Classification of Breast Cancer using Histology images: Handcrafted and Pre-Trained Features Based Approach

Jyoti Kundale¹
Rajiv Gandhi Institute of Technology, India¹
D.Y. Patil deemed to be University, Ramrao Adik Institute of Technology, India¹
jyojadhav04@gmail.com¹

Sudhir Dhage²
Sardar Patel Institute of Technology, India²
sudhir_dhage@spit.ac.in²

Abstract: Breast cancer has become a critical disease in women. The number of patients with breast cancer is quite high in India. It is of paramount importance to detect the disease in advance. Digital histopathology is one of the most advanced techniques for detection using machine learning. Artificial intelligence is going to be like a sunrise in the field of medicine. Deep neural networks have been successfully applied to the problem under consideration in the past. As, we know the feature extraction is one of the essential and crucial steps in case of classification. In this paper, we compare two approaches, first is feature extraction using traditional Handcrafted based and other is Transfer Learning based model (Pre-trained) for multiclass classification of Breast Cancer using Convolutional Neural Network (CNN) as a classifier. The models are trained using handcrafted features like Seeped Up Robust Features (SURF) and Dense Scale Invariant Feature Transform (DSIFT) techniques, later these extracted features are encoded by Locality Constrained Linear Coding method (LLC). In pre-trained model we have used VGG16, VGG19, ResNet50, GoogLeNet for feature extraction. The maximum accuracy for “SURF+CNN” is 92.88% for Handcrafted feature and in case of Pre-trained “GoogLeNet+ CNN” model gives 94%, both for 400X magnification factor.

Keywords: Histopathology Images, Locality Constrained Linear Coding, De-Novo Model, Pre-Trained Model, Handcrafted Feature, Convolutional Neural Network, Magnification Factor, Deep Learning

1. Introduction

Even though there have been outstanding advanced developments in the diagnosis and treatment of cancer, it has become a dangerous threat to human life. Cancer is a disease in which division of abnormal cell takes place in an uncontrollable fashion destroying other healthy body tissue finally resulting in a tumor. Breast cancer is a rampant and serious issue among females all over the world compared to the cancer of the lungs, brain or liver. Demographic aging and the adoption of bad daily routines by the people at large especially in the industrialized nations have become major contributing factors leading to an increase in the occurrence of disease-causing mortality rates to rise around the world by cancer. Breast cancer is leading in case of woman [1]. Around 9.6 million deaths occurred in 2018 due to cancer as reported by World Health Organization. In 2015, WHO reported 1.7 million (11.3%) victims of breast cancer and it is estimated that the number of new Breast cancer patients will grow by 70% in the next two decades. It is a well-known established fact that primary and exact diagnosis has a very significant role to play in the treatment thereby in-creasing the survival count of people suffering from Breast cancer by around 40% (World Health Organization) [2]. After the lung cancer Breast cancer is commonly found in now days in younger age group women. There are many reasons for Breast cancer like delay in child, alcohol consumption, hormonal changes, and obesity [3]. The foremost reason is being a woman and her day today life style. Now a day’s women are very busy in her daily routine. She is managing her job as
well as maintains social network also. She does not care about herself. In India, according to the International Agency for Research on Cancer, cancers of the lips, oral, lungs, neck and throat occur more in the males while females have to deal with breast, ovarian and cervical cancers. More females in the age group of 20 to 59 and in the age group of 59 plus die due to breast cancer. The trend is very disturbing as the number of patients is increasing in the younger age group more than in the older age group [4]. Hence there is a need for having a domain which will mainly focus on the early detection and cure of Breast Cancer among females. The main aim is to build systems that will detect and diagnose cancer an early stage so that proper treatment could be given to the patient. A complete cure of breast cancer is possible only if it is detected in an early stage. The rate of survival will improve by 95% if early detection of breast cancer is done for a patient. Thereby early detection can reduce the mortality rate automatically. Tumors in the breast are broadly divided into two types benign (i.e. ones that are not cancerous) and malignant (i.e. ones that are cancerous). Invasive and Non-Invasive also have subcategories which should be analyzed separately as every subcategory will need a unique course of cure. There are several ways for breast cancer detection like Ultrasound, Magnetic Resonance Imaging MRI, and Clinical examination, Mammograms, Biopsy and Histopathology [5]. Detection of cancer by the use of imaging is of prime significance because it can be commonly adapted and it is an effective and convenient way of detecting cancer when compared to other biopsy techniques such as Surgical (open) Biopsy (SOB), Vacuum Assisted Breast Biopsy (VABB), Core Needle Biopsy (CNB) and Fine Needle Aspiration (FNA) [6].

