Abnormality Classification and Localization Using Dual-Branch Whole-Region-Based CNN Model with Histopathological Images

Olaide N. Oyelade[1], Absalom E. Ezugwu[1], Hein S. Venter [2], Seyedali Mirjalili[3], Amir H. Gandomi[4]

[1]School of Mathematics, Statistics, and Computer Science, University of KwaZulu-Natal, King Edward Avenue, Pietermaritzburg Campus, Pietermaritzburg, 3201, KwaZulu-Natal, South Africa
[2]Department of Computer Science, University of Pretoria, Pretoria 0028, South Africa
[3]Centre for Artificial Intelligence Research and Optimization, Torrens University, Australia.
[4]Faculty of Engineering and Information Technology, University of Technology Sydney, Australia

Authors email: Olaide N. Oyelade (email: oyeladeo@ukzn.ac.za), Absalom E. Ezugwu (email: ezugwua@ukzn.ac.za), Hein S. Venter (venterh@cs.up.ac.za), Seyedali Mirjalili (seyedali.mirjalili@griffith.edu.au), Amir H. Gandomi (gandomi@uts.edu.au)

Corresponding authors: Amir H. Gandomi (gandomi@uts.edu.au)

Abstract:
The task of classification and localization with detecting abnormalities in medical images is considered very challenging. Computer-aided systems have been widely employed to address this issue, and the proliferation of deep learning network architectures is proof of the outstanding performance reported in the literature. However, localizing abnormalities in regions of images that can support the confidence of classification continues to attract research interest. The difficulty of using digital histopathology images for this task is another drawback, which needs high-level deep learning models to address the situation. Successful pathology localisation automation will support automatic acquisition planning and post-imaging analysis. In this paper, we address issues related to the combination of classification with image localization and detection through a dual branch deep learning framework that uses two different configurations of convolutional neural networks (CNN) architectures. Whole-image based CNN (WCNN) and region-based CNN (RCNN) architectures are systematically combined to classify and localize abnormalities in samples. A multi-class classification and localization of abnormalities are achieved using the method with no annotation-dependent images. In addition, seamless confidence and explanation mechanism is provided so that outcomes from WCNN and RCNN are mapped together for further analysis. Using images from both BACH and BreakHis databases, an exhaustive set of experiments was carried out to validate the performance of the proposed method in achieving classification and localization simultaneously. Obtained results showed that the system achieved a classification accuracy of 97.08%, a localization accuracy of 94%, and an area under the curve (AUC) of 0.10 for classification. Further findings from this study revealed that a multi-neural network approach could provide a suitable method for addressing the combinatorial problem of classification and localization anomalies in digital medical images. Lastly, the study's outcome offers means for automating the annotation of histopathology images and the support for human pathologists in locating abnormalities.

Keywords: classification, localization, breast cancer, deep learning, convolutional neural network, CNN

1. Introduction

The use of computer-aided systems for the diagnosis of diseases gave the possibility of huge advances in research in the medical field. Recently, there has been a shift in the trend in the medical imaging area, from diagnosis to classification, which has motivated the design of novel techniques for image classification. Interestingly, the use of deep learning has attracted a lot of research efforts, which led to the re-emergence of several neural network architectures. This has demonstrated state-of-the-art performance in classifying image samples [1], [2], [3], [4], [5], [6], and other computer-aided diagnostic methods [7] [8]. Considering the progress made in the classification research, the need for detection and localization in images has also seen advances with significant achievements [9]. The procedure for localization of objects or abnormalities within an image often requires performing a classification process as well, hence increasing computational cost. The same is observed with detection, which aims to classify and detect all objects in any given image. Both localization and detection discriminate objects within an image using a

bounding box, thereby isolating the object for identification. Advances in deep learning have been providing support in addressing the challenge of image classification, localization, and detection, to achieve a great research momentum. Localization of abnormalities in medical imaging is a significant innovation to boost confidence in the classification output. It is even more interesting when both processes are automated and jointly reported using a system. While we recognize the performance of convolutional neural networks (CNN) on image classification to have reached human performance [10], gaps still exist in blending the task of classification with localization to revolutionize medical image analysis [11]. Widening research efforts have addressed the challenge of both localization and detection of objects and anomalies in images, which is motivated by the high cost of generating expert annotations on medical images. It is well known that the unavailability of expert annotated images poses a strong limitation for supervised machine learning methods in medical imaging [9]. As a result, the need to automate the detection and localization of abnormality is increasing the research in object detection and image localization, which benefit image classification [12]. Localization of an object in an image often follows either of two major algorithm categories that are based on deep learning models: the one-stage and the two-stage algorithms. The one-stage algorithms are represented in YOLO model [13], while models such as R-CNN, Faster R-CNN, Mask R-CNN, and Cascade R-CNN [14], [15], [16], [17] represent the application of the two-staged algorithms, which often select regions within an image for the object detection process. Some of these models have been proposed as replacements for image augmentation, a task that has been sufficiently handled in previous research, using generative adversarial networks (GANs) as seen in [18], [19], [20]

A significant number of publicly available datasets were labeled by human experts, which is a costly procedure and often laced with errors resulting in weakly annotated datasets [21]. Although medical experts that perform complex visual and perceptual tasks have reported they can categorize an image as normal or abnormal in a single glance and even detect lesions in chest radiographs and mammograms with only a quarter of a second look at the image [22], this is only a self-sale as several deep learning models deflated human experts [23]. To support this claim is the state-of-the-art performance reported of deep learning models in image annotation, analysis, acquisition, feature extraction and interpretation. For instance, it is reported that subtle textual abnormalities that present difficulty for human physicians to isolate have been accurately detected and classified using deep learning models [24]. Another justification for this claim is the difficulty associated with massive data from pathology and radiology images, genomic data, and other text-based medical data, which often overwhelm humans, leading to increased true and false positive rates [25]. Therefore, this establishes the difficulty of relying only on human experts in curating annotated samples and as well for image classification, localization, and detection. Considering this limitation, deferring the task to machines is not a non-trivial problem from the design perspective of models. A more problematic task lies in designing deep learning models capable of solving classification and detection in a combinatorial approach. Doing these simultaneously with the hope of identifying abnormal regions in medical images can be simultaneously crucial and difficult [26]. Detection is an important problem in image analysis that still does not have a satisfactory solution. Anomaly localization is an important problem in computer vision, which involves localizing anomalous regions within images with applications in industrial inspection, surveillance, and medical imaging [27]. However, several challenges still blight the performance of deep learning models as it relates to classification and localization. For instance, a CNN trained with weakly annotated datasets presents a non-trivial problem for the following reasons:

- overfitting may occur with those CNNs trained with small but strongly annotated datasets [28]
- increasing complexity of object detection framework and CNNs; the need to incorporate localization of abnormalities in medical images as part of the confidence mechanism for classification to ascertain its uncertainty or certainty or explanation remains a challenge [29]
- availability of not sufficient ground truth and annotated samples, which are often wrongly annotated due to high intra-person and inter-person variations [30]
- small pixel coverage of the location of the abnormality for medical images in real-world scenarios [27]
- localization techniques fail to focus on the relevant image region that fully contains the anomaly [31]

A considerable number of studies have been focused on addressing some of the problems listed earlier. In addition to those of YOLO, R-CNN, Faster R-CNN, Mask R-CNN and Cascade R-CNN models [13] [14] [15] [16] [17], we found other works which have reported localization and detection models based on deep learning models. The study in [32] applied a deep convolutional network (DCN) architecture and ensemble models that successfully localize abnormalities on chest X-Ray. In a related work, authors in [33] explored the performance of different saliency map techniques on the RSNA Pneumonia dataset, discovering that the GradCAM was the most sensitive and accurate for localization of abnormalities. Similarly, the authors in [34] proposed a deep learning model for the detection and

diagnosis in weakly supervised learning, based on multiple instance learning (MIL) for the classification and localization of abnormalities in the image. A study has reported the localization of mass abnormalities by using a weakly and semi-supervised framework that detected the presence of breast cancer [28]. The use of a dual branch deep learning architecture for the classification and localization in mammograms, including the presence and severity of abnormalities, has been proposed in [9]. A convolutional latent variable to localize abnormalities in images based on a convolutional adversarial variational autoencoder with Guided Attention (CAVGA) has also been leveraged to address a similar problem [27]. Besides, the work in [35] explored the capability of the global average pooling layer to achieve localization. Although a significant problem associated with localization and object detection has been addressed in these studies, we found that most studies largely rely on human-aided annotated samples and anomalous training images to classify and localize anomalies. Some other approaches have coupled both weakly and semi-supervised methods to achieve the localization of abnormities or detection of objects, a method that suffers from inconsistency in performance.

To address these challenges, this study proposes a novel dual branch, whole-image and region-based deep learning framework for improved classification and localization of multiple abnormalities in digital histopathology images retrieved from samples obtained from the pathologist. We combined the task of classification and localization of abnormalities using the whole image and region-based patches for whole-image convolutional neural network (WCNN) and region-based convolutional neural network (RCNN) architectures. Localization in the WCNN follows the method of sensitivity of the classifier to image occlusion, whereas localization with the RCNN finds anomalies in several regions of an image using the region and image-level class-label variation method to identify the existence of abnormalities. Meanwhile, the proposed framework in this study provides a mechanism for mapping probability maps and localizations achieved with WCNN and RCNN to provide further confidence to the classification tasks. The proposed method was tested using the digital histopathology image samples, presenting a comprehensive and comparative analysis of the method with other state-of-the-art solutions. The following highlights the contribution of the study:

i. The design of a new dual-branch deep learning model based on CNN architectures for solving the combinatorial problem of classification and localization.
ii. Proposal of a novel probability map comparator mechanism to compare the output from the WCNN and RCNN models.
iii. Image-level and region-level abnormality classification methods are investigated using image occlusion and region-to-label classifier methods.

The remaining part of the paper is organized as follows: Section 2 presents a summary of all the recent studies related to the domain of interest of this paper, where both an overview of what has been achieved so far and the gap that this study intends to fill are presented. In Section 3, a detailed presentation of the proposed architecture that combines WCNN and RCNN is presented and discussed. Dataset preparation and system configuration details are presented in Section 4, and detailed results are presented with a discussion of findings in Section 5. The paper is concluded in Section 6.

2. Related Works

In the literature, several research efforts have been focused on the application of deep learning models to the task of classifying abnormalities in digital medical images. Interestingly, the use of digital mammography in the detection of cancerous lesions leading to breast cancer has been widely researched. However, recent studies are now emphasizing and looking at the need to include also the localization of lesions in the image samples. In this section, we focus the review on recent studies using different methods to achieve localization of abnormalities in medical images, with emphasis on digital mammography and histopathology images.

Bilen *et al.*, proposed the use of a weakly supervised learning method to tackle the challenge of detecting objects in images using deep CNNs, pre-trained on a large-scale image level [35]. The study compared their approach with similar methods, which have applied Pascal visual object classes (VOC) data to the classification of region-level objects in an entire image. The outcome of the comparison showed that their deep learning region-level classification in whole images outperformed standard data augmentation and fine-tuning techniques. In a related study, Wang *et al.* investigated the performance of a deep learning system for the detection of the presence of metastatic breast cancer in whole slide images [36]. The method successfully localized the presence of the tumor in image samples that yielded an area under the receiver operating curve (AUC) of 0.925 and 0.7051 for classification and tumor localization tasks, respectively. This method demonstrated an 85% reduction in the human error rate compared with human experts.

Abnormality localization in chest X-rays was the focus of the study presented by Islam *et al.*, which adapted a shallow DCN architecture with ensemble models for this task. The method successfully localized spatially spread-out abnormalities like cardiomegaly and pulmonary edema that yield the highest accuracy on chest X-Ray abnormality detection at an improvement of 17% [32]. The use of a shallow network as seen in the works of [32] cast aspersion on the performance of the model, while those of [37] and [36] reported a low performance in the classification accuracy.

The use of attention maps, saliency maps and other mapping techniques have shown to produce a good performance in the localization of abnormalities, tumors, and objects in images; although these techniques have also been applied to non-medical images, Arun *et al.*, investigated the performance and effectiveness of these techniques in medical images [33]. The study particularly explored different saliency map techniques on the RSNA Pneumonia dataset to establish how credible the methods are in addressing the challenge of localizing an abnormality. The study's outcome showed that GradCAM was the most sensitive to the model parameter and label randomization, and highly agnostic model architecture. Choukroun *et al.*, advanced on the challenge of localizing abnormality in medical images using digital mammography [34]. This was achieved using a novel deep learning model for the detection and diagnosis in weakly supervised learning, with local annotation of the sample images. The deep learning model is based on the multiple instance learning (MIL) paradigm, so the severity score and localization of abnormalities in the image are made possible. Using multi-center mammography cohort and INbreast datasets, the proposed method discriminated and localized patches in the mammograms which contain abnormalities. Using the same INbreast dataset, Ribli *et al.*, proposed an object detection framework based on Faster R-CNN to localize abnormalities [10]. Achieving an AUC of 0.85, the method performed well in the classification of the image samples and yielded high sensitivity in the localization of abnormalities. Shin *et al.* used a weakly and semi-supervised framework for the task of classification and localization of mass abnormalities in medical images to detect the presence of breast cancer [28]. Considering the availability of weakly annotated image samples, the authors investigated the effectiveness of their approach to successfully localize and classify masses. These combined only a few strongly annotated images with more weakly annotated images to give a 95% confidence interval.

Bakalo *et al.*, proposed the use of a dual branch deep learning architecture for the classification and localization of mammograms to identify the presence and severity of abnormalities [9]. A distinguishable feature in the proposed method is the ability to allow localization of the different types of findings. The proposed method combined both region-level classification and region ranking in the process of localization of abnormalities in image samples. An evaluation of their approach showed that it has more advantages of the proposed method over a previous weakly-supervised strategy. In another study, Bakalo *et al.* applied two deep learning branches to multi-class classification and localization of abnormalities in digital mammogram images [21]. While the first branch was applied to the region classification, the other branch was only applied to region detection and ranking regions relative to one another. Using an objective function supporting the use of local annotations, the study showed that much time and effort is reduced through the weakly supervised method that relies on only a subset of locally annotated data. Venkataramanan *et al.*, proposed a convolutional latent variable to localize abnormalities in images based on a convolutional adversarial variational autoencoder with guided attention (CAVGA) [27]. The aim of the proposed approach is to support the preservation of spatial information and to focus the model on normal regions in the image through an attention expansion loss mechanism. The weakly supervised component of their proposed model uses a complimentary guided attention loss to guide the attention map to focus on all normal regions. The proposed CAVGA was experimented with using the MNIST, CIFAR-10, Fashion-MNIST, and MVTec Anomaly Detection (MVTAD), modified ShanghaiTech Campus (mSTC) and Large-scale Attention-based Glaucoma (LAG) datasets.

Object localization has also been researched in other domains other than medical images. Zhou *et al.*, for instance, explored the capability of the global average pooling layer in conventional CNN to achieve localization. The proposed concept can produce a generic localizable deep representation, even when using image-level labels. The authors claimed the use of the method for localization in images other than medical images achieves a 37.1% top-5 error for object localization and discriminates image regions on a variety of image samples [35]. Image localization and object detection has also proven relevant in pedestrian detection. For example, Liu *et al.* argued that the need to discriminate false positive objects from hard negative is necessitating the use of high-level semantic vision cues. To address this, the authors proposed the use of a Faster R-CNN detection framework as a region-based method for the localization of pedestrians within the vehicular traffic space. Integration of a branch of the network for semantic image segmentation was made into the proposed framework to combine both the semantic features with convolutional features for achieving good detection accuracy for pedestrians at different scales [38]. Comprehensive reviews on deep learning and image analysis in medical fields can be respectively found in the recent studies of [39] and [40].

