE Access, vol. 7, pp. 36433–36445, 2019, doi: 10.1109/ACCESS.2019.2904245.

[13] R. K. Samala, H. P. Chan, L. Hadjiiski, M. A. Helvie, C. D. Richter, and K. H. Cha, "Breast cancer diagnosis in digital breast tomosynthesis: Effects of training sample size on multi-stage transfer learning using deep neural nets," IEEE Trans. Med. Imaging, vol. 38, no. 3, pp. 686–696, 2019, doi: 10.1109/TMI.2018.2870343.

[14] M. Byra, K. Dobruch-Sobczak, Z. Klimonda, H. Piotrzkowska-Wroblewska, and J. Litniewski, "Early Prediction of Response to Neoadjuvant Chemotherapy in Breast Cancer Sonography Using Siamese Convolutional Neural Networks," IEEE J. Biomed. Heal. Informatics, vol. 25, no. 3, pp. 797–805, 2021, doi: 10.1109/JBHI.2020.3008040.

[15] Y. Wang et al., "Deeply-Supervised Networks with Threshold Loss for Cancer Detection in Automated Breast Ultrasound," IEEE Trans. Med. Imaging, vol. 39, no. 4, pp. 866–876, 2020, doi: 10.1109/TMI.2019.2936500.

[16] "jiao2018.pdf." .

[17] D. Ribli, A. Horváth, Z. Unger, P. Pollner, and I. Csabai, "Detecting and classifying lesions in mammograms with Deep Learning," pp. 1–7, 2018, doi: 10.1038/s41598-018-22437-z.

[18] Y. U. Shaode, L. I. U. Lingling, W. Zhaoyang, D. A. I. Guangzhe, and X. I. E. Yaoqin, "Transferring deep neural networks for the differentiation of mammographic breast lesions," vol. 62, no. 3, pp. 2018–2019, 2019.

[19] D. Muduli, R. Dash, and B. Majhi, "Biomedical Signal Processing and Control Automated breast cancer detection in digital mammograms : A moth flame optimization based ELM approach," Biomed. Signal Process. Control, vol. 59, p. 101912, 2020, doi: 10.1016/j.bspc.2020.101912.

[20] D. M. Vo, N. Nguyen, and S. Lee, "Classification of breast cancer histology images using incremental boosting convolution networks," Inf. Sci. (Ny)., vol. 482, pp. 123–138, 2019, doi: 10.1016/j.ins.2018.12.089.

[21] S. Wang, Y. Wang, D. Wang, Y. Yin, and Y. Wang, "An improved random forest-based rule extraction method for breast cancer diagnosis," Appl. Soft Comput. J., vol. 86, p. 105941, 2020, doi: 10.1016/j.asoc.2019.105941.

[22] P. Kaur, G. Singh, and P. Kaur, "Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification," Informatics Med. Unlocked, vol. 16, no. January, p. 100151, 2019, doi: 10.1016/j.imu.2019.01.001.

[23] R. S. Bressan, P. H. Bugatti, and P. T. M. Saito, "Breast cancer diagnosis through active learning in content-based image retrieval," Neurocomputing, vol. 357, pp. 1–10, 2019, doi: 10.1016/j.neucom.2019.05.041.

[24] Y. Benhammou, B. Achchab, F. Herrera, and S. Tabik, "BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights," Neurocomputing, vol. 375, pp. 9–24, 2020, doi: 10.1016/j.neucom.2019.09.044.

[25] T. Pang, J. H. D. Wong, W. L. Ng, and C. S. Chan, "Deep learning radiomics in breast cancer with different modalities: Overview and future," Expert Syst. Appl., vol. 158, p. 113501, 2020, doi: 10.1016/j.eswa.2020.113501.

[26] 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, doi: 10.1016/j.bspc.2019.101789.

[27] H. Li, S. Zhuang, D. ao Li, J. Zhao, and Y. Ma, "Benign and malignant classification of mammogram images based on deep learning," Biomed. Signal Process. Control, vol. 51, pp. 347–354, 2019, doi: 10.1016/j.bspc.2019.02.017.

[28] Y. Fang, J. Zhao, L. Hu, X. Ying, Y. Pan, and X. Wang, "Image classification toward breast cancer using deeply-learned quality features," J. Vis. Commun. Image Represent., vol. 64, p. 102609, 2019, doi: 10.1016/j.jvcir.2019.102609.

[29] Z. Zhu et al., "Deep learning analysis of breast MRIs for prediction of occult invasive disease in ductal carcinoma in situ," Comput. Biol. Med., vol. 115, no. October, p. 103498, 2019, doi: 10.1016/j.compbiomed.2019.103498.

[30] F. Mohanty, S. Rup, and B. Dash, "Automated diagnosis of breast cancer using parameter optimized kernel extreme learning machine," Biomed. Signal Process. Control, vol. 62, no. July, p. 102108, 2020, doi: 10.1016/j.bspc.2020.102108.

