Machine Learning with Applications 6 (2021) 100154

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

An investigation of XGBoost-based algorithm for breast cancer classification

Xin Yu Liew *, Nazia Hameed, Jeremie Clos

University of Nottingham, Jubilee Campus, Wollaton Road, Nottingham, NG8 1BB, United Kingdom

ARTICLE INFO

Keywords:
Deep learning
Extreme gradient boosting
XGBoost
Machine learning
Computer-aided diagnosis
Breast cancer
Histopathology images
Classification

ABSTRACT

Breast cancer is one of the leading cancers affecting women around the world. The Computer-Aided Diagnosis (CAD) system is a powerful tool to assist pathologists during the process of diagnosing cancer, which effectively identifies the presence of cancerous cells. A standard CAD system includes processes of pre-processing, feature extraction, feature selection and classification. In this paper, we propose an enhanced breast cancer classification technique called Deep Learning and eXtreme Gradient Boosting (DLXGB) on histopathology breast cancer images using the BreaKHis dataset. This method first applies data augmentation and stain normalization for pre-processing, then pre-trained DenseNet201 will automatically learn features within an image and combine with a powerful gradient boosting classifier. The proposed classification technique is designed to classify breast cancer histology images into binary benign and malignant, and additionally one of eight non-overlapping/overlapping categories: i.e., Adenosis (A), Fibroadenoma (F), Phyllodes Tumour (PT), And Tubular Adenoma (TA) Ductal Carcinoma (DC), Lobular Carcinoma (LC), Mucinous Carcinoma (MC), And Papillary Carcinoma (PC). With DLXGB, we have obtained an accuracy of 97% for both binary and multi-classification improving the exiting work done by researchers using the BreaKHis dataset. The results indicated that this method could produce a powerful prediction for breast cancer image classification.

1. Introduction

The human body is formed of trillions of cells. ‘‘Cancer’’ is a term used when a cell divides abnormally or uncontrollably and can happen on various parts of the body. The disease type is categorized based on which part of the body cancer occurs. This situation will lead to death as it progresses to spread to other parts of the body. Amongst the distinct types of cancer, the most common type of cancer diagnosed by most women in the world is breast cancer. According to the World Health Organisation (WHO), breast cancer is the most frequent cancer among women, affecting 2.1 million women each year. In the United Kingdom, there are around 55,200 newly diagnosed breast cancer cases every year, which is about 150 every day from 2015 to 2017 (Cancer Research UK, 2014).

Statistically, if the disease is detected and diagnosed at an early stage, all (98%) patients will survive the disease for five years or more, compared to around 1 in 4 (26%) people when the disease is diagnosed at the latest stage (Cancer Research UK, 2014). Therefore, early detection of breast cancer can significantly increase the chances of a successful treatment plan and ensure the long-term survival of the patients (Sizilio et al., 2012). According to the most common procedure, ‘Two-week wait’ is the procedure to diagnose breast cancer (Cancer Research UK, 2014). This is because the standard procedure to diagnose breast cancer by pathologists usually requires extensive microscopic assessment, therefore takes an extensive amount of time to examine. Thus, developing an automated solution like a Computer-aided Diagnosis (CAD) system contributes to an easier diagnostic process for pathologists and reduces the subjectivity in diagnosis.

The term ‘‘breast cancer’’ refers to a malignant tumour that has developed from cells in the breast that are considered cancerous and cause danger to the health of the patient. Distinct stages of breast cancer determine the risk of the disease. In cases where cancer is detected, but no cancer cells are visible in the lymph glands, breast cancer is of lower risk. When spreading occurs, it carries a substantial risk of death, meaning that the cancer cells from the breast tissue have broken away, which can be carried to nearby lymph nodes by the lymph fluid (fluids that gather waste products and drain into veins to be removed) (Breast Cancer Organization, 2016). Breast tumours can be distinguished as benign (non-cancerous) and malignant (cancerous/metastatic) tumours. Benign tissue refers to changes in normal tissue of breast parenchyma, which does not relate to the development of malignancy (Alom et al., 2019). Contrarily, malignant tissue can be categorized into two types, in-situ carcinoma, and invasive carcinoma.

Recent innovations in machine learning have allowed for improved, faster processing of raw medical data. Embedding such advanced tools in the medical diagnosis process can provide useful knowledge to help pathologists in analysing large amounts of medical data more efficiently and more effectively. Robertson et al. (2018). This potentially speeds

* Corresponding author.
E-mail addresses: xinyu.liew@yahoo.com (X.Y. Liew), nazia.hameed@nottingham.ac.uk (N. Hameed), jeremie.clos@nottingham.ac.uk (J. Clos).

https://doi.org/10.1016/j.mlwa.2021.100154
Received 30 April 2021; Received in revised form 23 August 2021; Accepted 1 September 2021
Available online 8 September 2021

up the process because machine learning can process large data much faster than manual diagnosis by pathologists (Robertson et al., 2018).

Recently, deep learning methods were introduced and quickly became popular due to their automated feature extraction and representation learning ability (Lecun et al., 2015). This method can extract information from the images automatically as part of the learning process and performs more accurately in classification problems, leading to significantly improved performance (Lecun et al., 2015). In many research studies, deep learning was shown to outperform state-of-art methods in many fields of medical imaging analysis tasks like classification, detection, segmentation, and computer-aided diagnosis (Alkassar et al., 2021; Budak et al., 2019; Gandomkar et al., 2018; George et al., 2019; Toğaçar et al., 2020).

The Convolutional Neural Network (CNN) is a common variant of the Deep Neural Network (DNN), it has been used to develop various applications in computer vision and recognized for its characteristics and capability of weight sharing and local connectivity (Han et al., 2017; Spanhol et al., 2016a, 2016b). However, training a CNN from scratch to define the weights requires a large dataset and it takes more time and effort in adjusting the parameters to produce a reliable performance and robust model (Irsoy & Alpaydin, 2020; Shin et al., 2016). Transfer learning is a better approach when it comes to using a deep learning model as a baseline method or feature extractor. On the other hand, the eXtreme Gradient Boosting (XGBoost) is a scalable machine learning technique using tree boosting to avoid overfitting, it has recently gained the attention of researchers because it outperformed many existing traditional classifiers in machine learning (Chen & Guestrin, 2016). However, to the best of our knowledge, this technique has not been fully explored for image classification tasks, especially in the breast cancer histology dataset.

Each model has its advantages and disadvantages, this work aims to combine those two state-of-the-art methods in such a way that they overcome each model's weaknesses. It is only fair if each weakness of a model is considered and covered with a solution by using the strength of another model. Thus, the main contributions of this paper are:

- A CAD expert system capable of classifying breast cancer into the benign category of Adenosis (A), Fibroadenoma (F), Phyllodes Tumour (PT), and Tubular Adenoma (TA); and the malignant category of Ductal Carcinoma (DC), Lobular Carcinoma (LC), Mucinous Carcinoma (MC), And Papillary Carcinoma (PC) to assist the feasibility for breast cancer detection in medical practice.
- A technique combining Deep Learning and XGBoost (DLXGB) is applied to breast histopathology image classification by replacing the Fully Connected layer of a CNN with XGBoost, which can potentially reduce fine-tuning on parameters and still produce reliable performance. This proposed framework will help diagnose breast cancer in the preliminary stages.
- Implementing improved mechanism into our proposed framework that solves the common weaknesses of the previous work like data imbalance issue, bias, and variance in the machine learning model.
- A new potential approach combining DenseNet201 and XGBoost that has not been applied on breast cancer histology images based on previous work. To the best of our knowledge, there are very few research papers exploring a similar method for breast cancer, even these existing researches are exploring different pre-trained networks and experimenting on mammography images only.

The outline of the paper is provided as follows. Section 2 reviews the existing work on breast cancer CAD expert systems and the research gaps are identified. The proposed method is discussed in Section 3 which includes the experimental data, approach overview, data-pre-processing and augmentation, deep learning feature extraction, DenseNet feature extractor and extreme gradient boosting classifier. In Section 4, we evaluate the performance of the proposed approach and how the parameters are calculated. The proposed classification technique is also compared with existing work in this section. Lastly, we discuss the conclusion and future work in Section 5.

2. Literature review

Conventionally, there are several popular machine learning algorithms applied to classification problems that include Nearest neighbour (Witten et al., 2016), Decision Trees (Quinlan, 1986), Artificial Neural Network (Amato et al., 2013), Support Vector Machine (SVM) (Cortes & Vapnik, 1995)l, Ensemble Learning (Ho, 1998), Convolutional Neural Network (CNN) (LeCun et al., 2010) and many more. These algorithms are designed to obtain the highest classification accuracy. However, these methods might be lacking in overcoming issues like imbalanced costs of misclassification within many classes. For instance, when misclassification occurs from these methods, it produces a higher chance of misclassifying non-cancerous patients as cancerous or vice versa. This could cause heavy consequences like incurring a more expensive cost for further pathologist analysis. Therefore, deep learning was introduced to potentially be a more powerful approach in tackling issues of misclassification.

Bayramoglu et al. proposed two different CNN architectures; single task CNN is used to predict malignancy and multi-task CNN is used to predict both malignancy and image magnification level simultaneously with the highest accuracy of 83.13% (Bayramoglu et al., 2016). Gandomkar et al. proposed a two-step classification in which they first used a deep residual network (ResNet) with 152 layers to trained for classifying patches from the images as benign or malignant for each magnification factor (Gandomkar et al., 2018). Then, they used the same pre-trained model to further classify the breast cancer sub-classes with 96.25% accuracy. Toğaçar et al. proposed a novel method called BreastNet of using CNN model architecture that adopted a multi-layer perception (MLP) as a classifier with an accuracy of 98.8% (Toğaçar et al., 2020). Mahmood et al. perform a score-level fusion of Resnet-50 and Densenet-201 for classification and achieved a Precision of 87.6%, Recall of 84.1%, and F1-measure of 85.8% (Mahmood et al., 2020). Alirezazadeh et al. proposed a representation learning-based unsupervised domain adaptation method, which distinguishes benign extracted feature vectors from those of malignant ones by learning a domain invariant space, they achieved an average classification rate of 88.5% on BreaKHis dataset (Alirezazadeh et al., 2018). Vesal et al. investigate transfer learning on two pre-trained networks, Google's Inception-V3 and ResNet50, their results have shown that Inception-V3 achieved an accuracy of 97.08%, outperforming ResNet50 with an accuracy of 96.66% (Vesal et al., 2018).

Vo et al. proposed a model called Inception-ResNet-v2 that combines CNNs of Inception and ResNet to train and extract visual features from multiscale images to achieve both global and local features from breast tumours and feed it into a gradient boosting classifier, they achieved 99.5% and 96.4% accuracies (Vo et al., 2019). Alom et al. proposed a binary and multi-classification for breast cancer methods using the Inception Recurrent Residual Convolutional Neural Network (IR-RCNN) model and achieved 99.05% (for binary) and 98.59% (for multi) classification (Alom et al., 2019). Han et al. proposed a method class structure-based deep convolutional neural network (CSDCNN) based on GoogLeNet for eight-class classification of breast histopathological slides have shown that their accuracy was higher for fine-tuning in comparison with training from scratch with 93.2% accuracy (Han et al., 2017). Besides, Bardou et al. experimented with a second approach to apply a CNN model, and their results have shown that the deep learning approach has outperformed handcrafted features (Bardou et al., 2018). The research by Rakhlin et al. have applied ResNet-50, InceptionV3 and VGG-16 models for feature extraction and a gradient boosting tree as a classifier in their proposed method and achieved accuracy 93.8%, AUC 97.3%, and sensitivity/specificity 96.5/88.0% at the high-sensitivity operating point (Rakhlin et al., 2018).

