Summer 2019

Using Feature Extraction From Deep Convolutional Neural Networks for Pathological Image Analysis and Its Visual Interpretability

Wei-Wen Hsu
Old Dominion University, wxhsu001@odu.edu

Follow this and additional works at: https://digitalcommons.odu.edu/ece_etds

Part of the Artificial Intelligence and Robotics Commons, and the Bioimaging and Biomedical Optics Commons

Recommended Citation
Hsu, Wei-Wen. "Using Feature Extraction From Deep Convolutional Neural Networks for Pathological Image Analysis and Its Visual Interpretability" (2019). Doctor of Philosophy (PhD), Dissertation, Electrical & Computer Engineering, Old Dominion University, DOI: 10.25777/bmrq-kf64
https://digitalcommons.odu.edu/ece_etds/168

This Dissertation is brought to you for free and open access by the Electrical & Computer Engineering at ODU Digital Commons. It has been accepted for inclusion in Electrical & Computer Engineering Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.
USING FEATURE EXTRACTION FROM DEEP CONVOLUTIONAL NEURAL NETWORKS FOR PATHOLOGICAL IMAGE ANALYSIS AND ITS VISUAL INTERPRETABILITY

by

Wei-Wen Hsu
B.S. July 2008, National Central University
M.S. July 2011, National Taiwan University

A Dissertation Submitted to the Faculty of Old Dominion University in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

ELECTRICAL AND COMPUTER ENGINEERING

OLD DOMINION UNIVERSITY
August 2019

Approved by:

Chung-Hao Chen (Director)
Gene J. Hou (Member)
Jiang Li (Member)
Chunsheng Xin (Member)
ABSTRACT

USING FEATURE EXTRACTION FROM DEEP CONVOLUTIONAL NEURAL NETWORKS FOR PATHOLOGICAL IMAGE ANALYSIS AND ITS VISUAL INTERPRETABILITY

Wei-Wen Hsu
Old Dominion University, 2019
Director: Chung-Hao Chen

This dissertation presents a computer-aided diagnosis (CAD) system using deep learning approaches for lesion detection and classification on whole-slide images (WSIs) with breast cancer. The deep features being distinguishing in classification from the convolutional neural networks (CNN) are demonstrated in this study to provide comprehensive interpretability for the proposed CAD system using the domain knowledge in pathology. In the experiment, a total of 186 slides of WSIs were collected and classified into three categories: Non-Carcinoma, Ductal Carcinoma in Situ (DCIS), and Invasive Ductal Carcinoma (IDC). Instead of conducting pixel-wise classification (segmentation) into three classes directly, a hierarchical framework with the multi-view scheme was designed in the proposed system that performs lesion detection for region proposal at higher magnification first and then conducts lesion classification at lower magnification for each detected lesion. The majority voting scheme was adopted to improve the error tolerance of the system in lesion-wise prediction. For all collected 186 slides, the slide-wise prediction accuracy rate strikes to 95.16% (177/186) in binary classification to predict carcinoma (malignant) or non-carcinoma (benign), and the sensitivity for cases with carcinoma reaches 96.36% (106/110). In multi-classification, the accuracy rate is 92.47% (172/186) when predicting Non-Carcinoma, DCIS, and IDC for each slide. Most importantly, the interpretability for the mechanism of the proposed CAD system is provided from the pathological perspective. The experimental results show that the morphological characteristics and co-occurrence properties
learned by the deep learning models for lesion detection and classification meet the clinical rules in diagnosis. Accordingly, the pathological interpretability of the deep features not only enhances the reliability of the proposed CAD system to gain acceptance from medical specialists, but also facilitates the development of deep learning frameworks for various tasks in pathology.
This dissertation is dedicated to my parents Ming-Sheng Hsu and Tsai-Hsi Ho.
ACKNOWLEDGMENTS

I would like to express my special gratitude to my advisor Dr. Chung-Hao Chen as well as my mentors Dr. Xiang Gao and Dr. Yu-Ling Hou who gave me the guidance in research and provided all possible resources to help me complete this work. More importantly, I would like to thank Dr. Yanhong Tai, who dedicated lots of time and efforts on exploring AI potential in pathological applications with me.
TABLE OF CONTENTS

LIST OF TABLES ................................................................................................................. ix

LIST OF FIGURES ................................................................................................................ x

Chapter

1. INTRODUCTION .................................................................................................................. 1
   1.1 PROBLEM STATEMENT .................................................................................................. 2
   1.2 PURPOSE OF THE STUDY ............................................................................................ 3
   1.3 RESEARCH QUESTIONS ............................................................................................... 3
   1.4 SIGNIFICANCE OF THE STUDY .................................................................................. 4
   1.5 ORGANIZATION ........................................................................................................... 4

2. BACKGROUND ..................................................................................................................... 5
   2.1 NEURAL NETWORKS ..................................................................................................... 5
   2.2 CONVOLUTIONAL NEURAL NETWORKS ........................................................................ 7
   2.3 ALEXNET ..................................................................................................................... 11
   2.4 RESNET ....................................................................................................................... 12
   2.5 TEMPLATE MATCHING ............................................................................................... 13
   2.6 VISUAL INTERPRETABILITY OF MODELS .................................................................... 16
   2.7 SPATIAL CORRESPONDENCES OF OBJECT PARTS LEARNED FROM DATA ............... 19

3. RELATED WORK .................................................................................................................. 22
   3.1 CAD SYSTEM FOR PATHOLOGICAL IMAGE ANALYSIS ............................................. 22
   3.2 VISUAL INTERPRETABILITY FOR THE CAD SYSTEMS ............................................. 26

4. METHODOLOGY .................................................................................................................. 31
   4.1 MATERIALS ................................................................................................................... 31
   4.2 EXPERIMENTS FOR LESION DETECTION ..................................................................... 32
   4.3 CAD SYSTEM (LESION DETECTION AND LESION CLASSIFICATION) ....................... 33
   4.4 VISUALIZATION OF THE DEEP FEATURES .................................................................. 38

5. RESULTS AND DISCUSSION ............................................................................................... 41
   5.1 EXPERIMENTS FOR LESION DETECTION ..................................................................... 41
   5.2 CAD SYSTEM (LESION DETECTION AND LESION CLASSIFICATION) ....................... 53
   5.3 VISUAL INTERPRETABILITY FOR LESION CLASSIFICATION .................................... 63

6. CONCLUSION AND FUTURE WORK ............................................................................... 76

REFERENCES ......................................................................................................................... 78
VITA ................................................................. 82
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Number of Images in the Training Dataset and the Testing Dataset</td>
<td>17</td>
</tr>
<tr>
<td>2. Transfer Learning Using the Pretrained Models</td>
<td>17</td>
</tr>
<tr>
<td>3. Comparisons among Four Different Classifiers</td>
<td>43</td>
</tr>
<tr>
<td>4. Performance Before and After Feature Reduction</td>
<td>48</td>
</tr>
<tr>
<td>5. Performance Before and After Feature Selection</td>
<td>52</td>
</tr>
<tr>
<td>6. System Performance in Patch-Level, Region-Level, and Slide-Level through 5-Fold Cross-Validation</td>
<td>53</td>
</tr>
<tr>
<td>7. The Slide-Wise Predicting Performance on the Testing Set</td>
<td>55</td>
</tr>
<tr>
<td>8. Feature Weights for Each Class Learned from the Training Data (Positive $W_0$)</td>
<td>66</td>
</tr>
<tr>
<td>9. Feature Weights for Each Class Learned from the Training Data (Positive $W_1$)</td>
<td>69</td>
</tr>
<tr>
<td>10. Feature Weights for Each Class Learned from the Training Data (Positive $W_2$)</td>
<td>74</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Bio-Inspired Design</td>
<td>6</td>
</tr>
<tr>
<td>2. Multilayer Perceptron (MLP)</td>
<td>7</td>
</tr>
<tr>
<td>3. The Structure of LeNet-5 [18]</td>
<td>9</td>
</tr>
<tr>
<td>4. An Example of Optical Character Recognition (OCR) Task Using CNN</td>
<td>10</td>
</tr>
<tr>
<td>5. The structure of AlexNet [1]</td>
<td>12</td>
</tr>
<tr>
<td>7. An Example of Template Matching. (a) The Original Image. (b) Template to Find from the Original Image. (c) Response Map. (d) The Template Was Found and Enclosed Within a Bounding Box</td>
<td>15</td>
</tr>
<tr>
<td>8. Discriminative Features in Recognizing Cats</td>
<td>18</td>
</tr>
<tr>
<td>9. Discriminative Features in Recognizing Dogs</td>
<td>19</td>
</tr>
<tr>
<td>10. Spatial Correspondences of Object Parts Encoded in the DCNN Models [24]</td>
<td>19</td>
</tr>
<tr>
<td>11. The Puzzle Game Design. (a) The Interface on Mobile Application. (b) One of the Puzzles: The Original Image (Left) and the Randomly Initial Arrangement of Patches (Right). (c) The Activation of Detecting Goose Was Increased When the Patches Have a Better Spatial Arrangement During the Steps in Solving Puzzles</td>
<td>21</td>
</tr>
<tr>
<td>16. The Discriminative Patches with High Activation for Each Unit (Each Row) in Two Categories for the Diagnosis of Colon Cancer [35]</td>
<td>28</td>
</tr>
<tr>
<td>17. Class Activation Mapping (CAM) Analysis to Visualize the Salient Parts of Object in Classification [33]</td>
<td>29</td>
</tr>
</tbody>
</table>
18. Projecting Class Activation Analysis for Five Different Predicted Types of Colorectal Polyps [7]. ...........................................................................................................................................30

19. Annotations of Lesions and Training a Deep Model for Lesion Detection. (a) Precise Delineations of DCIS Lesions on WSIs. (b) The Training Procedure in the Deep Learning Framework for Lesion Detection. ..................................................................................................................33

20. The Designed Framework for the Diagnosis of WSIs with Breast Cancer. .........................34


22. Patch Sampling in Testing .........................................................................................................35

23. Procedures in Slide Assessment. (a) Results of Lesion Detection. (b) Proposed Candidate Regions After Filtering by Size. (c) Results of Lesion Classification. .................................36

24. One of the Training Set and Testing Set In 5-Fold Cross-Validation ........................................37

25. The Mappings Between Neurons and Their Corresponding FOVs. ....................................38

26. The Patch (Left) with the Highest Activation Unit in Channel No. 49 and Its Corresponding Activation Map (Right). ..................................................................................................................39

27. An Input Patch (Left) and Its Corresponding Activation Map Overlaid on the Input Patch (Right) ........................................................................................................................................39

28. The Structure of the End-To-End AlexNet Model ....................................................................42

29. The Fully Connected Layers in AlexNet Were Replaced by Other Classifiers (LR/SVM/RF) for the Control Group ..............................................................................................................42

30. The Activation Map Was Generated from the Results of Forward Propagation Through the Convolutional Layers in One of the Channels .................................................................44

31. The Activation Maps Reflect the High Response Regions. Most of the Learned Filters in DCNN can Work as Morphological Descriptors to Detect Specific Cells and Tissues. ...........45

32. Cells and Tissues Are Arbitrarily Distributed in the Sampling Patches .........................46

33. The Modified Model Using 13x13 Average Pooling Layer to Discard Spatial Information. 47

34. Visualization of the Deep Features Using the Modified Model in Section 5.1.3. The High-Response Regions Reflect the Distribution of A Specific Cell or Tissue .............................49

