International Journal of Scientific Research in Computer Science and Engineering
Vol.10, Issue.6, pp.09-21, December (2022)

Research Paper

E-ISSN: 2320-7639

Breast Histopathology Images Multi-Classification using Ensemble of Deep Convolutional Neural Networks

S.P. Akinrinwa[1*], O. Olabode[2], O.C. Agbonifo[2], K.G. Akintola[3]

[1]Department of Information Technology, Federal University of Technology, Akure, Nigeria
[2]Department of Information Systems, Federal University of Technology, Akure, Nigeria
[3]Department of Software Engineering, Federal University of Technology, Akure, Nigeria

[*]*Corresponding Author: spakinrinwa@futa.edu.ng Tel.: +234 803 634 1134*

Available online at: www.isroset.org | DOI: https://doi.org/10.26438/ijsrcse/v10i6.921

Received: 15/Sept/2022, Accepted: 13/Oct/2022, Online: 31/Dec/2022

***Abstract*—** Breast cancers have constituted a major health challenge as a leading cause of mortality in women. This has led to several interventions in the diagnosis and treatment of the disease. The digital classification and analysis of breast histopathology images provides a means for computerized-clinical diagnosis of breast cancers. In this study, three separate models based on the ensembles of Convolutional Neural Networks (CNNs) for the analysis and classification of histopathological images of breast tissue are presented. The ensembles make use of majority voting, averaging, and stacking techniques. The ensemble models seek to extend the performance of existing CNNs by combining AlexNet, VGGNet and ResNet using majority voting, averaging, and stacking ensemble rules. Furthermore, a model for classification of breast histopathology images was developed called the SaduNet model. This model was developed with few numbers of convolutional neural network layers to reduce computational cost in terms of memory and improve computation time. All the models were trained with histopathology images dataset collected at the Federal Teaching Hospital, Ido Ekiti, Ekiti state, Nigeria, with ethical clearance to ensure that the research is applicable locally. Different learning parameters were used in the different convolutional neural networks developed to ensure that they obtain optimal performances of the models on the histopathology images classification task. The comparative analysis performed showed that the developed models performed as well as those found in literature judging by the accuracies achieved. The ensemble methods also performed better in the terms of the sensitivity and predictability than the individual base models. This is shown in the high prediction and recall values obtained by the ensemble models. During testing, the base models generated the following accuracies: AlexNet: 92.91%, VGG16: 96.28% and ResNet: 99.25%. When tested with the FTH breast histopathology data, the averaging ensemble has accuracy of 99.47% while the majority voting ensemble has accuracy of 99.30% and stacking ensemble model has accuracy of 97.86%. The SaduNet model also achieved an accuracy of 75.89%.

***Keywords*—** Classification, Convolutional Neural Network, Histopathology Images, Deep learning, Ensemble

I. INTRODUCTION

Breast cancer is a worldwide public health problem affecting many women, therefore, there are concerns on the debilitating effects of breast cancer. Breast cancer amounted to about 800,000 deaths among women [1, 2]. In Nigeria, the health care sector suffers from the lack of substantial number of trained pathologists and over 50% of the medical laboratory test are not done or verified by pathologists. This leaves a lot of diseases undiagnosed. Low number of pathologists can lead to a reduced the number patholog1ical diagnosis [3]. Currently, available solutions make use of high-resolution whole slide imaging systems that are computationally expensive. Thus, this research is motivated by the need to provide a system for histopathology images classification that makes use of local dataset and has less computational cost because of the use of lower resolution histopathology images.

During a breast disease diagnosis, clinical investigations, Imaging and laboratory tests and pathological investigations are done to establish the presence of disease and the type of disease present. Clinical investigation usually involves collecting patient history records including the list of medications and medical conditions, key personal information, recent major lifestyle changes, family history of cancer diagnosis and physical examination which involves physically checking the breasts and lymph nodes in the armpit for palpable lesions or abnormalities. Diagnosis is done with techniques used for breast assessment and disease detection which include Imaging and laboratory tests with X-ray mammography, Magnetic Resonance Images, Ultrasounds, Radionuclide Breast Images techniques, and breast pathology [4]. Some disadvantages are associated with the use breast imaging-based laboratory test in the detection of breast diseases. For example, mammography often cannot detect the

differences between benign and malignant breast disease and cannot give the size of the detected lesion. Also, X-ray images are not useful in breast disease diagnosis after surgery or radiotherapy has been done on the affected breast, because of architectural distortion in the breast, a repeated scan may lead to an increased X-ray dose [4]. Breast Ultrasonography technique is operator/observer dependent and therefore error prone, and also cannot detect calcifications in the breast [5]. Magnetic resonance imaging is sensitive but not specific, therefore, it cannot be used to prove the presence or absence of cancer. It is also expensive and unavailable for use in breast cancer screening [4]. Other breast imaging equipment have high cost and lack medical centers and oncologists with the required experience of the imaging techniques and generally cannot be used identify small lesions or breast tissues close to the chest wall [4].

Breast histopathology involves the analysis of digital images obtained from tissues collected from patient's breast, to detect diseases with a computer system. Pathology diagnosis is performed by the human pathologist who observes the tissue sample fixed on a glass slide using a microscope [6]. A histopathology image is a digital image of the biopsy sample that is captured by a camera attached to the microscope [7]. The pathologist is tasked to physically examine the cell structure and tissue distribution to detect cancer, and to determine the malignancy stage. The diagnosis is entirely dependent on the expertise of the examining pathologist and may be affected by the mental or physical state of the pathologists, and also conditions such as fatigue and attention to details [8, 9, 10].

Improved computational power and tools have made digital processing of histopathological images a possibility. This merges computer image analysis with microscopy and provides suitable alternative to the error prone manual scoring of histopathology samples by the human pathologist. Neural networks have been explored in various image recognition research since the 1980s [11, 12]. A major tool that aids computerized-clinical analysis of histopathology samples is Convolutional Neural Networks (CNNs). CNNs are neural networks with many hidden layers stacked on each other, where the input signals in some of the hidden layers are convolved with kernels to enable them to learn increasingly abstract features that provide a representation of the input data. CNNs are used mostly in image and video processing tasks. Each level transforms its input into a more abstract feature representation of that input which is then passed to the next level. CNNs use these multiple layers of artificial neurons to emulate learning [13]. Ensemble methods perform better and gives greater accuracy when compared to the single CNN models [14, 15]. An ensemble is a technique often used in machine learning to combine base models to produce another model that possesses optimal predictive abilities.

Given the successful application of deep learning to image analysis [10, 16, 17], the objective of this research work is to develop an ensemble of AlexNet, ResNet and VGGNet Convolutional Neural Network models for breast histopathology images classification in breast cancer detection. The deep CNN architectures used in the ensemble are: AlexNet [18] has eight learned layers; VGGNet-16 [19] has a deeper network of 16 weighted layers; ResNet [20] provides a residual learning framework to ease the training of deeper networks. ResNet34 has 33 convolutional layers, 1 fully connected layer and 2 pooling layers.

The rest of the paper is organized as follow: Section II discusses some relevant literature, Section III describes the research methodology, AlexNet, ResNet, VGGNet and Ensemble performances on breast histopathology dataset is evaluated and the results and recommendations are discussed in Section IV while Section V has the conclusion and future scope.

II. LITERATURE REVIEW

Machine learning offers the methods and resources for resolving diagnostic and prognostic issues in a variety of domains relevant to medicine. [21]. Several research have been done to explore the use of machine learning in analysis of histopathological images.