Breast Cancer for the female population has a huge bearing in the context of general public health; hence the urgent need to provide tools that will automatically classify Breast Cancer using digitized images from histological slides. High level expertise and more than one pathologist are required to give the final classification decision, which is a very tedious and time-consuming task. In the early stages of Breast cancer, misdiagnosis is a serious issue that could ultimately cost a precious human life.

There are various imaging modalities are present for detection of Breast cancer like Ultrasound, Magnetic Resonance Imaging (MRI), Digital Mammography (DM), Histopathology. Digital mammography images are efficient to store and Process, Cost effective and Works well with Computer Aided Device. Drawback of this Digital Mammography is it provides less spatial resolution and more preprocessing is required [7] [8]. MRI gives more details inner tissues and further it can be sent to biopsy. Disadvantage of MRI is not recommended to pregnant women and Image of MRI can be enhanced using some chemical agent which may cause allergy to patient [9]. Pros of Ultrasound imaging is not risky for pregnant women where images are captured in every angle in real time so reduces false negative. Patient is never exposed to radiation. Cons of Ultrasound Images are poor quality of images are not able to distinguished between Malignant and Benign tissue. Highly expertise are required to diagnose lump in Breast [8]. Images form Histology can be used as Whole Slide Imaging (WSI) or ROI (Region of Interest). Performance is better than other image modalities. Using histopathology, it is able to diagnose different types of cancer Images can be stored for long time for future analysis. Another option to histopathological diagnosis is computer- aided diagnosis (CAD) for breast cancer classification. Exact diagnosis can be done using the system of CAD as it is reasonable, easily acquired, quick and dependable. Drawback: Highly expertise and knowledgeable pathologist are required, as it is very time consuming and tedious task[10].

In the era of recent advances in technology, histopathological image analysis is used for the detection and diagnosis of breast cancer by most of the pathologist. Final classification is done labelling by pathologist. It needs more concentrationa and very time-consuming task. Loss of attention may lead to misclassification of images. Automatic classification model is required to reduce mortality rate and to get proper treatment to patient.

The rest of the paper is organized as follows: The Background study of the proposed research. After that proposed methodology with two different feature extraction-based approaches basically handcrafted using SURF and DSIFT. In Experimental Analysis results are discussed with existing method followed by Conclusion and Future Scope is given.
2. Background

This section gives background study of various methods used by other researchers. The multi-class classification is done by Machine Learning method, De-Novo Model and Transfer Learning Based model. Based on pattern recognition and machine learning method handcrafted features are used for multiclass classification. Researcher have used ROI where features are selected from nuclei segmentation and used as for classification [11] [12]. In studies, it has been found that “texture” is used as feature for classification. There are various features descriptor like Local Binary Pattern (LBP), Scale Invariant feature transform (SIFT), Speeded UP Robust Transform (SURF). Gray Level co-occurrence Matrix (GLCM), Histogram of Gradient (HOG) are used for feature extraction [13] [14]. Oriented fast and rotated brief method is later used which covered drawback of SIFT and SURF, which requires less computational power and gives better performance [15]. One class kernel principal component analysis (KPCA) model is used with different features for every image [16].

The proposed system makes use of handcrafted feature for multiclass classification, which will be helpful for pathologist. In past year, many researchers have been done using conventional convolutional model (De-Novo Model). De-Novo Model are model are developed from scratch. This type of model consists of convolutional layer, pooling layers. The variant of fully connected depends on the number of classes that has to be classified as output. Dabeer et al [17] used Nuclei structure, cytoplasmic quantity as feature extraction and compression as preprocessing technique for binary classification on BreakHis dataset. S. Margret et. has proposed filter and wrapper method for feature extraction for binary classification by using normalization as preprocessing [18].