35. The Deep Learning Models Are Able to Reveal the Co-Occurrence Properties of Patterns That Were Learned from the Training Dataset. (a) The Learned Filter Targets on Cancerous Nuclei. (b) The Units Get High Activations on Those Regions of Cytoplasmic Clearing Around Cancerous Nuclei ........................................................................................................50
36. An Example of the Unrecognizable Feature.................................................................52
37. Structure of Lobules. (a) Normal Lobules. (b) Lobular Hyperplasia.............................56
38. Benign Lesions. (a) Benign Epithelial Hyperplasia. (b) Usual Ductal Hyperplasia (UDH) ...57
39. Lesions of DCIS. (a) Low-Grade DCIS. (b) High-Grade DCIS..................................58
40. Lesions of IDC. (a) IDC with Cord Growth Patterns. (b) IDC with Peritumoral Retraction Spaces..........................................................59
41. False Negative Cases with Lesions of (a) Low-Grade DCIS and (b) Low-Grade IDC ........61
42. False Alarms Happened Due to (a) the Invasive-Growth-Like Patterns in the Lesions of Lobular Cancerization and (b) Deformed Cells and Tissues from Cautery Artifacts. ........62
43. Activations between Non-Carcinoma and Carcinoma Were Well-Separated Within Feature No.1134. .............................................................................................................................63
44. The Supportive Features for the Category of Non-Carcinoma: Lobules. .......................64
45. The Supportive Features for the Category of Non-Carcinoma: Wavy Collagen Fibers. ......65
46. The Supportive Features for the Category of Non-Carcinoma: Benign Epithelial Hyperplasia. .............................................................................................................................65
47. The CAM Analysis for the Category of Non-Carcinoma....................................................67
48. The Supportive Features for the Category of DCIS: Comedo-Like Necrosis..................68
49. The Supportive Features for the Category of DCIS: Cribriform Growth Pattern.............68
50. The Supportive Features for the Category of DCIS: Rounded Configuration of Lesions. ....69
51. Lesions of UDH with Peripheral Cribriform Patterns and Rounded Configuration.........70
52. The CAM Analysis for the Category of DCIS.................................................................71
53. The Supportive Features for the Category of IDC: Cord Growth Pattern......................72
54. The Supportive Features for the Category of IDC: Irregular Clusters of Tumor Cells. ......73
55. The Supportive Features for the Category of IDC: Peritumoral Retraction Spaces..........73
56. The CAM Analysis for the Category of IDC....................................................................74
57. The Well-Developed Framework for the Diagnosis of Proliferative Lesions.................77
CHAPTER 1
INTRODUCTION

The framework of deep convolutional neural networks (DCNN) has achieved outstanding performance on many applications of computer vision since 2012 [1]. Many studies have shown that the classification results with the features extracted from deep convolutional networks, known as deep features, outperform the results with the conventional approaches using hand-crafted features [2-5]. Accordingly, the deep learning framework has been widely adopted for pathological image analysis [6]. In pathology, the registered pathologists are in high demand because the pathological examination is laborious and time-consuming. This dissertation proposes a computer-aided diagnosis (CAD) system using deep learning framework that assists pathologists in assessment to alleviate their workload and prevent assessments from subjectivity in diagnosis [6]. More importantly, it plays a role as a “second reader” system to reduce missing inspections and misinterpretations.

However, it is hard to be accepted by medical doctors for such CAD systems with deep learning approaches since the deep learning model is an end-to-end structure that takes raw images as the inputs and derives the outcomes directly [7]. It is deficient in theoretical explanations about the mechanism for such systems with deep learning approaches so that many experts in the medical field see it a “black box.” The contribution of this study is to provide the visual interpretability for the deep features to explain the mechanism of the proposed CAD system from the pathological perspective, increasing acceptability from medical experts. Meanwhile, with a better understanding of the mechanism, it can also facilitate the development of the deep learning framework for several tasks in pathology.
1.1 PROBLEM STATEMENT

Although a vast amount of literature has demonstrated the state-of-the-art computer-aided detection and diagnosis systems in digital pathology, most of their work did not meet the clinical needs [5]. That is, the fully automated diagnosis system that takes whole-slide images (WSIs) as the inputs and gives predictions for diagnosis in both lesion-level and patient-level to assist pathologists in clinical practice is demanded. Taking breast histopathology images for example, experiments [4, 8-11] were performed using the open-access dataset BreaKHIs [12] or the dataset from Bioimaging2015 challenge [13], in which microscopic images in each class were cropped from the WSIs beforehand for learning and testing, rather than taking WSIs directly. On the other hand, the proposed systems [2, 3, 14-17] took WSIs directly but only focused on the detection for a specific type of lesions rather than providing comprehensive diagnosis results. It is not until the Grand Challenge on Breast Cancer Histology images (BACH) [5] was launched in early 2018 that pixel-wise classification (semantic segmentation) into three classes was performed for the whole-slide breast histology images in one of the tasks. The 3-category segmentation results can be very useful as a second opinion system for pathologists in clinical practice; nonetheless, the performance evaluation of segmentation in the challenge may be less meaningful for clinical purposes than the assessment in lesion-level and slide-level. Accordingly, the current study aimed to integrate the algorithms of lesion detection and classification that makes the proposed CAD system be able to take WSIs directly and give lesion-wise and slide-wise predictions, working as a second opinion system for pathologists in clinical practice.

Besides, most developers of the state-of-the-art CAD systems with deep learning approaches treat deep learning models like a “black box” and only focus on the outcomes. As a result, the key features or the morphological characteristics learned by the deep learning models
that contributed to the outcomes in lesion detection and classification have rarely been discussed, especially for breast cancer histopathology. A comprehensive mechanism of feature extraction in the deep learning framework for lesion classification using pathological explanations is focused in this study. Therefore, the reliability of the proposed CAD system can be enhanced, gaining acceptance from medical experts at the same time.

1.2 PURPOSE OF THE STUDY

The main purpose of this study is to develop a fully automated CAD system that meets the clinical needs in pathological assessments for breast cancer. That is, the system takes WSIs directly and provides lesion-wise and slide-wise multi-class predictions for diagnosis. Moreover, the interpretability of the proposed CAD system using deep learning approach is addressed in this dissertation by explaining deep features that are crucial in tasks of lesion detection and lesion classification with pathological knowledge. The features that were explored by the deep models from the training dataset can reveal some clinical insights which may inspire the scientific researchers to discover new clinical markers in diagnosis in the future.

1.3 RESEARCH QUESTIONS

The key research questions are: (i) Does the proposed CAD system meet the clinical needs that can really assist pathologists in diagnosis? (ii) Whether the morphological characteristics in pathology can be captured by the deep learning models to distinguish different types (levels) of lesions (carcinoma)? The evaluation methods for the proposed CAD system will be laid out in detail in Chapter 4, and the system performance on diagnosis will be presented in Chapter 5.
1.4 SIGNIFICANCE OF THE STUDY

This study will contribute to the development of a fully automated CAD system for breast histology images analysis that (1) fills the clinical requirements and (2) provides the interpretability of the extracted deep features from the pathological perspective. The proposed CAD system is mainly aimed to assist pathologists in diagnosis to reduce missing inspections and misinterpretations effectively. In this dissertation, a comprehensive mechanism of a CAD system using the deep learning framework will be focused to establish the reliability and gain acceptance from medical specialists. By observing the deep features and the co-occurrence properties that were explored by the deep learning models, those outcomes are anticipated to be recognized by experienced pathologists and hopefully can inspire them to discover the new clinical markers in diagnosis.

1.5 ORGANIZATION

Following this chapter, Chapter 2 reviews the background information of deep convolutional neural networks (DCNN) with its applications in computer vision. Chapter 3 reviews and discusses recent published articles and literature on similar CAD systems for histology images and their methods for evaluation. Chapter 4 presents the methodology of developing the CAD system, including lesion detection, lesion classification, and the visualization of the deep features that contribute to the outcomes in detection and classification. Chapter 5 analyzes the performance of the proposed CAD system and discusses the deep features and learned properties that were used for diagnosis with pathological explanations. Chapter 6 concludes the dissertation and narrates the future work of this study.
CHAPTER 2

BACKGROUND

This section provides the background information of deep convolutional neural networks (DCNN) in the tasks of image classification. In addition, recent research work related to the interpretability of the deep features extracted from DCNN will be discussed.

2.1 NEURAL NETWORKS

It is very similar to the way we design the airplane which is inspired by the structure of birds, the structure of neural networks, at the very beginning, is also a bio-inspired model, taking reference to humans’ central nervous systems, as shown in Fig. 1. In machine learning, the multilayer perceptron (MLP) is a very popular model used in supervised learning, which is a feed-forward neural network model that consists of multiple layers of nodes (units), with each layer fully connected to the next one. The structure of MLP is shown as Fig. 2. MLP consists of an input layer and an output layer on both sides, and hidden layers in between. The parameter \( w(i) \) can be derived by the back-propagation algorithm, using the gradient descent method iteratively.

The depth of the architecture in neural networks refers to the number of hidden layers and an output layer. The decision of depth in networks can be made depending on the learning task of interest. For example, the shallow structure of the neural networks, such as MLP, is very useful in the classification problems with inputs of hand-crafted features (features extracted in advance with preprocessing). The shallow network is powerful enough (aggregation scheme with many hypotheses of linear perceptron) and efficient in training (fewer parameters to train compared with
deep networks). On the contrary, the deep network takes more time in the training process, and the model of complexity may become too high, resulting in a severe over-fitting problem.

![Deep Network Diagram](image)

**Fig. 1. Bio-Inspired Design**

Fortunately, with the hardware improvements (faster computing ability of GPU, multi-core CPU) and many regularization methods have been proposed to prevent the system from over-fitting, the drawbacks of deep networks can be resolved, and there are needs for certain applications that motivates us to “go deep” in neural networks.
2.2 CONVOLUTIONAL NEURAL NETWORKS

In convolutional neural networks (CNN), there are two important designs: (1) Local receptive fields with weight sharing scheme and (2) Sub-sampling. Different from the fully connected networks that ignore the topology of inputs, variables are highly correlated if they are close in the temporal or spatial domain for speech signals or images. Therefore, in CNN, one unit (neuron) in the next layer is only connected with the adjacent input units within the receptive field, instead of all units in the previous layer. In addition, all the units share the same weight set in a feature map. From the perspective of computer vision, the design of local receptive fields with weight sharing scheme is equivalent to applying a filter on an image for template matching. The receptive fields in the convolutional layers are to extract useful features and detect if certain patterns exist in the input image.

Another important design is to reduce the spatial resolution of feature maps by sub-sampling. The reduction of spatial resolution is necessary for two reasons. The first reason is to compose the larger feature elements by combining those smaller local features extracted from the previous convolutional layer. In the first convolutional layer, the small size of the receptive field (also called
filter, mask, kernel in image processing) was chosen so that it can extract elementary visual features like oriented edges, end-points, and corners in the input image. To obtain the larger-scale features of object parts that comprise these elementary local features, it can be achieved by either enlarging the size of the receptive field or retaining the smaller size of the receptive field with the shrinking size of the feature maps in the spatial domain. However, enlarging the size of receptive field means there are more connections between input units and the neurons, and there will be more weights to be obtained in training, which becomes a heavy burden in computation. Secondly, to make the prediction results more robust and invariant to shifts and distortions, the information of an object’s precise location should be ignored. In tasks of object recognition, only the approximate orientations of object parts are relevant, and the use of the exact position lets the classification system lose the general property. Therefore, down-sampling the spatial resolution can reduce the preciseness of the object parts’ locations to achieve shift-invariant property for the system. Both reasons make sub-sampling operation very important in CNN.