In another study, the need to investigate the use of deep learning models over machine learning models was advocated in handling cancer detection. The study aimed to overcome the difficulty associated with handcrafted feature extraction by adapting the deep learning model to automate the same task to improve classification accuracy. Using a hybrid of AlexNet-gated recurrent unit (AlexNet-GRU) model, the study was able to detect cases of breast cancer that has spread to the lymph node and using the extracted features, classification of the disease was accomplished. The performance of the hybrid model was investigated with similar hybrids such as CNN-GRU and CNN long short-term memory (CNN-LSTM. The results showed that their proposed model outperformed other methods to confirm the relevance of deep learning in feature extraction [41]. The outcome reported in the existing study aligns well with the argument considered in the current study as relating to the use of the deep learning model in feature extraction. However, the current study advocates for localization rather than simply doing classification alone. In a related work, Belay *et al.*, applied the hybrid model of CNN-LSTM to chickpea disease detection in crop production through feature extraction and then disease classification. Using well-known image pre-processing techniques such as Gaussian filter (GF), median filter (MF), and softmax as the classifier, the study revealed that the deep learning-based approach yielded outstanding classification performance when compared with other methods. Interestingly, we find the issue of using deep learning models for classification to be replete in literature and that further procedure might be required to explain the classification process [42]. Still, on the plant disease classification issue, Alshammari *et al*. study, which used a deep learning model to diagnose olive leaf disease, showed a good classification accuracy. The study applied an adaptive genetic algorithm to optimize the selection and combination of hyperparameters to obtain optimal combinations for improved performance. The outcome of their approach also confirmed the relevance of using a deep learning model for feature extraction leading to classification [43]. Another application of deep learning to plant disease detection has been reported in [44]. Using three variants of the popular Inception-v3 architecture, the study successfully classified cases of abnormal maize plant infection.

Among several disease detection and classification studies reported using deep learning, it appears that cancer disease classification has received much research attention. Saber *et al.*, reported how deep learning and transfer learning have been used to improve the procedure for feature extraction and classification of breast cancer. Using the popular mammographic image analysis-society (MIAS) dataset, the study investigated some pre-trained models: Inception V3, ResNet50, Visual Geometry Group networks (VGG)-19, VGG-16, and Inception-V2 ResNet to extract features from the image samples. However, the study was only scoped to the task of feature classification leading to breast cancer disease diagnosis [45]. We suggest that new research contributions must be advanced beyond classification alone as there are other clinical and pathological procedures often carried out in disease isolation and can also be automated using deep learning algorithms. To further support our observation on the need to advance research beyond the classification task alone, we reported Arooj et al.'s work, which focused on the classification of benign and malignant breast cancer cases. Using a combination of deep learning and transfer learning, the study applied both ultrasound and histopathology images samples to the task of breast cancer classification on the benchmarked CNN-AlexNet model [46]. Remarkably, transfer learning has already been widely reported in the literature to benefit research efforts in classification. Therefore, we advocate for a combinatorial solution in addressing combinatorial problems relating to disease diagnosis. In related work, a study has also demonstrated the relevance of deep learning models, particularly CNN models, to address the problem of feature extraction from image samples leading to tongue cancer diagnosis. Using a privately reconstructed dataset, the study successfully applied the CNN model to detect and classify malignant cases from the image samples. Results obtained from the study showed that both classification and sensitivity were improved using their method, leading to acceptable performance for tongue cancer diagnosis [47].

Image segmentation has also received substantial research interest in the area of deep learning. Shuai *et al.*, have reported successful use of deep leering models to locate and extract lesion presence from endoscopy image samples intelligently. The lesion segmentation aims to enable them to diagnose the presence of gastrointestinal tract diseases through image classification. The study was motivated by the need to improve the regular handcrafted feature extraction and representation method. The approach, an end-to-end lesion segmentation using a multi-scale context-guided deep network (MCNet), aimed to fully identify features from image samples covering local and global spatial space against down-sampling operations. The MCNet combined models which extract global structure and high-level semantic context from samples. This method can help localize abnormality and as well collect global information. The output from the model is feature maps which are combined to achieve the task of lesion segmentation [48]. Yu et al., applied a back-propagation (BP) neural network to the task of image segmentation for the retinal vessel. Using four different image enhancement methods: Hessian matrix filtering, adaptive histogram equalization (AHE), Gaussian matched filtering, and morphological processing, image samples from the DRIVE and STARE datasets were

applied to the hybrid model to achieve complete retinal blood vessel segmentation. In addition, the proposed neural network model successfully segmented connected vessel stems and terminals [49]. The approach demonstrates a high level of characterization and localization of abnormality through the lesion segmentation procedure. This is similar to what is proposed in the current study, which combines classification with localization for improved disease diagnoses.

The use of deep learning-based attention or localization mechanism for the detection of suggestive features has also received focus in domains outside medicine. For instance, Qing et al. demonstrated the possibility of applying a pre-trained CNN-based model, namely XM-CNN, for extracting semantic relations in the information extraction process. The deep learning model was combined with word-embedding and position embedding methods. The authors argued that the deep learning-based method proved relevant in addressing the problem of entity-aware attention mechanisms [50]. The novel coronavirus disease has been widely investigated using the computer-aided diagnostic systems (CADs) approach. The deep learning method has been proven to have outperformed most of the CADs systems reported in the literature. Su *et al.*, combined Multi-Verse Optimizer (MVO) and multilevel thresholding image segmentation (MTIS) approaches for the segmentation of chest x-ray. The performance of the hybrid model was compared with other optimization algorithms to investigate the enhancement received on the MTIS method [51]. In a different study, the authors proposed multiple visual fields cascaded convolutional neural network MVF-CasCNN to identify the features leading to the classification of breast cancer. Combining both whole-image-level and region-level classification, the study applied the CNN model to localize the presence of tumor from the input dataset Camelyon16 [52].

Meanwhile, in our recent studies, we have also investigated the use of deep learning in the classification of breast cancer using different datasets. In [20], we proposed a new CNN architecture for detecting breast cancer using DDSM+CBIS datasets with a data augmentation technique. We demonstrated that image augmentation and preprocessing techniques are relevant in reducing false positive errors. In another related study [1] which we carried out, we showed that using a hybrid random-grid model will support selecting optimal hyperparameters of CNN model, thereby improving classification performance. To further improve the model's performance in [1], we investigated the capability of metaheuristic algorithms in guiding the process of selecting optimal hyperparameters for a combinatorial optimization problem [3]. Considering the outstanding performance of five optimization algorithms when used with the CNN model, we improved the classification accuracy of diagnosing breast cancer in digital mammography. More so, in a recent study, we investigated the use of a novel optimization algorithm [5] in evolving a new CNN model capable of solving the classification problem in histopathology images for the detection of breast cancer [53]. The outstanding performance demonstrated by the neural architectural search (NAS) model showed that new neural architecture can be auto-designed using metaheuristic algorithms. In another study [6], we showed that the use of a wavelet-based activation function as a replacement for the popular RELU function is suitable for supporting feature detection and extraction procedure in convolutional-pooling operations. Results obtained in the study confirmed that our proposed wavelet function outperformed RELU in yielding better extraction of discriminant features of breast cancer in digital mammography. Meanwhile, we have also investigated the applicability of deep learning models, CNN, in classification tasks on the novel COVID-19 disease [4].

Based on the review in this current section, many studies have proposed different methods using deep learning models to address the challenge of classifying and detecting objects or localization abnormalities in the images. However, considerate was shown that while the methods reported in the literature have demonstrated good performance on medical images, there is a need for an improved method capable of localizing abnormalities in histopathology images with anomalies suggestive of breast cancer. This is motivated by the view that this class of medical image presents a difficult task of image analysis for the purpose of classification and detection of abnormalities. In addition, the multi-location and multi-abnormalities existing in a single image also demand a high-level model for localizing the presence of such abnormalities. As a result, this study aims to further close the gap in image localization by addressing some of these unresolved challenges through a novel method.

3. Methodology

The overview of the approach proposed in this study is described in this section. First, the pipeline of the flow of the process in the system is presented. Following this architectural design of the pipelining that shows the flow of the processes, there is a discussion on each component in the system. Furthermore, we present an algorithmic design of the system to support the implementation process. Meanwhile, a detailed description of the architectures of the CNNs that were applied for training in the proposed system is also presented.

3.1 Overview of the Approach

The architecture shown in Figure 1 summarizes the approach that is proposed in this study. The two-level medical image abnormality detection, classification and localization features that are included in the proposed system consist of the following major components:

i. The WCNN architecture for feature extraction in the whole image
ii. The classification-occlusion subsystem
iii. The probability map comparator
iv. Whole-image-to-region-based image extractor
v. The RCNN architecture for feature extraction in region-based images
vi. Fully connected layers
vii. Multiclass classifier of classification purpose
viii. Multiclass classifier for detection of abnormalities
ix. Region-to-label-probability mapper

The subsystems listed above are discussed in detail in the following paragraphs. The procedure for executing figure 1 follows the approach of parallel feature extraction, classification and localization. The implication is that the upper branch of the model is not dependent on the lower branch since both the WCNN and RCNN models consequently apply their computational procedure to the inputs supplied to the complete architecture.

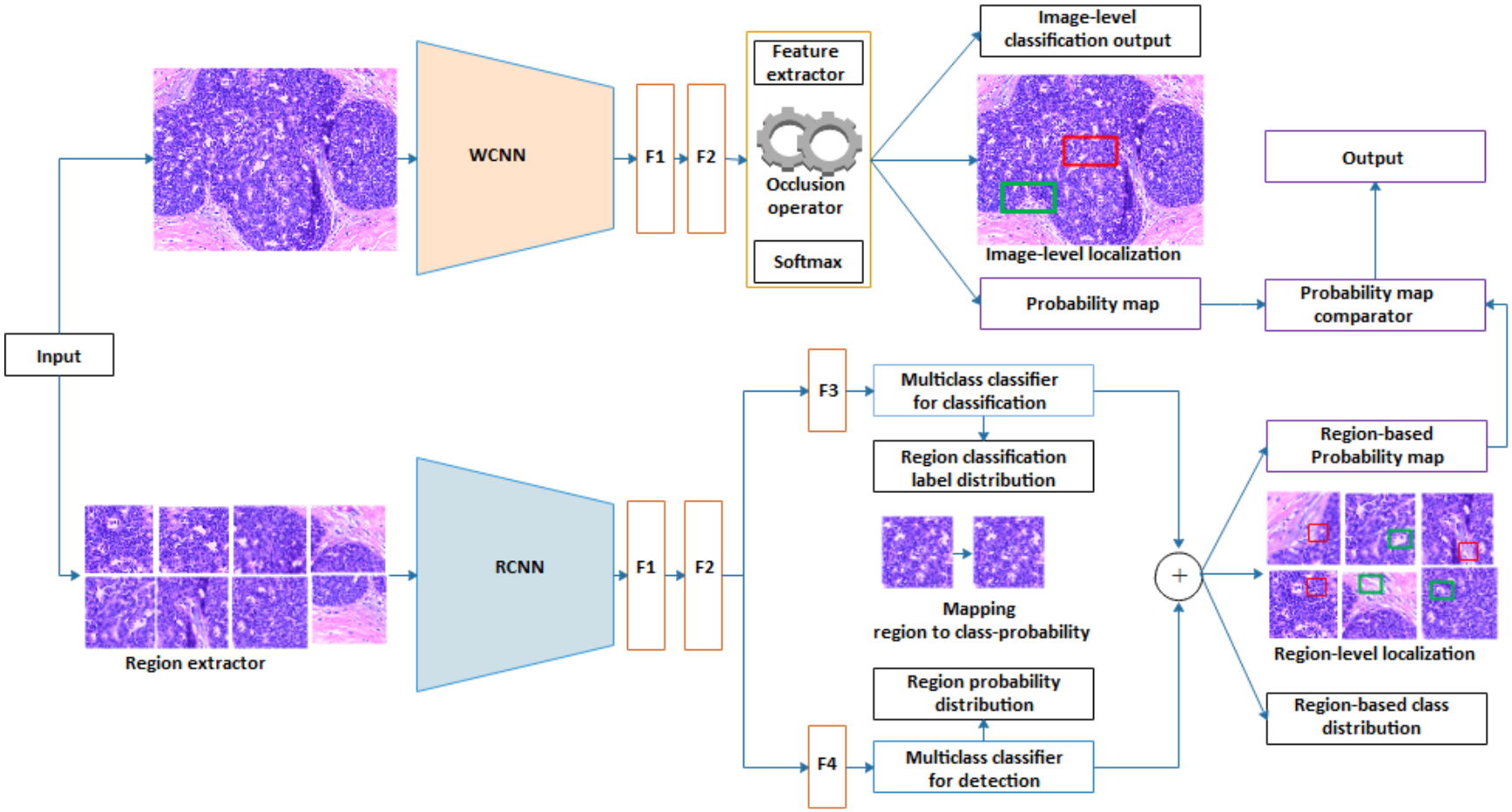


Figure 1: An architectural layout of the WCNN-RCNN for image-level and region-level characterization and localization of abnormalities in digital histopathology images. The approach combines the method of image occlusion and region-based detection for the localization of abnormalities.

A. The WCNN architecture for feature extraction in the whole image

Image inputs are passed into a WCNN, which applies the combination of convolution and pooling operations for the task of feature extraction. Image samples are first applied to some preprocessing techniques to support and enhance the feature extraction process. The resulting pre-processed image samples are then passed into the WCNN. Consider that a batch size $\boldsymbol{b}$ of image samples $X = \{x_1, x_2, \dots x_b\}$ were passed as input to the WCNN for the feature extraction procedure. The convolutional-pooling operation follows in Equation (1), the convolutional operation is represented based on filter specification and the outcome on all filter count for the convolutional layer is computed in Equation (2). In Equation (3), a pooling operation is applied to reduce the dimensionality of the previous output. The implication of using the convolution-pool operations over all the convolutional layers in WCNN is shown in Equation (4).

$$x = actFunc(\sum_{i=0}^{n}\sum_{j=0}^{m}(f * x_{i,j} \,.\, w_{i,j}) + b\,) \tag{1}$$

$$xfeaturemap = f(x;\ \omega_1 \dots \omega_h) \tag{2}$$

$$poolx = apool([\max|avg],\ \text{xfeaturemap}) \tag{3}$$

$$features = WCNN[layern(\dots layer2(layer1(poolx)))] \tag{4}$$

where $\boldsymbol{f}$ denotes the filter, $\boldsymbol{h}$ is the filter count, $\max|avg$ indicate that either the max-pooling or average-pooling operation is chosen, $\omega_1 \dots \omega_h$ denotes filter 1 to $\boldsymbol{h}$, and $\boldsymbol{x}$ is an instance of an image sample in $\boldsymbol{X}$.

Note also that $actFunc()$ represent the activation function (e.g RELU) used to achieve non-linearity. We defer the design and architectural composition of the WCNN to sub-section 3.2 for discussion.

B. The classification-occlusion subsystem

Localization of the presence and position of abnormality in image samples at the image level was achieved using the occlusion technique. The technique relies on the sensitivity of the classifier (in this case, the softmax function) to respond to the outcome of image classification, which is based on the image region covered during the feature detection procedure. The implication of this is that a section of the image is masked in a manner to overwrite the actual pixel representation in that region. Therefore, this renders that region of the whole image unresponsive to the classifier. Consequently, to find the region of an image where abnormality exists, the occlusion operator is applied on patches of the image, which is repeated until the image is covered. Meanwhile, the occlusion operator is supported by a feature extractor which allows for obtaining features for a given section of the image from the entire feature obtained from $\boldsymbol{F1}$ and $\boldsymbol{F2}$ layers. We describe the procedure for occlusion in Equation (5) and the application of the softmax function on the occluded image in Equation (6).

$$\vec{x} = occlude(xfeature,\ (w * h, 0)) \tag{5}$$

$$\sigma(\vec{x})_i = \frac{e^{\vec{x}_i}}{\sum_{j=1}^{n} e^{\vec{x}_j}} \tag{6}$$

where $xfeature$ represents the vector representation of feature from $F2$, tuple $(w * h, 0)$ represents the width and height of the occlusion region which is being cascaded over the image and 0 is an indication that a black mask is used for the occluding region in the whole image. The output $\vec{x}$ is the vector representation of occluded (or covered away) which is passed to the softmax function. e^{x_i} is the standard exponential function applied to each element of the input vector, and n is the number of classes in the multi-class classifier.

The output from the occlusion mechanism results in three objects: the image-level classification indicating the predicted class of abnormality or class of *normal* that the image belongs to in Equations (7) and (8); the probability map generated for the image, and image-level localization of abnormality detected in it.

$$argmax\big(\sigma(\vec{x})\big) = argmax\left(\begin{bmatrix} prob_1 \\ \vdots \\ prob_n \end{bmatrix}\right) = \begin{bmatrix} 1|0 \\ \vdots \\ 1|0 \end{bmatrix} \tag{7}$$

$$class(x) = \{c[i] \mid argmax(\sigma(\vec{x}_i)) = 1 \wedge i \in 1,2,\dots n\} \quad (8)$$

where $prob_1 \ \dots \ prob_n$ represents the probabilities of prediction on image sample(s) supplied as input, and the $argmax()$ translates the output into a format to support mapping the class listing in $c[i]$.