CNN can be implemented using pre-trained model. There are various pre-trained models are available like GoogLeNet and AlexNet which are developed on non-medical images such as image-net dataset. The knowledge of these pre-trained networks is used and transfer to other medical images. On such type of pertained network parameters like weights, learning rate and layers of the networks are tuned is called parameter tuning. VGG16, VGG19, Inception and ResNet are already available pertained networks. Depends on the requirement of classification model fine-tuning of layers and weights are done. The authors [19] have used VGG16, ResNet pre-trained models and data augmentation as preprocessing techniques for multiclass and binary classification. In [20], binary classification is done by a pre-trained CNN (AlexNet) was used for binary classification using BreakHis Dataset. Mahboubeh Jannesari et al. have used ResNet V1150 and ResNetV1152 for classification of histopathological images. The classification is done for binary as well as multiclass of both benign and malignant. In preprocessing images are normalized, adjusted to specific to height and width, cropped means variation is done. They have use transfer learning techniques to improve performance of model using various version 1 through 4 and ResNet (V1 50, V1 101 and V1 152) frameworks. It is found that in all Cancer datasets examined, deep ResNet frameworks found to be resilient and accurate than Inception [21]. Amirreza Mahbod et al. has proposed the Fine-Tuned Deep Network Fusion model using ResNet 50 and ResNet 101 models. The two datasets have been used for experiment that is ICIAR 2018 grand challenge dataset and BioImaging 2015 challenge yielding for four classes’ classification that is benign, normal, invasive carcinoma and in situ carcinoma. Preprocessing is one of the very essential steps before classification is done. In preprocessing normalization in order to reduce color variation and resizing is done using bi-cubic interpolation. Instead of using CNN from scratch single fine-tuned ResNet model gives better result [22]. Hafiz Mughees Ahmad et al. have proposed model based on transfer learning using AlexNet, GoogLeNet and ResNet for classification based on multiple cellular and nuclei con- figurations. Pre-processing of images are done using Macenku’s method. Initially patch wise classification is done in order to reduce size of database. All of separate samples with 512 X 512 size. In next step, whole slide is categorized depending on majority voting. ResNet50 is more accurate for both sample and whole slide wise categorization [23]. In [9], the authors have used a novel deep network named as class structure–based deep convolutional neural network (CSDCNN) in which discerning features are learnt in a hierarchy from at the various level. The performance is compared with the LeNet and AlexNet. The accuracy they have got 92 to 95% for BreakHis Dataset for various magnification factor. Yao Guo et al. Have used hybrid
CNN model which is based on GoogLeNet model. Voting tactics and bagging strategies are used in order to lessen bugs and improve the performance. The hybrid CNN is used on the BACH gives Benign, In situ, Invasive, Normal classes with accuracy 87.5% [24]. Zeya Wang et.al have proposed CNN with cascading loss for classification. There are four classes are defined like benign lesion, normal tissues, invasive carcinoma, in situ carcinoma. The patch wise classification takes place using VGG16 network. The loss classifies images into particular category effectively. Other important factor is for improve classification is global image pooling. Such type of structure helps to global information from high resolution feature map [25].

3. Proposed Methodology

This section gives idea about Handcrafted and Pre-trained model used as feature extractor for histopathology image dataset for multiclass classification using CNN as a classifier.

3.1 Handcrafted Feature Based Model

![Magnification Independent Handcrafted Feature Based Classification Model](image)

3.2 Pre-Trained Model as a feature Extractor Based Model

![Magnification Independent Pre-Trained Feature Based Classification Model](image)

For both the model Step1 (Input given), Step2 (Preprocessing) and Step4 (Convolutional Neural Network as classifier) are same. In Step3, two different feature extraction techniques are discussed.

Steps are follows:
Step 1: Input given as histopathology image dataset

Dataset from which images were taken as an input: BreakHis [26]
The BreakHis Dataset shown in Table 1 is used for experimental analysis. This dataset consists total 7909 images. This dataset contains benign and malignant images captured using a microscope. The benign categories include Tubular Adenoma (TA), Phyllodes Tumor (PT), Fibroadenoma (F), Adenosis (A). The malignant categories include Papillary Carcinoma (PC), Mucinous Carcinoma (MC), Lobular Carcinoma (LC) and Ductal Carcinoma (DC). The dataset contains images with various zooms specifically 40 times, 100 times, 200 times, 400 times.

Table 1. BreaKHis Dataset Details

| Classes      | Sub-classes       | 40X   | 100X   | 200X   | 400X   | Total |
|--------------|-------------------|-------|--------|--------|--------|-------|
| Benign       | Adenosis          | 114   | 113    | 111    | 106    | 444   |
|              | Fibroadenoma      | 253   | 260    | 264    | 237    | 1014  |
|              | Phyllodes Tumor   | 109   | 121    | 108    | 115    | 453   |
|              | Tubular Adenoma   | 149   | 150    | 140    | 130    | 569   |
| Malignant    | Papillary Carcinoma | 145 | 142    | 135    | 138    | 560   |
|              | Ductal Carcinoma  | 864   | 903    | 896    | 788    | 3451  |
|              | Mucinous Carcinoma| 205   | 222    | 196    | 169    | 792   |
|              | Lobular Carcinoma | 156   | 170    | 163    | 137    | 626   |
| Total        |                   | 1995  | 2081   | 2013   | 1820   | 7909  |

Initially a pathologist starts extracting ROI at 40X magnification level and keeps zooming upto 400X magnification level to get more detailed information. It also provides detail classification of benign and malignant subclasses.