The designs of (1) local receptive fields with weight sharing scheme and (2) sub-sampling in CNN are to extract different levels of features through optimization. And the forward propagation in the convolutional layers can be regarded to perform hierarchical template mating for classification.

In LeNet-5 [18], shown as Fig. 3, C1 means that the first hidden layer is the convolutional layer and filtering operation (cross-correlation) was performed. The number of feature maps indicates the number of different filtering masks used in C1, and each feature map is the result of an image filtered by a certain mask. Those masks obtained by training are to detect if certain features exist in the input image. S2 means that the second hidden layer is a sub-sampling layer; it reduces the spatial dimension of data from the previous layer, i.e., feature maps after C1.
In the first convolutional layer, i.e., C1, there are six weight sets (six filtering masks) obtained from the training process, and they extract six local features from the image. Since the size of the mask is relatively small compared with the whole image, the elementary local features like end-points, corners, and oriented-edges, can be captured in this layer. The spatial size of feature maps after C1 has been shrunken after S2, so the spatial size of feature maps after C3 (the third hidden layer as well as the second convolutional layer) becomes about half size in each dimension compared with C1. However, the size of masks in C3 remains the same as in C1, which means it can detect the larger-scale features of object parts in the image after C3.

Taking optical character recognition (OCR) as an example, as shown in Fig. 4, there were five kernels obtained by training that extracted five elementary local features from the image. Those five features may be horizontal edges, vertical edges, edges with 45°, edges with 135°, and corners (including end-points). After sub-sampling, in the second convolutional layer, larger elements that comprise the combined features from the first convolutional layer can be captured. The combined features of horizontal edges, edges with 45°, and edges with 135° can compose a larger element, like a curve-upward or a curve-downward shape, which is a crucial pattern to detect the number digit like 0, 2, 6, 8 and 9. Similarly, the combined feature of vertical edges, edges with 45°, and edges with 135° can form the curves to the right and to the left to detect the number digit.
like 0, 3, 5, 6, 8, and 9. And the combined feature of lines and corners can comprise the part in number 1, 4, 5, and 7.

In LeNet-5, the spatial size of each feature map after C5 is 5 by 5, which means the input image was divided into $5 \times 5 = 25$ regions for information of orientations in classification. For example, to recognize 5 and 7, a high response that shows the existence of a right angle corner roughly on the left top of the feature maps can be expected, and the detection of a curve-shape feature or a line-shape feature in the lower part of the feature maps can be used to distinguish them.

In LeNet-5, the feature maps after the last convolutional layer summarize the spatial distribution of the learned patterns in the input image. Before C5, the convolutional layers and sub-sampling layers are responsible for feature extraction from raw image data. After convolutional layers, the fully connected neural networks, also known as MLP, are to encode the spatial relations of object parts for classification. Instead of relying on hand-crafted features based on field-knowledge, the convolutional layers automatically discover the representations needed for classification from raw data and make machine learning algorithms work effectively.

Fig. 4. An Example of Optical Character Recognition (OCR) Task Using CNN.
2.3 ALEXNET

In 2012, the model of AlexNet [1], as shown in Fig. 5, was proposed. It was modified from the model of LeNet-5 and firstly used to classify objects in color images with large image data set (ImageNet) [19]. In the challenge, 1.2 million high-resolution images were classified into 1000 different objects, and alexnet largely improved the accuracy in classification. The strategy of parallel computing with two GPUs in AlexNet made the high computational network model feasible in training and testing, and it aroused an upsurge of deep learning in many applications.

In AlexNet, it is similar to LeNet-5 in feature extraction but with several different settings for improvement. Firstly, rather than using separate layers for sub-sampling in LeNet-5, they reduced the spatial resolution by methods of output striding and max-pooling. In max-pooling, only the neuron with the highest response in the receptive field was used to represent the existence of a certain feature. Secondly, since there were more than one million color images with 1000 categories of objects in the challenge, the feature elements were much more variant and complicated, compared with the task of optical character recognition. Therefore, all features extracted from the previous layers should be combined instead of the certain combinations of features used in LeNet-5. And the number of convolutional layers and the number of feature maps in each layer should also increase to have higher model complexity for the variant and complicated inputs (since there are 1000 categories of objects). Meanwhile, with higher model complexity for the learning system (many parameters to be obtained in training), the regularization method of dropout [1,5] was adopted to prevent the system from over-fitting in AlexNet effectively.
2.4 RESNET

When convolutional neural networks go deeper, unfortunately, the performance dropped, and there are some problems that we may encounter: (1) gradients vanishing or exploding and (2) the degradation of training accuracy.

In the ResNet model [20], the number of layers increased largely compared with the model of AlexNet. Therefore, the problems mentioned above should be well-addressed, or we cannot get benefits from the deeper models. To tackle the problem of vanishing/exploding gradients, the algorithm for weights initialization [21] was adopted, and batch normalization [22] was applied right after each convolution and before activation in ResNet.

For degradation problem, a smaller training error should be expected with a deeper model due to its higher model complexity, theoretically, but the deeper model led to a higher training error contrarily. That is, increasing the depth of a network leads to a decrease in performance on training data since accuracy gets saturated. Accordingly, the deep residual learning block, as shown in Fig. 6, was built in ResNet to tackle such degradation problems.
In Fig. 6, the design of skip connection performs identity mapping so that the training error will not increase when the accuracy gets saturated. In the implementation, it learned the residual representation functions instead of the signal representation directly.

![Diagram of Residual Learning Block in ResNet50](image)

Fig. 6. The Residual Learning Block in ResNet50 [20].

The model of ResNet was the winner of ILSVRC 2015 in image detection, and localization, as well as the winner of MS COCO 2015 detection, and segmentation. It is worth noting that the ensemble of six ResNet models achieved 3.57% top-5 error, which surpasses human-level performance (5.1% top-5) in the task of classification.

### 2.5 TEMPLATE MATCHING

The filter operation using cross-correlation can be used to align two images as template matching or image registration. In Formula (1), the value $M(i,j)$ with the highest value means the most matching position of the template $T$ on the image $I$.

\[
M(i,j) = \sum_{p=0}^{m-1} \sum_{q=0}^{n-1} [I(i + p, j + q) - \overline{I_{i,j}}] [T(p, q) - \overline{T}]
\]

(1)

\[\overline{I_{i,j}}\] computes the mean of intensities within the mask size $(m, n)$ starting at $I(i,j)$, as shown in Formula (2), and $\overline{T}$ is the mean of the template, as shown in Formula(3).
Followed the distributive property of summation, Formula (4) is the extended expression of Formula (1).

\[
M(i, j) = \sum_{p=0}^{m-1} \sum_{q=0}^{n-1} [I(i + p, j + q)] [T(p, q) - \bar{T}] - \bar{I}_{i,j} \sum_{p=0}^{m-1} \sum_{q=0}^{n-1} [T(p, q) - \bar{T}] \tag{4}
\]

Since the term \([T(p, q) - \bar{T}]\) is the zero-centered template, the summation of the mask with zero-mean distribution leads to zero, as shown in Formula (5).

\[
\sum_{p=0}^{m-1} \sum_{q=0}^{n-1} [T(p, q) - \bar{T}] = 0 \tag{5}
\]

As a result, Formula (1) can be simplified to Formula (6), and the template matching can be done by applying the zero-centered template, i.e. \([T(p, q) - \bar{T}]\), on the original image.

\[
M(i, j) = \sum_{p=0}^{m-1} \sum_{q=0}^{n-1} [I(i + p, j + q)] [T(p, q) - \bar{T}] \tag{6}
\]

Fig. 7 shows an example of template matching using cross-correlation operation. The template, shown as Fig. 7-(b), is used as the mask, i.e., \(T\) in Formula (1). The brightest spot indicates the position with the highest response after spatial filtering, as shown in Fig. 7-(c). And
Fig. 7-(d) shows the position where has the highest response is the most matching point of the template in the image.

Fig. 7. An Example of Template Matching. (a) The Original Image. (b) Template to Find from the Original Image. (c) Response Map. (d) The Template Was Found and Enclosed Within a Bounding Box.
In convolutional neural networks, the forward propagation of convolutional layers can be treated as a hierarchical template matching through each layer, and the trained parameters in each layer are filtering masks that look for certain templates (features). Then, the positions with high response indicate the learned features were found in the input image.

2.6 VISUAL INTERPRETABILITY OF MODELS

There is a story about the misinterpretation of neural networks [23]. The researchers collected 100 photos of camouflaged tanks in trees and another 100 photos of trees without tanks for supervised learning, and the dataset was split into a training set and a testing set evenly. The researchers trained a neural network on the training dataset and tested the remaining 100 photos for performance evaluation. The testing results were surprisingly successful that all images with camouflaged tanks were detected correctly without any false alarm. However, the trained neural networks model failed to distinguish pictures with or without camouflaged tanks in real-world testing.

It turned out that, in the researchers' data set, photos of camouflaged tanks had been taken on cloudy days, while photos of the plain forest had been taken on sunny days. The neural network had learned to distinguish cloudy days from sunny days, instead of distinguishing camouflaged tanks from the empty forest. This story may not really happen, but it is good motivation to understand what features had been learned by the neural networks. In the story, the trained networks overfitted the features that were caused by the data sources and lost the generalization properties of learning. It is very likely to happen when there are different data sources, and differences exist between them that were learned by the deep learning models. Therefore, in this study, what features were learned by the deep learning models will be discussed
and verified if they follow the domain knowledge, rather than overfitting the properties due to a specific data source.

Taking the image classification task on the dataset of “cats and dogs” as an example, as shown in Table 1, the discriminative features that recognized cats and dogs in the images were presented and discussed to verify those properties learned by the deep learning model if they followed the prior knowledge on classifying cats and dogs.

**Table 1**

**Number of Images in the Training Dataset and the Testing Dataset**

<table>
<thead>
<tr>
<th>Category</th>
<th>Training Set</th>
<th>Validation Set</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>cats</td>
<td>8750</td>
<td>3750</td>
<td>12500</td>
</tr>
<tr>
<td>dog</td>
<td>8750</td>
<td>3750</td>
<td>12500</td>
</tr>
</tbody>
</table>

Transfer learning was adopted that the pre-trained models of AlexNet and ResNet50 were used in the classification for performance comparisons, and the classification results are shown as Table 2. In the trail, the out-sample accuracy from ResNet50 outperformed AlexNet in classification, and the model of ResNet50 was visualized for feature analysis.