C. The probability map comparator

The localization obtained from both WCNN and RCNN are compared using the probability map comparator to boost the confidence level on the outcome of the system. To achieve this, the probability map from the two CNN models is compared to investigate if the error rate is below a threshold. When the condition is satisfied, we consider that the localization achieved from both models and the final classification is correct. The probability value $prbd(x)_i$ from the probability distribution $prbd(x)$, which corresponds to the classification result that matches the original label of x, is used to compute the probability similarity (ps) in Equation (9). For those in $prbd(x)$, which does not match with the original label of x, we apply Equation (10) to compute (ps).

$$ps = \frac{prbd(x)_i}{1 + prbd(x)_i} \quad (9)$$

$$ps = \frac{\max(prbd(x))}{1 - prbd(x)_i} \quad (10)$$

where $ps < 0.5$, tally and count, then compare the number of counts of WCNN with RCNN. The (ps) for all x instances in all images X used in WCNN and RCNN are compared.

D. Whole image to region-based image extractor

In the case of WCNN, we use the whole image for the feature extraction process whereas, in the RCNN, we applied regions extracted from the image for the feature extraction process leading to classification. For every image sample x, regions $r_i \in R$ are extracted, so that Equation (11) follows:

$$R = \delta(x, (\mathrm{w} * \mathrm{h})) \text{ yields } \{r_1, r_2 \dots r_m\} \quad (11)$$

where δ is the extractor function, contrary to the method used in [21], our regions do not overlap, being evenly extracted from x. These regions are passed into the RCNN for feature extraction and classification. Note that this means that each r_i is operated on as x in WCNN.

E. The RCNN architecture for feature extraction in region-based images

The operation of RCNN on the samples passed into it as input is comparable to that described in WCNN. We detail the design and discussion of RCNN in section 3.2. Whereas WCNN uses an occlusion mechanism on the whole image, the RCNN is appended with a dual-stream mechanism for detection of features leading to localization of region and the other stream for classification purposes.

F. Fully connected layers

The WCNN and RCNN architectures are followed by two fully connected layers namely F1 and F2. This allows for applying linear functions to the outcome of the convolution-pooling blocks to obtain a flattened output. This is achieved by mapping from $\mathbb{R}^{\mathrm{m}}$ to $\mathbb{R}^{\mathrm{n}}$ to return an output of the form $y_i = \emptyset\ (w_1 x_1 \ \mathrm{x1} + \cdots + w_m x_m)$. The duo F1 and F2 output are passed to the corresponding fully connected layers in the following dual branches.

G. Multiclass classifier of classification

The upper stream in the dual branch from RCNN describes the classification process of the regions that are extracted from each image. First, the features extracted at the region-level (f_i, r_i) are passed to a multiclass classifier, namely softmax in Equation (12). The application of $\sigma(f_i, r_i)$ yields a region-based classification $c(c_i, r_i)$ where each c_i belongs to a class of all abnormalities obtainable in the dataset.

$$\sigma_{cls}(\vec{r})_i = c[\max\left(\frac{e^{\vec{r}_i}}{\sum_{j=1}^{m} e^{\vec{r}_j}}\right), r_i] \quad (12)$$

H. Multiclass classifier for detection and localization of abnormalities

Region-level localization is achieved using the lower branch of the RCNN architecture. The chosen method is the one that detects the presence of abnormalities in the regions to determine what number of abnormalities (or abnormities) exist at the image-level through agglomeration of region-level abnormality. It should be noted that for this study, abnormality detection is not restricted to only malignant or only benign, rather, we narrow down the class of abnormalities to include: adenosis (A), fibroadenoma (F), phyllodes tumor (PT), and tubular adenona (TA) as benign; and carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC) and papillary carcinoma (PC) as malignant. We also note that since almost all r_i, as seen in Equation (13) will belong to the normal class. We ignore the processes such as r_i further from this level to reduce the computational cost, and focus on those abnormalities in the benign and malignant categories.

$$\sigma_{prob}(\vec{r})_i = \frac{e^{\vec{r}_i}}{\sum_{j=1}^{m} e^{\vec{r}_j}} \tag{13}$$

where r_i represents an arbitrary region in position i.

I. Region-to-label-probability mapper

The probability distribution of each region r_i extracted from image x is mapped to the class label as obtained by the upper branch of the architecture. To map a region to its class and localize the abnormality accordingly, we apply the map function in Equation (14). The function accepts the probability distribution of each r_i, obtain the highest probability value and then obtain the corresponding class label for the region r_i . Furthermore, all r_i in R are sorted in a manner as to compute the most suggestive overall class label for the image. Meanwhile, all r_i are agglomerated to recover the image while maintaining the local region-level localization obtained so far. We note that the localization of each r_i is categorically restricted to abnormalities in the categories of benign and malignant.

$$\partial_{localized}(\vec{r})_i = \partial(\sigma_{prob}(\vec{r})_i \rightarrow \sigma_{cls}(\vec{r})_i), \qquad x_{localized} = map(\partial_{localized}(\vec{r})_i, \vec{r}) \tag{14}$$

where $\partial_{localized}(\vec{r})_i$ represents region-level localization and $x_{localized}$ represents the agglomeration to the outcome of region-level to image-level agglomeration localization.

Algorithm 1: WCNN and RCNN image classification and localization

```
Result: classification and localization on image samples
Input: WCNN, RCNN, dataset, (w, h)
Output: imgloc, imgcls, pmap, confidence
traindx, evaldx, testdx = load(dataset, (75, 15, 10);
trainr, evalr, testr = extractor(traindx, evaldx, testdx);
wcnnmdl=train(traindx, evaldx, WCNN);
i=0;
while testdx = ∅ do
    occludx = Eq5(testdx[i], w, h)
    imgloc, imgcls, pmap ←predict(wcnnmdl, occludx)
    ++i
end
rcnnmdel=train(trainr, evalr, RCNN);
i=0;
while testr = ∅ do
    regiondx = Eq11(testr[i], w, h)
    imgcls2, pmap2 ← Eq12(rcnnmdl, regiondx)
    pmap2 ← Eq13(rcnnmdl, regiondx)
    for r ←rg in regiondx do
        imgloc2 ← Eq14(imgcls2, pmap2, r)
    end
```

```
    ++i
end
confidence, pmap ← Eq9(pmap, pmap2)
return imgloc, imgcls, confidence, pmap
```

In Algorithm 1, the summary of the procedure previously described is presented. The algorithm is expected to return the image localization of abnormalities at the image- and region-levels for comparison. To boost the confidence in the acceptability of the classification and localization tasks, the algorithm also compares the probability map of image samples that were obtained from the WCNN and RCNN, and the confidence value is the output. Also, the image classification results are returned to confirm what abnormality exists in the input samples. Meanwhile, the algorithm accepts as input the design of the WCNN and RCNN architectures, the dataset and a tuple representing the region extraction size. The computational analysis of the algorithm is computed, and the derived notation of $O(n)$ was achieved. The following subsection presents a detailed design of the two convolutional neural networks.

3.2 Design of CNN architectures

The architectural composition of both WCNN and the RCNN convolutional networks are presented in detail following the outcome of our recent study [53]. We build on the result of the study, which auto-designed the CNN network, which demonstrated, through experimentation, to be the optimal architecture to have learnt the classification problem in the histopathology datasets in [54] and BreakHis [55], [56]. The top-performing network is adopted for use as the WCNN, while the minimal top-performing network is adopted for the RCNN in this study. These designs of the networks for WCNN and RCNN are illustrated in Figures 2 and 3.

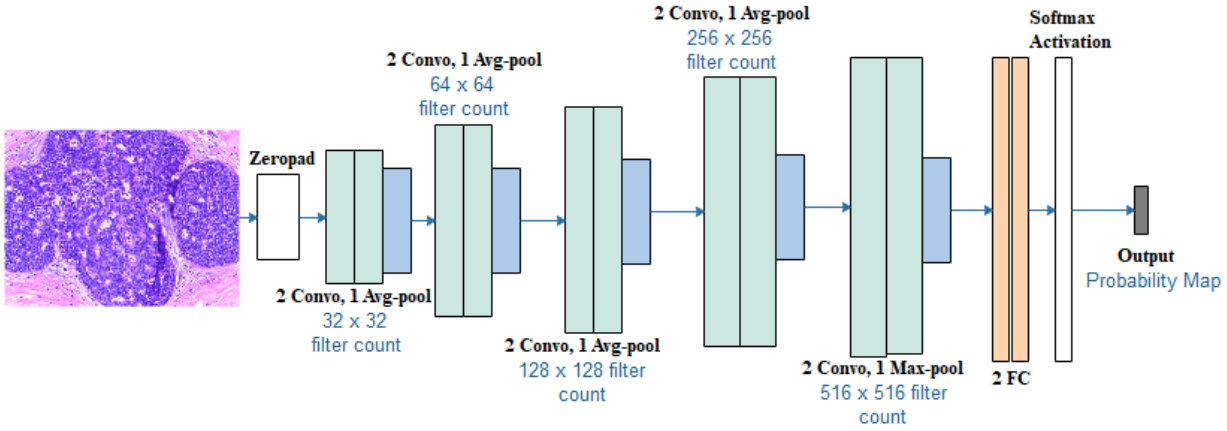


Figure 2: CNN architecture for the WCNN which was applied for the feature extraction and classification with focus on image-level localization

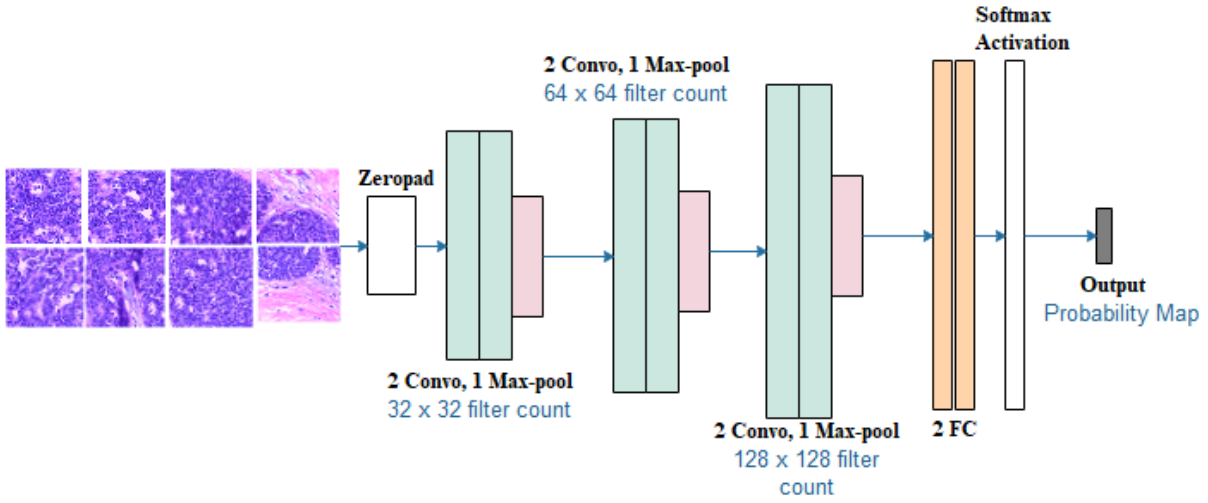


Figure 3: Convolutional neural network architecture for the RCNN which was applied for the feature extraction, detection, and classification of extracted regions with a focus on region-level localization

The architectural design of the WCNN consists of five blocks of convolutional-pooling layers. The first block comprises two convolutional layers with a 3×3 filter size and 32×32 filter count with a RELU activation function. These two convolutional layers were followed by the max-pooling layer. The second convolutional block has two convolutional layers using the same RELU function, followed by an average-pooling layer. Each of the convolutional layers uses a 64×64 filter count and a 3×3 filter size. In the third block of convolutional-pooling layers, the network is composed of two convolutional layers with a 128×128 filter count and 3×3 filter size with each layer, followed by the L1 regularization technique. This pattern is repeated for blocks four and five with 256×256 and 512×512 filter counts, respectively. Average-pooling layers follow the convolutional layers. The design of the feature extraction layers is followed by two fully connected layers. In the dense layer, a dropout layer was added with a drop rate of 0.45. The second neural network, RCNN, applied for the task of region-based detection and classification, consists of three blocks of convolution-pooling layers, with the first block having only two convolutional layers. The convolutional layers use the RELU activation function, a 32×32 filter count and a 3×3 filter size. The regularization technique was used in the convolution layer, and the layer is followed by an average-pooling layer with a filter size of 2×2. The second block consists of two convolutional layers that use the same RELU function and apply a 64×64 filter count with the same L1 regularization technique. Similarly, the third block follows the same pattern though, using 128×128 filter counts. Like the WCNN, the convolution layer is followed by a max-pooling layer. We experimented with these two networks and tweaked their configurations to achieve the best performance in both architectures. In the following section, we describe parameter and hyperparameter settings used in training both networks.

4. Experiments

The description of the implementation of the approach designed in the previous section and parameter configuration for experimentation of the architecture is presented in this section. In addition, a discussion on the datasets and related metadata on the datasets used for experimentation are presented in this section.

4.1 Implementation

The complete architecture of the design of the approach proposed in this study was implemented in Python, with the implementation of the two CNN architectures being achieved using Python 3.7.3 and all supporting libraries such as Tensorflow, Keras, and other dependent libraries. A personal computer with the following configuration was used for the implementation: Intel (R) Core i5-4200, CPU 1.70GHz, 2.40GHz; RAM of 8 GB; 64-bit Windows 10 OS. This same system with the indicated configuration was used for testing the trained model. Complete training and evaluation of the WCNN and RCNN architectures were achieved on Google's Colab environment with the following configurations: 2-core Intel(R) Xeon(R) CPU @ 2.30GHz, 13GB memory and 33GB hard drive; and GPU Tesla P100-PCIE-16GB. Trained models were stored for prediction or testing and used subsequently.

The chosen hyperparameters for training the WCNN use the learning rate of 0.005, though further experimentation was carried out using the values of 0.00001 and 0.0001. The Adam optimizer was used for the training while keeping parameter values for beta = 0.5, beta2 = 0.999, and epsilon = 1e-08. The network was experimented with using a batch size of 32. Meanwhile, the adopted network for WCNN was initially trained using a learning rate of 0.005, Adagrad optimizer with initial accumulator value = 0.1, epsilon = 1e-07. We experimented with both the initial configuration and the configuration described at the beginning of this paragraph. For the RCNN architecture, the adopted network uses the learning rate of 0.05, and RMSprop optimizer with rho = 0.9, momentum, epsilon = 1e-07, centered = False. However, we also experimented with RCNN using an Adam optimizer with a learning rate of 0.0001. the batch size of 32 was also used for the RCNN architecture.

4.2 Datasets

Two sets of datasets were combined to obtain a larger collection of samples for training the WCNN and RCNN architectures. The publicly available benchmark datasets, namely BACH [54] in Figure 4 and BreakHis [55], [56] in Figure 5 were sourced, preprocessed and applied for the experimentation phase. The image samples obtained from the BACH and BreakHis datasets were further resized to sizes 224 × 224 to allow for input to the WCNN architecture. The same preprocessed and resized samples were used for region extraction to train the RCNN architecture. About 400 samples from the BACH database are originally of 2048 × 1536 pixels. Over 7900 samples were obtained from the BreakHis dataset with an image size of 700X460 pixels. Combining these samples from the two sources provided over 8500 samples were used for the experimentation. The class labels reported with the BACH samples are normal, benign, malignant ***in situ***, and malignant invasive carcinoma. On the other hand, the BreakHis samples were reported with the following class labels: benign adenosis (A), benign fibroadenoma (F), benign phyllodes tumor (PT), benign tubular adenona (TA), malignant carcinoma (DC), malignant lobular carcinoma (LC), malignant mucinous carcinoma (MC) and malignant papillary carcinoma (PC). The following outlines the distribution of samples per class category: 100 samples with normal label, 2580 samples with benign label, 100 samples with *in situ* carcinoma label, and 100 samples with invasive carcinoma label, and 5429 samples with other malignant types.