Step 2: Preprocessing

To avoid the problem of limited size of data and unbalanced data, augmentation technique is used as preprocessing. This type of preprocessing technique helps to reduce the over fitting issue in case of CAD system. In data augmentation techniques images are flipped, cropped, rotated, translated and interpolated in case of medical imaging. Here we have used, rotation and horizontal flip technique as data augmentation. It is possible to loss informative feature if other augmentation technique is used because histology images are rotation and reflection in nature. We have rotated images by 90, 180, 270 degrees to increase the size of training dataset. In horizontal flip helps to double cardinality of images.

Step 3: Feature extraction

Feature extraction is most important step while doing classification. Because every feature carries important information. There may be morphological or texture feature which are related to nuclei or other tissue of breast. The given dataset consists of different magnification factors like 40X, 100X, 200X, 400X and each magnification shows various discriminative features which is helpful for multiclass classification.

a. Magnification Independent Handcrafted Feature Based Classification Model
In this, we have extracted features using Seeped Up Robust Feature (SURF) and Dense Scale Invariant Feature Transform (DSIFT) techniques. After extracting features encoding is done by local descriptor like locality constrained linear coding (LLC).

i. **SURF**: It is robust and faster than SIFT. There are three steps for SURF computation in first step given input is represented in scale space, in second step key points are detected by hessian matrix and orientation to that key point is assign. Finally, SURF descriptor is assigned to each detected key point. This feature descriptor generates 64-dimension feature vector. A rectangular area centered on the Interesting point is established and then the haar sum Wavelet reaction is performed on the point of interest into 4X4 sub region.

ii. **DSIFT**: It is opposite to SIFT technique. It gives 128 feature vectors by using histogram gradient method on 16X16 region. This descriptor uses grid in uniform way without identifying intersecting point. DSIFT gives more data/information in real time.

iii. **Feature Encoding**: While working with local descriptor it is necessary to do feature encoding to represent image for classification. Most of classifier accept input as fixed vector length. So, it is necessary to convert into fixed length vector to represent local descriptor into image. Here we are using locality constrained linear coding (LLC) for aggregating the local features to form entire image. LLC coding is variation of sparse coding where local coordinate of each descriptor combined using max pooling approach. LLC coding more efficient than Bag of words method in terms of its coding speed and representation.

b. **Magnification Independent Pre-Trained Feature Based Classification Model**

In 2012, AlexNet architecture is introduced for ImageNet Challenge having error rate of 16%. Later various variations of AlexNet with denser network are introduced. Here features are extracted using various variation of CNN model VGG16, VGG19, ResNet50, GoogLeNet (pre-trained) models. Each model has its own structure and gives variant vital feature used for classification.

i. **VGG architecture**: In 2014, Oxford Visual Geometry Group implemented VGG architectures to compete in the ILSVRC contest. There are two variants of VGG architecture one is VGG16 and VGG19. The key difference between two structures is number of convolutional layers. VGG16 possess 13 and VGG19 possess 19 convolutional layers. VGG19 is denser than VGG16. VGG architecture consists of five convolutional blocks. In first and second convolutional block consists of two convolutional layers with 64 and 128 filters each with 3X3 filter size. Third block of convolutional of VGG16 and VGG19 consists of three and four convolutional layers with 256 filters which are 3X3 in size. Fourth and fifth convolutional block also consists three and four convolutional layers with 256 filter in 3X3 size. These five convolutional blocks are separated by max pooling layer. VGG16 is 23 layers in depth with 138 million parameters and VGG19 is 26 layers in depth with 143 million parameters.

ii. **ResNet**: As we increase the number of layers in deep learning network then vanishing gradient become zero or too large. Other problem with that error rate is also increases between train and test dataset. In 2015, in order to avoid vanishing gradient problem Microsoft researcher has introduced new architecture called as Residual Network. This model uses a technique called skip connections where
it skips few layers and directly connects to output. There are four stages in ResNet50 architecture. In stage 1, three residual blocks with 3 layers each. To perform convolution operation in three layers of 1st block it uses 64, 64 and 128 kernel size with stride size is 2. The overall channel width gets doubled where size of the input decreased to half, when it switches from one stage to other.