Using features extracted from images' texture representation attributes, a pattern recognition system was created by [9] to distinguish between benign and malignant tumors. o train the classifiers, texture descriptors like Local Binary Patterns (LBP), Completed LBP (CLBP), Local Phase Quantization (LPQ), Gray-Level Co-Occurrence Matrix (GLCM), Threshold Adjacency Statistics (TAS), Parameter-Free Threshold Adjacency Statistics (PFTAS), and a key point descriptor called Oriented FAST and Rotated BRIEF are used. One-nearest neighbor (1-NN), quadratic linear analysis (QDA), support vector machines (SVM), and random forests of decision trees (RF) were the four machine learning models that the classification system was based on.

Feature learning rather than feature extraction is used in the class structure-based deep convolutional neural network (CSDCNN) identification technique for multi-class breast cancer classification that was presented in [10]. By using the intra-class and inter-class labels of breast cancer as prior information, some feature space distance constraints were developed and integrated into the CSDCNN to control the feature similarity of the various classes of the histopathology images. By using structured formulation and prior knowledge of class structure, the CSDCNN achieved learning-based methodology and consequently automatically learnt hierarchical feature representations.

Using histopathology pictures, [20] suggested a clustering approach for the automated categorization of intraductal breast lesions. In the photos of the histopathology tissues stained with hematosin and eosin, the clustering

algorithm recognized cell areas. An immersion simulation-based watershed technique was used for individual cell segmentation. Cell size, shape, and nucleoli appearance were considered as morphological features, while statistical features including mean, standard deviation, median, and mode were computed and utilized to train the binary classifier.

[23] studied the classification of histological images of breast cancer using convolutional neural networks with a tiny SE-ResNet module. The aim was to design a CNN with fewer parameters for image classification. The small SE-ResNet model developed combined residual modules and Squeeze-and-Excitation blocks. The BreaKHis dataset was used to evaluate the developed breast cancer histopathology image classification network.

[24] investigated a multi-layer feature fusion technique for categorizing breast cancer histopathology images with the goal of creating a framework that is only partially dependent on the other methods already in use. The research fine-tuned the ResNet architecture and used it as a feature extractor while SVM was used in the classification. [25] employed local clustering and deep learning methods to categorize slides of histological breast cancer. The goal of the study was to use the structural and statistical data from the images to categorize the BreakHis dataset. Image feature extraction was done with clustering algorithms K-means and Mean-Shift algorithm. Image analysis was carried out using CNN, LSTM architecture, and a hybridized CNN and LSTM. [16] created a technique for classifying breast cancer histopathological images, motivated by the need to facilitate accurate detection of tumor sub-types with reliable and economical machine learning approaches. The Inception and ResNet CNN architectures were used.

[26] examined the categorization of malignant and benign tumors in biopsy scans using an inception v3 convolutional neural network that has been pre-trained and fine-tuned. The research concentrated on extracting discriminative information from the data without manually creating features or adding supplemental textual descriptions by subject-matter experts. Transfer learning was done with GoogleNet Inception-v3 model on the breakHis data. In addition, [27] examined the classification of histological pictures for the breast carcinoma diagnosis and contrasted handcrafted features-based classification with CNN techniques on binary and multiple-class classification tasks. Incorrect diagnoses and differences in diagnostic outcomes from various pathologists served as the motivation for the study. The first method employed hand-crafted features extracted with local descriptors called Dense Scale Invariant Feature Transform (DSIFT) features and Speeded-Up Robust Features (SURF), encoded with two models (locality constrained linear coding and bag of words coding), and SVM was used to classify images. The second method used a CNN with five convolutional layers and two fully linked layers.

[28] suggested a histological image-based categorization of breast cancer using inception recurrent residual CNN. The requirement to automatically diagnose breast cancer using deep learning methods from histology pictures served as the driving force behind the project. Two convolutional layers, four Inception Recurrent Residual Convolutional Neural Network IRCNN blocks, transition blocks, a fully connected layer, and a softmax layer made up the proposed framework for Breast Cancer (BC) classification.

[29] conducted research utilizing CNN to classify breast cancer histology images. The research aimed to improve on conventional feature extraction-based classification of biopsy images. The images were normalized using Distance Weighted Discrimination (DWD) and this converted the colour in the images to an optical density (OD) using a logarithmic transformation. After that, the OD tuples were subjected to Singular Value Decomposition (SVD) in order to identify the 2D projections with the highest variance. After applying the resulting color space transform to the original image, a normalized image was produced. CNN and CNN+SVM classifiers were used to classify each patch. The image class was determined using a patch probability fusion method. The three-patch probability fusion techniques used were maximum probability, sum of probabilities, and majority voting.

[30] used discriminative feature-oriented dictionary learning to create a system for classifying histopathology images. The goal of the study was to create a flexible histopathology image classification system using discriminative, class-specific dictionaries that could automatically find features from training image examples. The variability of the manual inspection of histology images using morphological criteria by trained pathologists served as the inspiration for the study. The training images had healthy patches and diseased patches that were used by the Discriminative Feature-oriented Dictionary Learning (DFDL) technique. The threshold θ was learned using a SVM so that healthy images had a proportion of healthy patches larger or equal to θ and diseased images had a proportion of diseased patches less than θ. The percentage of healthy patches τ in a test image is calculated and put up against the threshold θ. If $\tau \geq \theta$ it is categorized as healthy and if $\tau < \theta$ it is diseased.

[31] created a computer-aided diagnosis method that uses histological pictures to categorize breast intraductal lesions. The necessity to prevent needless procedures and inadequate therapies that resulted in invasive carcinoma served as the driving force for the research. Nine shape and intensity features, twelve Haralick's features, and twelve graph features were collected from each Region of Interest (ROI) of the intraductal lesions and utilized as the input to the SVM to categorize the lesion types of the ROI.

This research seeks to improve on some of the limitations of research on histopathology images classification which

include unacceptably high percentages of false positive and negative results in classifications and remove the need for pathologist to make biomarkers of the cancerous region on the histopathological images used during training. This research provides a technique for multi-classification of histopathology images with the use of a balanced locally collected dataset to avoid the use of imbalanced histopathological images dataset.