[31] O. N. Oyelade and A. E. S. Ezugwu, "A State-of-the-Art Survey on Deep Learning Methods for Detection of Architectural Distortion from Digital Mammography," IEEE Access, vol. 8, pp. 148644–148676, 2020, doi: 10.1109/ACCESS.2020.3016223.

[32] X. Zhang et al., "Classification of whole mammogram and tomosynthesis images using deep convolutional neural networks," IEEE Trans. Nanobioscience, vol. 17, no. 3, pp. 237–242, 2018, doi: 10.1109/TNB.2018.2845103.

[33] H. Chougrad, H. Zouaki, and O. Alheyane, "Multi-label transfer learning for the early diagnosis of breast cancer," Neurocomputing, vol. 392, pp. 168–180, 2020, doi: 10.1016/j.neucom.2019.01.112.

[34] M. A. Al-antari, S. M. Han, and T. S. Kim, "Evaluation of deep learning detection and classification towards

computer-aided diagnosis of breast lesions in digital X-ray mammograms," Comput. Methods Programs Biomed., vol. 196, p. 105584, 2020, doi: 10.1016/j.cmpb.2020.105584.

[35] F. J. Pérez-Benito et al., "A deep learning system to obtain the optimal parameters for a threshold-based breast and dense tissue segmentation," Comput. Methods Programs Biomed., vol. 195, pp. 1–10, 2020, doi: 10.1016/j.cmpb.2020.105668.

[36] F. F. Ting, Y. J. Tan, and K. S. Sim, "Convolutional neural network improvement for breast cancer classification," Expert Syst. Appl., vol. 120, pp. 103–115, 2019, doi: 10.1016/j.eswa.2018.11.008.

[37] P. Gomathi, C. Muniraj, and P. S. Periasamy, "Breast thermography based unsupervised anisotropic- feature transformation method for automatic breast cancer detection," Microprocess. Microsyst., vol. 77, 2020, doi: 10.1016/j.micpro.2020.103137.

[38] R. Shen, K. Yan, K. Tian, C. Jiang, and K. Zhou, "Breast mass detection from the digitized X-ray mammograms based on the combination of deep active learning and self-paced learning," Futur. Gener. Comput. Syst., vol. 101, pp. 668–679, 2019, doi: 10.1016/j.future.2019.07.013.

[39] A. Al Nahid and Y. Kong, "Histopathological breast-image classification using local and frequency domains by convolutional neural network," Inf., vol. 9, no. 1, 2018, doi: 10.3390/info9010019.

[40] S. M. L. de Lima, A. G. da Silva-Filho, and W. P. dos Santos, "Detection and classification of masses in mammographic images in a multi-kernel approach," Comput. Methods Programs Biomed., vol. 134, pp. 11–29, 2016, doi: 10.1016/j.cmpb.2016.04.029.

[41] W. Peng, R. V. Mayorga, and E. M. A. Hussein, "An automated confirmatory system for analysis of mammograms," Comput. Methods Programs Biomed., vol. 125, pp. 134–144, 2016, doi: 10.1016/j.cmpb.2015.09.019.

[42] J. Wang, X. Yang, H. Cai, W. Tan, C. Jin, and L. Li, "Discrimination of Breast Cancer with Microcalcifications on Mammography by Deep Learning," Sci. Rep., vol. 6, no. May, pp. 1–9, 2016, doi: 10.1038/srep27327.

[43] M. Mullooly et al., "Application of convolutional neural networks to breast biopsies to delineate tissue correlates of mammographic breast density," npj Breast Cancer, vol. 5, no. 1, 2019, doi: 10.1038/s41523-019-0134-6.

[44] Q. Hu, H. M. Whitney, and M. L. Giger, "A deep learning methodology for improved breast cancer diagnosis using multiparametric MRI," Sci. Rep., vol. 10, no. 1, pp. 1–11, 2020, doi: 10.1038/s41598-020-67441-4.

[45] Y. Feng, L. Zhang, and J. Mo, "Deep Manifold Preserving Autoencoder for Classifying Breast Cancer Histopathological Images," IEEE/ACM Trans. Comput. Biol. Bioinforma., vol. 17, no. 1, pp. 91–101, 2020, doi: 10.1109/TCBB.2018.2858763.

[46] T. Mahmood, J. Li, Y. Pei, F. Akhtar, A. Imran, and K. Ur Rehman, "A brief survey on breast cancer diagnostic with deep learning schemes using multi-image modalities," IEEE Access, vol. 8, pp. 165779–165809, 2020, doi: 10.1109/ACCESS.2020.3021343.

[47] M. Hosni, I. Abnane, A. Idri, J. M. Carrillo de Gea, and J. L. Fernández Alemán, "Reviewing ensemble classification methods in breast cancer," Comput. Methods Programs Biomed., vol. 177, pp. 89–112, 2019, doi: 10.1016/j.cmpb.2019.05.019.