iii. GoogLeNet: This architecture is discovered during ImageNet competition with rate of error is 6.65% in 2014.In this model different convolutional layer having different kernel size run in parallel with one max polling layer. Final output is combined into single layer. Parallel structure helps to gives fast output without increasing the more depth. It consists of 23.8 million parameters with 159 in depth having filter sizes with 1X1, 3X3 and 5X5.

From each of these pre-trained model informative features are extracted. For VGG16 and VGG19 has output feature dimension size is 63X133X3. For Resnet50 and GoogLeNet, output Size is 126X266X3, 63X133X3 respectively.

**Step 4: CNN as Classifier**

Here, we have done classification using CNN as classifier is De-Novo Model, where the number of layers, size of filter and stride is decided at initial only. Such type of model is known as De-Novo Model. Performance of classifier can be checked by varying layer, filter size and activation function. CNN model from scratch gives better performance if it designs and trained efficiently as it can work on small size of datasets. The over fitting problem occurs if a greater number of levels are imposed on small set of datasets when training and testing is done. To optimize the overall performance, system requires huge number of computational resources and its very time-consuming task. We have experimented dataset by using convolution2dLayer with dimension (3,256), maxpooling2dLayer with stride size 2 and filter size 3X3. Relu as activation function is used as its give’s better performance than other activation function like sigmoid, tanh in real time scenario. As we have done multiclass classification so that in fully connected layer, we have given total eight classes for classification. The final layer is the classification layer. This layer uses the probabilities returned by the softmax activation function for each input to assign it to one of the mutually exclusive classes.

4. Experimental Analysis

This section provides systematic approaches that represent experimental results with two proposed methods. Experiments are implanted in MATLAB 2018b with 8GB RAM. The various performance parameters are considered while evaluating the performance of the model like Accuracy, Precision, Recall, F1- Score. One of the metrics which is commonly used for model performance is confusion matrix which is plot of actual label vs predicted labels, based on following parameters False Positive (F_P): A sample without breast cancer i.e. negative is wrongly identified as positive. False Negative (F_N): A sample with breast cancer is wrongly identified as negative. True Positive (T_P): This indicates a woman with a cancer of the breast. True Negative (T_N): This indicates patient is negative in the test. There is other measure like specificity, sensitivity, precision, accuracy, AUC, F-Measure and volume under the ROC surface through which model performance can be measure.

- Accuracy: It indicates correctly classified total instances. It basically indicates that the number of patients is correctly predicted without cancer and the number of patients is correctly diagnosed with cancer.

\[ A_c = \frac{(T_P + T_N)}{(T_P + T_N + F_P + F_N)} \]
• Sensitivity (S_n) or Recall (R_c) This test reveals just how many numbers of the overall positive cases are accurately predicted. The value of recall should be high as possible. It indicates that the number of patients having cancer correctly from the total number patient those having cancer. If the value of sensitivity is low patient having cancer are treated as normal which leads to threat to normal patient life.

\[ S_n = \frac{T_P}{T_P + F_N} \]

• Specificity (S_p) It indicate the number of patients not having cancer is classified as correctly as well as the number of patients is having cancer correctly. The value of specificity is should more but not more than Sensitivity (Sn) which has large impact in medical diagnosis.

\[ S_p = \frac{T_N}{T_N + F_P} \]

• Precision (P_r): This indicate that total number of patients for the positive predictions are accurately. This clearly depicts how many numbers of patient having cancer is predicted accurately. The value of recall and precision should more to diagnose patient correctly.

\[ P_r = \frac{T_P}{T_P + F_P} \]

• F1 Measure: This give combine impact of recall and precision. With the help of F-measure metric two models can be compare with high Sn and low Pr and vice versa.

\[ F1 = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \]

4.1 Results Analysis: Features extracted using SURF and SIFT of existing methods and proposed method

This section gives result analysis based on extracting handcrafted features using SURF and SIFT and applying CNN as classifier for 40X, 100X, 200X and 400X magnification factor.