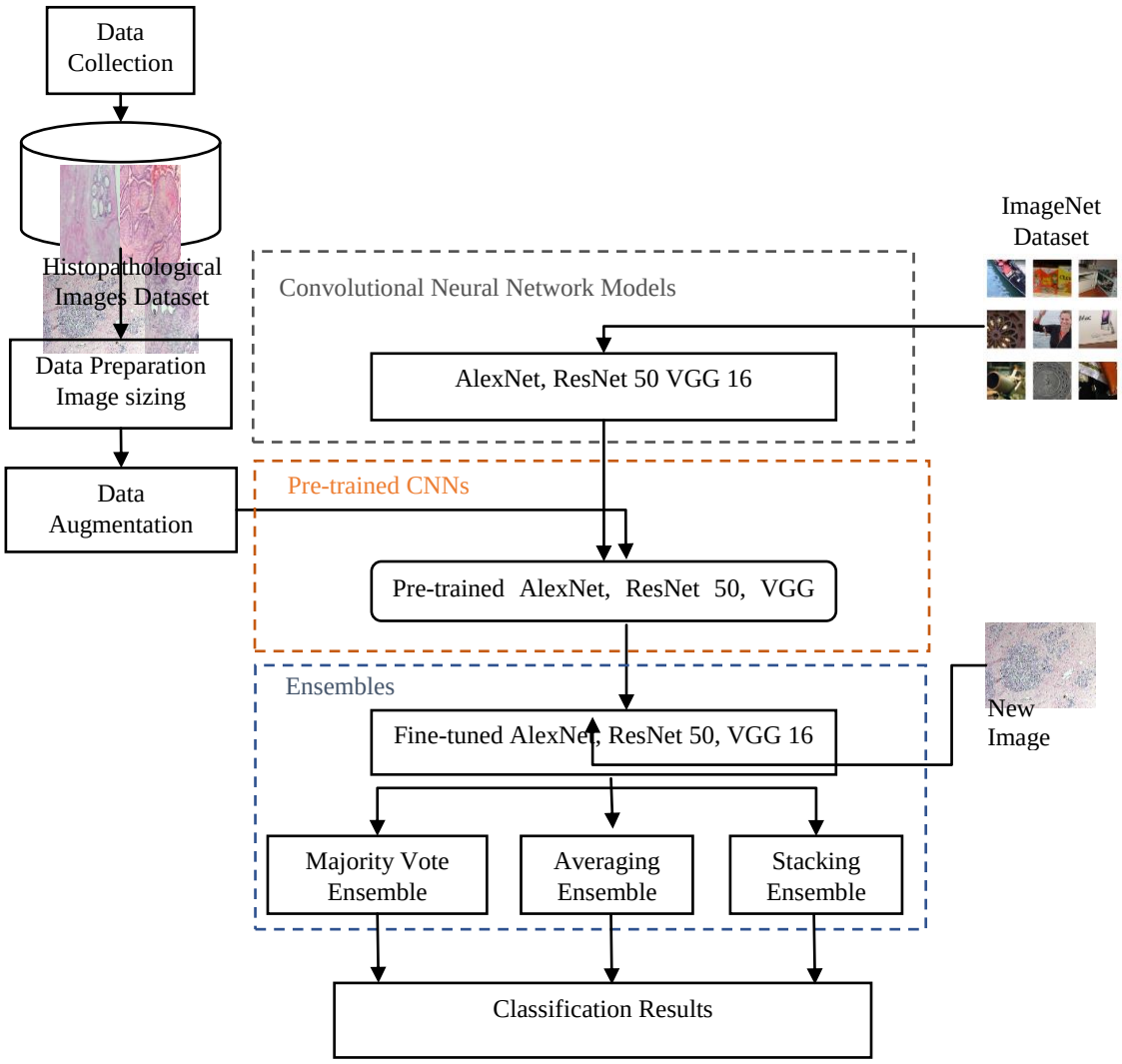

Figure 1. Architecture of the Proposed Ensembles for Breast Histopathology Images Classification

III. Methodology

Breast histopathology images classification with deep convolutional neural networks for cancer detection is achieved with the application of deep learning model to a clinical dataset of breast biopsy samples.

The research is carried out using these basic steps: Image collection, image preprocessing, data augmentation, ensemble model. The conceptual architecture of the proposed CNNs ensemble for breast histopathology images classification is shown in Figure 1.

A. Data Collection

The process taken to acquiring the histopathology images is shown in Figure 2 and described as follows:

1) Tissue Biopsy: Breast biopsy tissue sample are collected from patient's breast.

2) Fixation: Fixation is done to preserve the tissue. The fixative employed in the pathology is constituted using 10% formaldehyde in sodium chloride solution. The biopsy sample is fixed in the fixative to prevent post-mortem changes such as decomposing.

3) Tissue processing: There are series of steps done in the preparation of the tissue for microscopy, stated as follows:

a) Reception: The tissues samples are labelled with patient's information during reception.
b) Grossing / Cut up: This involves cutting of the tissue into cassettes.
c) The Automatic Processing Machine (APM) processes the tissue cassette in stages using some chemicals and materials. The stages are described as follows:

i) Fixation: The tissue is fixed in 10% Formal saline
ii) Dehydration: The tissue is dehydrated in absolute alcohol. This process is done to remove water from the tissue.
iii) Clearing and De-alcoholization: This is done to remove all the alcohol. Xylene is used as the clearing agent
iv) Infiltration / Impregnation: The tissue is impregnated with paraffin wax which serves as a medium to support the tissue internally when sectioning and slicing.
v) Embedding: The tissue is embedded in paraffin wax to give external solid support when sectioning and slicing

d) Microtomy: This involves cutting of tissue into thin sections. Microtomes are used to cut thin slices of tissues to be examined using a microscope. The slices can be put on a glass microscope slide and are thin enough to let light pass through.
e) Staining: The Haematoxylin and Eosin (H&E) stain is applied to the tissue on the glass slide. Cell nuclei are stained blue or purple by haematoxylin, while all other tissue structures are stained in various colors of pink by eosin. The stain is to enable the pathologist to measure the Nucleo-cytoplasmic ratio to make a diagnosis.
f) Dehydrate, Clear and Mount (DCM): The stained slides are firstly dehydrated and cleared, covered, and mounted unto a microscope.

4) Glass-Slide Imaging: Once mounted on the glass slide image scanning device, a digital version of the tissue sample is collected. The histopathology images collected are stored in a database alongside the pathological diagnosis provided. The laboratory setup used for the glass slides image scanning in this research is shown in Figure 3. It consists of a Nikon DS-Fi2 camera that is mounted upon a Biological Microscope ECLIPSE Ci-L and connected to a computer. The histopathology images are captured using the microscopy software installed on the computer.

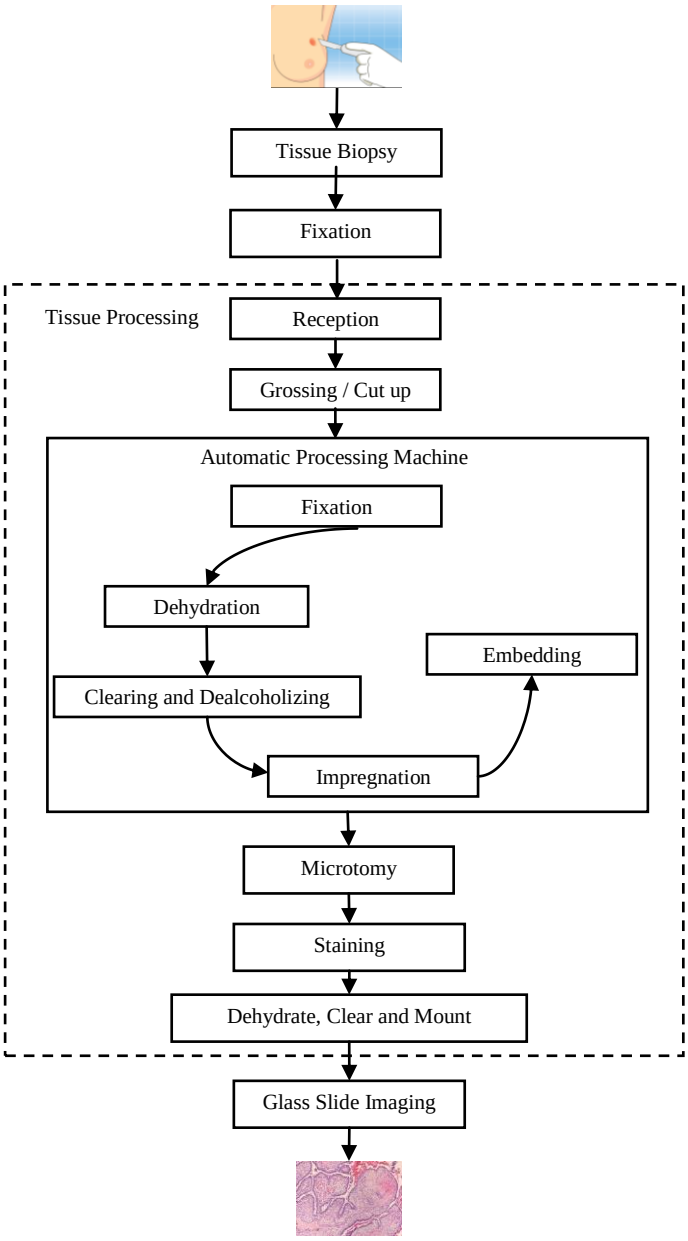

Figure 2. The histopathology tissue collection and preparation process

Figure 3. The histopathology glass slide imaging laboratory setup

B. Histopathological Images Dataset

A total of 98 glass slides of breast lesions obtained from patients that underwent breast biopsy or surgical removal of the breast tissue were obtained from the Federal Medical Centre (FMC) Ido Ekiti, Ekiti State, Nigeria. The tissues went through the histopathology tissue collection and preparation process described in Figure 2. The lesions were assessed and classified into Adenosis, Fibro Adenoma, Ductal Carcinoma, Mucinous Carcinoma and Tubular Adenoma by medical laboratory scientists and pathologists of the FMC Ido Ekiti, Ekiti State, Nigeria. Imaging and scanning of the image slides were performed with the laboratory device in Figure 3.