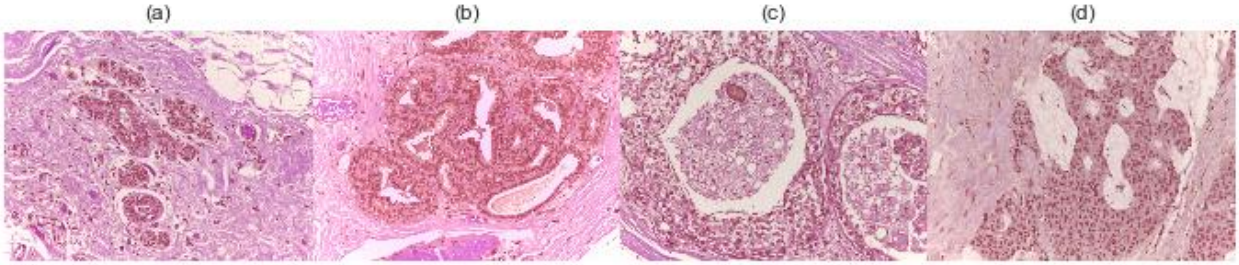


Figure 4: Sample images from the BACH datasets showing (a) normal (b) benign (c) malignant *in situ* carcinoma and (d) malignant invasive carcinoma [54]

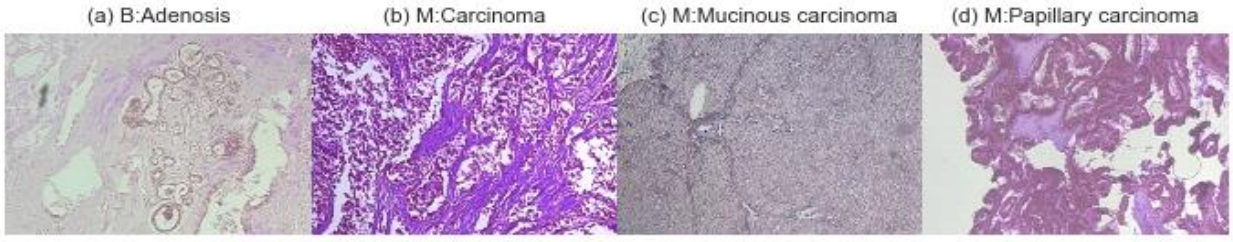


Figure 5: Samples from BreakHis datasets with 40X magnifications and containing Hematoxylin (dark purple) and Eosin (light pink) stain in the nuclei with and showing (a) benign adenosis, (b) malignant ductal carcinoma, (c) malignant mucinous carcinoma, and (d) malignant papillary carcinoma [55], [56].

The image samples were preprocessed using the image normalization operations based on Reinhard [57] and Macenku [58] methods. The magnification of BACH and BreakHis datasets as used for experimentation are datasets is 200X and 40X, respectively. We, however, preprocessed the images to allow for resizing and eliminating potential errors arising from stains on the raw inputs. We applied the basic operations of reduction of background noise and image enhancement. The outcome of applying these datasets to the implementation described earlier is discussed in the next section, which outlines the results and discussion on the findings of the study.

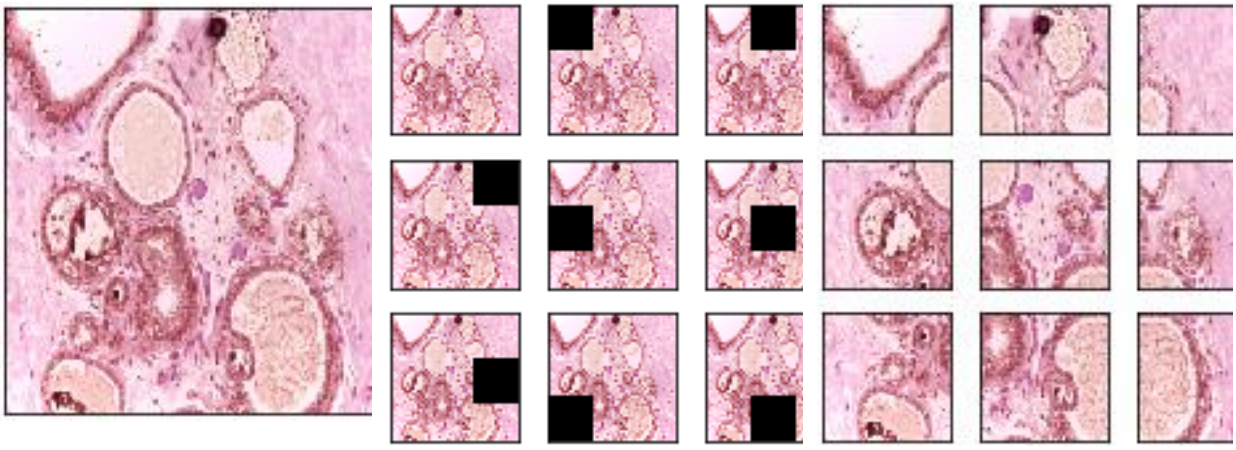
4.3 Image occlusion and region extraction

The implementation of occlusion of image samples for use in localization of samples using WCNN and the region extraction from RCNN are described in this section. We experimented with different tuple specifications for both occlusion and region extraction from an image sample. For instance, consider that we have an image sample say img and a tuple say A. Then when img and A are passed to the occlusion processor $\delta(img\ A)$ and the image extractor $\vartheta(img\ A)$ so that we obtain img with an occluded region or an extracted region from img as seen in Equation (14).

$$img_{region} = \begin{cases} \delta(img, A) & if\ wcnn \\ \vartheta(img, A) & if\ rcnn \end{cases} \tag{14}$$

where A represents the tuple (w, h) which describes the region of interest in img.

Interestingly, this notation allowed us to vary the region of interest (ROIs) size. For instance, the tuples $(40{,}40)$ and $(80{,}80)$ could be used to perform different experiments on WCNN and RCNN to investigate the effect of focusing the models on different regions. Meanwhile, the outcome of applying both processor $\delta(img\ A)$ and $\vartheta(img\ A)$ can be seen in Figures 6 (a-c).



(a) (b) (c)

Figure 6: An illustration of occluding and extracting regions in an image sample where (a) is the original image sample, (b) is the outcome of $\delta(img)$ on img (c) is the outcome of $\vartheta(img)$ on img, using a tuple of A = (80,80).

A large number of image samples were extracted from both the BACH and BreakHis data sources to ensure that sufficient data is made available for training the models, thereby eliminating the possibility of overfitting. In the following section, we report the sensitivity of both WCNN and RCNN to the variation of A in determining classification accuracy and localization.

5. Result and Discussion

The performance of the WCNN and RCNN architectures are investigated and reported based on rigorous experiments carried out. WCNN and RCNN were trained for 1000 epochs to allow full training while the trained models were saved for use. The datasets were partitioned into three sets, namely the training, validation and testing datasets using the approach of 0.70, 0.25, and 0.05, respectively. This data partitioning allows for sufficient data during training and validation stages to overcome the overfitting of the model and further support the model's generalisation. The localization result at the image-level and region –level is also reported, and the corresponding probability map obtained from the two networks is presented and discussed. We evaluated the performance of WCNN and RCNN, while under training, using the metrics of accuracy and loss values obtained for the number of epochs. We compared the outcome of different methods resulting from the variation of parameters and hyperparameters in this study.

The WCNN model was first trained and evaluated based on the datasets discussed in the last section. The fully trained model was then used for prediction and localization. Interestingly, we found that the WCNN architecture has elegantly learnt the classification problem, as can be seen in Figure 7(a-b). The accuracy of the training and evaluation datasets on the model showed a progressive rise in value to an impressive range above 90%. Similarly, the loss values obtained

for both training and evaluation demonstrated unbiased learning with the model since it could generalize well. The loss values dropped significantly for both the training and evaluation curve to signify the classification problem as obtained in samples in the datasets learnt by the WCNN model. The implication of this outcome suggests that discriminant features pointing to each abnormality represented by the samples have been correctly patterned by the model. This, therefore, allows for a better prediction when new samples are supplied. Much more to this is the ability of the model to localize the abnormalities in the sample, as indicated by the learning curve. This confirms that the classification problem shadows the localization problem, and the latter can be efficiently handled when the former is addressed. Considering the outcome of the classification of the model, which is significant, we proceed to investigate the performance of the WCNN on prediction and, most interestingly, localization.

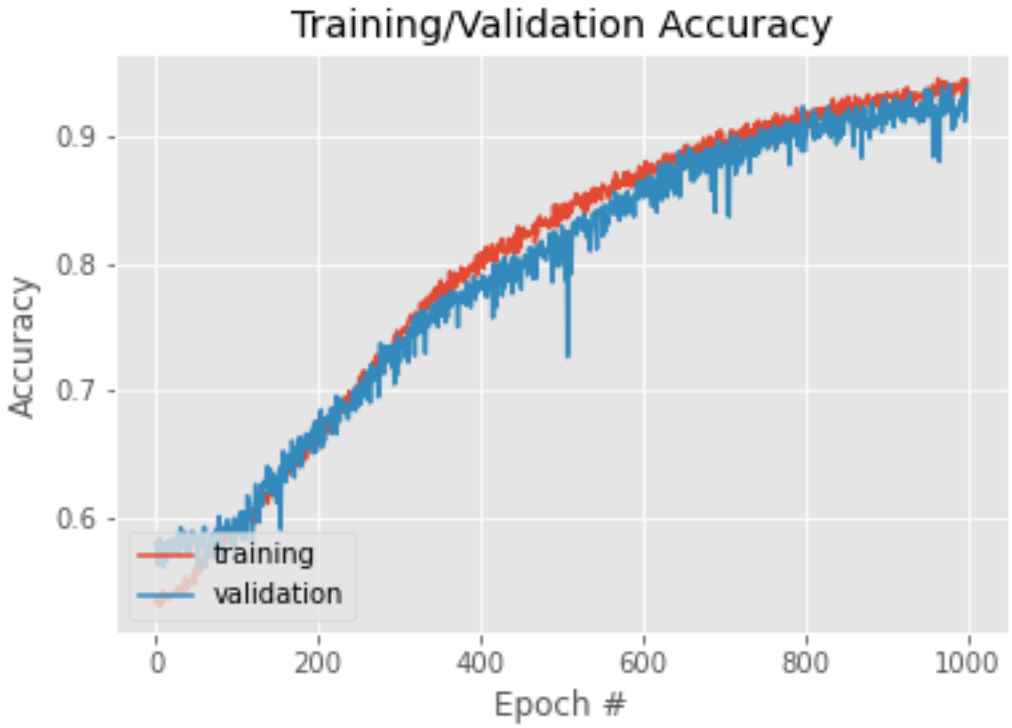


Figure 7(a): Plot of the accuracy for the training of the WCNN architecture. Training of WCNN using the Adam optimization configuration and learning rate of 0.00001

Figure 7(b): Plot of the loss values for the training of the WCNN architecture. Training of WCNN using the Adam optimization configuration and learning rate of 0.00001

The RCNN first proved difficult to train effectively but was stabilised through skilful parameter tuning. As observed in the performance of the WCNN, we found that the RCNN also has yielded a good learning curve. The pattern of the curve for both accuracy and loss values, as seen in Figures 8 (a-b), confirmed that results for both training and evaluation are relevant. Recall that, unlike the WCNN that is fed with a complete image sample, RCNN is attenuated to accept regions of the sample as input. The learning of both WCNN and RCNN is necessary to achieve a good localization of both the image and region levels. Although it can be considered the classification problem is now addressed in a relevant manner, both from image to region levels, using WCNN and RCNN, respectively, since it is important for the different employees, and this localization remains a matter of study.

Figure 8(a): Plot of the accuracy for the training of the RCNN architecture. Training of RCNN using the Adam optimization configuration and learning rate of 0.00001

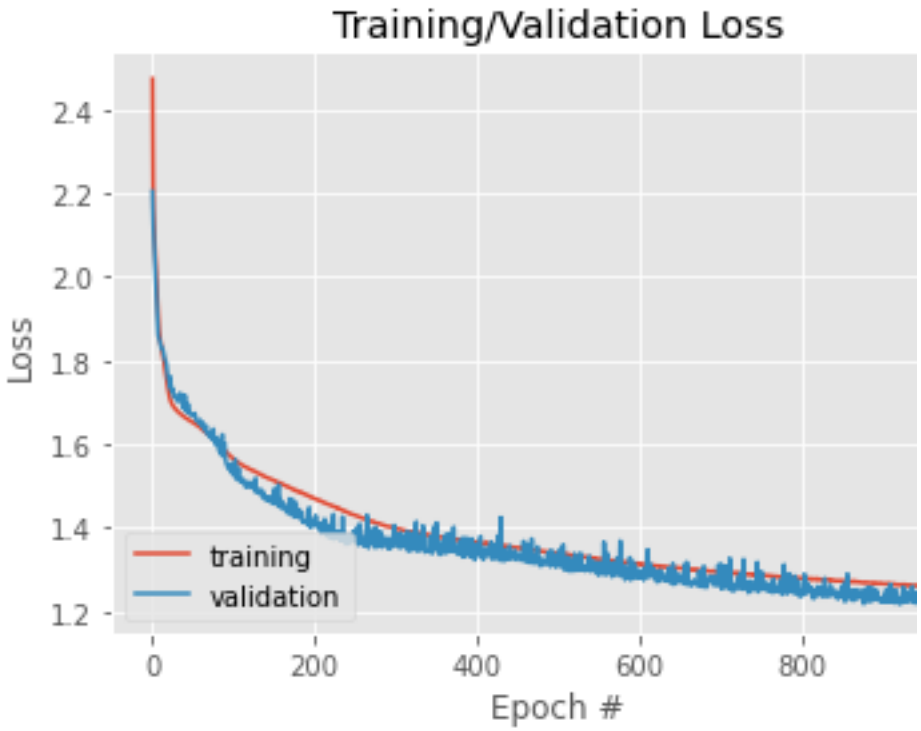


Figure 8(b): Plot of the loss values for the training of the RCNN architecture. Training of RCNN using the Adam optimization configuration and learning rate of 0.00001

In Tables 1 and 2, the outcome of evaluating the performances of WCNN and RCNN with respect to metrics is listed.

This evaluation became necessary, considering there is the need to assess multi-label classification problems using different performance measures to authenticate the model. Hence, test samples were applied to the fully trained models, being evaluated using the classification accuracy, specificity, sensitivity, precision, Cohen's Kappa (CK), Matthew's coefficient (MCC), Jaccard score (JS), F1-score, recall, and prediction time. Traditionally, most studies consider classification accuracy for assessing the validity of their model. Following this convention, we evaluated both WCNN and RCNN and found their classification accuracy to have yielded 97.08% and 52.57%, respectively, as listed in Tables 1 and 2, respectively. These values demonstrate the high performance of the models in understanding and learning the classification problem of abnormalities in histopathology images. Evaluation for the precision of WCNN in Table 1 and RCNN in Table 2 returned 97.08% and 52.57%, respectively, to show that the proportion of positive samples identified by both models was correct to that degree. Similarly, Tables 1 and 2 listed the number of actual positive cases that WCNN and RCNN were able to identify correctly. We evaluated the metric on recall in both Tables 1 and 2, which returned 97.08% and 52.57%, respectively. A striking balance between the performance of WCNN and RCNN was achieved using the F1 score, which showed 97.077% and 52.57%, respectively, which reveals that the performance is based on precision and recall.

Table 1: Performances comparison of the prediction in the classification of abnormalities on fully trained WCNN on samples at their respective image-level and region-level on some selected metrics

CNN Model	Classification Accuracy	Specificity	Sensitivity	F1	Recall	Precision	CK	MCC	JS	Predict time
WCNN	0.9708	1.0	1.0	0.9708	0.9708	0.9708	0.9622	0.9625	0.9432	162.22

Table 2: Performances comparison of the prediction in the classification of abnormalities on fully trained RCNN on samples at their respective image-level and region-level on some selected metrics

CNN Model	Classification Accuracy	Specificity	Sensitivity	F1	Recall	Precision	CK	MCC	JS	Predict time
RCNN	0.5257	0.6545	0.4607	0.5257	0.5257	0.5257	0.3172	0.3383	0.3566	41.57

During experimentation, samples were distributed equally based on class labels into training, evaluation, and test datasets. This, however, cannot eliminate class imbalance which exists in most publicly accessibly datasets and often biases classification accuracy. As a result, more metrics were evaluated on WCNN and RCNN for further performance measures. Cohen's Kappa (CK) and Matthew's correlation coefficient (MCC) metrics have shown to be robust in handling imbalanced class problems. The values of 0.9622 and 0.3172 were obtained for WCNN and RCNN, respectively, for CK showed a wider view of the performance of the classification models compared to accuracy. Similarly, the values of 0.7426 and 0.3383 obtained for WCNN and RCNN, respectively, for MCC demonstrate the wellness and quality of the models in prediction performance on new samples. Jaccard similarity or score (JS) describes the magnitude of intersection between predicted and true labels divided by the size of their union. This metric implies that it can show how relevant the model's prediction on the test dataset is. In Table 1, the value of 0.9432 was obtained for WCNN; in Table 2, the value of 0.3566 was obtained for RCNN. The values demonstrate how good the model's prediction is on the test dataset. The value of 0.94 showed a very high degree of good performance of the WCNN model during the prediction task, while the value of 0.3566 demonstrates an average performance by the RCNN in prediction with the trained model. Meanwhile, the sensitivity of the models to detecting true positive samples and their specificity in determining true negative samples were evaluated. Interestingly, we found both WCNN and RCNN to have proven effective in isolating samples with malignant from those with normal or benign. This is revealed from the results of specificity and sensitivity, which gave 1.0 and 1.0 for WCNN and 0.6545 and 0.4607 for RCNN. This study looked at the time taken to predict both image-level and region-level, discovering that WCNN took approximately 162.22x60 seconds while RCNN took approximately 41.57x60 to predict the same samples at the region-level.