**Table 2**

**Transfer Learning Using the Pretrained Models**

<table>
<thead>
<tr>
<th>Pretrained Models</th>
<th>In-Sample Accuracy</th>
<th>Out-Sample Accuracy</th>
<th>Difference</th>
<th>Elapsed Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNet</td>
<td>99.54%</td>
<td>97.52%</td>
<td>2.02%</td>
<td>14.03s</td>
</tr>
<tr>
<td>ResNet50</td>
<td>99.87%</td>
<td>99.01%</td>
<td>0.86%</td>
<td>91.82s</td>
</tr>
</tbody>
</table>
For feature analysis, the discriminative features for each category were presented to verify those learned features followed by the prior knowledge in recognizing cats and dogs. Fig. 8 shows the top 10 discriminative features that supported the system to recognize cats in the images. As we can observe from Fig. 8, the characteristics of a cat’s face with big eyes and small nose were used by the deep learning model to distinguish cats from dogs effectively. Besides, the narrow stripes that were seen more often on cats than dogs became a discriminative feature in recognizing cats as well. Moreover, even though dogs have whiskers, the whiskers on cats look more prominent because cats are usually less hairy than dogs. Lasts but not the least, the shape of the cat’s eyes was also distinctive from the dog’s eyes and can be very useful in classification.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Big eyes and small nose</strong></td>
<td><img src="image1" alt="Images" /> <img src="image2" alt="Images" /> <img src="image3" alt="Images" /> <img src="image4" alt="Images" /> <img src="image5" alt="Images" /> <img src="image6" alt="Images" /> <img src="image7" alt="Images" /> <img src="image8" alt="Images" /> <img src="image9" alt="Images" /> <img src="image10" alt="Images" /></td>
</tr>
<tr>
<td>957(1), 1739(2), 744(3), 1701(6),</td>
<td></td>
</tr>
<tr>
<td>327(8)</td>
<td></td>
</tr>
<tr>
<td><strong>Tabby (narrow stripes)</strong></td>
<td><img src="image11" alt="Images" /> <img src="image12" alt="Images" /> <img src="image13" alt="Images" /> <img src="image14" alt="Images" /> <img src="image15" alt="Images" /> <img src="image16" alt="Images" /> <img src="image17" alt="Images" /> <img src="image18" alt="Images" /> <img src="image19" alt="Images" /> <img src="image20" alt="Images" /></td>
</tr>
<tr>
<td><strong>Prominently long whiskers</strong></td>
<td><img src="image21" alt="Images" /> <img src="image22" alt="Images" /> <img src="image23" alt="Images" /> <img src="image24" alt="Images" /> <img src="image25" alt="Images" /> <img src="image26" alt="Images" /> <img src="image27" alt="Images" /> <img src="image28" alt="Images" /> <img src="image29" alt="Images" /> <img src="image30" alt="Images" /></td>
</tr>
<tr>
<td><strong>Cat’s eyes</strong></td>
<td><img src="image31" alt="Images" /> <img src="image32" alt="Images" /> <img src="image33" alt="Images" /> <img src="image34" alt="Images" /> <img src="image35" alt="Images" /> <img src="image36" alt="Images" /> <img src="image37" alt="Images" /> <img src="image38" alt="Images" /> <img src="image39" alt="Images" /> <img src="image40" alt="Images" /></td>
</tr>
</tbody>
</table>

Fig. 8. Discriminative Features in Recognizing Cats.

On the other hand, the discriminative features in recognizing dogs are listed in Fig. 9. The most distinctive features were various kinds of big noses on dogs. In addition, the characteristics of long limbs and lop ears can also be used to distinguish dogs from cats. Interestingly, since dogs need more outdoor activities than cats, it can be observed that there were many pictures of dogs
with the green grass background from the dataset. The deep learning model also noticed this property of co-occurrence from the data and utilized it in classification.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face with big nose</td>
<td><img src="image1" alt="Examples of face with big nose" /></td>
</tr>
<tr>
<td>850(1), 992(2), 1262(4), 145(6),</td>
<td><img src="image2" alt="Heatmaps of face with big nose" /></td>
</tr>
<tr>
<td>1972(8), 464(9)</td>
<td><img src="image3" alt="Examples of faces with big nose" /></td>
</tr>
<tr>
<td>Long limbs</td>
<td><img src="image4" alt="Examples of long limbs" /></td>
</tr>
<tr>
<td>1948(3)</td>
<td><img src="image5" alt="Heatmaps of long limbs" /></td>
</tr>
<tr>
<td>Lop ears</td>
<td><img src="image6" alt="Examples of lop ears" /></td>
</tr>
<tr>
<td>1515(5)</td>
<td><img src="image7" alt="Heatmaps of lop ears" /></td>
</tr>
<tr>
<td>Green grass (co-occurrence)</td>
<td><img src="image8" alt="Examples of green grass" /></td>
</tr>
<tr>
<td>1062(7)</td>
<td><img src="image9" alt="Heatmaps of green grass" /></td>
</tr>
</tbody>
</table>

Fig. 9. Discriminative Features in Recognizing Dogs.

2.7 SPATIAL CORRESPONDENCES OF OBJECT PARTS LEARNED FROM DATA

The spatial correspondences of object parts are encoded in the DCNN models. For example, the head pattern of a horse with its adjacent neck pattern detected can increase the confidence of recognizing the object “horse” for the system, as shown in Fig. 10. That is, the spatial correspondence of head pattern and neck pattern of a horse can be very useful for the classifier in object detection.

Fig. 10. Spatial Correspondences of Object Parts Encoded in the DCNN Models [24].
When it comes to the processing of contextual information, quantitative analysis is the most challenging part and becomes a bottleneck. With the mature development of deep learning framework, the semantic features can be extracted in the deeper layer, which largely improved the ability of machines in the tasks of classification and recognition. Besides, the spatial correspondences of object parts are encoded in the DCNN models, and it is highly related to the spatial arrangement in the puzzle games. In [25], the well-trained DCNN model was adopted to reflect the level of how puzzle players used global information in deciding steps.

In puzzle games, only some parts of objects can be seen in one piece so that the information may not be sufficient to have correct results for object recognition, and sometimes that also makes it very challenging for players to recognize objects. After several steps, many pieces were put together to construct a bigger (more complete) configuration of the object, and in the meanwhile, more features of that object were collected by the deep networks so that the higher activation can be obtained for the correct object category. In DCNN, the activations can reflect the confidence of the classification results. As a result, the activations become a crucial measurement, and the category with the greatest activation shows the maximum likelihood in the classification results. Such a classification scheme is very similar to what human do in playing puzzles.

The activations from the trained DCNN models can be utilized to reflect how strong the contextual information can be derived from a single piece or some combinations. Fig. 11 shows the puzzle game design for the experiment. With more pieces put together properly, the activation for correct category became higher, as shown in Fig. 11-(c), which means the arrangement of puzzle patches agrees with the spatial correspondences that were encoded by the deep learning models.
Fig. 11. The Puzzle Game Design. (a) The Interface on Mobile Application. (b) One of the Puzzles: The Original Image (Left) and the Randomly Initial Arrangement of Patches (Right). (c) The Activation of Detecting Goose Was Increased When the Patches Have a Better Spatial Arrangement During the Steps in Solving Puzzles.
CHAPTER 3
RELATED WORK

This chapter reviews relevant research and comments on the similarities and differences to the current study. To make the review more structured, this chapter focuses on investigating recent literature by categorizing these papers into two topics. The first topic covers similar CAD systems for pathological image analysis. The second topic describes recent literature, which focuses on the visual interpretability for state-of-the-art CAD systems. The methodology and study results from the related literature, as well as similarity and differences to the current research, will be discussed.

3.1 CAD SYSTEM FOR PATHOLOGICAL IMAGE ANALYSIS

The pathological examination has been the gold standard for diagnosis in cancer. It plays an important role since cancer diagnosis and staging help determine treatment options. In pathology, the manual process of slide assessment is laborious and time-consuming, and wrong interpretations may happen due to fatigue or stress in specialists. Besides, there has been an insufficient number of registered pathologists, as a result, the workload for pathologists turns heavier, becoming a problem in pathology. Recently, the techniques of image processing and machine learning have significantly advanced, and the computer-aided detection/diagnosis (CADe/CADx) systems were developed to assist pathologists in slide assessment. Working as a second opinion system, it is designed to alleviate the workload of pathologists and avoid missing inspections.

In digital pathology, glass slides with tissue specimens were digitized by the whole-slide scanner at high resolution, becoming whole-slide images (WSIs) [26]. The analysis of WSIs is
non-trivial because it involves a large amount of data (gigapixel level) to process and visualize [27]. In the preliminary stage, experiments [4, 10-13] were performed on open-access datasets from BreaKHIs [12] or Bioimaging 2015 challenge [13] / BACH [5], in which the microscopy images for each class were cropped from WSIs beforehand for training and testing, as shown in Fig. 12, instead of taking WSIs as the inputs directly. However, their approaches did not meet the clinical needs since cropping several regions of interests (ROIs) manually and sending them into the CAD system are almost infeasible in real clinical practice. To fill the clinical requirements, a fully automated diagnosis system that takes WSIs as the inputs and provides diagnostic assessment in both lesion-level and slide-level to assist pathologists in clinical practice is highly demanded.

![Multi-classification for Microscopic Images with Breast Cancer](image)

**Fig. 12.** Multi-classification for Microscopic Images with Breast Cancer in One of the Tasks from BACH Challenge [5].

For the CAD systems that process WSIs directly, the developers worked on the dataset from CAMELYON challenge [28] that detected breast cancer metastases on WSI of lymph node sections in [16, 17]. The detection results over WSI are shown in Fig. 13, and its detection performance was evaluated in both lesion-level (with FROC) and slide-level (with ROC). However, when it comes to the detection problem in WSIs, especially for lesion-level evaluation, the data information of WSIs may be too large for the manual examination that it is hard to avoid missing inspections for some small lesions. Many lesion regions were reported missing from the forum of the challenge after the competition. From the technical perspective, the CAD systems are
expected to detect all suspicious cancerous regions. On the other hand, one detected lesion can determine the diagnosis result in clinical practice. The methods of performance assessment should be further discussed to make system evaluation suited for clinical use.

![Detection of Breast Cancer Metastases in CAMELYON Challenge](image1)

**Fig. 13.** The Detection of Breast Cancer Metastases in CAMELYON Challenge [16].

In the work by Bejnordi et al. [15], the clustering algorithm was adopted to detect ductal carcinoma in situ (DCIS) for WSIs with breast cancer, as shown in Fig. 14. Even though the system performance was evaluated in both lesion-level and slide-level rigorously in their work, the developed system that simply detects lesions with DCIS may be inadequate for clinical use. The detection for invasive breast cancers may be more demanded from the clinical perspective since they are more severe and tend to have a poorer prognosis.

![Clustering Algorithm Was Adopted for the Detection of DCIS](image2)

**Fig. 14.** Clustering Algorithm Was Adopted for the Detection of DCIS [15].
The deep learning framework was applied in [14] to detect invasive breast cancer in WSIs. The system performance was evaluated by the Dice Coefficient, which is a common assessment method in tasks of segmentation (pixel-wise classification). Because it is hard to manually define precise and consistent contours for the invasive lesions, fair performance evaluation in lesion-level became almost infeasible. However, such an assessment of segmentation is more meaningful from the technical perspective than the clinical perspective. Besides, as it is shown in Fig. 15, the manual ground truth annotations for delineating IDC lesions may not be precise enough compared with the contours from the detected regions by the system. As a result, the assessment method in segmentation seemed not fair in performance evaluation.

Fig. 15. The Detection of IDC Using Deep Learning Approach [14].
From late 2017 to early 2018, the Grand Challenge on BreAst Cancer Histology images (BACH) [5] was launched, and the pixel-wise labeling in whole-slide breast histology images with three classes (benign, in situ carcinoma, and invasive carcinoma) was performed in one of the tasks. The 3-category segmentation results can be very useful as a second opinion system for pathologists in clinical practice; nonetheless, the performance evaluation of segmentation in the challenge may be less meaningful for clinical purposes than the assessments in lesion-level and patient-level. Accordingly, to meet the clinical requirements, the performance of the proposed CAD system in both lesion-level and slide-level will be evaluated in this study.