C. Data Preparation, Image Sizing and Normalization

In preparing the images for the convolutional neural network, the following activities are performed: All the images captured are stored in separate folders according to the diagnostic information given by the pathologists; the pixel values in images are normalized to prevent the pixel values from going outside a range of value which may result from glare during the slide scanning. The aspects of the glass slide that hold no tissue related information are also trimmed off. The ImageNet statistics (mean and standard deviation) was used during the image normalization.

D. Data Augmentation

The use of augmented datasets has been proven to improve the networks' ability to generalize while reducing training errors [18]. To augment the dataset, some steps were taken: The histopathology images were firstly divided into training and test datasets with ratio 70:30. This was done to ensure that the training and test sets both used different images to avoid over-fitting which might result from exposing the model to the test images. The images were sliced into patches of sizes 256 × 256, thereby increasing the size of the dataset. The resulting dataset had 1,876 training images: 356 Adenosis, 310 Fibro Adenoma, 395 Ductal Carcinoma, 430 Mucinous Carcinoma, and 385 Tubular Adenoma, and 564 test images: 107 Adenosis, 93 Fibro Adenoma, 119 Ductal Carcinoma, 129 Mucinous Carcinoma, and 116 Tubular Adenoma. The training dataset were also further augmented during training with rotation, flipping and randomized cropping that allowed for the creation of random-sized cropped versions of the original image with randomized aspect ratios. The crops were resized to the input size that was applicable to the CNN architecture.

E. CNN Models

CNNs have multiple layers of neural networks that learn visual patterns through convolving the image to extract features from the image. As shown in Figure 4, the CNN designs are made up of convolution and pooling layers that are alternated, followed by fully linked layers. The histopathology dataset is split into two subsets. The model is trained using the first subset, which comprises 70% of the histopathology dataset, and tested using the second subset, which comprises 30% of the histopathology dataset. A validation set of 20% of the training set is applied which is used during training for tuning of the parameters to achieve an optimized model. The test dataset is passed into the trained CNN classifier after training had been done to check the model's performance.

The image convolution, image normalization, image pooling, and Rectified Linear Units activation are used to generate feature representations that are learned and used in classifying images.

Convolution is done by multiplying kernel weights with corresponding pixel value of the receptive field in the input image. Convolution at a pixel point is described by the following mathematical expression.

$$O(i,j) = k * I(i,j) = \sum_{j=-q}^{q} \sum_{i=-q}^{q} k(i,j) * I(i,j) \qquad 1$$

where k is the convolution filter or kernel, $k(i,j)$ is the coefficient of the kernel at position i,j,
$I(i,j)$ refers to the position i,j on the input image I, the result obtained is the pixel value for the output image at the vertical and horizontal coordinates (i,j). Convolution is therefore the sum of element-wise matrix multiplication between the kernel and the region covered by the kernel on the input image. The kernel is slide throughout the input image till it covers the entire input image to generate an output image.

Image Normalization is an operation carried out on the data dimensions of the image to bring each dimension to approximately the same scale. For a layer with d-dimensional input $x = (x^{(1)} \ldots x^{(d)})$,$x$ is of the size N, *where* N is number of data, normalization is achieved in each dimension $x^{(k)}$ using equation 2,

$$\hat{x}^{(k)} = \frac{x^{(k)} - \mu[x^{(k)}]}{\sigma[x^{(k)}]} \quad k = 1, 2, \ldots, n \qquad 2$$

where $\mu[x^{(k)}]$ is the mean x value in the dimension $x^{(k)}$, and $\sigma[x^{(k)}]$ is the standard deviation, calculated using equation 3 and 4.

$$\mu[x^{(k)}] = \frac{1}{N} \sum_{i=1}^{N} x_i{}^{(k)} \qquad 3$$

$$\sigma[x^{(k)}] = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i{}^{(k)} - \mu[x^{(k)}])^2} \qquad 4$$

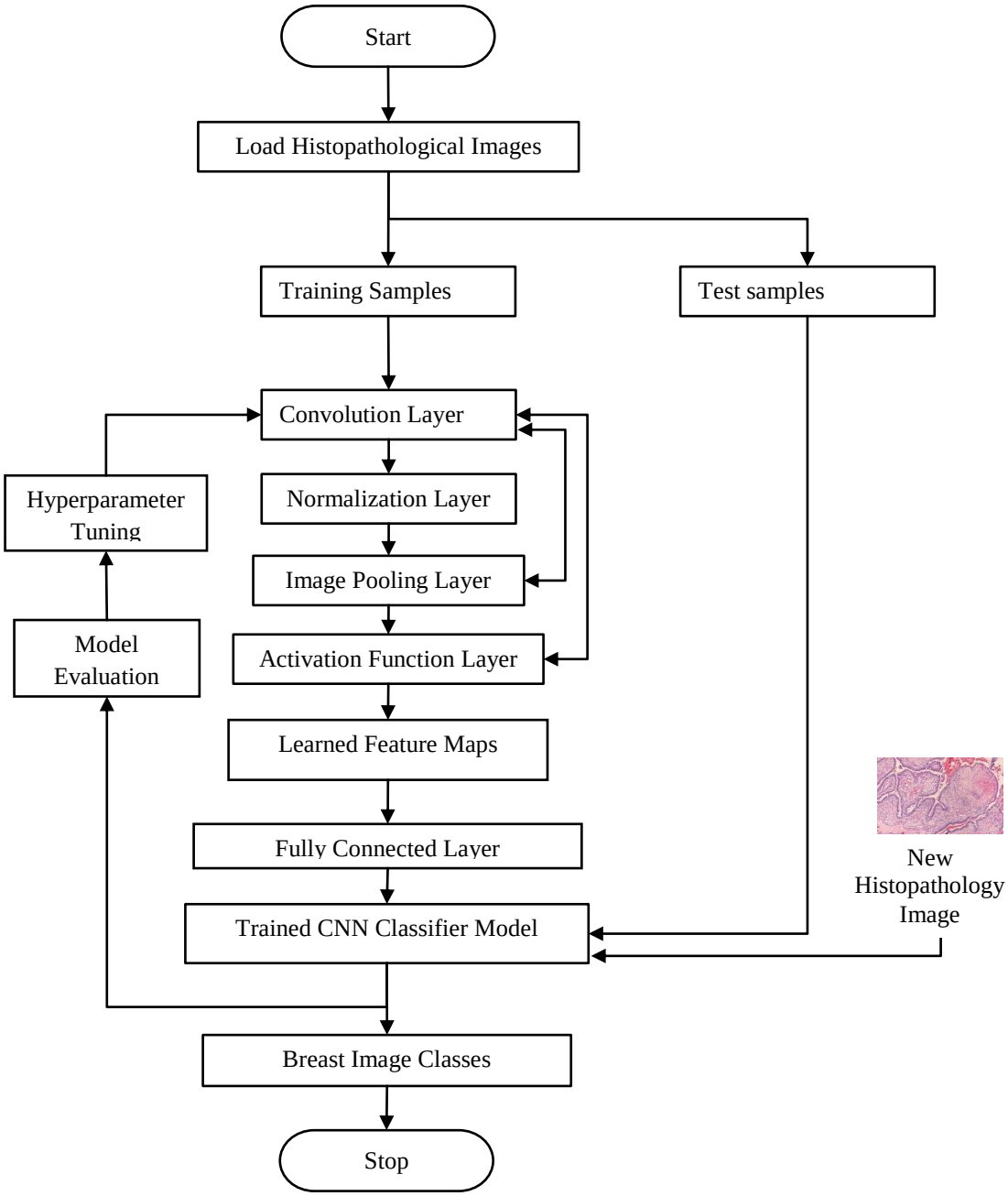

Figure 4: Logical Flow of the Convolutional Neural Network Image Classification Process

The ImageNet [32] statistics: the mean and standard deviation are used in the normalization of images because they perform better than that computed over the training dataset. The normalization values are used when implementing the base models.