Cai et al. adopted modified faster-RCNN (Regional Convolutional Neural Network) for detecting mitosis cells using Resnet-101 network pre-trained on ImageNet database to extract features for classification (Cai et al., 2019). Budak et al. proposed a novel method that

uses a fully convolutional network (FCN) transform from AlextNet as an encoder for high-level feature extraction, the output of the FCN will then be transformed to a one-dimensional sequence for classification using Bi-LSTM, they achieved a 91.90% accuracy (Budak et al., 2019). Munien et al. applied EfficientNet architecture for the classification of breast cancer histology images and obtain the highest accuracy of 98.33% (Munien & Viriri, 2021). Laxmisagar and Hanumantharaju have adopted an efficient lightweight neural network called MobileNet2.10ex for feature extraction and designed a fully connected deep neural network as a classifier, they obtain the highest accuracy of 88.92% (Laxmisagar & Hanumantharaju, 2021).

Yari et al. applied deep learning feature extraction based on DenseNet121 pre-trained weights with some fine-tuning and data augmentation, they obtain up to 97% image-level accuracy for binary classification and 100% for multi-classification (Yari & Nguyen, 2020). Nedjar et al. proposed a method based on consensus, a hard voting method, oriented by InceptionV3, ResNet50 and MobileNet, their method achieved the highest accuracy of 91.06% (Nedjar et al., 2019). One of the most promising developed deep learning models was the Lymph Node Assistant (LYNA) algorithm based on Inception-v3 by the researchers of Naval Medical Center San Diego and Google AI (Artificial Intelligence) (Liu et al., 2019). They adopted the Inception-v3 network because this model has been shown to achieve greater than 78.1% accuracy on Stanford's ImageNet dataset. Their results have successfully achieved a receiver operating characteristic area under the curve (AUC) of 99% and a tumour-level sensitivity of 91% at 1 false positive per patient (Liu et al., 2019). Kushwaha et al. adopted the method of pre-trained neural network DenseNet-201 for the classification task and obtained an accuracy of 97.05% (Kushwaha et al., 2021). Al-Haija & Adebanjo proposed a ResNet-50 convolutional neural network and achieved 99.10% accuracy in binary classification (Al-Haija & Adebanjo, 2020). Boumaraf et al. proposed a deep neural network ResNet-18 and achieved 98.42% on binary classification and 92.02% on multi-classification (Boumaraf et al., 2021). The state of the art of existing work is summarized in Table 1.

As we observe from Table 1 that compares the existing work for breast histology images, it can be noticed that the XGBoost algorithm is not yet explored for classifying breast cancer in this type of breast sample. In this research work, we have presented the design of a machine learning computer-aided diagnosis expert system for breast cancer on histopathology images using a deep learning approach. Publicly available histopathology datasets mentioned above like BreaKHis (Spanhol et al., 2016a, 2016b), Bioimaging Challenge 2015 (Araujo et al., 2017), BACH (Aresta et al., 2019), CAMELYON (Litjens et al., 2018), PCam (Veeling et al., 2018), MITOS-12 (Roux et al., 2013), MITOS-ATYPIA-14 (MITOS-ATYPIA-14 Grand Challenge, 2014), TUPAC16 (Veta et al., 2019) have also captivated the interest of many developers and researchers in exploring the possibilities in diagnosing breast cancer. Applying machine learning techniques to CAD systems have shown promising performance in comparison to a microscope examination performed by a pathologist. The recent improvement of deep learning has also remarkably outperformed the traditional approach to learn and extract image features. However, the process to analyse medical images also produce a challenging task because of the complexity of histology images. Developing a CAD system has given potential opportunities to pathologists by improving the medical image analysis with better performance and time efficiency. Nevertheless, there are a few research gaps to be mentioned while developing a real-world applicable breast cancer CAD system.

First, there are data limitations when it comes to developing a good performing algorithm for the CAD expert system. To elaborate on this, medical data are challenging to work around because it requires a high processing power and memory storage for training. Besides, the performance of a machine learning model highly depends on the amount of data being feed into as input. This will directly affect the dependability of the CAD expert system when the model only learned a limited amount of knowledge. Although there are various publicly available datasets, they are considered small datasets, which creates doubt on the applicability of a CAD expert system in a real-world environment, like hospitals. Besides, bias in a model and imbalance classes in a dataset can lead to undesired classification errors during diagnosing. Bias to a specific class occurs when a model is trained on a dataset with imbalance classes, it affects the reliability of a CAD system due to an increase in the rate of the wrong classification. Several approaches like oversampling, under-sampling, and algorithm level methods are suggested to tackle this problem (He & Garcia, 2009). Even so, there are limited research and investigation applied to present the significance and improvement of data balancing.

There are also hardly any available domain-specific models pre-trained on medical images in the medical field. Most existing work mentioned that adopted pre-trained networks for transfer learning are networks that are trained on large image datasets of non-medical related images. This could potentially lead to more resources and effort needed to modify the network to classify medical images. To conclude, it is essential to produce a robust technique in producing a CAD expert system to overcome all underlying variations mentioned. In this paper, we modify the deep learning model by exploring the most recent development of the XGBoost algorithm in developing a breast cancer diagnosis system.

3. Methodology

3.1. Dataset

The dataset that we are using for experimenting with the proposed method is the Breast Cancer Histopathological Image Classification (BreakHis) (Spanhol et al., 2016a, 2016b). This dataset provides 4 different magnification levels of 40×, 100×, 200×, and 400× histology images of size 752 × 582 pixels. It consists of a total number of 7909 images acquired from a clinical study from January 2014 to December 2014 in P&D Laboratory, Brazil by 82 patients. For binary classification, there are two categories of benign and malignant to determine cancer or non-cancerous. There are 1995 images (625 benign and 1370 malignant) in 40× magnification level; 2081 images (644 benign and 1437 malignant) in 100× magnification level; 2013 images (623 benign and 1390 malignant) in 200× magnification level; and 1820 images (588 benign and 1232 malignant) in 400× magnification level in the dataset. To further perform multiclassification, the dataset contains four distinct types for each breast tumours. The category benign type of breast tumour consists of adenosis (A), fibroadenoma (F), phyllodes tumour (PT), and tubular adenoma (TA). While, the malignant type of breast tumour consists of ductal carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC), and papillary carcinoma (PC). This dataset is the most used dataset by many researchers for CAD breast cancer in histopathology images (Alkassar et al., 2021; Alom et al., 2019; Bardou et al., 2018; Bayramoglu et al., 2016; Budak et al., 2019; Chan & Tuszynski, 2016; Gandomkar et al., 2018; George et al., 2019; Han et al., 2017; Kassani et al., 2020; Murtaza et al., 2020; Spanhol et al., 2016a, 2016b; Sudharshan et al., 2019; Toğaçar et al., 2020; Vo et al., 2019). This dataset can be obtained from http://web.inf.ufpr.br/vri/breast-cancer-database.

BreaKHis datasets are compiled from histopathology (histology) images, which are samples of breast lesions obtained through either needles or surgical operations. These collected samples will later be processed and allocated to a glass slide to undergo a staining process to acquire a histology image for further analysis. The development in the computation of whole slide digital scanners enabled the digitalization of histopathological tissue sections and produced the development of digital pathology into routine practice (Pantanowitz et al., 2011). This type of sample contains a vast amount of information for pathologists to perform a quantitative analysis (Ghaznavi et al., 2013). These images are in three-channel red, green blue (RGB). Currently, histopathological

Table 1
Comparison of existing work done on the breast cancer classification.

Ref	Dataset	Methods	Results
Bayramoglu et al. (2016)	BreaKHis	Single task CNN (malignancy); Multitask CNN (magnification level)	Prr: 83.13%. Prr: 80.10%
Han et al. (2017)	BreaKHis	Custom CSDCNN (Class Structure-based Deep Convolutional Neural Network) based on GoogLeNet	Accuracy: 93.2%
Rakhlin et al. (2018)	BACH	ResNet-50, InceptionV3, VGG-16 and Gradient boosted trees	Accuracy: 87.2% (for binary) and 93.8% (for multi) AUC: 97.3%; Sensitivity: 96.5; Specificity: 88.0%
Gandomkar et al. (2018)	BreaKHis	ResNet CNN	Accuracy: 98.77% (Binary); Prr: 96.25% (Multi class)
Alirezazadeh et al. (2018)	BreaKHis	Representation learning-based unsupervised domain adaptation	Accuracy: 88.5%
Vesal et al. (2018)	BACH 2018	Inception-V3 and ResNet50	Accuracy: 97.08% and 96.66%
Vo et al. (2019)	BreaKHis	Inception and ResNet CNN (IRRCNN) and Gradient boosting trees	Accuracy: 93.8%–96.9%
Alom et al. (2019)	Bioimaging Challenge 2015	Inception Recurrent Residual CNN (IRRCNN)	Accuracy: 99.05% (for binary) and 98.59% (for multi)
Cai et al. (2019)	MITOS-ATYPIA-14 and TUPAC-16	Modified faster-RCNN	Precision: 76%; Recall: 72%; F1 score: 73.6%
Budak et al. (2019)	BreaKHis	FCN (Fully Convolutional Network) based on AlexNet and Bi-LSTM (Bidirectional Long Short-Term Memory)	Accuracy: 91.90%; Sensitivity: 96.8%; Specificity:91%
Nedjar et al. (2019)	BreaKHis	Consensus oriented by InceptionV3, ResNet50 and MobileNet	Accuracy: 91.06%
Liu et al. (2019)	Camelyon Challenge and Dataset from Naval Medical Center San Diego	Lymph Node Assistant (LYNA) algorithm based on Inception-v3	AUC: 99%; Sensitivity: 91%
Toğaçar et al. (2020)	BreaKHis	CNN features with MLP (Multi-Layer Perceptron)	Accuracy: 98.80%
Mahmood et al. (2020)	MITOS-12 and MITOS-ATYPIA-14	Faster-RCNN and a score-level fusion of Resnet-50 and Densenet-201 CNNs	Precision: 87.6%; Recall: 84.1%; F1-measure: 85.8%
Yari and Nguyen (2020)	BreaKHis	DenseNet121	Accuracy: 97%–100%
Al-Haija and Adebanjo (2020)	BreaKHis	ResNet-50	Accuracy: 99.10%
Munien and Viriri (2021)	ICIAR2018 dataset	EfficientNet	Accuracy: 98.33%
Laxmisagar and Hanumantharaju (2021)	Bioimaging Challenge 2015	MobileNet2.10ex	Accuracy: 88.92%
Kushwaha et al. (2021)	BreaKHis	DenseNet201	Accuracy: 97.05%
Boumaraf et al. (2021)	BreaKHis	ResNet-18	Accuracy: 92.03–98.42%

images play a vital role in breast cancer as well as in other types of cancer diagnosis because of the large amount of information it provides for medical image analysis (Khan et al., 2019). Therefore, in this research work, we have chosen to study stained images of breast cancer containing breast tissues in developing a DLXGB based CAD system. Fig. 1 presents the sample of images in the BreaKHis dataset.