3.2 VISUAL INTERPRETABILITY FOR THE CAD SYSTEMS

In machine learning, many studies focused on the development of classifiers during the early years. However, data scientists found feature extraction for data representation the bottleneck of performance in tasks of classification and detection. Therefore, feature engineering that concentrates on the methods to extract features and make machine learning algorithms work effectively becomes more and more critical for better performance. In representation learning, scientists aim to develop the techniques that allow a system to automatically discover the representations needed for classification or detection from raw data. Since 2012 [1], the framework of Deep Convolutional Neural Networks (DCNN) has achieved outstanding performance on many applications of computer vision. Many studies have shown that the classification results with the features extracted from deep convolutional networks, known as deep features, outperform the results with the conventional approaches using hand-crafted features [2, 3]. Accordingly, the deep learning framework has been widely adopted for pathological image analysis. Nonetheless, such CADe/CADx systems with deep learning approaches are hard to be accepted by medical specialists
since the deep learning framework is an end-to-end fashion that takes raw images as the input and derives the outcomes directly. It is deficient in theoretical explanations about the mechanism for such systems with deep learning approaches. Most developers solely focused on the efficacy of outcomes, without explaining why their proposed frameworks can work effectively [7]. Consequently, many medical specialists treat deep learning framework as a “black box” and doubt about the feasibility of such systems in clinical practice.

In DCNN, it comprises convolutional layers and fully connected layers to perform feature extraction and classification respectively during the process of optimization. In convolutional layers, local features such as colors, end-points, corners, and oriented-edges are collected in the shallow layers. These local features in the shallow layers are integrated into larger structural features like circles, ellipses, and specific shapes or patterns when the layer goes deeper. Afterward, these structural patterns or shapes constitute the high-level semantic representations that describe feature abstraction for each category [29]. On the other hand, in fully connected layers, it takes the extracted features from the convolutional layers as the inputs and works as a classifier. These fully connected layers can encode the spatial correspondences of those semantic features and convey the co-occurrence properties between patterns or objects [30].

Many studies have worked on the visual interpretability of deep learning models on the datasets of natural images [29, 31-34] and showed the mechanism of deep learning frameworks follows the prior knowledge for each category in classification. The process of the classification system is concordant with humans’ intuitions in tasks of image classification [24]. However, for pathological image analysis, there has been insufficient research for explanations about the mechanism of systems with deep learning approaches, and the feasibility of such systems keeps being questioned by medical specialists.
For visual interpretability of DCNN framework, the most instinctive method is to collect those input patches with high activations from CNN units (neurons) and observe the patterns that caused high responses in the images [29]. For pathological images, visualization of deep features was applied to observe the characteristics that were learned from large image dataset automatically in Xu’s work [35] for colon cancer. As it is shown in Fig. 16, similar patterns can be observed in the selected patches from the same category. The visualization results show the learned features by the deep learning models yield biologically meaningful insights that are recognized by pathologists.

Fig. 16. The Discriminative Patches with High Activation for Each Unit (Each Row) in Two Categories for the Diagnosis of Colon Cancer [35].
In Xu’s study [35], it has shown promising results on providing visual interpretability of the DCNN models for pathological images analysis. That is, the local patterns learned by the deep learning models agree with the morphological characteristics in pathology. However, more features and properties can be explored from the dataset by the deep learning models if deeper models and samples in different scales are used. Besides, except for the analysis of discriminative features, the analysis that takes trained weights for each feature into consideration to observe the salient parts of objects in classification was proposed in [7, 33, 34].

Zhou et al. [33] proposed the Class Activation Mapping (CAM) analysis that the trained weights are applied on the corresponding activation map in each channel from the last convolutional layer, and the results were resized and overlaid on the original image to observe the crucial parts that impacted the classification results for each category. In Fig. 17, the salient parts of the object that contribute to the correct classification class are highlighted, and the highlighted parts should be accordant to the prior knowledge in classification so that the validity of the deep learning models can be established.

![Class Activation Mapping (CAM) Analysis to Visualize the Salient Parts of Object in Classification [33]](image)

Fig. 17. Class Activation Mapping (CAM) Analysis to Visualize the Salient Parts of Object in Classification [33].
And the similar analysis was adopted in the work by Korbar et al. [7] on the pathological images for multi-classification of colorectal polyps. As it is shown in Fig. 18, the influential regions are highlighted to identify the histopathological characteristics that contribute to the results for each class. The visualization results from their work convey some clinical insights that were discovered by the deep learning models; however, it is lack of explanations from the pathological point of view to verify the mechanism of deep learning framework can really work for pathological analysis.

Fig. 18. Projecting Class Activation Analysis for Five Different Predicted Types of Colorectal Polyps [7].

In this dissertation, a deep learning framework that (1) integrates lesion detection and lesion classification using multi-view scheme and (2) offers the prediction accuracy in lesion-level and slide-level for performance evaluation will be proposed to fill the clinical requirements. More importantly, the deep features learned by the deep learning models for both lesion detection and lesion classification will be discussed, and the mechanism of the framework will be explained by the domain knowledge of pathology.
CHAPTER 4
METHODOLOGY

This chapter (i) introduces the deep learning framework for pathological image analysis that performs lesion detection and lesion classification, (ii) describes the performance evaluation methods to verify the efficacy of the CAD system, and (iii) demonstrates the methods of networks visualization and feature analysis to provide visual interpretability for the framework.

4.1 MATERIALS

In the implementation of the proposed CAD system, there are two main phases: (1) Lesion Detection and (2) Lesion Classification. In order to verify the detection algorithm is valid, 27 H&E stained samples of breast tissues with Ductal Carcinoma in Situ (DCIS) lesions were collected, and most of the lesions were annotated precisely by pathologists. In lesion classification, a total of 186 H&E stained samples of breast tissues were collected from three categories: Non-Carcinoma, Ductal Carcinoma in Situ (DCIS), and Invasive Ductal Carcinoma (IDC). Lesion regions were labeled with three classes (Non-Carcinoma, DCIS, and IDC) by an experienced pathologist in 109 slides for cross-validation, and the other 77 cases were classified by the slide category for testing. All labeled regions were double-checked by the second registered pathologist, and the ground truth classification for each slide was determined by the diagnosis assisted with immunohistochemical results. All samples were digitized to the format of Whole-Slide Images (WSIs) by the scanner of ZEISS Axio Scan.Z1 at high resolution (x40).
4.2 EXPERIMENTS FOR LESION DETECTION

In the experiment of lesion detection, 27 H&E stained samples of breast tissues with DCIS were collected and digitized to the format of WSIs. All lesions of DCIS were delineated in yellow by a registered pathologist and confirmed by the second registered pathologist, as shown in Fig. 19-(a). The original dataset was split into two sets: 15 cases for training and the remaining 12 cases for testing. To perform lesion detection through WSIs, many small patches were sampled under high magnification (x40), called patching [36, 37]. There are two kinds of sampling sets: a positive set and a negative set. The positive set collected the patches with tumorous cells by sampling inside the annotated regions. On the other hand, the patches with normal cells or normal tissues were sampled outside the annotated regions, comprising the negative set. There were about 200k patches sampled from the training dataset and about 85k patches from the testing dataset with balanced class distribution in the initial stage of data acquisition. The training procedure in the deep learning framework for lesion detection is shown as Fig. 19-(b).

In the designed experiments, the pre-trained AlexNet [1] model on the ImageNet dataset was used to perform transfer learning [38]. Within the pre-trained AlexNet model, the feature size for each patch is 9216 by 1 in classification. Since the classifiers of Support Vector Machine (SVM) and Random Forests (RF) were used to replace the fully connected layers to achieve decomposition of the end-to-end DCNN framework, the dataset would be too large for SVM and RF if all 200k sampling patches were used in training. Therefore, to shrink the size of the dataset to make training feasible, 20k patches (positive: negative = 1:1) were randomly selected from the original training dataset, becoming the real training dataset to fine-tune the deep learning model [16]. For performance evaluation, 10k patches (positive: negative = 1:1) were also randomly
collected from the original testing dataset as the real testing dataset to compute the out-sample accuracy in patch classification.

Fig. 19. Annotations of Lesions and Training a Deep Model for Lesion Detection. (a) Precise Delineations of DCIS Lesions on WSIs. (b) The Training Procedure in the Deep Learning Framework for Lesion Detection.

4.3 CAD SYSTEM (LESION DETECTION AND LESION CLASSIFICATION)

A fully automated CAD system that provides complementary and objective assessments to assist pathologists in diagnosis is the main purpose of this study. To achieve that, a multi-view scheme was applied in the design that patches were sampled at different resolutions for analysis. Under higher magnification, the morphological characteristics of nuclei are focused for lesion detection. On the contrary, the spatial arrangement of cells or the co-occurrence properties of tissues is suited for observation at lower magnification. Accordingly, a hierarchical framework
was designed, as shown in Fig. 20, for the CAD system to (1) perform lesion detection at high magnification for region proposal firstly and (2) classify each proposed region into Non-Carcinoma, DCIS, and IDC at intermediate to low magnification for lesion classification afterward.

![Diagram](image)

**Fig. 20.** The Designed Framework for the Diagnosis of WSIs with Breast Cancer.

In the implementation, patches were sampled at different resolutions from the 109 slides with lesion regions labeled. For region proposal, a total of 130k patches were sampled at higher magnification from all 109 cases to fine-tune a pre-trained AlexNet-like model. For lesion classification, 82k patches were sampled at intermediate to low magnification for training and testing through 5-fold cross-validation using the pre-trained model of ResNet50 [20]. Fig. 21 shows the training procedure in the deep learning framework for lesion classification.

For slide assessment, the detection model was applied to perform lesion detection first. Then, the detected regions were filtered by the minimum size limit of 1 mm for region proposal since the case of microinvasive carcinoma was not considered in this study. The detection of microinvasion requires the framework with more precise localization algorithm rather than a patch-base method (patching). For each candidate region, several patches were randomly sampled from the region at lower magnification for lesion classification, as shown in Fig. 22. The number
of sampling patches depends on the area of the lesion. Afterward, the predictions of these sampling patches from the classification model determined which category the region belongs to using majority voting. Finally, the slide-wise classification can be achieved by analyzing the types of lesions that exist in the WSI. The above procedures are shown in Fig. 23.

Fig. 21. The Training Procedure in the Deep Learning Framework for Lesion Classification.

Fig. 22 Patch Sampling in Testing.
Fig. 23. Procedures in Slide Assessment. (a) Results of Lesion Detection. (b) Proposed Candidate Regions After Filtering by Size. (c) Results of Lesion Classification.
The cross-validation set comprised 109 slides with lesion regions labeled. We computed the classification accuracy rates in patch-level, region-level, and slide-level through 5-fold cross-validation for performance analysis. Each WSI in the cross-validation set was unseen by its corresponding testing model for assessment, and all cases were executed in the testing procedure to derive the overall accuracy rate of classification. Fig. 24 shows one of the training set and testing set in 5-fold cross-validation. For the evaluation in region-level, the detected regions after size filtering were further classified into three categories (Non-Carcinoma, DCIS, and IDC) by two registered pathologists separately. If a region was labeled differently, the final decision was made after they reached a consensus.

To further validate the slide-wise performance of the proposed CAD system, all sampling patches from the cross-validation set (109 cases) were combined together for training. Then, the trained model was applied to the 77 cases in the testing set for slide assessment.