Rectified Linear Units (ReLU) is used to transform an input into an acceptable range of value to aid convergence of gradient decent. The ReLU takes the form of the equation 5

$$y = \max(0, x) \quad (5)$$

In CNNs, image pooling is a crucial component because it aims to minimize the dimensionality of an input representation through down-sampling in order to make assumptions about the features present in the image. Image pooling is performed after the ReLU layer. While average

pooling returns the average value for each patch of the feature map, maximum pooling (Max-pooling) returns the greatest value for each patch.

Convolutional and pooling layers produce high-level feature representations of the input image as their output. Based on the training dataset, the fully connected layer classifies the input photos into the various breast image classes using these features. All neurons in the layer are totally linked to all activation in the layer below in the fully connected layer. To ensure that the output probabilities add up to 1, the Softmax activation function is applied to the output of the final layer of the Fully Connected layer. A vector of real numbers is provided as input to the Softmax function, which normalizes it into a probability distribution with *K* probabilities between zero and one that are proportional to the exponential of the input values. Some of the input values may be negative or greater than one before applying Softmax, but all of the resultant values are between 0 and 1 after applying Softmax, and all of their components add up to 1 so that they can be read as probabilities. The mathematical definition of Softmax is given in equation 6 [31].

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad 6$$

where $\vec{z}$ is the input vector made up of values $(z_0, \ldots, z_k)$, all the values z_i are elements of the vector, e^{z_i} is the exponential of each vector and *K* represents the number of classes in the classifier.

In the classification task, the last neural network layer will produce raw prediction values for the image which will be transcribed into the probability that the image is from a class.

Model evaluation is carried out using the Cross-entropy loss. As the anticipated probability distances from the actual class label, cross-entropy loss grows. For each class label per observation, a distinct loss value is computed, and the result is added across the training samples using equation 7.

$$CEL = -\sum_{i=1}^{M} y \log(p_i) \quad 7$$

where *M* denotes the number of breast tissue classes (Adenosis, Ductal Carcinoma, Fibro Adenoma, Mucinous Carcinoma and Tubular Adenoma); y is the expected probability returns 1 for the target class label and 0 for other classes; p_i is the predicted probability output from the Softmax function for the i_{th} class; log represents the natural logarithm. This measures the distance between the target output and the Softmax values.

The loss value represents the difference in the prediction and the actual target values. During training, the model adjusts the weights to minimize the loss value and give outputs close to the target values. The neural network weights are updated with the learning rate value until the error is minimized, and the gradient of the loss function is calculated with respect to each weight.

F. CNN Ensemble

An ensemble method combines results from base learners, and based on the collective learning of these outcomes, create a final prediction[14]. The process is illustrated in Figure 5. Three different Convolutional Neural Network models AlexNet, VGG16, and ResNet 50, are used. The base models are created with the same training dataset and the individual models are combined using the majority voting ensemble method to provide improved classification results. During image classification, each CNN model produces a set containing tensor values as the final output from the fully connected layer Softmax activation. The neuron with the highest tensor value is activated as the output class. A tensor is the primary data structure for neural networks and holds a multi-dimensional array. The conceptual diagram for the ensemble model is shown in Figure 5.

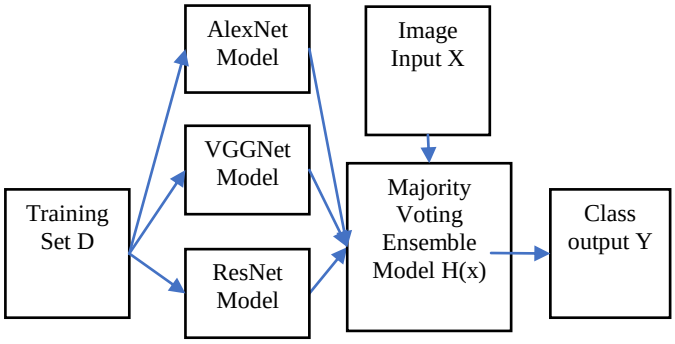

Figure 5. Conceptual Diagram of the Ensemble Model

G. Majority Voting Ensemble Model

The majority voting method was selected for this work because the majority voting method enhances the prediction abilities of the base models and produces stable and more robust model with improved prediction results while also capturing the intelligence of the base models.

Each base classifier is trained separately using the entire training set and each test case, each of these base models make predictions. The results of these CNN model's predictions are combined in the ensemble model. The final output from the ensemble is the most voted image class.

Algorithm 1: Majority Voting Ensemble

Input: Training set D = $\{(x_i, y_j)\}$ as an input with a set of images X = $\{x_i, \ldots, x_n\}$ with input size 224 × 224 × 3 and a set of image classes Y = $\{y_j, \ldots, y_m\}$, solving the problem of supervised classification to learn the model on the training set.
Output: A majority voting ensemble classifier h

Begin

Learn from the training data using the base classifiers

1. for i←1 to t do
2. Learn a base classifier h_i based on D

#To learn a new classifier h based on the modal output of the base classifiers

3. *count* =0
4. for test image *count_class_*y_j in parallel output:
5. if classifier h_i returns class y_j:
6. *count* += *count_class_*y_j
7. if *count_class_*y_j >count:
8. H(x)= y_j
9. **Return** H(x) = y_j = mode((h_1(x),(h_2(x) ,… (h_t(x))
10. End

H. Averaging Ensemble Model

The averaging ensemble model computes the average predictions of the base models for each test image and thereby produces a better classification than the individual base model. To achieve the averaging ensemble, each base classifier is trained separately using the entire training set and each test case, each of these base models make predictions. The results of these CNN model's predictions are combined in the ensemble model by averaging the output prediction values. The final output from the ensemble is the image class with the highest values after averaging.

Algorithm 2: Averaging Ensemble

Input: Training set D = {(x_i, y_j)} as an input with a set of images X = {x_i, …, x_n} with input size 224 × 224 × 3 and a set of image classes Y = {y_j, …, y_m}, solving the problem of supervised classification to learn the model on the training set.

Output: An averaging ensemble classifier h

Begin

Learn from the training data using the base classifiers

1. for i←1 to t do:
2. Learn a base classifier h_i based on D

To make predictions using a classifier h based on the mean of the outputs of the base classifiers learned

3. for test image x_t,
4. len =0
5. final_predictions = []
6. if classifier h_t returns prediction values prediction:
7. for column_num in prediction
8. prediction[column_num]+=prediction[column_num]
9. len +=1
10. H(x)=final_predictions.append(prediction[column_num]/len)
11. Return H(x)

I. Stacking Ensemble Model

The stacking ensemble model use the base model predictions values for each test image as input to the Meta model to produce its own prediction. To achieve the stacking ensemble, each of the base classifier are firstly trained individually on the whole training data and for each of the test instances, each of these base models make predictions. The resulting tensor values prediction from the training model are used as features that the meta model use to finalize the ensemble model's predictions on the test set. The final output from the ensemble is the image class with the maximum tensor values in the stacked tensor values. This is illustrated as in Figure 6.