3.2. Approach overview

In this section, the proposed Deep Learning with eXtreme Gradient Boosting (DLXGB) on histopathology breast cancer images will be described in detail. This architecture is composed of two major parts of feature extraction and classifier, each with different layers. The initial steps taken in this framework is to perform pre-processing on images to ensure all images in the dataset contain a consistency in colour, size, and variation. Our proposed method has approached this using stain normalization, resizing, data resampling and data augmentation. All the steps have been taken to ensure the quality of images before feeding them into the machine learning model. Then, a pre-trained network called DenseNet201 plays a vital role in extracting all the key features from an image to gather many knowledges. This chosen pre-trained network allows a deeper exploration of all pixels within an image. The last step of the proposed framework trains all these extracted features using an eXtreme Gradient Boosting (XGBoost) classifier. This boosting algorithm is a highly effective and scalable method to train our proposed network in classifying distinct types of breast histology images. Fig. 2 demonstrates the methodology of the proposed DLXGB for the breast cancer classification expert system.

3.3. Data pre-processing and augmentation

Image pre-processing is an essential route to prepare images from a dataset during the very first phase to ensure the raw data are transformed to a more consistent and normalized dataset to perform any further analysis. First, to avoid bias in our model due to class imbalance issues, we have resampled the original dataset to obtain a consistent number of samples for each class in binary and multi-classification tasks. Table 2 shows the number of images after we balance the data using resampling techniques.

To prepare our experimental setup, pre-processing step on final resampled data is performed before feeding the images into the proposed architecture. In this research, we have chosen histopathology breast samples that commonly uses colour normalization as part of pre-processing due to the colour variation that exists in this type of sample which can highly impact the performance of the machine learning model. Therefore, we applied colour normalization using a method proposed by Macenko et al. (2009) that uses the concept of fringe search algorithm (graph search algorithm) which traverse through nodes to

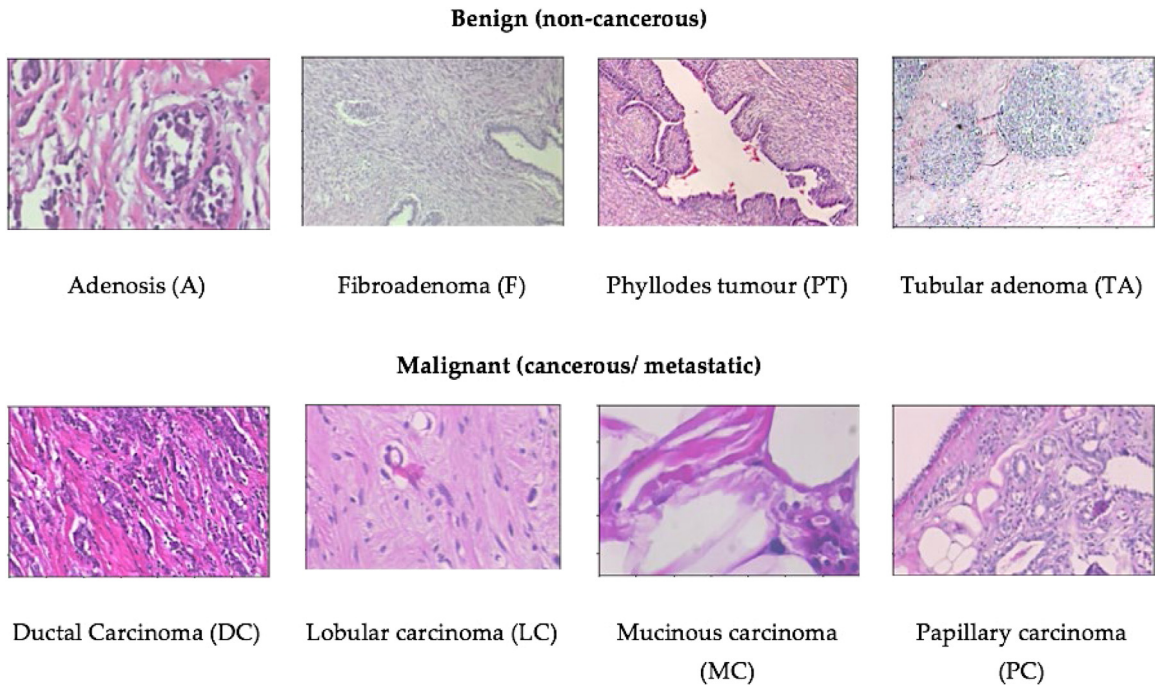


Fig. 1. Sample of distinct types of breast cancer histopathology images from BreaKHis dataset (Spanhol et al., 2016a, 2016b).

Table 2
Resampled experimental data.

Task	Label	Samples	Original	Resampled
Binary	0	Benign	2480	5429
	1	Malignant	5429	5429
		Total	**7909**	**10,858**
Multi-class	0	Fibroadenoma (F)	1014	1500
	1	Adenosis (A)	444	1500
	2	Tubular Adenoma (TA)	569	1500
	3	Phyllodes Tumour (PT)	453	1500
	4	Ductal Carcinoma (DC)	3451	1500
	5	Lobular Carcinoma (LC)	626	1500
	6	Mucinous Carcinoma (MC)	792	1500
	7	Papillary Carcinoma (PC)	560	1500
		Total	**7909**	**12,000**

search for a path with the minimal cost. To simplify this proposed method, they use the singular value decomposition method (SVD) to obtain the optical density of images to perform quantitative analysis-based colour normalization (Macenko et al., 2009). This method has been used by multiple researchers like (Alkassar et al., 2021; Araujo et al., 2017; Ciompi et al., 2017; Gandomkar et al., 2018; George et al., 2019; Kassani et al., 2020; Li et al., 2019; Liu et al., 2019; Rakhlin et al., 2018; Vo et al., 2019). Fig. 3 illustrates the histology image after performing colour normalization.

The images are later resized from $700 \times 460 \times 3$ into $224 \times 224 \times 3$ pixels to reduces the calculation dimensionality to support the training with a more efficient time. Then, the dataset is randomly shuffled to be split into training, testing, and validation set. This processing is to support the proposed network by predicting a small part of the dataset to show the dependability of the trained network. The ratio of splitting the dataset into training, testing, and validation is split by 80%, 16% and 4% respectively. Table 3 shows the experimental data in training, testing and validation after division.

Data augmentation is another commonly used technique in pre-processing for a small dataset. Thus, to produce a model with high accuracy and avoiding overfitting, we have also applied data augmentation on our limited dataset to produce a robust CAD system because a limited dataset is a common challenge for medical image analysis

Table 3
Division of experimental data in training, testing and validation dataset.

Task	Samples	Total samples	Image size
Binary	Training	8686	$224 \times 224 \times 3$
	Testing	1737	
	Validation	435	
	Total	**10,858**	
Multi-class	Training	9600	$224 \times 224 \times 3$
	Testing	1920	
	Validation	480	
	Total	**12,000**	

(Shorten & Khoshgoftaar, 2019). Data augmentation is a data-space solution to the problem of limited data by enhancing the size and quality of training datasets to generate a better and improved Deep Learning model (Shorten & Khoshgoftaar, 2019). This approach of pre-processing covers a set of techniques that augment the dataset by implying data wrapping and oversampling to artificially increase the sample size of the training dataset. This process can potentially strengthen the deep learning model, for example, it trains the algorithm to understand that rotation is not an important feature of the data. The main objectives we want to achieve in this process are: (1) Increase the sample size in training data, (2) Regularization, and (3) Solve imbalance classes in the dataset. Some techniques like flipping, cropping, rotation, shifting, scaling, zooming, and more have been applied in several previous research studies (Alom et al., 2019; Bayramoglu et al., 2016; Gandomkar et al., 2018; Kassani et al., 2020; Rakhlin et al., 2018; Toğaçar et al., 2020; Vo et al., 2019). Choosing the augmentation techniques is important because not all techniques are suitable to apply in histopathology images. Especially when histopathology images possess rotation and reflection symmetry (Veeling et al., 2018), selecting appropriate augmentation is important to avoid any loss on discriminating features or details. Here, we have applied several data augmentation with its values as shown in Table 4. Fig. 4 shows some examples of an image after applying data augmentation techniques.

Table 5 shows the total number of images generated by different data augmentation techniques at every epoch of training calculated by the respective number of batch size and training images.

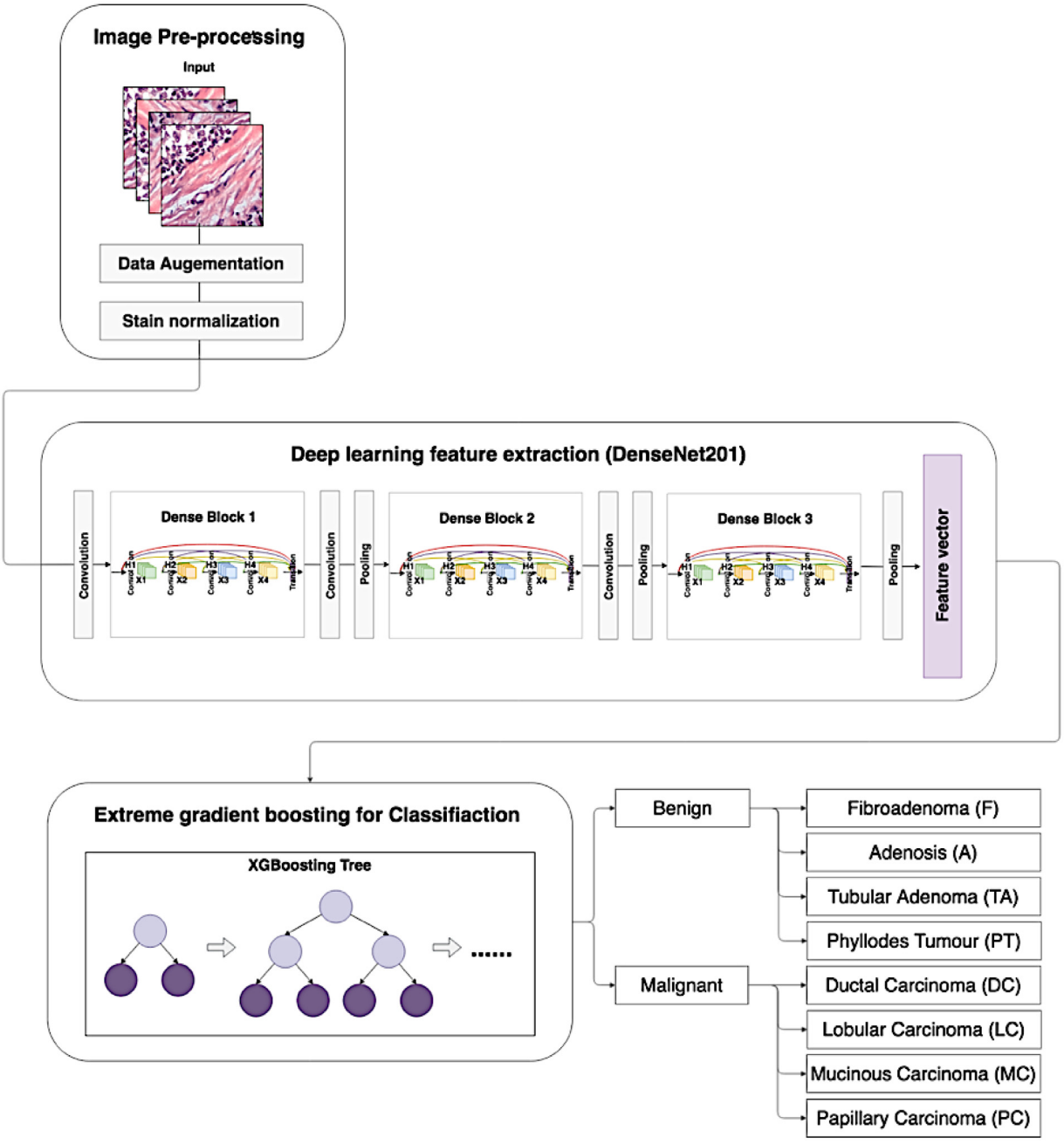
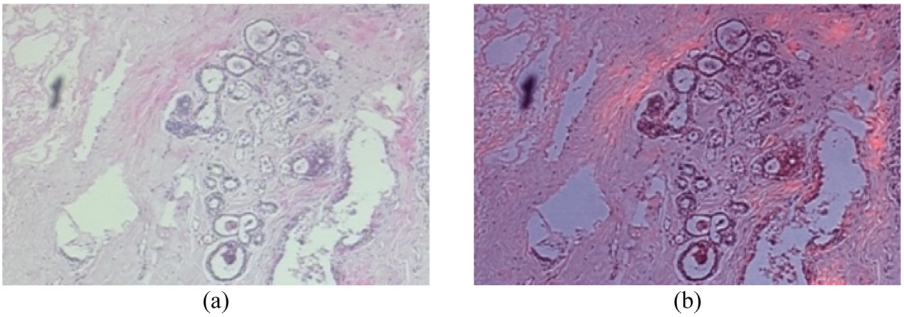


Fig. 2. Proposed DLXGB methodology for classifying breast tumours.