The performance of the classification models was further investigated to understand what abnormalities were adequately learnt by the deep learning models. Considering the difficulty associated with the classification problem using stained histopathology images, it became necessary to know how the WCNN and RCNN were able to navigate solutions to this problem on a class-level classification task. In Table 3, precision, recall and F1-score were adapted

to this task, and the obtained results are listed. Twelve (12) classes were applied for the experimentation, six (6) different types of malignant cases, five (5) types of benign cases, and one (1) normal category. Computed result for the precision, recall and F1-score shows that normal samples, on row 1, were correctly identified; benign samples, on rows 2-6, were more differentiated compared with normal samples; results for malignant samples, as seen on rows 7-12, were more discriminant considering that more samples were used. A number of samples used for prediction by WCNN at the class-level are seen in the Support column correspondingly with each class label.

Table 3: Class-based performance evaluation for precision, recall and F1-score using WCNN

Class label	Precision	Recall	F1-score	Support
Normal (N)	0.99	0.99	0.99	72
Benign (B	1.00	0.91	0.95	75
Adenosis (A)	1.00	1.00	1.00	342
Fibroadenoma (F)	1.00	1.00	1.00	845
Phyllodes tumor (PT)	1.00	1.00	1.00	351
Tubular adenona (TA)	1.00	1.00	1.00	472
In situ carcinoma (IS)	0.95	0.99	0.97	75
Invasive carcinoma (IV)	0.95	0.99	0.97	75
Carcinoma (DC)	0.95	0.99	0.97	2313
Lobular carcinoma (LC)	0.88	0.72	0.79	319
Mucinous carcinoma (MC)	0.99	0.92	0.96	378
Papillary carcinoma (PC)	1.00	0.97	0.98	261
Macro average	**0.98**	**0.96**	**0.96**	**5578**
Weighted average	**0.97**	**0.97**	**0.97**	**5578**

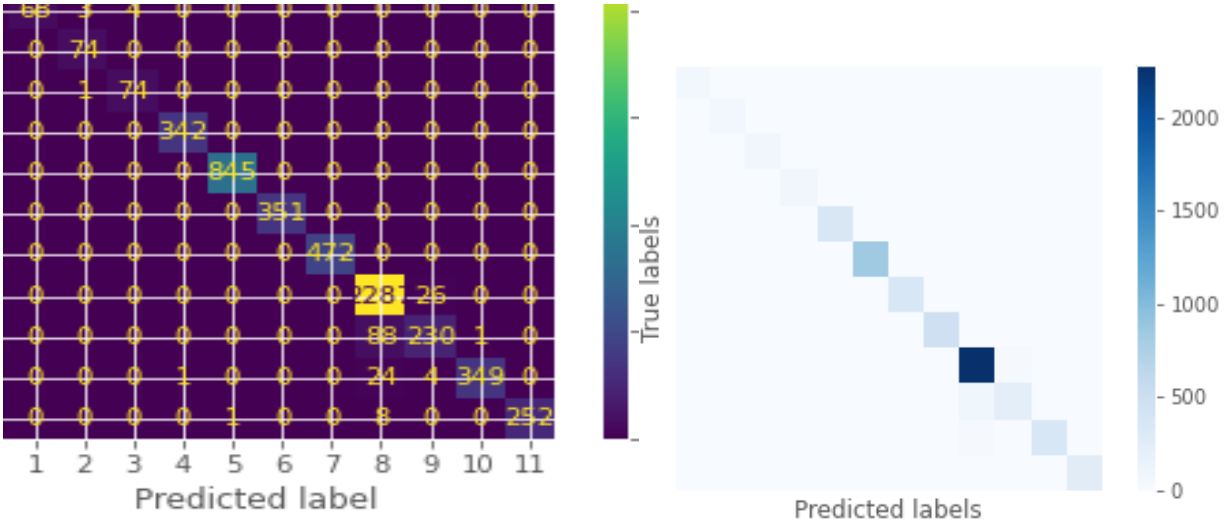
Similar to class-level computation carried out for WCNN, the same was evaluated for RCNN as seen in Table 4. Since the number of samples used for prediction with RCNN corresponds to region extraction from those used in WCNN, we expect the number of samples reported in the Support column to be higher. The performance evaluation for precision, recall and F1-score showed that the trained model significantly learnt discriminant features at the region level to help it classify samples accordingly. For malignant samples, seen on rows 7-12, the values obtained for precision, recall and F1-score were supportive of a good classification performance since most values rose above average. Those benign samples, listed on rows 2-6, were also significantly classified at the region level with values for precision, recall and F1-score clocking above 0.70. Meanwhile, the classification of normal samples is sustained as reported in the case of WCNN.

Table 4: Class-based performance evaluation for precision, recall and F1-score using RCNN

Class label	Precision	Recall	F1-score	Support
Normal (N)	0.26	0.16	0.20	684
Benign (B	0.31	0.14	0.19	684
Adenosis (A)	0.43	0.15	0.22	3240
Fibroadenoma (F)	0.33	0.48	0.39	8100
Phyllodes tumor (PT)	0.12	0.00	0.00	3348
Tubular adenona (TA)	0.44	0.23	0.30	4464
In situ carcinoma (IS)	0.29	0.20	0.24	648
Invasive carcinoma (IV)	0.49	0.71	0.58	720
Carcinoma (DC)	0.61	0.93	0.74	22248
Lobular carcinoma (LC)	0.55	0.35	0.43	3096
Mucinous carcinoma (MC)	0.23	0.01	0.01	3636
Papillary carcinoma (PC)	0.00	0.00	0.00	2520
Macro average	**0.34**	**0.28**	**0.28**	**53388**
Weighted average	**0.44**	**053**	**045**	**53388**

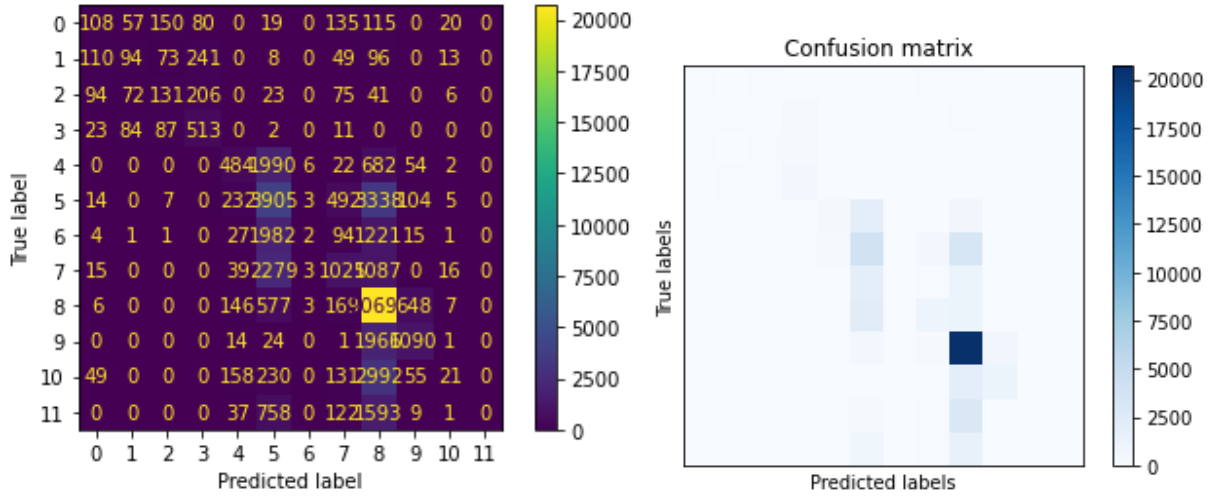
Visualization of the performance of the WCNN and RCNN models in prediction-based on the twelve classes are presented using confusion matrix. Here, we utilized two variants of confusion matrix to allow for appreciation of the values obtained for each class distribution. In Figure 9 (a-b) the confusion matrix demonstrating the performance of

WCNN on the N, B, IS, IV, A, F, PT, TA, DC, LC, MC, and PC class labels obtained in samples, are presented. The values 71, 68, 74, 74, 342, 845, 351, 472, 2287, 230, 349, and 252 were obtained for N, B, IS, IV, A, F, PT, TA, DC, LC, MC, and PC, respectively, showed the number of true positives. The implication of these values as it relates to the classification problem interprets that WCNN correctly classified a significant number of samples as distributed in their respective classes. This is further confirmed by the values obtained for false positives (class-label rows) and false negatives (class label columns) which often returned zero for all classes as can be seen in Figure 9(a) and outlined in Figure 9(b) through dimly colored cells and brightly colored diagonal cells. The values obtained for the true negative, represented by contiguous blocks of cells which are adjacent to the corresponding class-label in the diagonal cell, are also closely correlated with what is obtainable in the dataset. In the same vein, we investigated the performance of RCNN, since it relates to the specification of the confusion matrix and found that the values 108, 94, 131, 513, 4841, 905, 982, 1025, 69, 5090, 21, and 0 were obtained for N, B, IS, IV, A, F, PT, TA, DC, LC, MC, and PC, respectively. This result can be confirmed in Figures 10 (a-b), where Figure 10(a) shows these images in the diagonal cells and Figure 10(b) confirms the values through the variation of the colors (bright) of the diagonal cells. The outcome seen in the confusion matrix of both WCNN and RCNN as it relates to their performance in addressing the classification problem of breast cancer in histopathology images supports the viability of the models in addressing the localization problem.



(a) (b)

Figure 9: Confusion matrix showing the classification performance of the fully trained model of WCNN when applied to test datasets for prediction. (a) shows the values obtained for true positives and (b) illustrate the values using cell-color variation for all class-labels using WCNN model



(a) (b)

Figure 10: Confusion matrix showing the classification performance of the fully trained model of RCNN when applied to test datasets for prediction. (a) shows the values obtained for true positives and (b) illustrate the values using cell-color variation for all class-labels using the RCNN model

Receiver operating characteristic (ROC) represents a graphical approach for illustrating the classification accuracy of a model so that all classification thresholds are considered. Therefore, we subject the evaluation of WCNN and RCNN to the use of ROC to also observe how these models can effectively address the classification problem leading to the localization problem. This graphing is obtained through the computation of True Positive Rate (TPR) and False Positive Rate (FPR) values so that the summary of the performances of the models is seen over all possible thresholds. Another interpretation of the use of TPR and FPR would put sensitivity and (1−specificity) on the *y*-axis and *x*-axis, respectively. Figures 11(a-b) graph the results returned by WCNN and RCNN for the ROC plot so that each abnormality or class-label is accounted for. To analyze the curves, it is traditionally expected to have the diagonal line from (0,0) to (1,1) corresponding to a ROC curve of random chance, and a perfect ROC curve should run through the point (0,0) to (0,1) and (0,1) to (1,1). However, since all classification models are expected to run below this perfect ROC curve, we present the analysis of our curve using this gold standard. Figure 11(a) shows the distribution of AUC values for all class-labels and the ROC curves corresponding to each label. This figure shows that the overall performance of WCNN as seen in the ROC is nearly perfect, as the general classification curve (dotted-black-coloured) runs close to the gold standard. At the class level, ROC curves for the abnormalities PT, TA, DC, LC, MC, and PC are impressive since the curves rose above the random chance curve. Similarly, we observed that distinguishing malignant from benign is a problem well addressed by WCNN as the ROC curves for B, IS, IV, A, and F were distinctly drawn towering above the random chance curve as well. The performance of RCNN has also been observed here to note how the classification problem is being successfully addressed at the general and class-based levels. Like what is reported for WCNN, RCNN showed that the general performance classification performance is significant, as seen in the dotted-black-coloured ROC curve in Figure 11(b). It is also interesting to note that the region-level abnormality characterization was good in terms of classification performance.

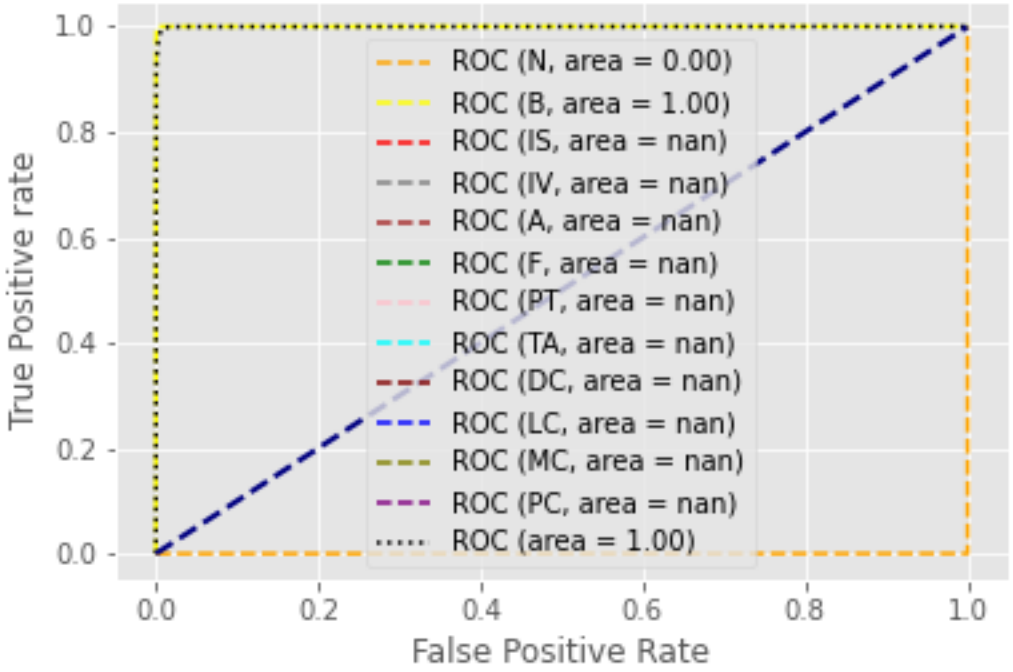


Figure 11(a): ROC curve for WCNN demonstrating the AUC value and curve obtained on an abnormality-based computation for all classes reported in the datasets

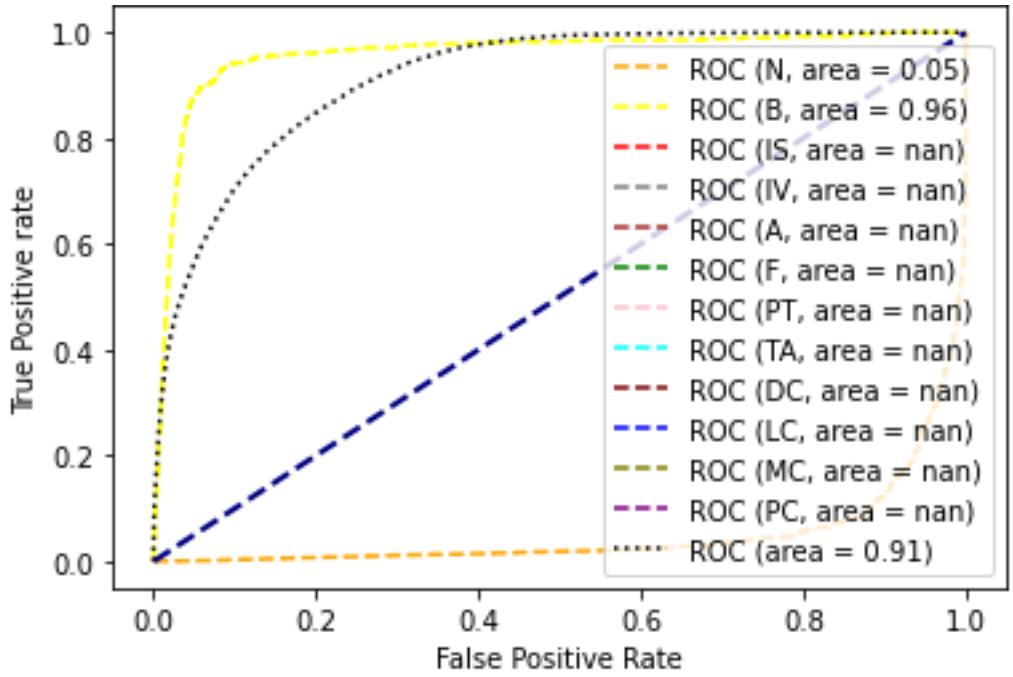


Figure 11(b): ROC curve for RCNN demonstrating the AUC value and curve obtained on an abnormality-based computation for all classes reported in the datasets

In addition to the evaluation of the models based on ROC curves, we observe the reported AUC values for the general classification problem and the class-based classification, as seen in Figures 11(a-b). The values correspond to the relevance of the entire two-dimensional area under the entire ROC curve. So, for instance, the AUC value is 1.00, which was reported for the general classification task in Figure 11(a), estimating the entire two-dimension area underneath the ROC curve represented by a dotted-black-coloured line. A good value for AUC is expected to range

between 0.5-1.0, with the upper bound corresponding to a perfect solution while the lower bound is the random chance solution. Also, the AUC reported for RCNN, which follows a similar performance, showed a result of 0.91. These all imply that classification problem is effectively addressed at both image- and region-levels.