Fig. 24. One of the Training Set and Testing Set In 5-Fold Cross-Validation.
4.4 VISUALIZATION OF THE DEEP FEATURES

To observe the influential patterns that were used in patch classification, the size of the Field-of-Views (FOVs) was computed to derive the mappings between the units (neurons) and their corresponding FOVs in the input image, as shown in Fig. 25. In DCNN models, the number of channels in the assigned convolutional layer indicates the number of learnable filters which represent particular features. The units (neuron) in each channel represents the spatial orientation with respect to its corresponding FOV in the input image. For the unit that gets high activation, it means the learned pattern has a strong response to the corresponding region (FOV) in the image, reflecting the matching level between them. For visualization [35], the activations of units in the assigned convolutional layer were recorded for all patches, and all patches were ranked by the units’ activations for each channel. Afterward, the patches with the units of top 100 activations for each channel were collected with the corresponding high-response region highlighted in a yellow bounding box, as shown in Fig. 26. Besides, the corresponding activation maps were resized to the same size as the input image to have a better observation of the learned pattern and its spatial distribution. Fig. 26 shows one of the examples that the learned pattern reflects the distribution of lymphocytes.

Fig. 25. The Mappings Between Neurons and Their Corresponding FOVs.
Similarly, for lesion classification, to observe the semantic features that were learned by the deep learning model in classification, the input patches with high activation units were collected from the selected channel [29, 35], and the corresponding activation maps were also generated to reflect the distribution of the matching patterns that the networks look for in the input images. Fig. 27 shows an example of an input patch with the highest activation unit in channel No.1134 and its corresponding activation map. From observations, the learned filter of no.1134 detects the distribution of lobules in the input patch, and it is distinctive to the category of Non-Carcinoma in lesion classification.
For feature analysis, the method of decision stump [39] was adopted to rank the top 100 discriminant features out of all 2048 features extracted by the model of ResNet50 for each category. The activations of all 2048 units from all sampling patches were computed and collected as a training dataset (#patches by 2048 features) for binary classification using decision stump. For each category in the classification, the classifier of decision stump derived an optimal threshold that led to the best classification results in every feature. If there strikes a high classification accuracy in a specific feature, it implies the feature is distinctive to the category so that the data within different classes can be well-separated. It infers what features are discriminative and contributory to classification. Furthermore, the trained weights of each feature for each class were also taken into consideration for analysis. The technique of Class Activation Mapping (CAM) [33] was applied to reflect the influential patterns in the input patch that are decisive in classification.
CHAPTER 5
RESULTS AND DISCUSSION

In this chapter, the patch-wise performance with various classifiers for lesion detection is evaluated for comparisons in section 5.1. For lesion classification, the rigorous assessments on patch-wise, lesion-wise, and slide-wise performance are presented in section 5.2. Finally, the feature analysis is demonstrated in section 5.3 to provide visual interpretability for the proposed CAD system.

5.1 EXPERIMENTS FOR LESION DETECTION

In this section, four experiments were conducted to study the properties of the deep features learned by the DCNN models in lesion detection. In section 5.1.1, the deep features are proved to be transferable and meaningful in classification. The visualization of the deep features in the section 5.1.2 shows that they work as morphological descriptors to detect specific cells and tissues, and the results are accordant to the category in pathology. In section 5.1.3, the original AlexNet model was modified based on prior knowledge to strike better performance in both efficacy and efficiency. Afterward, all features were ranked by importance to compare views between humans and machines in section 5.1.4.

5.1.1 FEATURE EXTRACTION IN DCNN

Motivation: Even though the deep learning model is an end-to-end structure, it, in fact, can be decomposed into two parts: convolutional layers for feature extraction and fully connected layers for classification. The goal of this experiment is to verify that the features extracted by the
deep learning models are meaningful in classification so that those features are capable of incorporating with other classifiers, rather than being exclusive to the neural networks.

**Hypothesis:** Features extracted from the convolutional layers are meaningful in classification and can work with other classifiers as well.

**Models:** The end-to-end pre-trained AlexNet [1] model was used in training and testing, and the structure is shown in Fig. 28. For the control group, the fully connected layers in AlexNet were replaced by other classifiers, including Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF), as shown in Fig. 29.

Fig. 28. The Structure of the End-To-End AlexNet Model.

Fig. 29. The Fully Connected Layers in AlexNet Were Replaced by Other Classifiers (LR/SVM/RF) for the Control Group.
Results and Discussion: The performance using different classifiers in training and testing are listed in the column of in-sample accuracy and out-sample accuracy respectively in TABLE 3. The testing results show tiny differences in accuracy rates among different classifiers. That means the features extracted from the convolutional layers are not restricted to the end-to-end neural networks. Those features are meaningful in classification and can incorporate with other classifiers as well. From TABLE 3, it is noteworthy that overfitting seemed to occur on the model trained with Random Forests, on the other hand, it achieved the highest out-sample accuracy rate on the model trained with Logistic Regression. It may imply the simpler model can somehow lead to better performance on the out-sample dataset due to its better property of generalization.

TABLE 3

COMPARISONS AMONG FOUR DIFFERENT CLASSIFIERS

<table>
<thead>
<tr>
<th>Model</th>
<th>In-Sample Accuracy</th>
<th>Out-Sample Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNet (9216)</td>
<td>0.999</td>
<td>0.978</td>
</tr>
<tr>
<td>CNN + Logistic Regression (9216)</td>
<td>1</td>
<td>0.980</td>
</tr>
<tr>
<td>CNN + SVM (9216)</td>
<td>1</td>
<td>0.974</td>
</tr>
<tr>
<td>CNN + Random Forests (9216)</td>
<td>1</td>
<td>0.966</td>
</tr>
</tbody>
</table>
5.1.2 VISUALIZATION OF MODEL

Motivation: From the previous experiment, the deep learning model has demonstrated its capability in distinguishing patches with or without lesions, and the deep features learned by the DCNN models are meaningful in classification. In this experiment, the patterns that contribute to the classifier in decision making are revealed to understand the mechanism of deep learning models from the pathological perspective.

Hypothesis: Most deep features learned by the DCNN models agree with the pathological rules in classification.

Model: The trained AlexNet model from section 5.1.1 was used for visualization, and forward propagation was performed through the convolutional layers for the input patch to derive its corresponding activation map in each channel, as shown in Fig. 30.

Fig. 30. The Activation Map Was Generated from the Results of Forward Propagation Through the Convolutional Layers in One of the Channels.

Results and Discussion: The sampling patches and the corresponding activation maps for the selected channels are presented and classified by the pathological categories in Fig. 31. From
observations, most of the learned filters in DCNN work as morphological descriptors to detect specific cells and tissues. And the activation maps reflect the spatial distribution of the learned patterns from the input patches. Interestingly, in this experiment, only the regions with lesions were manually labeled by the pathologists; however, we found the deep learning models are able to analyze the main components in the patches and categorize them by their characteristics. That is, in lesion detection, the deep learning models not only detect the distribution of tumor cells but also recognize lymphocytes, collagen fibers, and some other non-cell structural tissues such as luminal spaces, areas of necrosis and secretions. The results show that the deep features learned by the DCNN model agree with the clinical insights in pathology and the hypothesis holds.

<table>
<thead>
<tr>
<th>Category</th>
<th>Samples for selected channels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor Cells</td>
<td><img src="image1" alt="Channel No. 6" /> <img src="image2" alt="Channel No. 114" /> <img src="image3" alt="Channel No. 250" /></td>
</tr>
<tr>
<td>Lymphocytes</td>
<td><img src="image4" alt="Channel No. 5" /> <img src="image5" alt="Channel No. 131" /> <img src="image6" alt="Channel No. 132" /></td>
</tr>
<tr>
<td>Collagen Fibers</td>
<td><img src="image7" alt="Channel No. 1" /> <img src="image8" alt="Channel No. 75" /> <img src="image9" alt="Channel No. 77" /></td>
</tr>
<tr>
<td>Others</td>
<td><img src="image10" alt="Channel No. 10" /> <img src="image11" alt="Channel No. 15" /> <img src="image12" alt="Channel No. 34" /></td>
</tr>
</tbody>
</table>

Fig. 31. The Activation Maps Reflect the High Response Regions. Most of the Learned Filters in DCNN can Work as Morphological Descriptors to Detect Specific Cells and Tissues.
5.1.3 FEATURE REDUCTION

Motivation: In tasks of image classification on natural images, the spatial arrangement of object parts is an essential characteristic for the deep learning models in object recognition. For example, eyes are supposed to be detected above a nose or a mouth for a human face in the image. However, for pathological images, since patches were sampled at high magnification (x40), cells and tissues are arbitrarily distributed in the small sampling patches, as shown in Fig. 32. The information of objects’ spatial positions becomes meaningless and irrelevant in the task of patch classification here.

![Cells and Tissues Are Arbitrarily Distributed in the Sampling Patches.](Image)

Hypothesis: Characteristics of objects’ spatial orientations can be ignored in patch classification, and feature reduction can be applied to speed up the system.

Model: From the previous experiment, it has shown that the deep learning models could recognize tumor cells, lymphocytes, and collagen fibers. Most of the learned DCNN features can be regarded as detectors for these categories. Since we assume the information of spatial orientations for these elements could be ignored within the small sampling patches, the tasks of patch classification can be accomplished by checking if any lesion exists in the patch, without knowing its exact orientation. Accordingly, a 13 by 13 average pooling layer was adopted to replace the original 6 by 6 max pooling layer in AlexNet-Layer 5. The modified model is shown
in Fig. 33. As a result, the total number of features in classification will be reduced from 6x6x256 (9216) to 1x1x256 (256). The size of features became its 1/36 compared with the original structure.

Fig. 33. The Modified Model Using 13x13 Average Pooling Layer to Discard Spatial Information.

**Results and Discussion:** For comparisons, the performance before and after feature reduction are listed in TABLE 4. With the feature size that is 36 times smaller than the original one, the out-sample accuracy can still remain at the same level or even slightly better. That means the characteristics of objects’ spatial orientations is redundant and can be discarded within the small-size sampling patches, which proves the hypothesis. Since the model’s complexity drops after feature reduction, the results suggest that constraining the complexity of model somehow can trade a better generalization property to prevent the model from overfitting and achieve better out-sample accuracy. Moreover, after feature reduction from 9216 to 256, the system of lesion detection became 23% faster in execution. The performance was improved in both efficacy and efficiency using the modified model.
TABLE 4
PERFORMANCE BEFORE AND AFTER FEATURE REDUCTION

<table>
<thead>
<tr>
<th>Model</th>
<th>In-Sample Accuracy</th>
<th>Out-Sample Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNet (9216)</td>
<td>0.999</td>
<td>0.978</td>
</tr>
<tr>
<td>CNN + Logistic Regression (9216)</td>
<td>1</td>
<td>0.980</td>
</tr>
<tr>
<td>CNN + Logistic Regression (256)</td>
<td>0.985</td>
<td>0.979</td>
</tr>
<tr>
<td>CNN + SVM (9216)</td>
<td>1</td>
<td>0.974</td>
</tr>
<tr>
<td>CNN + SVM (256)</td>
<td>0.990</td>
<td>0.976</td>
</tr>
<tr>
<td>CNN + Random Forests (9216)</td>
<td>1</td>
<td>0.966</td>
</tr>
<tr>
<td>CNN + Random Forests (256)</td>
<td>1</td>
<td>0.978</td>
</tr>
</tbody>
</table>

5.1.4 FEATURE SELECTION

Motivation: After feature reduction, the same method of visualization in section 5.1.2 was used to observe the patterns learned from the modified model in section 5.1.3. The visualization results are summarized in Fig. 34. The original activation maps from the modified model were in size of 13x13 before resizing, and the corresponding size of FOV for each unit is about the same size as a cancerous nucleus in the patch. Therefore, the high-response regions in the activation maps reflect the distribution of tumor cells very well (more precise compared with the results using the original AlexNet model in section 5.1.2). Besides, we also found the deep learning models can reveal the co-occurrence properties of patterns by exploring the data. In Fig. 35, it shows that the deep learning models not only focused on the characteristics of cancerous nuclei but also noticed the effect of cytoplasmic clearing around those cancerous nuclei. In this experiment, the purpose of feature selection is to understand better how the trained model utilizes these 256 deep features from section 5.1.3.