(a) (b)

Fig. 3. (a) Original image (b) Stain normalized image.

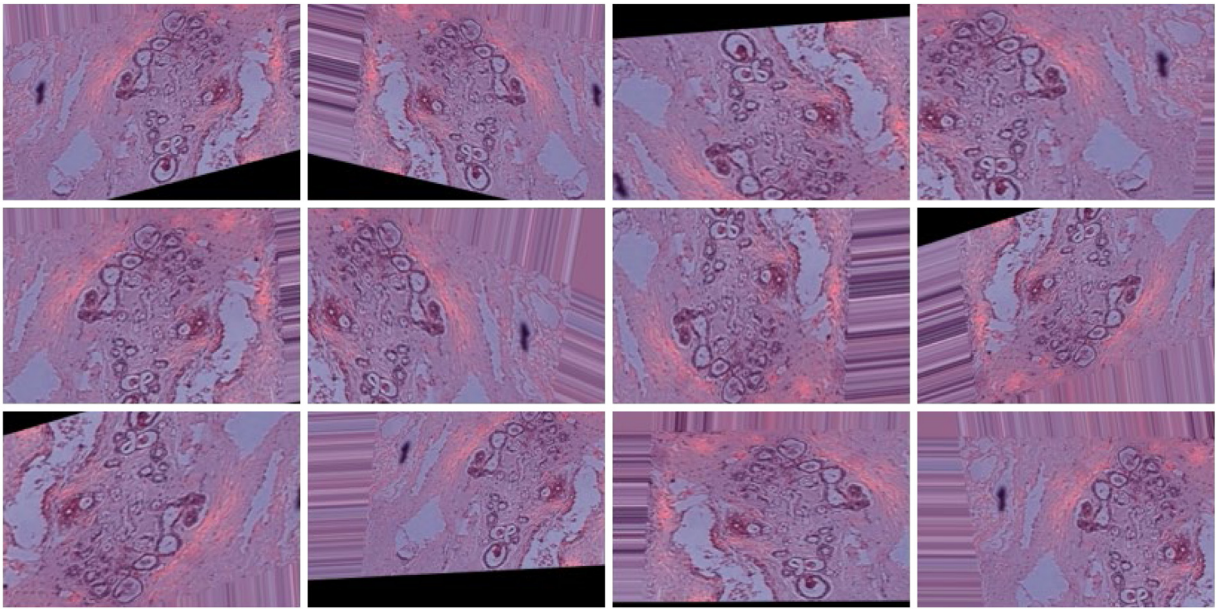


Fig. 4. Some examples of an image after data augmentation.

Table 4
Data augmentation techniques applied in the pre-processing step.

Data augmentation	Value
Rotation range	15
Shift range	0.2
Shear range	15
Horizontal flip	True
Vertical flip	True
Fill mode	'nearest'
Zoom range	0.2

Table 5
Number of images after applying data augmentation.

Classification task	Batch size	Number of images
Binary	10	869
Multi	10	960

3.4. Deep learning feature extraction

Feature extraction is an essential step to learn the features contain in breast cancer histopathology images and this process will highly influence the performance of the classification task. We have chosen to apply Deep Learning (DL) approach because the model will automatically figure out the features from an image based on each pixel from the image. Based on the idea of Convolution Neural Network (CNN), the convolution performs signal processing operation which can be easily computed as a discrete spatial processing operation (Wadkar et al., 2019). To implement deep learning feature extraction, there are two methods (1) Training from scratch and (2) Transfer Learning.

- **Training from scratch:** This approach will require an extensive amount of histology breast images to be input for training to be accurate and reliable. More effort and time will be needed to achieve this approach as it requires defining and fine-tuning the hyperparameters to obtain the best performance. This hyperparameter includes learning rate, number of layers, type of convolutional layer and much more. Besides the challenging task to tune the parameter, it also requires a high-performance GPU processing power to implement training (Bevilacqua et al., 2019).
- **Transfer Learning:** This approach provides a solution to problems like overfitting and a more generalized deep learning model by transferring knowledge learned from a source domain i.e., the ImageNet dataset that has a large amount of data to our target model. These pre-trained models can provide sufficient knowledge and prepare the small histology dataset in a deep learning model. This approach is aimed to use the knowledge learn from a set of data samples and apply that knowledge to any future samples that do not exist in the data. This way, the model will be able to transfer all the features/knowledge learn to make future predictions for new samples. learning has advantages such as speeding up the convergence of the network, reducing the computational power, and optimizing the network performance (Lu et al., 2019).

To conclude, instead of training from scratch to extract features/knowledge from input with randomly initialized weights, transfer learning provides a more efficient solution to exploit knowledge from cross domains. Hence, we adopted the state-of-art of transfer learning by using a pre-trained network in this process. Several popular deep learning-based models improved the CNN model for medical image classification tasks, such as AlexNet (Deng et al., 2009), VGGNet (Simonyan & Zisserman, 2015), GoogLeNet (Szegedy et al., 2015), Inception (Ioffe & Szegedy, 2015), DenseNet (Huang et al., 2017), Xception (Chollet, 2017) and ResNet (He et al., 2016).

3.5. DenseNet Feature extractor

In this expert system, we have chosen DenseNet as our final pre-trained feature extraction model to apply transfer learning. DesneNet is an improved version of ResNet that adopted the idea of using shortcut connections to eliminate at least two layers to avoid problems like gradient vanishing that limits a deeper exploration (Huang et al., 2017). This section for feature learning consists of several layers including the input layer, data-pre-processing layer, convolutional layer, and the last layer to reshape features vectors.

The traditional CNN architecture adds all the output features from the previous layer, however, the DenseNet architecture is designed to concatenate the feature maps sequentially, as formulated in Eq. (2). This process happens in the dense block of the network, Huang et al. describes the network as follows. The input of dense block can be expressed as $x_{\ell-1}$, then the output after two convolution layers can be expressed as $H_{\ell}(x_{\ell-1})$, this output will then be added with the shortcut

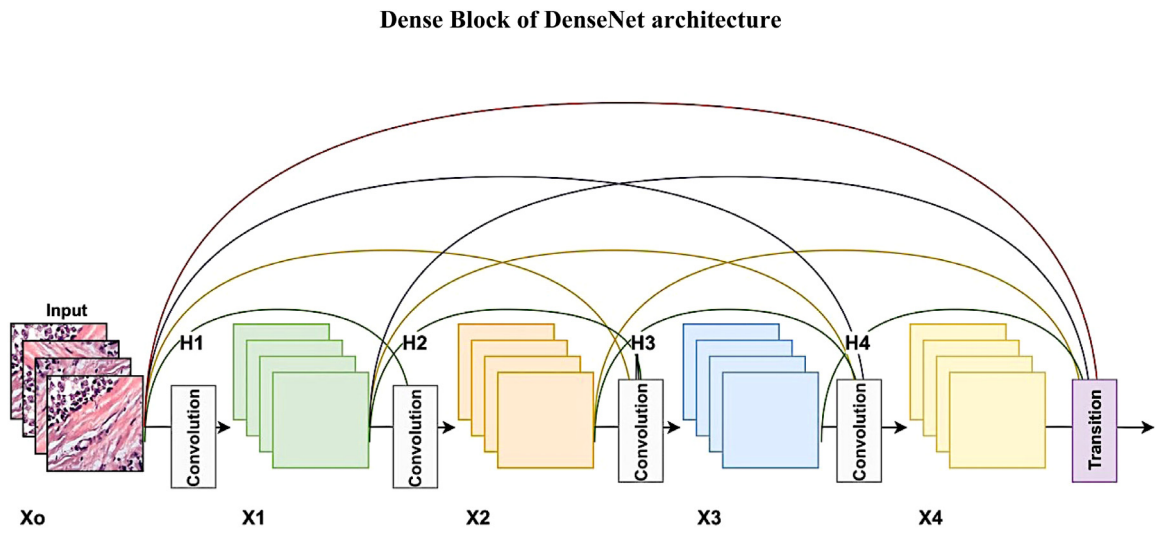


Fig. 5. Formation of a dense block with five layers (Huang et al., 2017).

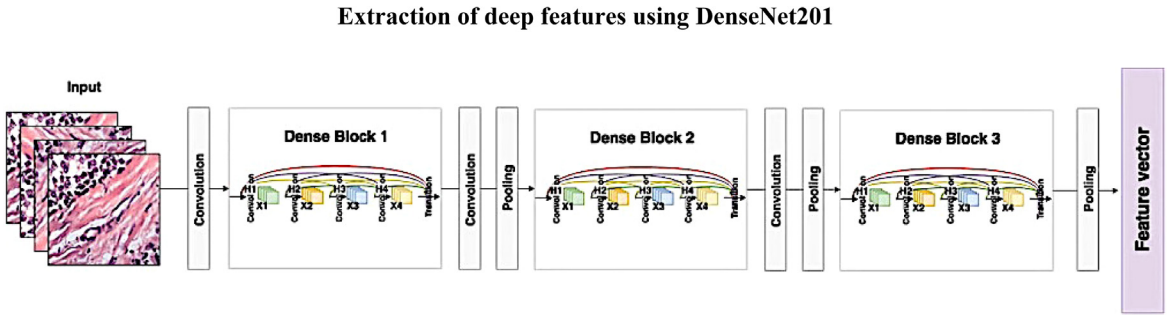


Fig. 6. Overview of network architecture for deep learning feature extraction (Huang et al., 2017).

to the input layer $x_{\ell-1}$. After the ℓ^{th} layer of dense block, the output will be added together, this process can be expressed in Eq. (1).

$$x_\ell = \mathrm{H}_\ell(x_{\ell-1}) + x_{\ell-1}, \tag{1}$$

$$x_\ell = \mathrm{H}_\ell([x_0, x_1, x_2, \ldots, x_{\ell-1}]) \tag{2}$$

where ℓ represents the layer index, H represents the non-linear operation, and x_ℓ describes the features extracted of the ℓ^{th} layer. Fig. 5 illustrates the dense block of DenseNet explained above. The first layer of the block has x_0 feature maps, the second layer has $x_0 + \mathrm{x}$ feature maps, plus the last layer has $x_0 + 4\mathrm{x}$ feature map. Every layer in the dense block will take in the previously extracted features maps as input (Huang et al., 2017). Fig. 6 illustrates the process of how transfer learning feature extraction is implemented through pre-trained DenseNet201 on breast cancer histology images by applying dense blocks.