Now that the CNN models have successfully learned the classification problem, which led to the identification of features suggesting abnormalities in histopathology images, we proceed to examine the performance of the models in achieving localization of the detected abnormalities. This represents a combinatorial problem that requires both classification and localization problems to be solved, so it could be able to annotate the samples correctly. To present the achievement of this task, we represent the outcome of localization leading annotation of samples using heatmaps, boundary boxes and inverted masked samples. The aim is to allow for the presentation of the localized abnormalities in an elegant and reinforced manner.

Recalling the localization strategy adopted in WCNN, it follows the occlusion of segments of samples to leverage the sensitivity of the softmax function used as a classifier. Hence, the visualization of the localization task performed by WCNN as described here follows the approach of occlusion. We focus on localization abnormalities in malignant samples, with little emphasis on benign samples, since the aim is to design deep learning models' similitude to human physicians with the improved performance desired here. Figures 12 (a-c) are presented in the localization of some selected samples. The bounding box localization of the abnormalities is represented in Figures 13(a-c). In Figures 14(a-c), a heatmap is applied to further analyze and present the localization outcome from the WCNN model. Furthermore, an investigation into the performance of the model with respect to localization accuracy and some other related metrics are listed in Figure 15, where each radar plot demonstrates a comparison of the distribution of values per abnormality for metrics such as accuracy, recall, precision, sensitivity, specificity, F1-score, Cohen Kappa, Mathew's coefficient, and Jaccard index. We observed that the localization accuracy for all abnormalities investigated through localization showed an impressive performance. A similar trend is reported for other metrics across the selected abnormalities.

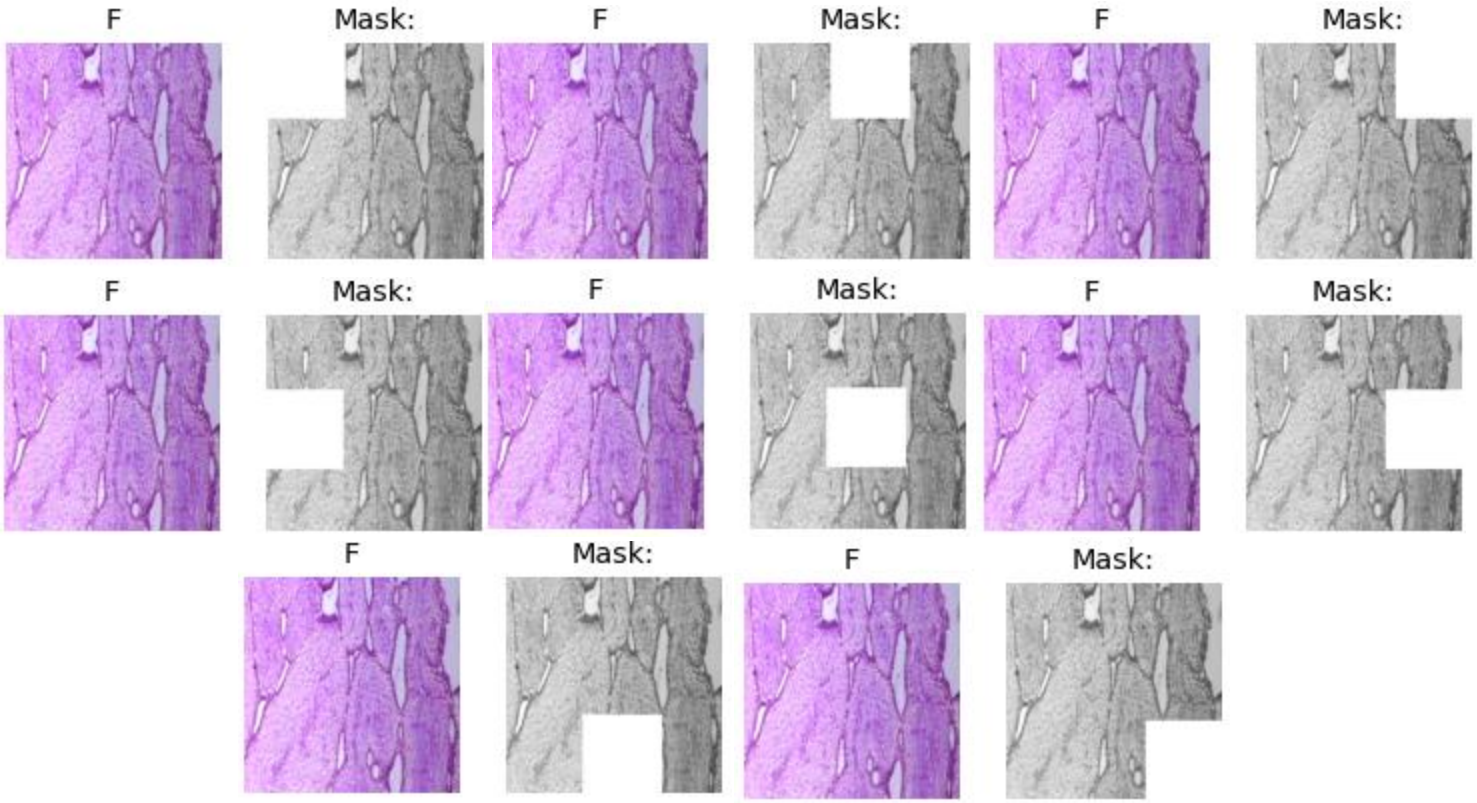
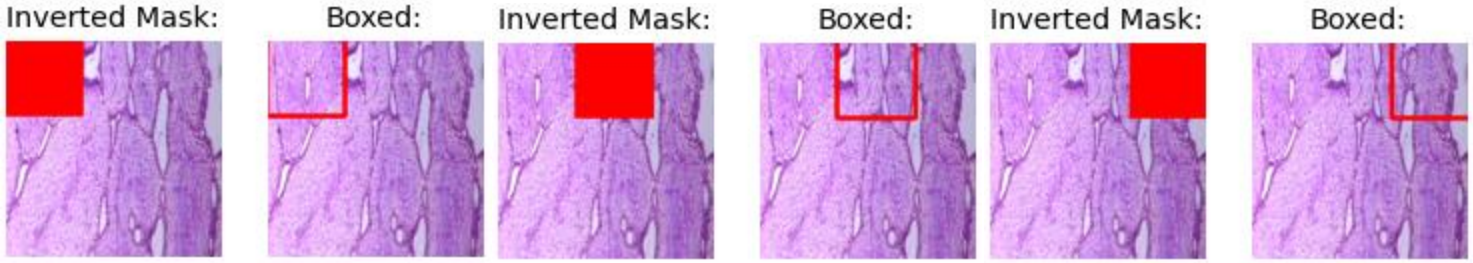


Figure 12: The localization using an inverted mask for fibroadenoma (F) shows all occluded regions that returned negative with the classifier and the missing region that returned positive with the classifier.



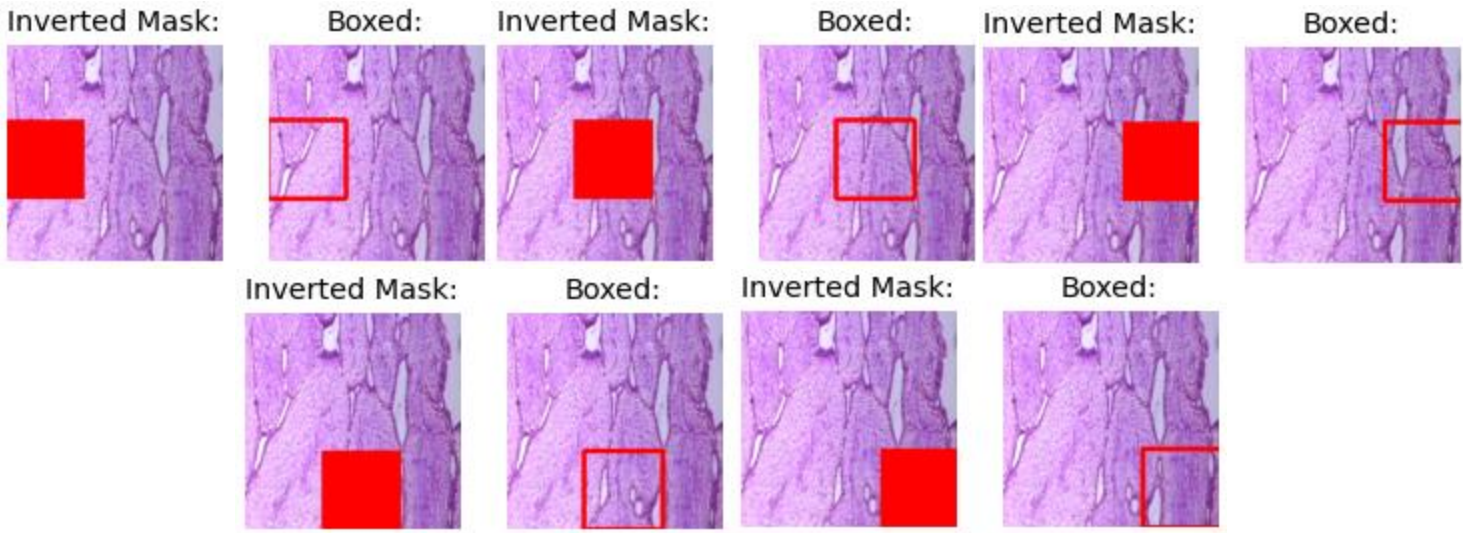
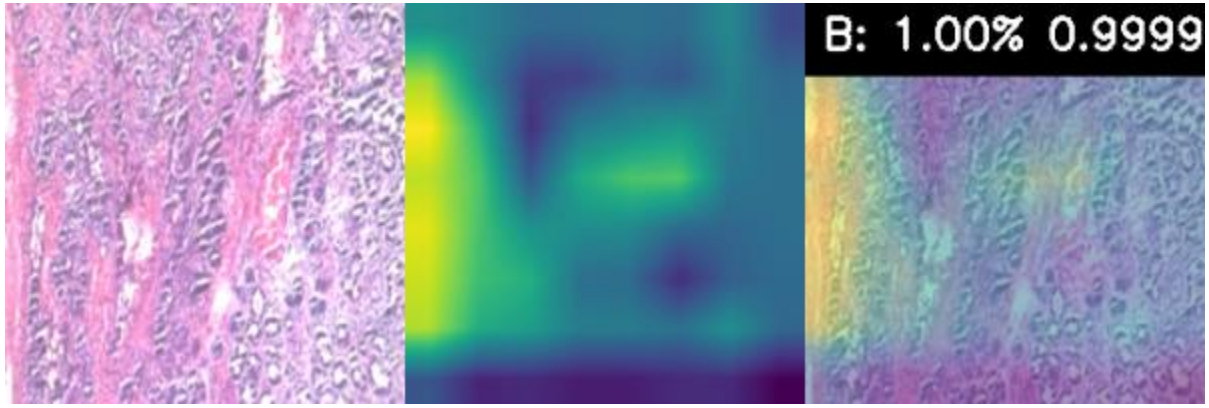
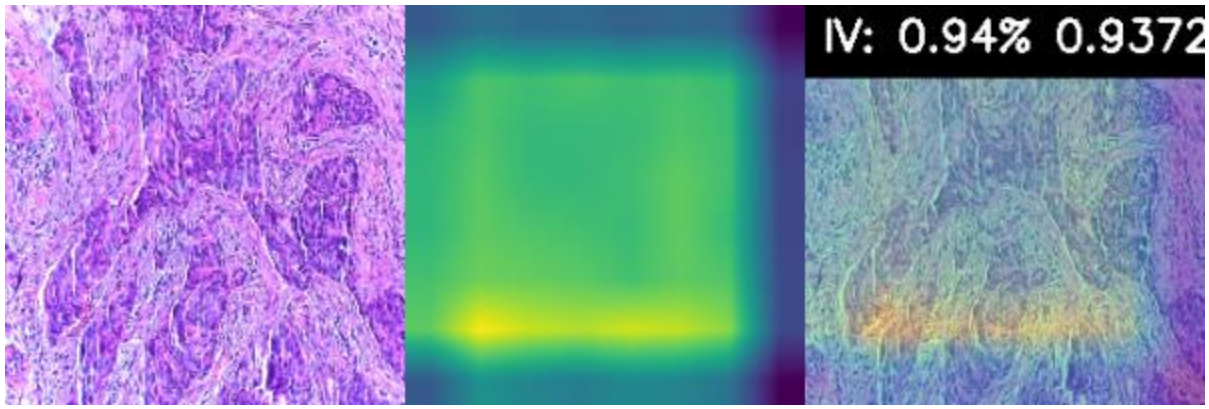


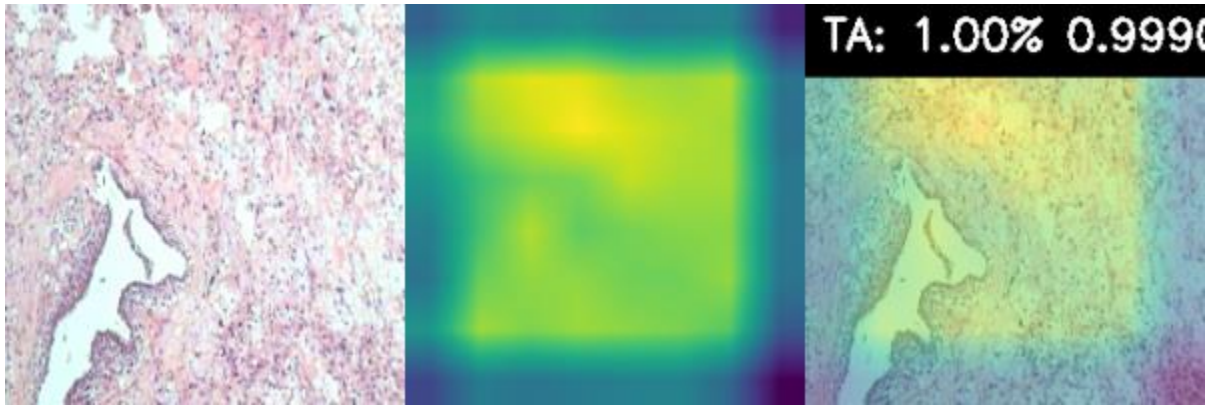
Figure 13: The localization using bounding box for mask for fibroadenoma (F) showing all occluded regions which returned negative with the classifier, and the missing region showing positive with the classifier