Method: All 256 features from the section 5.1.3 were partitioned into two groups. One group is to collect the features that can convey clinical insights, which means the features can work
as detectors for specific cells or tissues, like those features collected in Fig. 31 and Fig. 34, reported as “recognizable features” here. On the other hand, the rest of the features that cannot be correspondent with a specific category in pathology belonged to another group and were reported as “unrecognizable features.” Fig. 36 shows an example of the unrecognizable feature. From observations, 151 features (out of all 256 features) were categorized into the group of “recognizable features,” and 43 of them were cell-structure features, such as tumor cells or lymphocytes. These 43 cell-structure features were collected manually in this experiment to further reduce the feature size (from 256 to 43). Besides, another 43 features were randomly selected from the group of “unrecognizable features” as the control group for comparisons.

<table>
<thead>
<tr>
<th>Category</th>
<th>Samples for selected channels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor Cells</td>
<td><img src="image1.png" alt="Images" /></td>
</tr>
<tr>
<td>Lymphocytes</td>
<td><img src="image2.png" alt="Images" /></td>
</tr>
<tr>
<td>Collagen Fibers</td>
<td><img src="image3.png" alt="Images" /></td>
</tr>
<tr>
<td>Others</td>
<td><img src="image4.png" alt="Images" /></td>
</tr>
</tbody>
</table>

Fig. 34. Visualization of the Deep Features Using the Modified Model in Section 5.1.3. The High-Response Regions Reflect the Distribution of A Specific Cell or Tissue.
**Hypothesis:** In manual lesion inspection, the pathologists usually focus on different types of cells and then determine whether those cells are cancerous or not by the morphological properties. Similarly, I argue that if we further reduce the feature size by only selecting the cell-structure features, lesion detection should also be achieved. And the model trained with the cell-structure features is supposed to outperform the model trained with the “unrecognizable features” under the same feature size since the cell-structure features are more useful and important from the pathological perspective.

![Figure 35](image1)

(a) The Learned Filter Targets on Cancerous Nuclei.

(b) The Units Get High Activations on Those Regions of Cytoplasmic Clearing Around Cancerous Nuclei.

Fig. 35. The Deep Learning Models Are Able to Reveal the Co-Occurrence Properties of Patterns That Were Learned from the Training Dataset. (a) The Learned Filter Targets on Cancerous Nuclei. (b) The Units Get High Activations on Those Regions of Cytoplasmic Clearing Around Cancerous Nuclei.
Results and Discussion: In this experiment, the Random Forests classifier was used to have constant comparisons among all scenarios of performance starting from the first experiment. TABLE 5 lists the results in comparisons with the original model, after feature reduction, and after feature selection. In TABLE 5, the feature set of 43 cell-structure features from the group of “recognized features” is denoted as (43), and another feature set of 43 features randomly selected from the group of “unrecognized features” is denoted as (43). After feature selection, the results show that performance decreased for both models, compared with the model trained with all 256 features. And the model trained with the selected 43 cell-structure features outperformed the model trained with the 43 unrecognizable features. Surprisingly, the model trained with the 43 features randomly selected from the group of “unrecognizable features” can still strike the out-sample accuracy to 94% above. It implies that those features which are not intuitive to the specialists may still be useful for machines and discriminative in classification statistically. Accordingly, the top 43 important features ranked by the importance property from Random Forests out of all 256 features were further collected, and the feature set is denoted as (*43) in TABLE 5. As a result, the model trained with the top 43 important features outperformed the model trained with the 43 cell-structure features. Analyzing the members in the feature set of (*43), 33 features were from the group of “recognizable features,” in which 14 features were cell-structure and 19 features were related to collagen fibers or other tissues. And the rest 10 features were from the group of “unrecognizable features.” Fig. 36 is an example of the unrecognizable feature that was discriminative in patch classification. The activation maps in Fig. 36 show the learned filter drives high response to those cytoplasmic parts of the tumor cells near interstitial spaces. These discriminative but unrecognizable features extracted by the deep learning models are worth being
further studied in order to find out the reasonable correlations to pathological knowledge and may facilitate the research of new characteristics in diagnosis.

### TABLE 5

**PERFORMANCE BEFORE AND AFTER FEATURE SELECTION**

<table>
<thead>
<tr>
<th>Model</th>
<th>In-Sample Accuracy</th>
<th>Out-Sample Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNet (9216)</td>
<td>0.999</td>
<td>0.978</td>
</tr>
<tr>
<td>CNN + Random Forests (9216)</td>
<td>1</td>
<td>0.966</td>
</tr>
<tr>
<td>CNN + Random Forests (256)</td>
<td>1</td>
<td>0.978</td>
</tr>
<tr>
<td>CNN + Random Forests (43)</td>
<td>1</td>
<td>0.961</td>
</tr>
<tr>
<td>CNN + Random Forests (43)</td>
<td>1</td>
<td>0.947</td>
</tr>
<tr>
<td>CNN + Random Forests (*43)</td>
<td>1</td>
<td>0.974</td>
</tr>
</tbody>
</table>

![Fig. 36. An Example of the Unrecognizable Feature.](image)
5.2 CAD SYSTEM (Lesion Detection and Lesion Classification)

For lesion classification, the performance in patch-level, region-level, and slide-level through 5-fold cross-validation is provided for evaluation and analysis, as shown in TABLE 6. For 5-fold cross-validation, a set of 109 WSIs was divided into 5 datasets with slides evenly distributed in each category.

<table>
<thead>
<tr>
<th>5-fold cross-validation</th>
<th>Non-carcinoma</th>
<th>Carcinoma</th>
<th>Overall (109 cases)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.822 (20867/25394)</td>
<td>0.925 (52804/57093)</td>
<td>0.893 (73671/82487)</td>
</tr>
<tr>
<td>Patch-level</td>
<td></td>
<td>0.808 (13574/16808)</td>
<td>0.787 (31687/40285)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.893 (958/1073)</td>
<td></td>
</tr>
<tr>
<td>Region-level</td>
<td>0.970 (1373/1416)</td>
<td>0.847 (611/721)</td>
<td>0.804 (283/352)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slide-level</td>
<td>0.881 (37/42)</td>
<td>0.970 (65/67)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.862 (25/29)</td>
<td>0.974 (37/38)</td>
</tr>
</tbody>
</table>

In patch-wise training and testing, a total of 82k patches were sampled from three different labeled regions (Non-Carcinoma, DCIS, and IDC). In 5-fold cross-validation, there were five datasets, and each dataset was treated as a testing set sequentially. If one of the datasets was assigned to be the testing set, the other four datasets were merged together to train a model. As a result, every sampling patch was blindly tested by a trained model. For binary classification, the
models were used to predict whether there exists carcinoma in the input patch or not. For 3-category classification, the patches were categorized into Non-Carcinoma, DCIS, and IDC. In the experiment, the overall accuracy rate in patch-level reaches 89.3% for binary classification and 80.2% for 3-category classification.

In region-wise testing, approximately 2.5k regions were proposed from the phase of lesion detection after screening by the region size. All proposed regions were classified into 3 categories by two separate registered pathologists, and the agreement between them was 0.951 by Cohen’s kappa coefficient. For those regions with different labels, the final ground truth was determined after coming to an agreement. The binary classification accuracy rate in region-level is 93.7%, and the 3-category classification accuracy rate is 91.1%. It is noteworthy that the region-level accuracy was largely improved compared with the accuracy in patch-level, which should be attributed to the design of the hierarchical framework and the scheme of majority voting in the proposed system.

Instead of performing DCIS or IDC detection through WSIs directly, we designed the hierarchical framework that samples patches at different resolutions for lesion detection and classification for two main reasons: (1) The multi-view approach that lesions are observed under different power magnifications is highly accordant to the manual process of slide assessment in practice. The morphological properties are more apparent at higher magnification, while the architectural features are best seen under lower magnification. Accordingly, lesion detection and lesion classification were performed under different sampling views. (2) The scheme of majority voting in region-wise testing can achieve better error tolerance for the CAD system. Since not every sampling patch in testing can present the core properties of the lesion and the patch-wise predicting model is not perfect, applying majority voting for each proposed region can reduce the impacts of those misclassified patches effectively. Therefore, the error tolerance of the proposed system was
improved by reducing the influences of misinterpretations. The region-wise classification results in detecting the structure of lobules are shown in Fig. 37. The lesions of benign epithelial hyperplasia and Usual Ductal Hyperplasia (UDH) were also recognized by the system, as shown in Fig. 38. For malignant cases, Fig. 39 and Fig. 40 show different types of DCIS and IDC respectively that recognized by the system correctly. The results show the multi-view scheme and hierarchical framework can obtain better delineations of lesions and more consistent prediction results in classification.

Finally, the slide assessment can be accomplished by analyzing the lesion types that exist in each WSI. The slide-wise predicting accuracy rate is 93.6% for binary classification and 90.8% for 3-category classification through cross-validation. To further verify the robustness of the proposed CAD system for clinical practice, the total patches from the cross-validation set were used for training, and another 77 WSIs with confirmed slide labels were collected for testing. The prediction results are listed in TABLE 7. accuracy rates in slide-level are 97.4% and 94.8% for binary classification and 3-category classification respectively

<table>
<thead>
<tr>
<th>Testing</th>
<th>Non-carcinoma</th>
<th>Carcinoma</th>
<th>Overall (77 cases)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slide-level</td>
<td></td>
<td>DCIS</td>
<td>IDC</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.953</td>
<td>0.974 (75/77)</td>
</tr>
<tr>
<td></td>
<td>(34/34)</td>
<td>(41/43)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.727</td>
<td>0.969 (31/32)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(8/11)</td>
<td></td>
</tr>
</tbody>
</table>

TABLE 7

THE SLIDE-WISE PREDICTING PERFORMANCE ON THE TESTING SET
Fig. 37. Structure of Lobules. (a) Normal Lobules. (b) Lobular Hyperplasia.
Fig. 38. Benign Lesions. (a) Benign Epithelial Hyperplasia. (b) Usual Ductal Hyperplasia (UDH).
Fig. 39. Lesions of DCIS. (a) Low-Grade DCIS. (b) High-Grade DCIS.
Fig. 40. Lesions of IDC. (a) IDC with Cord Growth Patterns. (b) IDC with Peritumoral Retraction Spaces.
For all 186 cases in the experiment, there are four false negative cases that were misclassified by the system. Two of them are the cases with low-grade DCIS, and another two cases contain the lesions of low-grade IDC. Fig. 41 shows the examples of the misclassified lesions that caused false negatives in diagnosis. To tackle the issue of misinterpretations on cases with low-grade DCIS and low-grade IDC, more samples of these lesion types should be collected to enhance the training models regardless of their incidence rates. Also, the category of Atypical Ductal Hyperplasia (ADH) may need to be considered in lesion classification. Even though ADH shares many similarities with low-grade DCIS, many experts suggest it can be separated from low-grade DCIS by the quantitative criteria (2 mm in aggregate) [40, 41], and that can be done by the CAD system easily.