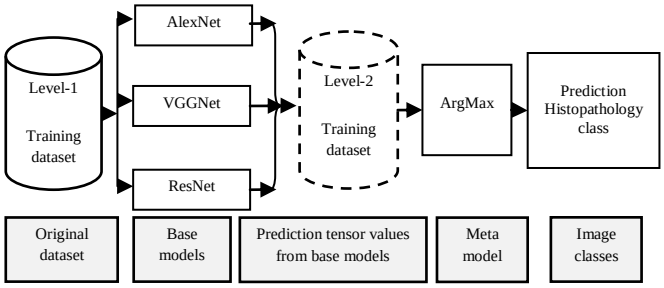

Figure 6: Stacking ensemble framework

J. The Proposed SaduNet Algorithm

To classify medical images, some convolutional neural network techniques have been employed. With the use of the Convolutional Neural Network Image Classification Process described in Figure 4, the SaduNet algorithm attempts to improve classification of histopathology images with the adoption of the modular convolutional neural network units. The algorithm operates as follows:

Algorithm 3: SaduNet

Input: Training set D = {(x_i, y_j)} as an input with a set of images X = {x_i, …, x_n} with input size 224 × 224 × 3 and a set of image classes Y = {y_j, …, y_m}, solving the problem of supervised classification to learn the model on the training set.

Output: A SaduNet classifier model

Begin

Step 1: SaduNet Block

To create a convolutional network block

1. Function Conv_Block_module (x, filter, kernel_size, stride, activation)
2. if conv_first:
3. x = conv(x)
4. if batch_normalization:
5. x = BatchNormalization()(x)
6. if activation is not None:
7. x = Activation(activation)(x)
8. if pooling:
9. x = AveragePooling2D()(x)
10. or
11. x = MaxPooling2D()(x)
12. else:
13. if batch_normalization:
14. x = BatchNormalization()(x)
15. if activation is not None:
16. x = Activation(activation)(x)
17. x = conv(x)
18. **return** x

Step 2: SaduNet Module

19. kernel_size = {7×7,5×5,3×3,1×1}
20. activation = {relu, softmax}
21. #To loop for the number of blocks needed using the Conv_Block_module function, t=number of SaduNet blocks,
22. for i ← 0 to t,
23. do
24. if i==0:
25. stride = (1,1)
26. else:


```
stride = (2,2)
filters = i + 1
x=Conv_Block_module(x, filters, kernel_size, stride, activation)
x = Flatten()(x)
x = Dense(classes)(x)
x = Activation(softmax)(x)
# Softmax activation generates output probabilities that classify image inputs to image classes Y; Y = {y_j, ..., y_m}
model = model(Input)
return model
end
```

128. A maximum pooling layer with kernel_size of 2 and stride of 2 was added to this layer.

The fourth and fifth convolutional layer – conv4 and conv5 have the same configurations with the second convolutional layer but with an output feature map of 256, and 512 respectively and a maximum pooling layer with kernel_size of 2 and stride of 2 was added after the fifth convolutional layer.

All layers including linear layers utilize the Rectified Linear Units activation function.

The Root Mean Square Propagation (RMSProp) neural network optimization algorithm is utilized. The schematic representation of the SaduNet model is shown in Figure 7.

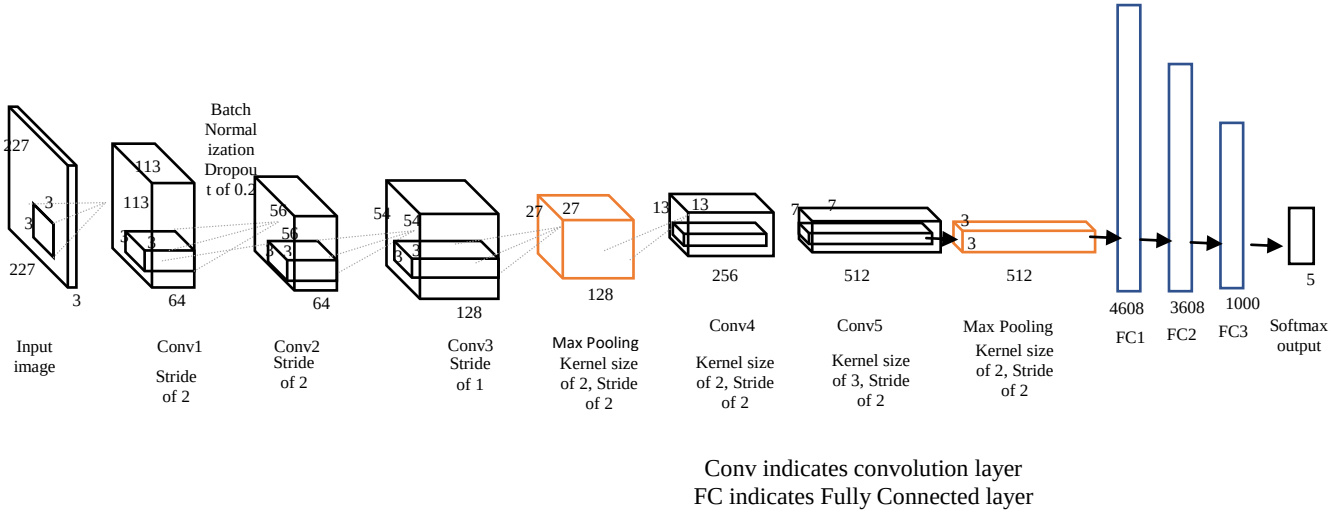

Figure 7: Schematic representation of the SaduNet Model

K. The SaduNet Model Design

The SaduNet has 5 Convolutional layers, 2 maximum pooling layers and 4 fully connected linear layers which sums to 11 layers with 9 weighted layers.

The first convolutional layer - conv1 of the SaduNet takes a RGB input image with size 227x227 and convolves with a kernel size of 3 and stride of 2 and then output a feature map with a depth of 64. In other to avoid covariant shift on the data, batch normalization technique is applied to the output of conv1 and then dropout with probability (0.2) is applied as a means to reduce overfitting.

The second convolutional layer – conv2 has an input of size 113x113 and convolves it with a kernel size of 3 and stride of 2 and then output a feature map with a depth of 64.

The third convolutional layer – conv3 accepts an input of size 56x56 and convolves this with a kernel size of 3 and a stride of 1, and outputs a feature map with depth size of

L. Implementation

The proposed model was implemented in three phases.

Phase 1: The base models were implemented on Google Colab device NVIDIA Tesla K80 GPU with 12 GB of RAM. This device is made freely accessible online at https://colab.research.google.com and provided the computation resources used to train the CNN models. The base models AlexNet, VGG-16 and ResNet-50 were implemented using Python programming language libraries Fastai. Fastai is a deep learning library that provides programmable components that can be used to develop models that solve machine learning problems. The base models were retrained using the histopathology images after being pre-trained using the ImageNet dataset. The training strategy used involved retraining the last layer of the models with the dataset before unfreezing. The settings on the parameters for training the base models are displayed in Table 1.

Phase 2: The ensemble was implemented using a personal computer with processor 6th Gen Intel Corei7 with speed ~2.6GHz (8 CPUs), a NVIDIA GeForce GTX 960M GPU

having 2GB of dedicated RAM, 16 GB RAM and 500GB hard disk, and a 64-bit version of Windows 10. The ensemble models are implemented using Python programming language libraries Flask (a microweb framework in Python). The ability of Flask to support extensions, which can add application functionalities that were not implemented in Flask, is the functionality used in this implementation. The CNN models created were ensemble using flask.

Phase 3: The SaduNet model was implemented on Google Colab device NVIDIA Tesla K80 GPU with 12 GB of GPU memory, accessible online at https://colab.research.google.com. This provided the computation resources needed to train SaduNet model. The implementation was done with Python programming language library PyTorch.