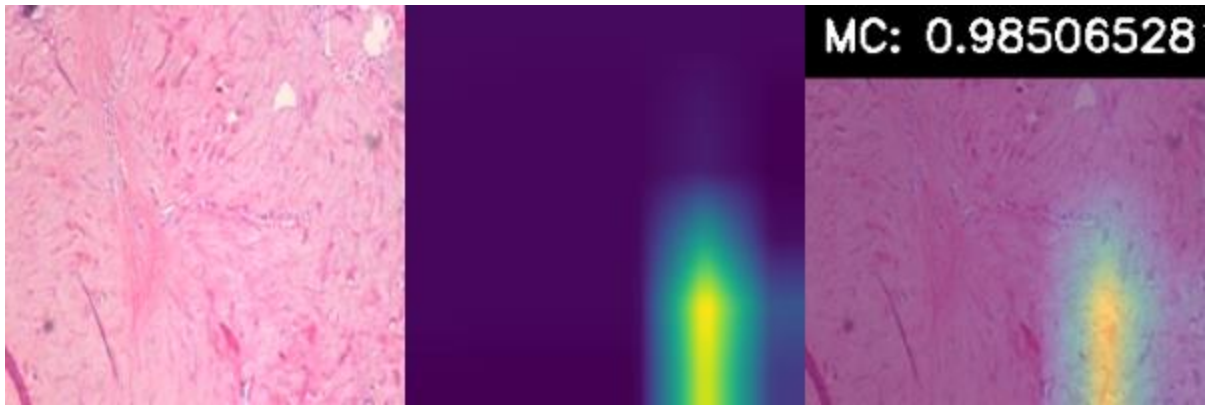
(a)



(b)

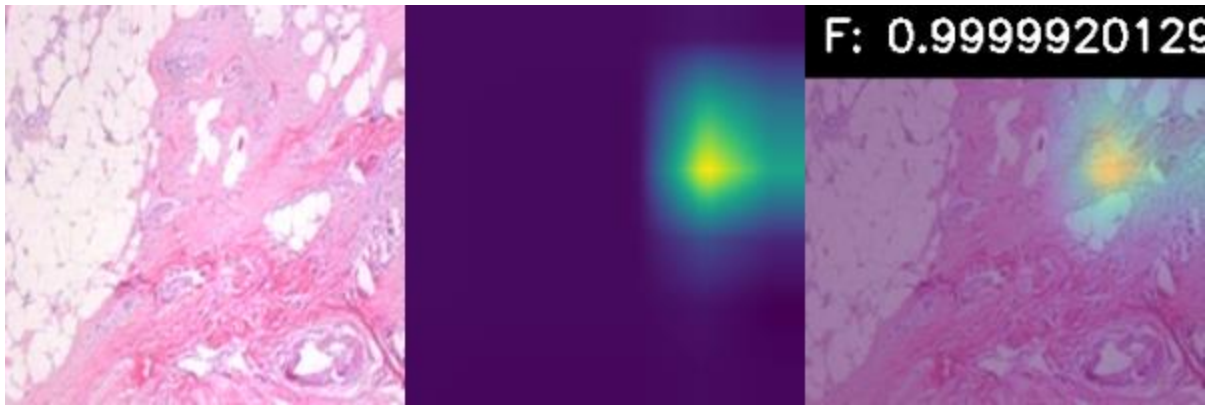


€



(d)

(e)



(f)

(g)

Figure 14: The localization using heatmaps for (a) benign [B] (b) malignant *invasive carcinoma* [IV] and (c) tubular adenona [TA], (d) mucinous carcinoma as malignant (MC), (e) phyllodes tumor as benign (PT), (f) fibroadenoma as benign (F), (g) in situ carcinoma (IS). The localized area is highlighted around the yellow-cloudy coverage of the histopathology images.

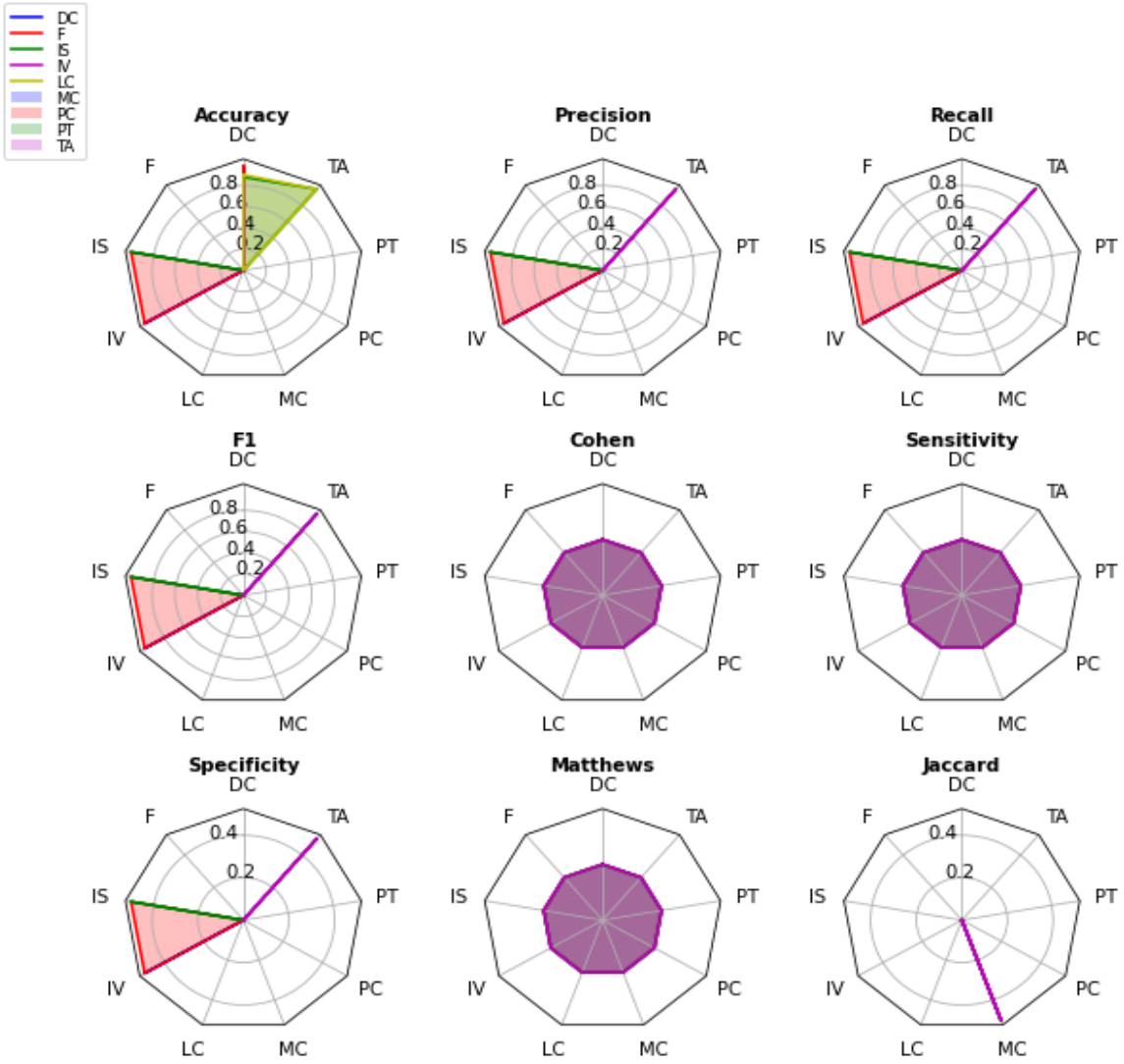


Figure 15: A radar plot for WCNN showing the distribution of performance of WCNN in localization tasks based on accuracy, recall, precision, F1-score, sensitivity, specificity, Jaccard index, Cohen Kappa, and Matthew's coefficient.

We summarize the findings of this study through a comparative analysis of the obtained performance for both WCNN and RCNN, with related studies which have used similar approaches or deep learning datasets to address the same problem. The outcome of the comparative analysis is presented in Table 5. This proposed approach presented the following average values of AUC: 1.00, accuracy: 0.9708, specificity: 1.00, sensitivity: 1.00, recall: 0.9708, precision: 0.9622, F-score: 0.9708, CK: 0.9622, MCC: 0.9625, and JS: 0.9432 were obtained. These values are then further compared with those obtained from other related works. For instance, the work of Nejad et al. [59] reported a detection rate of 0.775, whereas this study can achieve 0.93 and 0.91 for classification accuracy and localization accuracy, respectively. Compared with the work of Araújo et al. [60], which demonstrated the values of 0.778 and 0.956 for accuracy and sensitivity, this study reported 0.9708 and 1.00 and 0.94, 0.50 for classification accuracy, specificity, and localization accuracy and specificity, respectively. The works of Han et al. [61], Zhu et al. [62], and Xie et al. [63] returned classification accuracy of 0.932, 0.875, and 0.968, respectively, as compared with the 0.9708 and 0.94 obtained in this study for classification and localization. Furthermore, the values for precision, recall, F-score and AUC reported by the works of Saha et al. [64] and Hägele et al. [65] trail what was obtained in this study, where the values of 0.9622, 0.9708, 0.9708, and 1.00 were reported for the same metrics accordingly.

Table 5: Comparison of the performance of combined WCNN and RCNN models for classification and localization tasks as compared with other similar state-of-the-art CNN models addressing the same problems using histopathology images with breast cancer abnormalities.

Authors and References	Methods	Performance	Dataset
Nejad et al **[59]**	CNN + Data augmentation	Detection rate 77.5%	BreakHis database
Han et al **[61]**,	Structured Deep Learning Model+ Data augmentation	93.2% accuracy	BreakHis database
Zhu et al **[62]**	Squeeze-Excitation-Pruning (SEP)+CNN	Accuracy of 87.5%	BreaKHis and BACH dataset
Xie et al **[63]**,	Inception_V3 and Inception_ResNet_V2	Accuracy 96.84%	BreaKHis
This Study	WCNN and RCNN	AUC: 1.00 Accuracy: 0.9708 Specificity: 1.00 Sensitivity: 1.00 Recall: 0.9708 Precision: 0.9622 F-score: 0.9708 CK: 0.9622 MCC: 0.9625 JS: 0.9432	BreakHis and BACH databases

This performance obtained for the approach in this study confirms that deep learning models are sufficient and capable of addressing the combined challenge of classification and localization of abnormalities in medical images. This further indicates the need for a careful selection of CNN models suitable for the medical images, which should be analyzed to achieve the best performance. The study's outcome also indicates that achieving an acceptable localization of abnormalities in digital histopathology images must precede a successful classification of the same abnormalities in the image datasets. Whereas most studies using deep learning methods on digital medical images for detecting the presence of breast cancer often scope their work around classification, the result of this study is a compelling approach to motivate improved localization in both digital mammography and histopathology images with abnormalities leading to breast cancer. This is reinforced by the fact that radiologists leverage their skill to localize abnormalities, so their acceptance can get a boost, or otherwise, their conclusion would remain a guess. The proposed method and achieved results demonstrate the advancement of research in the use of histopathology images for the classification and localization of abnormalities in breast cancer. Therefore, this does not limit the applicability of the method to the problem represented by the selected dataset but allows for the generalization of the method to other datasets for different medical images.

This study has been demonstrated to have overcome some major drawbacks associated with using deep learning models in image analysis in the medical domain. The first is the challenge of overfitting the model, which was addressed by applying sufficient image samples for the model's training to achieve generalization. Moreover, the images were first preprocessed, and all traces of stain seen in the image samples were eliminated to ensure that the annotations and discriminant features were not rendered blurred during the

feature extraction process. Secondly, the core objective of this study is to achieve localization of abnormalities in addition to the classification task. From the results obtained through a rigorous experimental stage, we have reported that the proposed model achieves impressive classification and localization accuracy performance. We showed that the combinatorial approach adopted in this study could boost confidence in accepting the classification result. Thirdly, to overcome the challenge of a wrongly annotated dataset, this study utilized two major histopathology benchmarked datasets which have been widely used and reported in most experimental work reported in the literature. Fourthly, the dual-branch deep learning model, which uses both image-level and region-level inputs, enables the model to locate abnormalities with a small-pixel coverage adequately. Fifthly, we noted that the outcome of the localization process, as shown by the result obtained in this section, the model adequately isolated the region where the abnormality exists. These, therefore, showed that the result of this study has been able to address the drawbacks existing in literature which were previously outlined in the introduction section.

6. Conclusion

This study is focused on solving the combinatorial problem of classification and localization using histopathological images. Using a dual-branch CNN framework, we proposed the use of image- and region-level models, which are represented by WCNN and RCNN, respectively, so that the combined efforts of the two models are used to address the problem. WCNN leverages the sensitivity of the classifier to occlude samples in detecting the location where features suggesting the presence of abnormality exist. On the other hand, RCNN applied a unique probability function to map the regions of specific samples to agglomerate their classification output for decision at the image level. The dual-branch deep learning model was applied to well-preprocessed samples sourced from BreakHis and BACH datasets for training, evaluation, and testing. The trained models were exhaustively evaluated using different measures, which include classification accuracy and the progress of loss values obtained during training. Furthermore, the fully trained model was applied for prediction on some samples and performance was reported to have yielded impressive performance. Furthermore, the localization problem was addressed using the trained models to annotate the samples. Visualization for the localized abnormalities was demonstrated using an inverted mask, bounding boxes, and heatmaps with exemplification done with selected abnormalities containing malignancy. Also, the localization accuracy, precision, recall, sensitivity, and specificity of the localization task were reported. Finally, a comparative analysis of the approach reported in this study was carried out in other similar studies. The outcome of the comparison showed that the method demonstrated in this study yielded good performance. This, therefore, shows that the combinatorial problem of classification and localization in medical images can be effectively handled using a deep learning model. the concept may be leveraged in addressing similar problems across fields of medicine having to do with disease classification in digital images. In future, the research should focus on some interesting region selection models for the improvement of the RCNN so that suggestive regions with discriminant features are selected. In addition, the localization achieved by WCNN and that of RCNN can be compared further to corroborate the relevance of their underlying techniques. Furthermore, evolutionary computation and swarm intelligence algorithms are proposed for optimizing the two deep learning models for improved classification and localization performances.

Declaration of interests

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

List of abbreviations

Acronym	Description
A	Adenosis
AUC	Area under the curve
B	Benign
BACH	BreAst Cancer Histology
BreakHis	Breast Cancer Histopathological Image Classification
Cascade R-CNN	Cascade region-based convolutional neural network
CAVGA	Convolutional adversarial variational autoencoder with Guided Attention
CIFAR-10	Canadian Institute For Advanced Research 10
CK	Cohen's Kappa
CNN	Convolutional neural networks
CPU	Central processing unit
DC	Carcinoma
DCN	Deep convolutional network
F	Fibroadenoma

Faster R-CNN	Faster region-based convolutional neural network
FPR	False Positive Rate
GANs	Generative adversarial networks
GB	Gigabyte
GHz	Giga Hertz
GPU	Graphic processing unit
GradCAM	Gradient-weighted Class Activation Mapping
InceptionV3	Inception version 3
IS	In situ carcinoma
IV	Invasive carcinoma
JS	Jaccard similarity or score
L1	Lasso regression
LAG	Large-scale Attention-based Glaucoma
LC	Lobular Carcinoma
Mask R-CNN	Mask region-based convolutional neural network
MC	Mucinous carcinoma
MCC	Matthew's correlation coefficient
MIL	Multiple instance learning
MITOS-ATYPIA-14	Mitos and Atypa 2014 contest
MNIST	Modified National Institute of Standards and Technology
mSTC	modified ShanghaiTech Campus
MVTAD	Machine vision technology Anomaly Detection
MVTec	Machine vision technology
N	Normal
OS	Operating system
PC	Papillary carcinoma
PCIE	Peripheral component interconnect express
PT	Phyllodes tumor
RAM	Random access memory
RCNN	Region-based CNN
R-CNN	Region-based convolutional neural network
RELU	Rectified linear unit
ResNet_V2	Residual network version 2
RMSProp	Root Mean Squared Propagation
ROC	Receiver operating characteristic
ROIs	Region of interest
RSNA	Radiological Society of North America
SEP	Squeeze-Excitation-Pruning
TA	Tubular adenona
TPR	True Positive Rate
VOC	Visual object classes
WCNN	Whole-image based CNN
YOLO	You only look once

References

[1] O. N. Oyelade and A. E. Ezugwu, "A Comparative Performance Study of Random-Grid Model for Hyperparameters Selection in Detection of Abnormalities in Digital Breast Images," *Concurrency and Computation: Practice and Experience*, 2022.