Table 2. Performance analysis for Magnification Independent Handcrafted Feature Based Classification Model

| Feature + Classifier | Magnification Factor | Accuracy | Precision | Recall | F1-Score |
|----------------------|----------------------|----------|-----------|--------|---------|
| SURF+CNN             | 40X                  | 0.9214   | 0.6306    | 0.6092 | 0.6081  |
|                      | 100X                 | 0.9267   | 0.6531    | 0.6409 | 0.6379  |
|                      | 200X                 | 0.9228   | 0.6195    | 0.6270 | 0.6198  |
|                      | 400X                 | 0.9288   | 0.6639    | 0.6440 | 0.6478  |
| DSIFT+CNN            | 40X                  | 0.9161   | 0.6075    | 0.6435 | 0.6231  |
|                      | 100X                 | 0.9215   | 0.6317    | 0.6705 | 0.6485  |
|                      | 200X                 | 0.9167   | 0.6013    | 0.6367 | 0.6170  |
|                      | 400X                 | 0.9234   | 0.6416    | 0.6565 | 0.6456  |
| DSIFT+CNN [27]       | 40X                  | 60.58    |           |        |         |
|                      | 100X                 | 57.44    |           |        |         |
|                      | 200X                 | 70.00    |           |        |         |
| Feature + Classifier   | Magnification Factor | Accuracy | Precision | Recall  | F1-Score |
|------------------------|----------------------|----------|-----------|---------|----------|
| SURF+CNN [27]          | 400X                 | 46.96    |           |         |          |
|                        | 40X                  | 80.37    |           |         |          |
|                        | 100X                 | 63.84    |           |         |          |
|                        | 200X                 | 74.54    |           |         |          |
|                        | 400X                 | 54.70    |           |         |          |
| DSIFT + SVM [27]       | 40X                  | 44.54    |           |         |          |
|                        | 100X                 | 51.68    |           |         |          |
|                        | 200X                 | 44.30    |           |         |          |
|                        | 400X                 | 35.54    |           |         |          |
| SURF+SVM [27]          | 40X                  | 53.75    |           |         |          |
|                        | 100X                 | 44.30    |           |         |          |
|                        | 200X                 | 45.30    |           |         |          |
|                        | 400X                 | 40.88    |           |         |          |

From Table 2, it is clear that features extracted using SURF + CNN get maximum accuracy that is 92.88 for multiclass classification.

4.2 Results Analysis: Features extracted using Pre-Trained model such as VGG16, VGG19, ResNet50, and GoogleNet of existing methods and proposed method

Table 3. Performance analysis for Magnification Independent Pre-Trained Feature Based Classification Model

| Feature + Classifier | Magnification Factor | Accuracy | Precision | Recall | F1-Score |
|----------------------|----------------------|----------|-----------|--------|----------|
| VGG16+CNN            | 40X                  | 0.9247   | 0.6412    | 0.6057 | 0.6075   |
|                      | 100X                 | 0.9286   | 0.6453    | 0.6450 | 0.6388   |
|                      | 200X                 | 0.9246   | 0.6156    | 0.6240 | 0.6155   |
|                      | 400X                 | 0.9301   | 0.6609    | 0.6316 | 0.6377   |
| VGG19+CNN            | 40X                  | 0.9183   | 0.6540    | 0.6178 | 0.6199   |
|                      | 100X                 | 0.9249   | 0.6691    | 0.6622 | 0.6594   |
|                      | 200X                 | 0.9199   | 0.6413    | 0.6491 | 0.6409   |
|                      | 400X                 | 0.9261   | 0.6734    | 0.6590 | 0.6584   |
| ResNet50+CNN         | 40X                  | 0.9192   | 0.6397    | 0.6145 | 0.6108   |
| Feature + Classifier | Magnification Factor | Accuracy | Precision | Recall | F1-Score |
|----------------------|----------------------|----------|-----------|--------|----------|
|                      | 100X                 | 0.9256   | 0.6610    | 0.6509 | 0.6498   |
|                      | 200X                 | 0.9223   | 0.6424    | 0.6360 | 0.6351   |
|                      | 400X                 | 0.9266   | 0.6699    | 0.6534 | 0.6537   |
| GoogLeNet+CNN        | 40X                  | 0.9335   | 0.6028    | 0.5747 | 0.5724   |
|                      | 100X                 | 0.9386   | 0.6202    | 0.6199 | 0.6136   |
|                      | 200X                 | 0.9373   | 0.6043    | 0.6027 | 0.5995   |
|                      | 400X                 | 0.9400   | 0.6342    | 0.6147 | 0.6166   |
| CNN features + linear SVM [27] | 40X                  | 72.35    |           |        |          |
|                      | 100X                 | 67.68    |           |        |          |
|                      | 200X                 | 66.45    |           |        |          |
|                      | 400X                 | 64.95    |           |        |          |
| CNN features + RF [27] | 40X                  | 66.38    |           |        |          |
|                      | 100X                 | 65.12    |           |        |          |
|                      | 200X                 | 69.80    |           |        |          |
|                      | 400X                 | 67.96    |           |        |          |
| CSDCNN + augmented data [9] | 40X                  | 92.8 ± 2.1 |         |        |          |
|                      | 100X                 | 93.9 ± 1.9 |       |        |          |
|                      | 200X                 | 93.7 ± 2.2 |       |        |          |
|                      | 400X                 | 92.9 ± 1.8 |       |        |          |
| VGG19 + SVM (L, 1) [28] | 40X                  | 92.64    | 92.00     | 92.00  | 92.00    |
|                      | 100X                 | 91.25    | 91.00     | 91.00  | 91.00    |
|                      | 200X                 | 81.42    | 82.00     | 82.00  | 82.00    |
|                      | 400X                 | 80.84    | 82.00     | 81.00  | 81.00    |
| VGG16 + SVM (L, 1)[28] | 40X                  | 93.97    | 94.00     | 93.00  | 94.00    |
|                      | 100X                 | 92.92    | 92.00     | 91.00  | 91.00    |
|                      | 200X                 | 91.23    | 92.00     | 92.00  | 92.00    |
|                      | 400X                 | 91.79    | 92.00     | 91.00  | 91.00    |