Table 1. Parameter settings for Implementation of the base models

Algorithm	Parameter settings
AlexNet	Training epochs = 8, learning rate = 2.4e-2
VGG16	Training epochs = 8, learning rate = 2.2e-3
ResNet 50	Training epochs = 8, learning rate = 2.0e-2

IV. RESULTS AND DISCUSSION

A. *Results during Training and Validation and Testing of the Developed Models with the FTH Ido Histopathology Dataset*

Table 2 presents the results obtained during training and validation of the base models. The validation set is 20% of the training set amounting to 375 images that had been set aside to tune and validate the model. The AlexNet, VGGNet and ResNet models were each trained with the whole training set and no data folding was done. The comparison of the AlexNet, VGGNet and ResNet models shows that ResNet models possesses greater accuracy during validation than the other base models.

Table 2: Results of the base models during validation

Base models	Accuracy	Error Rate	Precision	Recall	False Positive Rate	F1 Score
AlexNet	83.73%	0.1627	0.8368	0.8606	0.0456	0.8485
VGG16	82.67%	0.1733	0.8249	0.8308	0.0494	0.8278
ResNet50	97.87%	0.0213	0.9773	0.9824	0.0053	0.9798

The Table 3 presents the result of the base and proposed models during testing with the test set. The test set is 30% of the whole histopathology dataset and has 564 images. The test sample was collected with a script that randomly selected images for testing from the data.

Table 3: Results of the proposed models and base models during testing

Base models	Accuracy %	Error Rate	Precision	Recall	False Positive Rate	F1 Score
AlexNet	92.91	0.0709	0.9329	0.9322	0.0186	0.9325
VGG16	96.28	0.0372	0.9640	0.9661	0.0096	0.9650
ResNet50	99.29	0.0071	0.9933	0.9930	0.0018	0.9931
Majority Voting Ensemble	99.30	0.0070	0.9935	0.9931	0.0018	0.9931
Averaging Ensemble	99.47	0.0053	0.9950	0.9945	0.0014	0.9947
Stacking Ensemble	97.86	0.0214	0.9785	0.9778	0.0054	0.9781
SaduNet	75.89	0.2411	0.7592	0.7619	0.0735	0.7605

The base model accuracy during testing is notably better than that achieved during validation. This may be attributed to the test sample collection process.

The comparison of all the models shows that Averaging Ensemble model possesses more accuracy during testing than all the other models. The False Positive Rate of the Averaging ensemble is lower than that of the other ensemble models, meaning that there are a smaller number of histopathological images classified as cancer that are indeed not cancer. When considering the Recall values, the averaging ensemble method provides a better recall. This means there are less false negative values with the averaging ensemble than the other models tested. Furthermore, judging from the higher F1 score and Precision values, the Averaging model provides better classification. This is because the Averaging model tries to minimize the false positives and false negative scores. It is therefore the better algorithm in classifying the histopathology images. inally, of all models developed, the SaduNet achieved the lowest accuracy of 75.89% when tested. Also, the lowest F1 score and Precision values were achieved by the SaduNet. This is attributed to the small number of training images.

B. *Results during Testing of the Developed Models with the BreakHis Dataset*

The models developed were tested with the 40x magnification size of the BreakHis dataset and compared with some existing research works in literature that made use of the BreakHis dataset.

Table 5: Comparison of Accuracies of Developed Models with Accuracies of other Research Works

Models	Accuracy %
IRRCN [28]	98.59
Emax VGGNet [34]	91.28
NDCNN – AlexNet [35]	82.7
BiCNN [36]	97.89
Feature Dimensionality Reduction [37]	84.0
Majority Voting Ensemble	**97.90**
Averaging Ensemble	**98.32**
Stacking Ensemble	**97.90**
SaduNet	**79.20**

In the comparison of the accuracies of models in literature that were evaluated on the BreakHis dataset and that of the models developed presented in Table 5, the model developed by [28], which used Inception Recurrent Residual CNN performed with 98.59% accuracy followed by the Averaging ensemble model developed which gave 98.32% accuracy. This proves that the developed Averaging ensemble model possess particularly good predictive capacity during testing. The other developed ensemble models – the Majority Voting and Stacking ensemble models also performed well with accuracies of 97.90% each. Also, the Binary classification CNN model developed by [36] provided the accuracy of 97.89%. The other models Emax by [34], the CNN features dimensionality reduction model by [37], and NDCNN by [35], gave accuracies of 91.28%, 84.0%, and 82.7% accuracies respectively. The SaduNet developed performed with accuracy of 79.2% on the BreakHis data.

V. CONCLUSION AND FUTURE SCOPE

Digital analysis of histopathology images has brought great improvements to breast disease diagnosis and presents a better alternative to the manual scoring of histology slides done by pathologists. With breast histopathology, the analysis of digital images obtained from tissues collected from patient's breast to detect diseases is done with the use of computerized technologies. The benefits of these computerized analysis include the avoidance of human errors in the diagnosis of histopathology images. This research work presents the classification of histopathological images with the use of deep CNN ensembles and with the use of a convolutional neural network model - SaduNet. The AlexNet, VGGNet and ResNet are highly effective base models and form the basis of the ensemble methods implemented. The voting, averaging, and stacking ensemble methods presented achieved comparably high accuracies similar to the accuracies achieved by the other models in literature. The comparative analysis performed also proves that the ensemble methods performed better in the terms of the sensitivity and predictability than the individual base models. This is shown in the high prediction and recall values obtained by the ensemble models. The averaging ensemble has accuracy of 99.47%, majority voting ensemble has accuracy of 9.30%, stacking ensemble model has accuracy of 97.86%, and SaduNet model has accuracy of 75.89% when tested with the FTH histopathology data. Also, the averaging ensemble has accuracy of 98.32%, the majority voting and stacking ensemble models has accuracies of 97.90% respectively, and the SaduNet model has accuracy of 79.2% when tested with the BreakHis dataset. The SaduNet was trained for 100 epochs. The accuracy achieved by the SaduNet can be improved with more epochs of training and more training data.

In future work, we will seek the deployment of ensemble models in real-time classification of histopathology images.

REFERENCES

[1] A. Mashi, "Assessing Breast Cancer Burden amongst Women at General Hospital Katsina, State Nigeria.," International Journal of Social and Humanities Sciences (IJSHS), vol. 4, no. 1, pp. **95-114, 2020.**

[2] L. Elfatimi and H. Boucheneb, "rain Tumor Detection using Cellular Automata based image Segmentation techniques," International Journal of Scientific Research in Computer Science and Engineering, vol. 10, no. 5, pp. **22-36, 2022.**

[3] J. Qu, N. Hiruta, K. Terai, H. Nosato, M. Murakawa and H. Sakanashi, "Gastric Pathology Image Classification Using Stepwise Fine-Tuning for Deep Neural Networks," Journal of Healthcare Engineering, pp. **1-13, 2018.**

[4] M. A. Alnafea, "Detection and Diagnosis of Breast Diseases," in Breast Imaging, IntechOpen, **2017.**

[5] M. H. Vorteile and G. d. Brust-Ultraschalldiagnostik, "Advantages and limitations of breast ultrasound," Gynakol Geburtshilfliche Rundsch, vol. 42, no. 4, pp. **185-190, 2002.**

[6] H. Mohan, Textbook of Pathology, New Delhi: JAYPEE BROTHERS MEDICAL PUBLISHERS (P) LTD, **2010.**

[7] G. Orchard and B. Nation, Histopathology, 2nd ed., Oxford University Press, **2017**.

[8] M. Kowal, P. Filipczuk, A. Obuchowicz, J. Korbicz and R. Monczak, "Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images," Computers in Biology and Medicine, **2013**.