3.6. Extreme gradient boosting classifier

The Extreme Gradient Boosting (XGBoost) is a new tree-based algorithm that has been increasing in popularity for data classification recently, that has been proved to be a highly effective method for data classification (Parashar et al., 2020). The XGBoost is a highly scalable end-to-end tree boosting system used in machine learning for classification and regression tasks (Chen & Guestrin, 2016). We have replaced the Fully Connected Layer (FCL) from the DenseNet201 with the XGBoost classifier. This is because the original FCL classifies on the ImageNet dataset which consists of non-medical related images. The authors that proposed this method, Chen and Guestrin have explained their concept of approach in detail. This method is new, so we have summarized the calculations and definitions in the remaining section.

First, a tree ensemble method of classification and regression trees (CARTs) with a set of $K_E^i | i \in 1 \ldots K$ nodes. The final prediction output of class label $\hat{y}_i$ is calculated based on the total prediction scores at a leaf node f_k for each tree k^{th}. As expressed in Eq. (3).

$$\hat{\boldsymbol{y}}_{\boldsymbol{i}} = \boldsymbol{\varphi}(\boldsymbol{x}_{\boldsymbol{i}}) = \sum_{k=1}^{K} \boldsymbol{f}_{\boldsymbol{k}}(\boldsymbol{x}_{\boldsymbol{i}}), \quad \boldsymbol{f}_{\boldsymbol{k}} \in \boldsymbol{F}, \tag{3}$$

where $\boldsymbol{x}_{\boldsymbol{i}}$ is the training set and F represents the set of all K scores for all CARTs. Then, a regularization step is applied to improve the results, as shown in Eq. (4).

$$\mathcal{L}(\varphi) = \sum_i \ell(\hat{y}_i, y_i) + \sum_k \Omega(f_k), \tag{4}$$

where ℓ represents the differentiable loss function, define by calculating the error difference between target y_i and predicted class labels $\hat{y}_i$. The second part performs penalization Ω on the model complexity to avoid over-fitting problems. The function for the penalty Ω is calculated by Eq. (5).

$$\boldsymbol{\Omega}(\boldsymbol{f}) = \boldsymbol{\gamma T} + \frac{\mathbf{1}}{\mathbf{2}} \boldsymbol{\lambda} \sum_{j=1}^{T} \boldsymbol{w}_j^2, \tag{5}$$

where γ and λ are configurable parameters to control the degree of regularization. T represents the leaves in the tree and w stores the value of weights for each leaf.

Then, Gradient Boosting (GB) is applied to effectively solve the classification problem along with the loss function and extended by a second Taylor expansion. The constant term will be removed to obtain

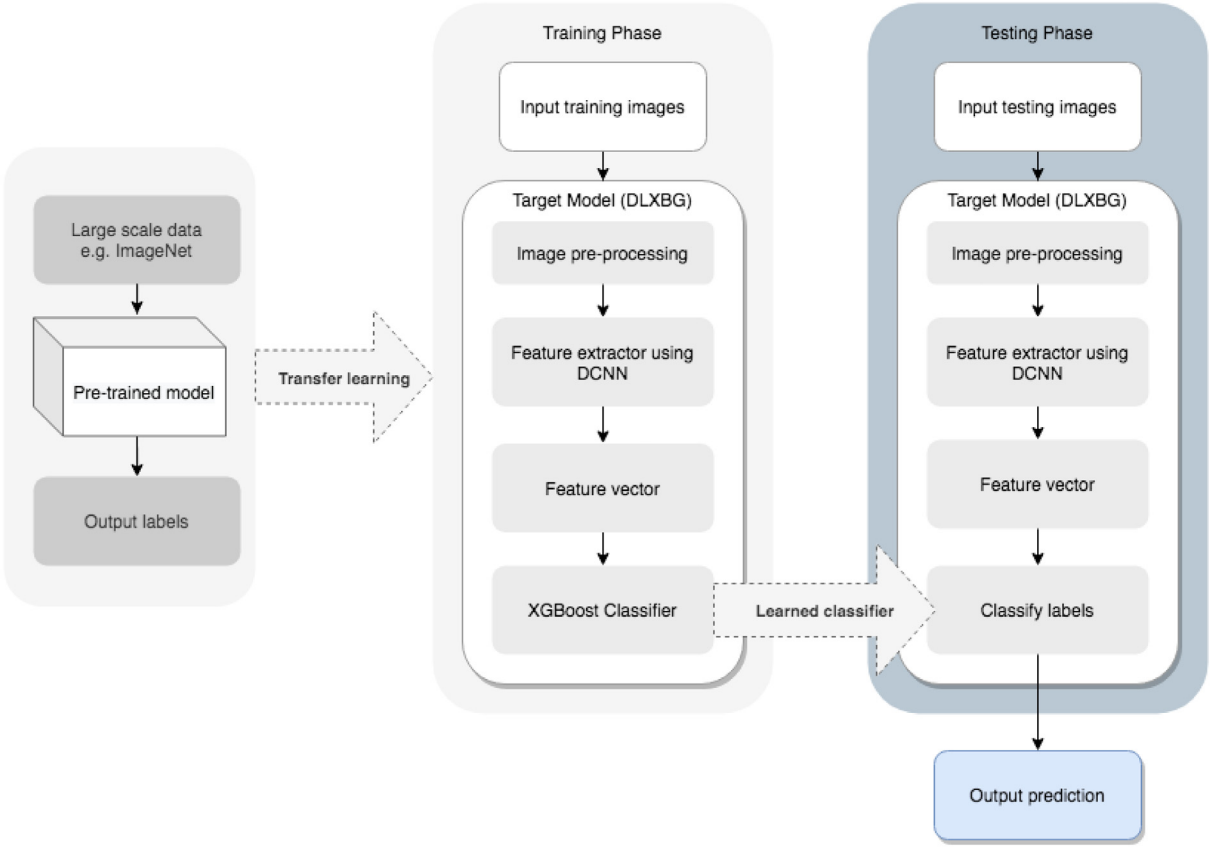


Fig. 7. Overview of the proposed method.

a simplified objective at step t, as calculated in Eq. (6)

$$\begin{aligned}\tilde{\mathcal{L}}^{(t)} &= \sum_{i=1}^{n}\left[g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)\right] + \Omega(f_t) \\ &= \sum_{i=1}^{n}\left[g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)\right] + \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2 \\ &= \sum_{j=1}^{T}\left[(\sum_{i\in I_j} g_i)w_j + \frac{1}{2}(\sum_{i\in I_j} h_i + \lambda)w_j^2\right] + \gamma T\end{aligned} \tag{6}$$

where $I_j = \{i|q(x_i) = j\}$ denotes the instance of leaf t, and the equation for first g_i and second h_i order gradient statistics of the loss function are defined in Eqs. (7)–(8).

$$g_i = \frac{\partial \ell(\hat{y}_i^{(t-1)}, y_i)}{\partial \hat{y}_i^{(t-1)}} \tag{7}$$

$$h_i = \frac{\partial^2 \ell(\hat{y}_i^{(t-1)}, y_i)}{\partial (\hat{y}_i^{(t-1)})^2} \tag{8}$$

The optimal weight w_j^* of leaf j can be then calculated by Eq. (9).

$$w_j^* = \frac{\sum_{i\in Ij} g_i}{\sum_{i\in Ij} h_i + \lambda} \tag{9}$$

A function to be used as a scoring function to measure the quality of a tree structure q, for a given tree structure $q(x_i)$ can be calculated by Eq. (10).

$$\tilde{\mathcal{L}}^{(t)}(q) = -\frac{1}{2}\sum_{j=1} T\frac{(\sum_{i\in Ij} g_i)^2}{\sum_{i\in Ij} h_i + \lambda} + \gamma T \tag{10}$$

Typically, to measure the split nodes by applying scoring in the instance set of left I_L and right I_R nodes after the splitting are done, the loss reduction after splitting is calculated in Eq. (11).

$$\mathcal{L}_{split} = \frac{1}{2}\left[\frac{(\sum_{i\in I_L} g_i)^2}{\sum_{i\in I_L} h_i + \lambda} + \frac{(\sum_{i\in I_R} g_i)^2}{\sum_{i\in I_R} h_i + \lambda} + \frac{(\sum_{i\in I} g_i)^2}{\sum_{i\in I} h_i + \lambda}\right] - \gamma \tag{11}$$

where $I = I_R \cup I_L$.

Table 6
Confusion matrix for binary classification obtained through the best results across 5-fold cross-validation.

	Benign	Malignant
Benign	1058	26
Malignant	46	1042

4. Experiments & results

In this research, we have chosen to focus on developing a breast cancer CAD expert system using Deep Learning and XGBoost. The proposed network is implemented with Python 3.7.9, image processing for colour normalization is implemented on MATLAB R2020a. Experiments are performed in Google Colab, a cloud-based environment with 25 GB of RAM (Random Access Memory). Fig. 7 illustrates the general overview of our proposed method.

4.1. Experimental setting

4.1.1. Configuration of DenseNet201

Training procedure To perform training of the model of DenseNet201 in our proposed system, we train the network using RMSprop with the learning rate starting from 1e-4 and gradually lowered to be closer to zero. The initial weights trained on the ImageNet dataset was used at the beginning of the network. Then, further fine-tuning

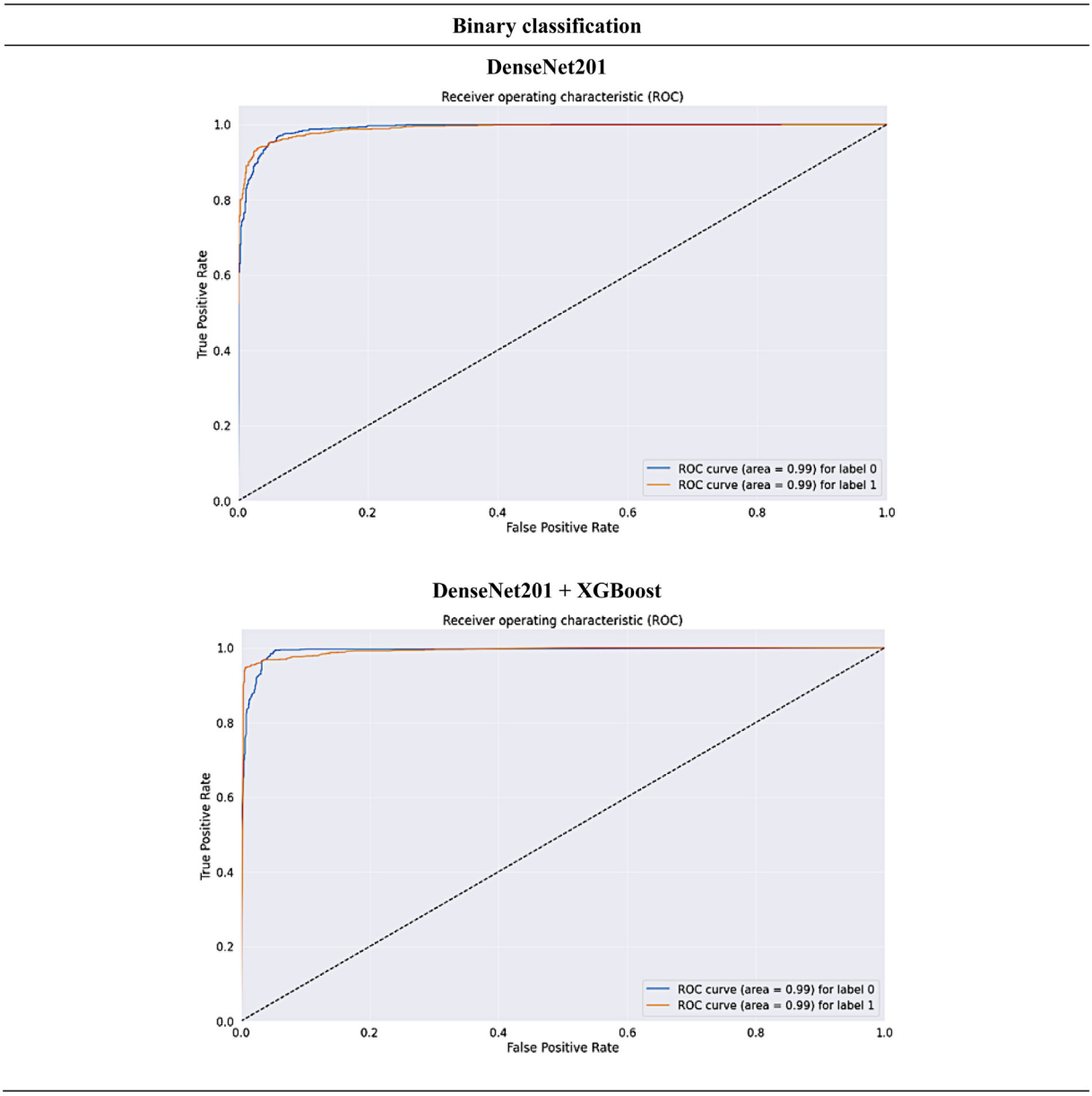


Fig. 8. ROC Curve of binary classification obtained through the best results across 5-fold cross-validation.