[2] O. Taiwo, A. E.-S. Ezugwu, O. N. Oyelade and M. S. Almutairi, "Enhanced Intelligent Smart Home Control and Security System Based on Deep Learning Model," *Wireless Communications and Mobile Computing*, vol. 7, pp. 1-22, 2022.

[3] O. N. Oyelade and A. E.-S. Ezugwu, "Characterization of abnormalities in breast cancer images using nature-inspired metaheuristic optimized convolutional neural networks mode," *Concurrency and Computation Practice and Experience*, vol. 34, no. 4, 2021.

[4] O. N. Oyelade, A. E.-S. Ezugwu and H. Chiroma, "CovFrameNet: An enhanced deep learning framework for COVID-19 detection," *IEEE Access*, p. 99, 2021.

[5] O. N. Oyelade and A. E. Ezugwu, "Ebola Optimization Search Algorithm: A New Nature-Inspired Metaheuristic Optimization Algorithm," *IEEE Access*, vol. 10, pp. 1-38, 2022.

[6] O. N. Oyelade and E. Absalom E., "A Novel Wavelet Decomposition and Transformation Convolutional Neural Network with Data Augmentation for Breast Cancer Detection Using Digital Mammogram," *Scientific Reports:*, 2022.

[7] O. Olaide, E. Aghiomesi, O. Najeem and A. A. Sambo, "A Semantic Web Rule and Ontologies Based Architecture for Diagnosing Breast Cancer Using Select and Test Algorithm," *Computer Methods and Programs in Biomedicine Update*, 2021.

[8] O. Olaide and A. E.-S. Ezugwu, "Enhancing reasoning through reduction of vagueness using fuzzy OWL-2 for representation of breast cancer ontologies," *Neural Computing and Applications*, 2021.

[9] R. Bakalo, R. Ben-Ari and J. Goldberger, "Classification and detection in mammograms with weak supervision via dual branch deep neural net," *IEEE International Symposium on Biomedical Imaging (ISBI), ,* 2019.

[10] D. Ribli, A. Horváth, Z. Unger, P. Pollner and I. Csabai, "Detecting and classifying lesions in mammograms with Deep Learning," *Scientific Reports*, vol. 8, no. 1, p. 4165, 2018.

[11] Z.-Q. Zhao, Z. P. Zheng, S.-t. Xu and X. Wu, "Object Detection with Deep Learning: A Review," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 99-111, 2019.

[12] B. Qiang, R. Chen, M. Zhou, Y. Pang, Y. Zhai and M. Yang, "Convolutional Neural Networks-Based Object Detection Algorithm by Jointing Semantic Segmentation for Images," *Sensors*, vol. 20, no. 5080, 2020.

[13] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection.," *In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA,*, p. 779–788, 2016.

[14] G. Shi, J. Suo, C. Liu, K. Wan and X. Lv, "Moving target detection algorithm in image sequences based on edge detection and frame difference," *In Proceedings of the 2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference, Chongqing, China,* p. 740–744, 2017.

[15] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal Machine Intelligence*, vol. 39, p. 1137–1149, 2017.

[16] K. He, G. Gkioxari, P. Dollar and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, p. 386–397, 2020.

[17] Z. Cai and N. Vasconcelos, "Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving into high quality object detection," *In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA*, p. 6154–6162, 2018.

[18] O. N. O. Ezugwu, A. E. Ezugwu, A. K. Saha, L. Abualigah, M. S. Almutari and H. Chiroma, "A generative adversarial network for synthetization of regions of interest based on digital mammograms," *Scientific Reports*, 2022.

[19] O. N. Oyelade and A. E. Ezugwu, "ArchGAN: A Generative Adversarial Network for Architectural Distortion Abnormalities in Digital Mammograms," in *Conference: 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, Cape Town, 2021.

[20] O. N. Oyelade and A. E. Ezugwu, "A deep learning model using data augmentation for detection of architectural distortion in whole and patches of images," *Biomedical Signal Processing and Control*, 2022.

[21] R. Bakalo, J. Goldberger and R. Ben-Ari, "Weakly and semi supervised detection in medical imaging via deep dual branch net," *Neurocomputing*, no. 421, pp. 15-25, 2021.

[22] K. K. Evans, D. Georgian-Smith, R. Tambouret, R. L. Birdwell and J. M. Wolfe, "The gist of the abnormal: Above-chance medical decision making in the blink of an eye," *Psychonomic*

Bulletin & Review, vol. 20, p. 1170–1175, 2013.

[23] A. Buetti-Dinh, V. Galli, S. Bellenberg, O. Ilie, M. Herold, S. Christel, M. Boretska, I. V. Pivkin, P. Wilmes, W. Sand, M. Vera and M. Dopson, "Deep neural networks outperform human expert's capacity in characterizing bioleaching bacterial biofilm compositio," *Biotechnology Reports,* vol. 22, 2019.

[24] O. N. Oyelade and A. E. Ezugwu, "A State-of-the-Art Survey on Deep Learning Methods for Detection of Architectural Distortion From Digital Mammography," *EEE Access, doi: 10.1109/ACCESS.2020.3016223,* pp. 148644-148676, 2020.

[25] S. S., S. V. and B. K., "A review on deep learning in medical image analysis," *Intenrational Journal of Multimedia Information Retrieval,* vol. 11, no. 1, pp. 19-38, 2022.

[26] M. Hajabdollahi, R. Esfandiarpoor, E. Sabeti, N. Karimi, R. S. S.M. and S. Samavi, "Multiple abnormality detection for automatic medical image diagnosis using bifurcated convolutional neural network," *Biomedical Signal Processing and Control* , vol. 57, no. 101792, 2020.

[27] S. Venkataramanan and A. Mahalanobis, "Attention Guided Anomaly Localization in Images," in *Computer Vision - ECCV 2020 Lecture Notes in Computer Science*, 2020.

[28] S. Shin, S. Lee, I. Yun, S. Kim and K. Lee, "Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images,," *IEEE Transactions on Medical Imaging* , vol. 38, no. 3, p. 762–774., 2019.

[29] D. Sarvamangala and R. Kulkarni, "Convolutional neural networks in medical image understanding: a survey.," *Evol. Intel. https://doi.org/10.1007/s12065-020-00540-3,* vol. 15, pp. 1-22, 2022.

[30] J. Kang, K. Oh and I.-S. Oh, "Accurate Landmark Localization for Medical Images Using Perturbations," *Applied Sciences,* vol. 11, pp. 1-14, 2021.

[31] S. Kashyap, A. Karargyris, J. Wu, Y. Gur, A. Sharma, K. C. L. Wong, M. Moradi and T. Syeda-Mahmood, "Looking in the Right place for Anomalies: Explainable AI through Automatic Location Learning," in *IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020.

[32] M. T. Islam, A. Aowal, A. T. Minhaz and K. Ashraf, "Abnormality Detection and Localization in Chest X-Rays using Deep Convolutional Neural Networks," *Computer vision and pattern recognition,* pp. 1-16, 2017.

[33] N. T. Arun, N. Gaw, P. Singh, K. Chang, K. V. Hoebel, J. Patel, M. Gidwani and J. Kalpathy-Cramer, "Assessing the trust worthiness of saliency maps for abnormality localization in medical imaging," *Radiology: Artificial Intelligence,* vol. 3, no. 6, 2021.

[34] Y. Choukroun, R. Bakalo, R. Ben-Ari, A. Akselrod-Ballin, E. Barkan and P. Kisilev, "Mammogram Classification and Abnormality Detection from Nonlocal Labels using Deep Multiple Instance Neural Network," *Eurographics Workshop on Visual Computing for Biology and Medicine,* 2017.

[35] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba, "Learning Deep Features for Discriminative Localization," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 27-30 June, 2016, Las Vegas, NV, IEEE,* p. 2921–2929, 2016.

[36] D. Wang, A. Khosla, R. Gargeya, H. Irshad and A. H. Beck, "Deep Learning for Identifying Metastatic Breast Cancer," *Harvard Medical School,* 2016.

[37] H. Bilen and A. Vedaldi, "Weakly Supervised Deep Detection Networks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),,* pp. 12846-2854, 2016.

[38] T. Liu and T. Stathaki, "Faster R-CNN for Robust Pedestrian Detection Using Semantic Segementation Network," *Frontiers Neurorobotics,* vol. 12, no. 64, 2018.

[39] M. Shehab, L. Abualigah, Q. Shambourd, M. A.Abu-Hasheme, M. K. Y. Shambour, A. I. Alsalibi and A. H.Gandomi, "Machine learning in medical applications: A review of state-of-the-art methods," *Computers in Biology and Medicine,* vol. 145, 2022.

[40] M. A. Azam, K. B. Khan, S. Salahuddin, E. Rehman, S. A. Khan, M. A. Khan, S. Kadry and A. H.Gandomi, "A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics," *Computers*

in Biology and Medicine, vol. 144, 2022.

[41] S. Ahmad, T. Ullah, I. Ahmad, A. AL-Sharabi, K. Ullah, R. Khan, S. Rasheed, I. Ullah, N. Uddin and S. Ali, "A Novel Hybrid Deep Learning Model for Metastatic Cancer Detection," *Computational Intelligence and Neuroscience,* no. 8141530, pp. 1-14, 2022.

[42] A. J. Belay, A. O. Salau, M. Ashagrie and M. B. Haile, "Development of a chickpea disease detection and classification model using deep learning," *Informatics in Medicine Unlocked,* vol. 31, no. 2022, pp. 1-12, 2022.

[43] H. Alshammari, K. Gasmi, M. Krichen, L. B. Ammar, M. O. Abdelhadi, A. Boukrara and M. A. Mahmood, "Optimal Deep Learning Model for Olive Disease Diagnosis Based on an Adaptive Genetic Algorithm," *Wireless Communications and Mobile Computing,* no. 8531213, pp. 1-13, 2022.

[44] A. Haque, S. Marwaha, C. K. Deb, S. Nigam, AlkaArora, K. S. Hooda, P. L. Soujanya, S. KumarAggarwal, B. Lall, M. Kumar, S. Islam, M. Panwar, P. Kumar and R. C.Agrawal, "Deep learning-based approach for identifcation of diseases of maize crop," *Scientifc Reports,* vol. 12, no. 6334, pp. 1-14, 2022.

[45] A. SABER, M. SAKR, O. M. ABO-SEIDA, A. KESHK and H. CHEN, "A Novel Deep-Learning Model for Automatic Detection and Classification of Breast Cancer Using the Transfer-Learning Technique," *IEEE Access,* vol. 9, no. 2021, pp. 71194 -71209, 2021.

[46] S. Arooj, Atta-ur-Rahman, M. Zubair, M. F. Khan, K. Alissa, M. A. Khan and A. Mosavi, "Breast Cancer Detection and Classification Empowered With Transfer Learning," *Front. Public Health,* vol. 10, no. 924432, pp. 1-18, 2022.

[47] J. Heo, J. H. Lim, H. R. Lee, J. Jang, Y. S. Shin, D. Kim, J. Lim, Y. M. Park, Y. W. Koh, Soon-HyunAhn, E. Chung, D. Lee, J. Seok and C. Kim, "Deep learning model for tongue cancer diagnosis using endoscopic images," *Scientifc Reports,* vol. 12, p. 6281, 2022.

[48] S. Wang, Y. Cong, H. Zhu, X. Chen, L. Qu, H. Fan, Q. Zhang and M. Liu, "Multi-Scale Context-Guided Deep Network for Automated Lesion Segmentation With Endoscopy Images of Gastrointestinal Tract," *IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS,* vol. 25, no. 2, pp. 514-525, 2021.

[49] S. Yu and T. Feifei, "Construction and verification of retinal vessel segmentation algorithm for color fundus image under BP neural network model," *The Journal of Supercomputing,* vol. 77, no. 1, pp. 1-15, 2021.

[50] Q. Li, L. Li, W. Wang, Q. Li and J. Zhong, "A comprehensive exploration of semantic relation extraction via pre-trained CNNs," *Knowledge-Based Systems,* vol. 194, no. 2020, pp. 1-7, 2020.

[51] D. Z. H. Su and H. Elmannail, "Multilevel threshold image segmentation for COVID-19 chest radiography: A framework using horizontal and vertical multiverse optimization," *Computers in Biology and Medicine,* vol. 146, no. 2022, 2022.

[52] H. Ni and H. Liu, "Multiple Visual Fields Cascaded Convolutional Neural Network for Breast Cancer Detection," in *PRICAI 2018: Trends in Artificial Intelligence. PRICAI 2018. Lecture Notes in Computer Science*, Cham.

[53] O. N. Oyelade and A. E. Ezugwu, "A bioinspired neural architecture search based convolutional neural network for breast cancer detection using histopathology images," *Sceintific Reports,* vol. 11, no. 19940, 2021.

[54] A. Polónia, C. Eloy and P. Aguiar, "BACH Dataset : Grand Challenge on Breast Cancer Histology images," *Medical image analysis,* 2019.

[55] F. Spanhol, L. S. Oliveira, C. Petitjean and L. Heutte, "Breast Cancer Histopathological Database (BreakHis)," [Online]. Available: https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-BreakHis/. [Accessed 20 April 2021].

[56] S. F., O. L.S., P. C. and L. Heutte, "A Dataset for Breast Cancer Histopathological Image Classification," *IEEE Transactions on Biomedical Engineering (TBME),* vol. 63, no. 7, pp. 1455-1462, 2016.

[57] E. Reinhard, M. Adhikhmin, B. Gooch and P. Shirley, "Color transfer between images.," *IEEE*

Comput. Graph. Appl., vol. 21, p. 34–41, 2001.

[58] M. M, N. M, M. JS, B. D, W. JT, G. X, C. Schmitt and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis.," *In: 2009 IEEE International Symposium on Biomedical Imaging. Boston, MA:,* p. 1107–10, 2009.

[59] E. M. Nejad, L. S. Affendey, R. B. Latip and I. B. Ishak, "Classification of Histopathology Images of Breast into Benign and Malignant using a Single-layer Convolutional Neural Network," *ICISPC 2017: Proceedings of the International Conference on Imaging, Signal Processing and Communication,* p. 50–53, 2017.

[60] 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," *PlosOne,* vol. 12, no. 6, 2017.

[61] 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," *Sceintific Report,* vol. 7, no. 4172 , 2017.

[62] C. Zhu, F. Song, Y. Wang, H. Dong, Y. Guo and J. Liu, "Breast cancer histopathology image classification through assembling multiple compact CNNs," *BMC Medical Informatics and Decision Making,* 2019.

[63] J. Xie, R. Liu, J. L. IV and C. Zhang, "Deep Learning Based Analysis of Histopathological Images of Breast Cancer," *Frontiers Genetics,* 2019.

[64] M. Saha, C. Chakraborty and D. Racoceanu, "Efficient deep learning model for mitosis detection using breast histopathology images," *Computerized Medical Imaging and Graphics,* vol. 64, pp. 29-40, 2018.

[65] M. Hägele, P. Seegerer, S. Lapuschkin, M. Bockmayr, W. Samek, F. Klauschen, K.-R. Müller and A. Binder, "Resolving challenges in deep learning-based analyses of histopathological images using explanation methods," *Scientific Reports,* vol. 10, 2020.

[66] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang and A. Madabhushi, "Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images," *IEEE Transactions on Medical Imaging,* vol. 35, no. 1, pp. 119-130, 2016.

[67] B. E. Bejnordi, G. Zuidhof, M. Balkenhol, M. Hermsen, P. Bult, B. v. Ginneken, N. Karssemeijer, G. Litjens and J. v. d. Laak, "Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images," *Journal of Medical Imaging,* 2017.

[68] Y. Zheng, Z. Jiang, F. Xie, H. Zhang, Y. Ma, H. Shi and Y. Zhao, "Feature extraction from histopathological images based on nucleus-guided convolutional neural network for breast lesion classification," *Pattern Recognition,* pp. 14-25, 2017.

[69] P. Khosravi, E. Kazemi, M. Imielinski, O. Elemento and I. Hajirasouliha, "Deep Convolutional Neural Networks Enable Discrimination of Heterogeneous Digital Pathology Images," *EBioMedicine,* vol. 27, pp. 317-328, 2018.

[70] I. Kandel and M. Castelli, "A Novel Architecture to Classify Histopathology Images Using Convolutional Neural Networks," *Applied Sciences,* vol. 10, no. 8, 2020.

[71] S. Venkataramanan, K.-C. Peng, R. V. Singh and A. Mahalanobis, "Attention Guided Anomaly Localization in Images," *arXiv:1911.08616,* 2020.