[9] F. A. Spanhol, L. S. Oliveira, C. Petitjean and L. Heutte, "A Dataset for Breast Cancer Histopathological Image Classification," IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, vol. 63, no. 7, pp. **1455-1462, 2016.**

[10] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li and S. Li, "Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model," Scientific Reports, vol. 7, p. Article No 4172, **2017**.

[11] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," Biol Cybern, vol. 36, no. 4, p. **193–202, 1980.**

[12] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition.," Proceedings of the IEEE, vol. 86, p. **2278–2324, 1998.**

[13] F. Chollet, Deep Learning with Python, New York: Manning Publications Co., **2018**.

[14] D. Li and J. C. Platt, "Ensemble Deep Learning for Speech Recognition," Redmond, **2014**.

[15] V. E. A. A. Adeyemo, N. JhanJhi, S. Mahadevan and A. O. Balogun, "Ensemble and Deep-Learning Methods for Two-Class and Multi-Attack Anomaly Intrusion Detection: An Empirical Study," International Journal of Advanced Computer Science and Applications, vol. 10, no. 9, pp. **520-528, 2019.**

[16] M. H. Motlagh, M. Jannesari, H. Aboulkheyr, P. Khosravi, O. Elemento, M. Totonchi and I. Hajirasouliha, "Breast Cancer Histopathological Image Classification: A Deep Learning Approach," biorxiv the preprint server for biology, 4 January **2018**.

[17] S. D. Raut and S. A. Thorat, "Deep Learning Techniques: A Review," International Journal of Scientific Research in Computer Science and Engineering, vol. 8, no. 1, pp. **105-109, 2020.**

[18] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," NIPS, **2012**.

[19] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in ICLR, **2015**.

[20] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, **2016**.

[21] G. D. Magoulas and A. Prentza, "Machine Learning in Medical Applications," in Machine Learning and Its Applications: Advanced Lectures, Berlin, Heidelberg, Springer Berlin Heidelberg, **2001**.

[22] M. M. Dundar, S. Badve, G. Bilgin, V. Raykar, R. Jain, O. Sertel and M. N. Gurcan, "Computerized Classification of Intraductal Breast Lesions using," IEEE Trans Biomed Eng., vol. 58, no. 7, p. **1977–1984, 2011.**

[23] Y. Jiang, L. Chen, H. Zhang and X. Xiao, "Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module," PLoS ONE, vol. 14, no. 3, **2019.**

[24] V. Gupta and A. Bhavsar, "Partially-Independent Framework for Breast Cancer Histopathological Image Classification," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), **2019.**

[25]A.-A. Nahid, M. A. Mehrabi and Y. Kong, "Histopathological Breast Cancer Image Classification by Deep Neural Network Techniques Guided by Local Clustering," BioMed Research International, **2018**.
[26]Y. Benhammou, S. Tabik, B. Achchab and F. Herrera, "A first study exploring the performance of the state-of-the-art CNN model in the problem of breast cancer," Morroco, 2018.
[27]D. Bardou, K. Zhang and S. M. Ahmad, "Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks," IEEE Access, pp. **24680-24693, 2018.**
[28]M. Z. Alom, C. Yakopcic, T. M. Taha and V. K. Asari, "Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network," **2018**.
[29]T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia and A. Campilho, "Classification of breast cancer histology images using Convolutional Neural Networks.," PLoS ONE, vol. 12, no. 6, **2017.**
[30]T. H. Vu, H. S. Mousavi, V. Monga, G. Rao and A. U. Rao, "Histopathological Image Classification using Discriminative Feature-oriented Dictionary Learning," IEEE TRANSACTIONS ON MEDICAL IMAGING, vol. 35, no. 3, p. 738–751, March **2016**.
[31]C.-C. Ko, C.-Y. Tsai, C.-H. Lin and L.-S. Liao, "A Computer-Aided Diagnosis System of Breast Intraductal Lesion Using Histopathological Images," in IVCNZ '14: Proceedings of the 29th International Conference on Image and Vision Computing New Zealand, **2014**.
[32]J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," Computer Vision and Pattern Recognition, IEEE, pp. 248-255, **2009**.
[33]S. Sharma, S. Sharma and A. Athaiya, "Activation Functions in Neural Network," International Journal of Engineering Applied Sciences and Technology, vol. 4, no. 12, pp. 310-316, **2020**.
[34]W. Zhi, H. W. F. Yueng, Z. Chen, S. M. Zandavi, Z. Lu and Y. Y. Chung, "Using transfer learning with convolutional neural networks to diagnose breast cancer from histopathological images," in International Conference on Neural Information Processing, **2017**.
[35]S. Akbar, M. Peikari, S. Salama, S. Nofech-Mozes and A. Martel, "Transitioning between convolutional and fully connected layers in neural networks," Deep Learning in Medical Image Analysis and Multimodal Decision Support, p. 143–150, **2017**.
[36]B. Wei, Z. Han, X. He and Y. Yin, "Deep learning model based breast cancer histopathological image classification," in Cloud Computing and Big Data Analysis (ICCCBDA), 2017 IEEE 2nd International Conference, **2017**
[37]S. Cascianelli, R. Bello-Cerezo, F. Bianconi, B. M. Fravolini, B. Palumbo and J. N. Kather, "Dimensionality reduction strategies for cnn-based classification of histopathological images," in International Conference on Intelligent Interactive Multimedia Systems and Services, **2017**.

AUTHORS PROFILE

S. P. Akinrinwa holds the Bachelor of Technology degree, Master of Technology degree and Doctor of Philosophy in Computer Science, obtained at the Federal University of Technology Akure, Nigeria in 2008, 2016 and 2021. She currently works as a Principal Network Engineer at the Department of Information Technology, Federal University of Technology, Akure, Nigeria. Her current research interests are in Deep Learning for Health tissue classification, focused on the use of Deep Convolutional Neural Network for the classification of Hematoxylin & Eosin (H&E) stained breast histopathology images. Her future interests are in Bioinformatics.

O. Olabode is a Professor of Computer Science from the Federal University of Technology, Akure. He holds a B.Tech degree in Industrial Mathematics, M.Tech. Degree in Computer Science, M.Tech in Statistics, MSc in Peace and Security Studies and PhD in Computer Science in 1991,1999, 2015, 2021 and 2005 respectively. He has over 80 publication in reputable Journals and Conference proceedings. His research interest is in AI and Machine learning. He is a reviewer of some academic and professional journals such as International Journal of Computer Systems and Applications, British Journal of Mathematics & Computer Science, Issues in Business Management and Economics.

O. C. Agbonifo obtained her B.Sc. degree in Computer Science from University of Ibadan, Ibadan, Nigeria in year 1995. She also received her M.Tech. and PhD degrees in Computer Science from Federal University of Technology, Akure, Nigeria in year 2005 and 2012 respectively. She is an Associate Professor in the Department of Information Systems. Federal University of Technology, Akure. She has authored/co-authored more than 50 research papers both at local and reputed international journals (including Scopus and Web of Science) and conference proceedings. She is a regular reviewer in local as well as international scientific/academic journals. She is a member of the NCS, CPN, IEEE, AACE and OWSD. Her research interest areas include: Personalised and Adaptive E-learning Systems, Digital Game-Based Learning Systems, Software Engineering and Artificial Intelligence. She has 16 years of teaching and research experience.

K. G. Akintola graduated from the Federal University of Technology with second class upper in Computer Science in 1995. He had his master's degree in the same Institution in 2004. In 2015, he bagged his PhD in Computer Science in the same institution. He also had a master's degree in Computer Information system from the University of Houston, Victoria, USA in 2011. His research interest include Software Engineering, Computer Vision and Machine Learning. He is currently a Reader in the Department of Software Engineering, Federal University of Technology, Akure, Nigeria.