was performed with the BreakHis dataset within the connection of the final layers of the network. Image augmentation, data resampling and image resize are implemented to normalize and balance the dataset. All images are resized to a width and height of 224 pixels. This has potentially reduced the computational time during training. The training of the model consist of 8 epochs and the size of batch size is 10.

Loss function The function used to compute loss is the cross-entropy loss, also known as log loss. This loss function measures the performance of a classification model using a probability computation between 0 and 1. The value of cross-entropy loss will gradually increase if the predicted probability differs from the actual label.

Early stopping: We have also included a learning rate reducer as early stopping criteria by implementing the 'ReduceLROnPlateau' function. This function will observe the validation loss and stops the training when the model does not further improve. The function was configured with the following parameter settings:

- monitor = 'val_loss'
- patience = 5
- verbose = 1
- factor = 0.2
- min_lr = 1e–7

4.1.2. Configuration of XGBoost classifier

After training the DenseNet201 model, we use the model to extract the key features within the training data, which will later be learned by the classifier. The default settings to train the eXtreme Gradient Boosting (XGBoost) Classifier:

- learning_rate = 0.6
- n_estimators = 1000
- max_depth = 5
- min_child_weight = 11
- gamma = 0.1
- subsample = 0.8
- colsample_bytree = 0.7
- objective = 'multi:softprob'
- n_jobs = −1

Table 7
Confusion matrix for multi-classification obtained through the best results across 5-fold cross-validation.

	Fibroadenoma (F)	Adenosis (A)	Tubular Adenoma (TA)	Phyllodes Tumour (PT)	Ductal Carcinoma (DC)	Lobular Carcinoma (LC)	Mucinous Carcinoma (MC)	Papillary Carcinoma (PC)
Fibroadenoma (F)	295	0	1	9	1	0	0	0
Adenosis (A)	1	300	0	1	0	0	0	2
Tubular Adenoma (TA)	3	0	294	0	1	0	1	0
Phyllodes Tumour (PT)	1	0	0	294	1	0	0	0
Ductal Carcinoma (DC)	0	0	0	0	261	18	20	2
Lobular Carcinoma (LC)	0	0	0	0	12	291	3	0
Mucinous Carcinoma (MC)	0	0	0	0	1	1	275	0
Papillary Carcinoma (PC)	1	0	2	0	0	0	0	318

Table 8
Evaluation of average result of classification task obtained in 5-fold cross-validation.

Task	Approach	Classes	Precision	Recall	F1-score
Binary	DenseNet 201	Benign	0.94	0.95	0.95
		Malignant	0.95	0.95	0.95
		Accuracy			**0.95**
	DenseNet 201 + XGBoost	Benign	0.96	0.98	0.97
		Malignant	0.98	0.96	0.97
		Accuracy			**0.97**
Multi-class	DenseNet 201	Fibroadenoma (F)	0.94	0.80	0.86
		Adenosis (A)	0.97	0.96	0.97
		Tubular Adenoma (TA)	0.95	0.97	0.96
		Phyllodes Tumour (PT)	0.86	1.00	0.92
		Ductal Carcinoma (DC)	0.79	0.82	0.81
		Lobular Carcinoma (LC)	0.94	0.84	0.89
		Mucinous Carcinoma (MC)	0.93	0.88	0.91
		Papillary Carcinoma (PC)	0.89	0.99	0.94
		Accuracy			**0.91**
	DenseNet 201 + XGBoost	Fibroadenoma (F)	0.98	0.96	0.97
		Adenosis (A)	1.00	0.99	0.99
		Tubular Adenoma (TA)	0.99	0.98	0.99
		Phyllodes Tumour (PT)	0.97	0.99	0.98
		Ductal Carcinoma (DC)	0.94	0.90	0.92
		Lobular Carcinoma (LC)	0.94	0.95	0.94
		Mucinous Carcinoma (MC)	0.95	0.99	0.97
		Papillary Carcinoma (PC)	0.99	0.99	0.99
		Accuracy			**0.97**

Table 9
Comparison of the proposed method with existing deep learning approaches in classification using the BreaKHis dataset. The results presented are the best accuracies obtained.

Reference	Classification	Method used	Results
Spanhol et al. (2017)	Multi-class	Modified AlexNet and DeCAF (Deep Convolutional Activation Feature)	81.5–86.3%
Han et al. (2017)	Multi-class	Custom CSDCNN (Class Structure-based Deep Convolutional Neural Network) based on GoogLeNet	93.2%
Bardou et al. (2018)	Multi-class	CNN	83.31%–88.23%
Vo et al. (2019)	Multi-class	Inception and ResNet CNN (IRRCNN) and Gradient boosting trees	93.8%–96.9%
Budak et al. (2019)	Binary	BiRNN and LSTM	91.90%
Nedjar et al. (2019)	Multi-class	Consensus oriented by InceptionV3, ResNet50 and MobileNet	91.06%
Yari and Nguyen (2020)	Multi-class	DenseNet121	97%–100%
Toğaçar et al. (2020)	Binary	CNN features with Multi-Layer Perceptron	98.80%
Al-Haija and Adebanjo (2020)	Binary	ResNet-50	99.10%
Alkassar et al. (2021)	Multi-class	Xception and DenseNet CNNs	92%
Kushwaha et al. (2021)	Binary	DenseNet201	97.05%
Boumaraf et al. (2021)	Binary and Multi-class	ResNet-18	92.03–98.42%
Proposed	**Binary and Multi-class**	**DenseNet201 and XGBoost**	**97%**

- scale_pos_weight = 1
- seed = 1234

4.2. Results and discussions

To show the effectiveness of a CAD system for breast cancer diagnosis, it is important to evaluate our approaches to understand the performance of the system quantitatively as well as inspecting the underlying problems to be improved. Commonly used metrics for diagnosis evaluation are F1-measure, Precision, Recall, ROC Curve and Accuracy. All experiments were performed using the balanced BreakHis dataset obtained through resampling. The final proposed DLXGB model was trained and tested on 10,858 images for binary classification, and 12,000 images for multi-classification. To further reduce bias in a limited size of the dataset and build a more generalized model, a K-fold cross-validation technique with a value of 5 was used during the training and testing of the model. The implementation of this technique allows the data to be randomly divided into k numbers of equal groups, and the holdout method will be repeated until all unique groups have been used as a test set. The results will be calculated through the average performance across all k trials.

All results shown are the average performance of 5-fold cross-validation. Tables 6 and 7 presents the confusion matrix of both binary and multi-classification based on the obtained True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). Table 8 compares the approach of general Deep Learning with DenseNet201 and our approach of combining DenseNet201 with XGBoost classifier on the binary and multi-classification task. From this table, we can

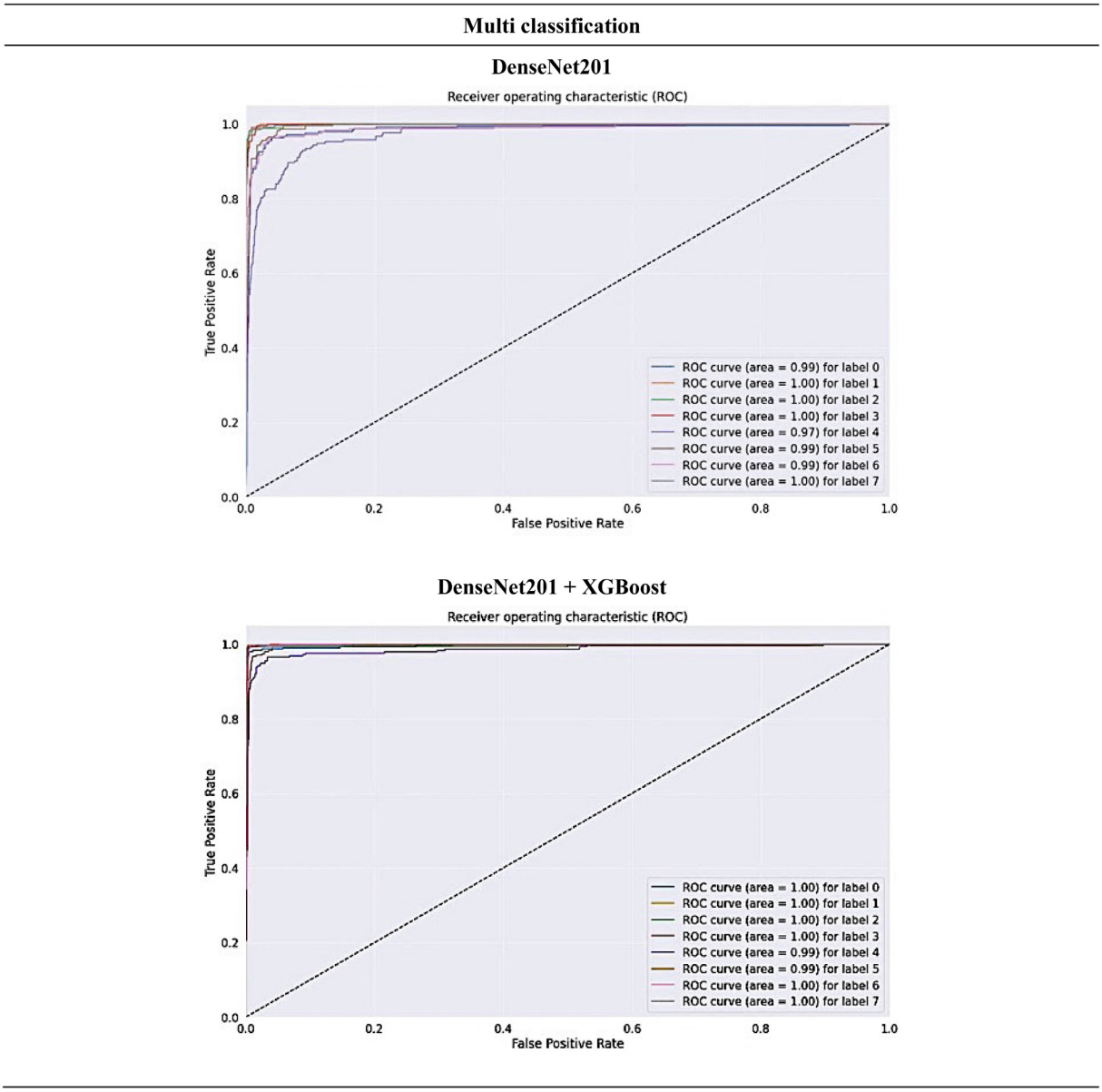


Fig. 9. ROC Curve of multi-classification obtained through the best results across 5-fold cross-validation.

observe that our approach of replacing the FCL of DenseNet201 to XGBoost classifier improves in both binary and multi-classification. Figs. 8 and 9 present the ROC curve comparison of the effect of XGBoost on Deep Learning on binary and multi-classification, respectively. Lastly, Table 9 summarizes the comparison of our approach and existing literature work on the BreaKHis dataset classification task.