From Table 3, it is observed that accuracy for GoogLeNet+CNN is 94% for 400X magnification factor which more than another pre-trained model and handcrafted feature model.
4.3 Confusion matrix

a. Confusion matrix for SURF+CNN for 40X, 100X, 200X and 400X

(i)  

(ii)  

(iii)  

(iv)  

b. Confusion matrix for SIFT+ CNN for 40X, 100X, 200X and 400X
(i)  (ii)  
(iii)  (iv)  
c. Confusion matrix for VGG16 + CNN for 40X, 100X, 200X and 400X
d. Confusion matrix for VGG19 + CNN for 40X, 100X, 200X and 400X
e. Confusion matrix for ResNet50 + CNN for 40X, 100X, 200X and 400X
(i) Confusion matrix for GoogLeNet+CNN for 40X, 100X, 200X and 400X

(ii) Confusion matrix for GoogLeNet+CNN for 40X, 100X, 200X and 400X

(iii) Confusion matrix for GoogLeNet+CNN for 40X, 100X, 200X and 400X

(iv) Confusion matrix for GoogLeNet+CNN for 40X, 100X, 200X and 400X

f. Confusion matrix for GoogLeNet+CNN for 40X, 100X, 200X and 400X
5. Conclusion and Future Scope

In this paper, we have compared handcrafted features and pre-trained model for multiclass classification for breast cancer classification using deep learning model. It has been observed that pre-trained model (GoogLeNet) gives better performance i.e 94% accuracy than handcrafted feature-based model.
Experiments are done using CNN from scratch which gives good results than other conventional classifier for all magnification factors. In future other variants of pre-trained model can be used as a feature extractor as well as classifier. CNN with fine tuning of layer can be employed to improve the accuracy. The complexity can be increased with increase in magnification level. Computational efficiency of model can be improving by using parallel processing and cloud-based approach, as it will reduce the overall training time.