While comparing the results of existing work to our proposed method in Table 9, it is a fair comparison as the dataset used is the same and the classification task is also consistent. Our proposed method is an improved method as we can observe from the results, it has achieved a promising result based on accuracy with similar conditions of existing work. Besides, the method can avoid overfitting and achieve generalization better than the existing literature. To further explain this, although the proposed method by Yari & Nguyen and Al-Haija & Adebanjo achieved an accuracy range of 97%–100%, their results were obtained through an imbalance dataset which is highly prone to overfitting and bias in the model. Yari and Nguyen mentioned that even though their precision and recall reaches 100%, it usually indicates some level of overfitting. Therefore, our method has improved by implementing data resampling to tackle this problem. The method proposed by Vo et al. have a high accuracy as our proposed method, however, their method suffers the same imbalance data issue. One of the future improvement that mentioned by the authors was to improve classification performance by implementing boosting tree classifier. In which, our proposed method have adapted their challenges and suggestions by using XGBoost classifier for improved classification performance. At this point, we know that the BreakHis dataset is widely used by many researchers, however, the dataset exist obvious data imbalance problems causing bias in the model. BreaKHis dataset was also described in one of the surveys by Benhammou Y. et al. to have a data imbalance issue (Benhammou et al., 2020). Nevertheless, the proposed method by Toğaçar et al. and Boumaraf et al. faces the same issue as mentioned, showing our method is potentially more robust and less biased. Besides, the approach by Boumaraf et al. deteriorated during multi-classification with a decrease in accuracy of 6% from their binary classification, in which our improved method did not deteriorate in different tasks. The approach by Kushwaha et al. not only has a data imbalance issue, but their method could potentially have invariance issues as well because they did not implement data augmentation, in which our method has improved with this implementation.

We can also notice that combining more than one model produces more superior results than just a single model for classifying breast histology images. The strength of this method is that it covers common classification problems in histology images like bias caused by imbalance classes, model variance, and inconsistent colour variance in images. Furthermore, it improves a pre-trained network that was trained from classifying objects by changing its classifier strength to classifying histology images. However, some limitations can be further improved in our method. First, the hyperparameter of the CNN backbone is barely fine-tuned, which we believe if further fine-tuning is performed, it could potentially improve the feature extraction deep learning model even better. Besides, due to computational resources and data limitations, we were not able to pre-trained a domain-specific model with reliable results for a CAD expert system.

5. Conclusion & future work

In this paper, we have presented an improved classification expert system for breast cancer using machine learning techniques. The presented paper aims to explore innovative approaches for breast cancer classification by adapting the latest machine learning mechanisms. To be the best of your knowledge, our approach has not been explored by any related work for breast cancer histology images, only a small amount of research is done that used XGBoost, even these existing works only investigate breast mammography images. This proposed method has achieved a high performance by combining deep learning feature extraction and XGBoost classifier for classification problems. The expert system can identify the classes of benign and malignant for binary classification. Additionally, identifying Adenosis (A), Fibroadenoma (F), Phyllodes Tumour (PT), And Tubular Adenoma (TA), Ductal Carcinoma (DC), Lobular Carcinoma (LC), Mucinous Carcinoma (MC), And Papillary Carcinoma (PC) of multi-classification. The method adopted Deep Learning techniques by using pre-trained network DenseNet201 to learn features from the histology images. Then, the XGBoost classifier performs the classification task based on the extracted feature vectors. This method is new and requires more exploration and investigation. Overall, we can conclude that this method is potentially effective in classifying medical images in developing a CAD expert system. We have achieved an average accuracy of 97% for both binary and multi-classification. Data augmentation and data balancing were implemented to avoid overfitting and bias.

The paper intends to investigate a new exploration on a combination of DenseNet201 and XGBoost. Due to time limitations, we are not able to investigate further on the speed, space, or footprint etc. Therefore, the manuscripts do not claim to improve for better performance within these scopes. In the future, some more spaces and directions can be explored to improve this method. These include investigating the time and memory efficiency of this method. Besides, evaluating this method on another similar dataset can also be performed to further confirm this method. Further experiments and fine-tuning could also be investigated for a potentially better parameter setting. We will also focus on developing the expert system to be mobile compatible, this is to make it more convenient to be accessible to all users, especially those that have limited resources in computer systems.

CRediT authorship contribution statement

Xin Yu Liew: Conceptualization, Methodology, Software, Data curation, Writing – original draft. **Nazia Hameed:** Supervision, Writing – review & editing. **Jeremie Clos:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Al-Haija, Q. A., & Adebanjo, A. (2020). Breast cancer diagnosis in histopathological images using ResNet-50 convolutional neural network. In *IEMTRONICS 2020 - International IOT, electronics and mechatronics conference, proceedings*. http://dx.doi.org/10.1109/IEMTRONICS51293.2020.9216455.

Alirezazadeh, P., Hejrati, B., Monsef-Esfahani, A., & Fathi, A. (2018). Representation learning-based unsupervised domain adaptation for classification of breast cancer histopathology images. *Biocybernetics and Biomedical Engineering*, *38*(3), http://dx.doi.org/10.1016/j.bbe.2018.04.008.

Alkassar, S., Jebur, B. A., Abdullah, M. A. M., Al-Khalidy, J. H., & Chambers, J. A. (2021). Going deeper: magnification-invariant approach for breast cancer classification using histopathological images. *IET Computer Vision*, *15*(2), 151–164. http://dx.doi.org/10.1049/cvi2.12021.

Alom, M. Z., Yakopcic, C., Nasrin, M. S., Taha, T. M., & Asari, V. K. (2019). Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network. *Journal of Digital Imaging*, *32*(4), http://dx.doi.org/10.1007/s10278-019-00182-7.

Amato, F., López, A., Peña Méndez, E. M., Vaňhara, P., Hampl, A., & Havel, J. (2013). Artificial neural networks in medical diagnosis. *Journal of Applied Biomedicine*, *11*(2), http://dx.doi.org/10.2478/v10136-012-0031-x.

Araujo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., Polonia, A., & Campilho, A. (2017). Classification of breast cancer histology images using convolutional neural networks. *PLoS ONE*, *12*(6), http://dx.doi.org/10.1371/journal.pone.0177544.

Aresta, G., Araújo, T., Kwok, S., Chennamsetty, S. S., Safwan, M., Alex, V., Marami, B., Prastawa, M., Chan, M., Donovan, M., Fernandez, G., Zeineh, J., Kohl, M., Walz, C., Ludwig, F., Braunewell, S., Baust, M., Vu, Q. D., To ..., M. N. N., & Aguiar, P. (2019). BACH: Grand challenge on breast cancer histology images. *Medical Image Analysis*, *56*, http://dx.doi.org/10.1016/j.media.2019.05.010.

Bardou, D., Zhang, K., & Ahmad, S. M. (2018). Classification of breast cancer based on histology images using convolutional neural networks. *IEEE Access*, *6*, http://dx.doi.org/10.1109/ACCESS.2018.2831280.

Bayramoglu, N., Kannala, J., & Heikkila, J. (2016). Deep learning for magnification independent breast cancer histopathology image classification. In *Proceedings - International conference on pattern recognition*. http://dx.doi.org/10.1109/ICPR.2016.7900002.

Benhammou, Y., Achchab, B., Herrera, F., & Tabik, S. (2020). BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights. *Neurocomputing*, *375*, http://dx.doi.org/10.1016/j.neucom.2019.09.044.

Bevilacqua, V., Brunetti, A., Guerriero, A., Trotta, G. F., Telegrafo, M., & Moschetta, M. (2019). A performance comparison between shallow and deeper neural networks supervised classification of tomosynthesis breast lesions images. *Cognitive Systems Research*, *53*, 3–19. http://dx.doi.org/10.1016/j.cogsys.2018.04.011.

Boumaraf, S., Liu, X., Zheng, Z., Ma, X., & Ferkous, C. (2021). A new transfer learning based approach to magnification dependent and independent classification of breast cancer in histopathological images. *Biomedical Signal Processing and Control*, *63*, http://dx.doi.org/10.1016/j.bspc.2020.102192.

Breast Cancer Organization (2016). *What is breast cancer? | Breastcancer.org* (pp. 1–19). Breastcancer.Org, https://www.breastcancer.org/symptoms/understand_bc/what_is_bc.

Budak, Ü., Cömert, Z., Rashid, Z. N., Şengür, A., & Çıbuk, M. (2019). Computer-aided diagnosis system combining FCN and Bi-LSTM model for efficient breast cancer detection from histopathological images. *Applied Soft Computing*, *85*, http://dx.doi.org/10.1016/j.asoc.2019.105765.

Cai, D., Sun, X., Zhou, N., Han, X., & Yao, J. (2019). Efficient mitosis detection in breast cancer histology images by RCNN. In *Proceedings - International symposium on biomedical imaging, 2019-April*. http://dx.doi.org/10.1109/ISBI.2019.8759461.

Cancer Research UK (2014). *Breast cancer statistics | Cancer Research UK*. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer#heading-Zero.

Chan, A., & Tuszynski, J. A. (2016). Automatic prediction of tumour malignancy in breast cancer with fractal dimension. *Royal Society Open Science*, *3*(12), http://dx.doi.org/10.1098/rsos.160558.

Chen, T., & Guestrin, C. (2016). *XGBoost*. http://dx.doi.org/10.1145/2939672.2939785.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings - 30th IEEE conference on computer vision and pattern recognition, CVPR 2017, 2017-January*. http://dx.doi.org/10.1109/CVPR.2017.195.

Ciompi, F., Geessink, O., Bejnordi, B. E., De Souza, G. S., Baidoshvili, A., Litjens, G., Van Ginneken, B., Nagtegaal, I., & Van Der Laak, J. (2017). The importance of stain normalization in colorectal tissue classification with convolutional networks. In *Proceedings - International symposium on biomedical imaging*. http://dx.doi.org/10.1109/ISBI.2017.7950492.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), http://dx.doi.org/10.1023/A:1022627411411.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, Kai, & Fei-Fei, Li (2009). ImageNet: A large-scale hierarchical image database. (pp. 248–255). http://dx.doi.org/10.1109/cvpr.2009.5206848.

Gandomkar, Z., Brennan, P. C., & Mello-Thoms, C. (2018). MuDeRN: Multi-category classification of breast histopathological image using deep residual networks. *Artificial Intelligence in Medicine*, *88*, http://dx.doi.org/10.1016/j.artmed.2018.04.005.

George, K., Faziludeen, S., Sankaran, P., & Paul, J. K. (2019). Deep learned nucleus features for breast cancer histopathological image analysis based on belief theoretical classifier fusion. In *IEEE region 10 annual international conference, proceedings/TENCON, 2019-October*. http://dx.doi.org/10.1109/TENCON.2019.8929539.