REFERENCES

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal, 2018 Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin;68(6):394–424.
[2] Organization WHO. Available from: http://www.who.int/en/
[3] M. C. Chun. Breast Cancer: Symptoms, Risk Factors, and Treatment, Medical News Today Available from: https://www.medicalnewstoday.com/articles/37136.php.
[4] D. S. Shah. Statistics of breast cancer in India, 2018. Available from: http://www.breastcancerindia.net/statistics/trends.html
[5] Altaf F, Islam SMS, Akhtar N, Janjua NK. 2019 Going Deep in Medical Image Analysis: Concepts, Methods, Challenges, and Future Directions. IEEE Access;7:99540–72.
[6] Murtaza G, Shuib L, Abdul Wahab AW, Mujtaba G, Mujtaba G, Nweke HF et al. 2020 Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges. Artif Intell Rev 53(3):1655–720. Available from: https://doi.org/10.1007/s10462-019-09716-5
[7] Elmore JG, Jackson SL, Abraham L, Miglioretti DL, Carney PA, Geller BM, et al. 2009 Variability in interpretive performance at screening mammography and radiologists’ characteristics associated with accuracy. Radiology 253(3):641–51.
[8] Sadaf A, Crystal P, Scaranello A, Helbich T. 2011 Performance of computer-aided detection applied to full-field digital mammography in detection of breast cancers. Eur J Radiol. 77(3):457–61. Available from: http://dx.doi.org/10.1016/j.ejrad.2009.08.024
[9] 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. Sci Rep. 7(1):1–10.
[10] Pantanowitz L, Farahani N, Parwani A. 2015 Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. Pathol Lab Med Int 23.
[11] Kowal M, Filipczuk P, Obuchowicz A, Korbicz J, Monczak R. 2013 Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images. Comput Biol Med.;43(10):1563–72.
[12] George YM, Zayed HH, Roushdy MI, Elbagoury BM. 2014 Remote computer-aided breast cancer detection and diagnosis system based on cytological images. IEEE Syst J.8(3):949–64.
[13] Alhindi TJ, Kalra S, Ng KH, Afrin A, Tizhoosh HR 2018 Comparing LBP, HOG and Deep Features for Classification of Histopathology Images. Proc Int It Conf Neural Networks ;2018-July:1–7.
[14] Belsare AD, Mushrif MM. 2015 Images using Texture Feature Analysis.
[15] Rublee E, Garage W, Park M. 2011 ORB : an efficient alternative to SIFT or SURF 2564–71.
[16] Zhang Y, Zhang B, Coenen F, Xiao J, Lu W. 2014 One-class kernel subspace ensemble for medical image classification. EURASIP J Adv Signal Process 2014(1):1–13.
[17] Dabeer S, Khan MM, Islam S. 2019 Informatics in Medicine Unlocked Cancer diagnosis in histopathological image : CNN based approach. Informatics Med Unlocked 16(August):100231. Available from: https://doi.org/10.1016/j.imu.2019.100231
[18] Anouncia SM. Knowledge Computing and Its Applications. Vol. 1.
[19] Xie J, Liu R, Luttrell J, Zhang C. 2019 Deep learning based analysis of histopathological images of
breast cancer. *Front Genet.* **10**(FEB):1–19.

[20] Spanhol FA, Oliveira LS, Petitjean C, Heutte L. 2016 Breast Cancer Histopathological Image Classification using Convolutional Neural Networks *International joint conference on neural networks (IJCNN)* IEEE Jul 24, pp. 2560-2567.

[21] Motlagh MH, Jannesari M, Aboulkheyr H, Khosravi P. 2018 Breast Cancer Histopathological Image Classification: A Deep Learning Approach.:1–8.

[22] Mahbod A, Ellinger I, Ecker R, Smedby Ö, 2018 Wang C. Breast Cancer Histological Image Classification Using Fine-Tuned Deep Network Fusion. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* p. 754–62.

[23] Ahmad HM, Ghuffar S, Khurshid K. 2019 Classification of Breast Cancer Histology Images Using Transfer Learning. In: *Proceedings of 2019 16th International Bhurban Conference on Applied Sciences and Technology, IBCAST 2019*. p. 328–32.

[24] Guo Y, Dong H, Song F, Zhu C, Liu J. 2018 Breast Cancer Histology Image Classification Based on Deep Neural Networks Breast Cancer Histology Image Classification Based on Deep Neural *Springer International Publishing* Available from: http://dx.doi.org/10.1007/978-3-319-93000-8_94

[25] Wang Z, Dong N, Dai W, Rosario SD, Xing EP. 2018 Classification of Breast Cancer Histopathological Images using Convolutional Neural Networks with Hierarchical Loss and Global Pooling Vol. 2. *Springer International Publishing* 745–753 p. Available from: http://dx.doi.org/10.1007/978-3-319-93000-8_84

[26] Spanhol FA, Oliveira LS, Petitjean C, Heutte L 2016. A Dataset for Breast Cancer Histopathological Image Classification. *IEEE Trans Biomed Eng.* **63**(7):1455–62.

[27] Bardou D, Zhang KUN, Ahmad SM. 2018 Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks. *IEEE Access.* **6**:24680–93.

[28] Sharma S, Mehra R 2020. Conventional Machine Learning and Deep Learning Approach for Multi-Classification of Breast Cancer Histopathology Images—a Comparative Insight. *J Digit Imaging* **33**(3):632–54.