Ghaznavi, F., Evans, A., Madabhushi, A., & Feldman, M. (2013). Digital imaging in pathology: Whole-slide imaging and beyond. In *Annual review of pathology: Mechanisms of disease, Vol. 8*. http://dx.doi.org/10.1146/annurev-pathol-011811-120902.

Han, Z., Wei, B., Zheng, Y., Yin, Y., Li, K., & Li, S. (2017). Breast cancer multi-classification from histopathological images with structured deep learning model. *Scientific Reports*, *7*(1), http://dx.doi.org/10.1038/s41598-017-04075-z.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), http://dx.doi.org/10.1109/TKDE.2008.239.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE computer society conference on computer vision and pattern recognition, 2016-December*. http://dx.doi.org/10.1109/CVPR.2016.90.

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(8), http://dx.doi.org/10.1109/34.709601.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings - 30th IEEE conference on computer vision and pattern recognition, CVPR 2017, 2017-January*. http://dx.doi.org/10.1109/CVPR.2017.243.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *32nd international conference on machine learning, ICML 2015, 1*.

Irsoy, O., & Alpaydin, E. (2020). Continuously constructive deep neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, *31*(4), http://dx.doi.org/10.1109/TNNLS.2019.2918225.

Kassani, S. H., Kassani, P. H., Wesolowski, M. J., Schneider, K. A., & Deters, R. (2020). Classification of histopathological biopsy images using ensemble of deep learning networks. In *CASCON 2019 Proceedings - Conference of the centre for advanced studies on collaborative research - Proceedings of the 29th annual international conference on computer science and software engineering*.

Khan, S. U., Islam, N., Jan, Z., Ud Din, I., & Rodrigues, J. J. P. C. (2019). A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recognition Letters*, *125*, http://dx.doi.org/10.1016/j.patrec.2019.03.022.

Kushwaha, S., Adil, M., Abuzar, M., Nazeer, A., & Singh, S. K. (2021). Deep learning-based model for breast cancer histopathology image classification. In *Proceedings of 2021 2nd international conference on intelligent engineering and management, ICIEM 2021*. http://dx.doi.org/10.1109/ICIEM51511.2021.9445319.

Laxmisagar, H. S., & Hanumantharaju, M. C. (2021). Design of an efficient deep neural network for multi-level classification of breast cancer histology images. In *Advances in intelligent systems and computing, Vol. 1172*. http://dx.doi.org/10.1007/978-981-15-5566-4_40.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), http://dx.doi.org/10.1038/nature14539.

LeCun, Y., Kavukcuoglu, K., & Farabet, C. (2010). Convolutional networks and applications in vision. In *ISCAS 2010-2010 IEEE international symposium on circuits and systems: Nano-bio circuit fabrics and systems*. http://dx.doi.org/10.1109/ISCAS.2010.5537907.

Li, Y., Wu, J., & Wu, Q. (2019). Classification of breast cancer histology images using multi-size and discriminative patches based on deep learning. *IEEE Access*, *7*, http://dx.doi.org/10.1109/ACCESS.2019.2898044.

Litjens, G., Bandi, P., Bejnordi, B. E., Geessink, O., Balkenhol, M., Bult, P., Halilovic, A., Hermsen, M., van de Loo, R., Vogels, R., Manson, Q. F., Stathonikos, N., Baidoshvili, A., van Diest, P., Wauters, C., van Dijk, M., & van der Laak, J. (2018). 1399 H & E-stained sentinel lymph node sections of breast cancer patients: The CAMELYON dataset. *GigaScience*, *7*(6), http://dx.doi.org/10.1093/gigascience/giy065.

Liu, Y., Kohlberger, T., Norouzi, M., Dahl, G. E., Smith, J. L., Mohtashamian, A., Olson, N., Peng, L. H., Hipp, J. D., & Stumpe, M. C. (2019). Artificial intelligence–based breast cancer nodal metastasis detection insights into the black box for pathologists. *Archives of Pathology and Laboratory Medicine*, *143*(7), http://dx.doi.org/10.5858/arpa.2018-0147-OA.

Lu, S., Lu, Z., & Zhang, Y. D. (2019). Pathological brain detection based on AlexNet and transfer learning. *Journal of Computer Science*, *30*, http://dx.doi.org/10.1016/j.jocs.2018.11.008.

Macenko, M., Niethammer, M., Marron, J. S., Borland, D., Woosley, J. T., Guan, X., Schmitt, C., & Thomas, N. E. (2009). A method for normalizing histology slides for quantitative analysis. In *Proceedings - 2009 IEEE international symposium on biomedical imaging: From nano to macro, ISBI 2009* (pp. 1107–1110). http://dx.doi.org/10.1109/ISBI.2009.5193250.

Mahmood, T., Arsalan, M., Owais, M., Lee, M. B., & Park, K. R. (2020). Artificial intelligence-based mitosis detection in breast cancer histopathology images using faster R-CNN and deep CNNs. *Journal of Clinical Medicine*, *9*(3), http://dx.doi.org/10.3390/jcm9030749.

MITOS-ATYPIA-14 Grand Challenge (2014). https://mitos-atypia-14.grand-challenge.org/.

Munien, C., & Viriri, S. (2021). Classification of hematoxylin and eosin-stained breast cancer histology microscopy images using transfer learning with EfficientNets. *Computational Intelligence and Neuroscience*, *2021*, http://dx.doi.org/10.1155/2021/5580914.

Murtaza, G., Shuib, L., Mujtaba, G., & Raza, G. (2020). Breast cancer multi-classification through deep neural network and hierarchical classification approach. *Multimedia Tools and Applications*, *79*(21–22), http://dx.doi.org/10.1007/s11042-019-7525-4.

Nedjar, I., Mahmoudi, S., & Chikh, M. A. (2019). Classification of breast cancer histology images using consensus oriented by three deep convolutional neural networks. *International Journal of Computer Assisted Radiology and Surgery*, *14*.

Pantanowitz, L., Evans, A., Pfeifer, J., Collins, L., Valenstein, P., Kaplan, K., Wilbur, D., & Colgan, T. (2011). Review of the current state of whole slide imaging in pathology. *Journal of Pathology Informatics*, *2*(1), http://dx.doi.org/10.4103/2153-3539.83746.

Parashar, J., Sumiti, & Rai, M. (2020). Breast cancer images classification by clustering of ROI and mapping of features by CNN with XGBOOST learning. *Materials Today: Proceedings*, http://dx.doi.org/10.1016/j.matpr.2020.09.650.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), http://dx.doi.org/10.1023/A:1022643204877.

Rakhlin, A., Shvets, A., Iglovikov, V., & Kalinin, A. A. (2018). Deep convolutional neural networks for breast cancer histology image analysis. In *LNCS: vol. 10882, Lecture notes in computer science (Including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. http://dx.doi.org/10.1007/978-3-319-93000-8_83.

Robertson, S., Azizpour, H., Smith, K., & Hartman, J. (2018). Digital image analysis in breast pathology—from image processing techniques to artificial intelligence. *Translational Research*, *194*, http://dx.doi.org/10.1016/j.trsl.2017.10.010.

Roux, L., Racoceanu, D., Loménie, N., Kulikova, M., Irshad, H., Klossa, J., Capron, F., Genestie, C., Naour, G., & Gurcan, M. (2013). Mitosis detection in breast cancer histological images an ICPR 2012 contest. *Journal of Pathology Informatics*, *4*(1), http://dx.doi.org/10.4103/2153-3539.112693.

Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, *35*(5), http://dx.doi.org/10.1109/TMI.2016.2528162.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1), http://dx.doi.org/10.1186/s40537-019-0197-0.

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *3rd international conference on learning representations, ICLR 2015 - Conference track proceedings*.

Sizilio, G. R. M. A., Leite, C. R. M., Guerreiro, A. M. G., & Neto, A. D. D. (2012). Fuzzy method for pre-diagnosis of breast cancer from the Fine Needle Aspirate analysis. *BioMedical Engineering Online*, *11*, http://dx.doi.org/10.1186/1475-925X-11-83.

Spanhol, Fabio A., Cavalin, P. R., Oliveira, L. S., Petitjean, C., & Heutte, L. (2017). Deep features for breast cancer histopathological image classification. In *2017 IEEE international conference on systems, man, and cybernetics, SMC 2017, 2017-January*. http://dx.doi.org/10.1109/SMC.2017.8122889.

Spanhol, Fabio A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016a). A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, *63*(7), http://dx.doi.org/10.1109/TBME.2015.2496264.

Spanhol, Fabio Alexandre, Oliveira, L. S., Petitjean, C., & Heutte, L. (2016b). Breast cancer histopathological image classification using convolutional neural networks. In *Proceedings of the international joint conference on neural networks, 2016-October*. http://dx.doi.org/10.1109/IJCNN.2016.7727519.

Sudharshan, P. J., Petitjean, C., Spanhol, F., Oliveira, L. E., Heutte, L., & Honeine, P. (2019). Multiple instance learning for histopathological breast cancer image classification. *Expert Systems with Applications*, *117*, http://dx.doi.org/10.1016/j.eswa.2018.09.049.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE computer society conference on computer vision and pattern recognition, 07-12-June-2015*. http://dx.doi.org/10.1109/CVPR.2015.7298594.

Toğaçar, M., Özkurt, K. B., Ergen, B., & Cömert, Z. (2020). BreastNet: A novel convolutional neural network model through histopathological images for the diagnosis of breast cancer. *Physica A: Statistical Mechanics and its Applications*, *545*, http://dx.doi.org/10.1016/j.physa.2019.123592.

Veeling, B. S., Linmans, J., Winkens, J., Cohen, T., & Welling, M. (2018). Rotation equivariant CNNs for digital pathology. In *LNCS: vol. 11071, Lecture notes in computer science (Including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. http://dx.doi.org/10.1007/978-3-030-00934-2_24.

Vesal, S., Ravikumar, N., Davari, A. A., Ellmann, S., & Maier, A. (2018). Classification of breast cancer histology images using transfer learning. In *LNCS: vol. 10882, Lecture notes in computer science (Including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. http://dx.doi.org/10.1007/978-3-319-93000-8_92.

Veta, M., Heng, Y. J., Stathonikos, N., Bejnordi, B. E., Beca, F., Wollmann, T., Rohr, K., Shah, M. A., Wang, D., Rousson, M., Hedlund, M., Tellez, D., Ciompi, F., Zerhouni, E., Lanyi, D., Viana, M., Kovalev, V., Liauchuk, V., Phoulady ., H. A., & Pluim, J. P. W. (2019). Predicting breast tumor proliferation from whole-slide images: The TUPAC16 challenge. *Medical Image Analysis*, *54*, http://dx.doi.org/10.1016/j.media.2019.02.012.

Vo, D. M., Nguyen, N. Q., & Lee, S. W. (2019). Classification of breast cancer histology images using incremental boosting convolution networks. *Information Sciences*, *482*, http://dx.doi.org/10.1016/j.ins.2018.12.089.

Wadkar, K., Pathak, P., & Wagh, N. (2019). Breast cancer detection using ann network and performance analysis with SVM. *International Journal of Computer Engineering and Technology*, *10*(3), http://dx.doi.org/10.34218/ijcet.10.3.2019.009.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data mining: Practical machine learning tools and techniques. In *Data mining: Practical machine learning tools and techniques*. http://dx.doi.org/10.1016/c2009-0-19715-5.

Yari, Y., & Nguyen, H. (2020). A state-of-the-art deep transfer learning-based model for accurate breast cancer recognition in histology images. In *Proceedings - IEEE 20th international conference on bioinformatics and bioengineering, BIBE 2020*. http://dx.doi.org/10.1109/BIBE50027.2020.00153.