The common misclassified lesions that caused false alarms are shown in Fig. 42. In Fig. 42-(a), invasive-growth-like patterns are present in the lesions of lobular cancerization (cancerization of lobules, COL) surrounded by lymphocytes and collagen that led to the misinterpretation. Because COL and IDC are morphologically different in detail, more patch samples from COL should be collected for training to capture the discriminative features. The area of cautery/fulguration artifacts [42, 43] is demonstrated in Fig. 42-(b). Since no training patches were sampled from these areas with cautery artifacts, these artifacts may not be recognized by the CAD system and become a potential cause of misinterpretation. Accordingly, the impacts of artifacts should also be considered in the development of CAD systems. In this case, those regions with cautery artifacts should have been excluded in the phase of region proposal by the detection model since the cautery artifacts contained no appreciable cellular architecture but heat-induced coagula [44].
Fig. 41. False Negative Cases with Lesions of (a) Low-Grade DCIS and (b) Low-Grade IDC.
Fig. 42. False Alarms Happened Due to (a) the Invasive-Growth-Like Patterns in the Lesions of Lobular Cancerization and (b) Deformed Cells and Tissues from Cautery Artifacts.
5.3 VISUAL INTERPRETABILITY FOR LESION CLASSIFICATION

Recently, deep learning frameworks have been widely adopted to perform high-stake tasks in many medical applications. However, without providing a robust interpretation for system’s mechanism, a lack of validity of the system will become a problem. In this dissertation, the highly discriminant features for each category and their corresponding matching patterns are presented to provide the visual interpretability for the proposed CAD system.

In the pre-trained model of ResNet50, a total of 2048 features were used in classification. A patch with high activation on a unit implies a specific pattern being detected in the input image, and its corresponding activation map shows the pattern’s spatial distribution. In the experiment, the most distinctive feature for the category of Non-Carcinoma took place on unit No.1134, denoted as feature No.1134 here. In the binary classification using decision stump, the feature No.1134 was used to distinguish the Non-Carcinoma patches from the patches with carcinoma lesions. Within deep feature No.1134, the activations between Non-Carcinoma and Carcinoma are well-separated, and its in-sample classification accuracy rate reached 83.0% by the classifier of decision stump (simply using one feature), as shown in Fig. 43.

Fig. 43. Activations between Non-Carcinoma and Carcinoma Were Well-Separated Within Feature No.1134.
5.3.1 SUPPORTIVE FEATURES IN NON-CARCINOMA

The patches with high activation on unit No.1134 and the corresponding activation maps are shown in Fig. 44. From observations, the feature No.1134 looks for lobules in the input patch. The learned weights of feature No.1134 for each category are listed in TABLE 8. The positive weight $W_0$ implies that the detection of lobules is advantageous to classify the patch to the category of Non-Carcinoma. On the other hand, if the patch has a high activation on unit No.1134, it means that lobules were detected in the patch and its negative weights $W_1$ and $W_2$ tended to avoid the system from classifying the patch to the categories of DCIS and IDC.

![Fig. 44. The Supportive Features for the Category of Non-Carcinoma: Lobules.](image)

Except for lobules, the deep features of wavy collagen fibers, as shown in Fig. 45, and the morphological characteristics of benign epithelial hyperplasia, as shown in Fig. 46, are also exclusive to the category of Non-Carcinoma. Similarly, their corresponding trained weights listed in TABLE 8 show positive $W_0$ and negative $W_1$, $W_2$ as the supportive features for the category of Non-Carcinoma.
Fig. 45. The Supportive Features for the Category of Non-Carcinoma: Wavy Collagen Fibers.

Fig. 46. The Supportive Features for the Category of Non-Carcinoma: Benign Epithelial Hyperplasia.
TABLE 8
FEATURE WEIGHTS FOR EACH CLASS LEARNED FROM THE TRAINING DATA

(POSITIVE $W_0$)

<table>
<thead>
<tr>
<th>Feature Number</th>
<th>$W_0$ (Non-Carcinoma)</th>
<th>$W_1$ (DCIS)</th>
<th>$W_2$ (IDC)</th>
<th>Matching Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. 1134</td>
<td>0.081</td>
<td>0.051</td>
<td>0.045</td>
<td>lobules</td>
</tr>
<tr>
<td>No. 1833</td>
<td>0.053</td>
<td>-0.018</td>
<td>0.04</td>
<td>wavy collagen fibers</td>
</tr>
<tr>
<td>No. 685</td>
<td>0.048</td>
<td>0.037</td>
<td>0.014</td>
<td>hyperplasia</td>
</tr>
</tbody>
</table>

In the deep learning model, the final classification decision was made by all 2048 features with trained weights taken into consideration. Fig. 47 shows two examples of patches from Non-Carcinoma using the CAM analysis to reflect the influential patterns that contribute to the classification results. In Fig. 47-(a), the feature of wavy collagen fibers may not be so influential in the decision when the structure of lobules is prominent. But the wavy pattern of collagen fibers becomes very crucial especially when the lesions of benign epithelial hyperplasia partially appeared in the input patch, as shown in Fig. 47-(b). In pathology, Tumor-Associated Collagen Signatures (TACS) are used to classify the distinctive patterns of collagen reorganization that occur during breast cancer progression. For normal and benign (TACS-1) cases, collagen is characterized to appear wavy and curly [45, 46]. The results show that deep learning model focused on not only the morphological characteristics of cells but also the co-occurrence properties from interstitial portions around lesions, mimicking an experienced pathologist.
5.3.2 SUPPORTIVE FEATURES IN DCIS

For the category of DCIS, the most distinctive feature took place on unit No. 1815, and it detects the distribution of comedo-like necrosis in the sampling patches, as shown in Fig. 48. With the classifier of decision stump, the in-sample classification accuracy rate achieves 84.3% in classifying DCSI patches or the others. Similarly, the learned weights for feature No. 1815 in TABLE 9 show that it is a discriminative feature for the category of DCIS (positive $W_1$ and negative $W_0, W_2$). In Fig. 49, the cribriform growth pattern for the lesions of cribriform DCIS was captured by the deep learning model. Besides, the detection of the rounded configuration of lesions shown in Fig. 50 can also contribute to the category of DCIS in classification. From observations, in general, the system recognizes the lesions of DCIS by checking if there exist cribriform growth.
patterns, intraluminal necrosis, or the solid nests with rounded configuration (different from lesions of microinvasion or IDC). The approach of the deep learning model in recognizing lesions of DCIS agrees with the clinical rules in histopathology.

Fig. 48. The Supportive Features for the Category of DCIS: Comedo-Like Necrosis.

Fig. 49. The Supportive Features for the Category of DCIS: Cribriform Growth Pattern.
Fig. 50. The Supportive Features for the Category of DCIS: Rounded Configuration of Lesions.

Different from the deep features mentioned in section 5.3.1 that the positive weights only happened on $W_0$ for the category of Non-Carcinoma and negative weights for other classes. For some supportive features of DICS listed in TABLE 9, positive weights were not exclusive to $W_i$. Even though the features like cribriform pattern and rounded configuration of lesions are very useful in detecting DCIS lesions, the statistical results show that similar patterns could also be observed in the lesions of Non-Carcinoma. Fig. 51 shows the lesions of UDH can exhibit peripheral cribriform patterns and rounded configuration.

TABLE 9
FEATURE WEIGHTS FOR EACH CLASS LEARNED FROM THE TRAINING DATA
(POSITIVE $W_i$)

<table>
<thead>
<tr>
<th>Feature Number</th>
<th>$W_0$ (Non-Carcinoma)</th>
<th>$W_1$ (DCIS)</th>
<th>$W_2$ (IDC)</th>
<th>Matching Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. 1815</td>
<td>-0.029</td>
<td><strong>0.087</strong></td>
<td>-0.05</td>
<td>necrosis</td>
</tr>
<tr>
<td>No. 1956</td>
<td><strong>0.008</strong></td>
<td><strong>0.016</strong></td>
<td>-0.046</td>
<td>cribriform pattern</td>
</tr>
<tr>
<td>No. 1402</td>
<td><strong>0.006</strong></td>
<td><strong>0.034</strong></td>
<td>-0.021</td>
<td>rounded configuration</td>
</tr>
<tr>
<td>No. 1819</td>
<td>-0.0408</td>
<td><strong>0.0487</strong></td>
<td><strong>0.0028</strong></td>
<td>solid nests</td>
</tr>
</tbody>
</table>
Even though the different categories of lesions may share the common deep features, the mechanism of the deep learning approaches can still distinguish different lesions and achieve lesion classification. In fact, the cribriform patterns in UDH and DCIS are morphologically different in detail, and it can be noticeable by the deep learning models. For example, the cribriform pattern in UDH reflects relatively weaker response than the one in DCIS within feature No. 1956, which constrains the likelihood of predicting a UDH lesion with cribriform patterns to the category of DCIS. On the other hand, for a cribriform lesion in DCIS, it is unlikely for the system to classify the lesion to Non-Carcinoma over DCIS since $W_1$ is larger than $W_0$ in feature No. 1956. More importantly, the deep learning models do not solely rely on one feature in decision making; instead, all features were taken into consideration in classification. The CAM analysis in Fig. 52 shows that the deep learning model recognizes the DCIS lesions by checking the co-existence of (a) solid nests and cribriform patterns or (b) solid nests and rounded configuration. If
the solid nests were detected in the input patch, the negative weight $W_0$ of feature No. 1819 tended to avoid the system from classifying the patch to the category of Non-Carcinoma.

![Image](image_url1)

(a)

![Image](image_url2)

(b)

Fig. 52. The CAM Analysis for the Category of DCIS.
5.3.3 SUPPORTIVE FEATURES IN IDC

The most distinctive feature for IDC is feature No.1261 in the experiment, as shown in Fig. 53, which detected the cord growth patterns in the sampling patches. The in-sample classification accuracy rate is 83.6% in recognizing IDC patches using the cord growth feature by decision stump.

In TABLE 10, the weights of feature No.1261 also indicate the pattern is exclusive to the category of IDC (positive weight only happened on $W_2$). Fig. 54 and Fig. 55 illustrate the detected irregular clusters of tumor cells and solid nests with retraction spaces surrounded, which are also the distinctive features for IDC.

Fig. 56 shows that the co-occurrence property of solid nests and peritumoral retraction spaces is very crucial for the system to distinguish the lesions of IDC from DCIS. Since the weights of feature No.107 shows the solid patterns are supportive features for both DCIS and IDC, the existence of peritumoral retraction spaces becomes very discriminative in classification. In pathology, the study [47] has shown the peritumoral retraction artifact could be used as an
inexpensive but effective invasive cancer marker. Such a phenomenon of tissue shrinkage of malignant tumor cells was also discovered by the deep learning models.

Fig. 54. The Supportive Features for the Category of IDC: Irregular Clusters of Tumor Cells.

Fig. 55. The Supportive Features for the Category of IDC: Peritumoral Retraction Spaces.
TABLE 10
FEATURE WEIGHTS FOR EACH CLASS LEARNED FROM THE TRAINING DATA
(POSITIVE $W_2$)

<table>
<thead>
<tr>
<th>Feature Number</th>
<th>$W_0$ (Non-Carcinoma)</th>
<th>$W_1$ (DCIS)</th>
<th>$W_2$ (IDC)</th>
<th>Matching Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. 1261</td>
<td>-0.013</td>
<td>-0.02</td>
<td><strong>0.031</strong></td>
<td>cord growth pattern</td>
</tr>
<tr>
<td>No. 1344</td>
<td>-0.023</td>
<td>-0.029</td>
<td><strong>0.085</strong></td>
<td>irregular clusters of tumor cells</td>
</tr>
<tr>
<td>No. 1180</td>
<td>-0.025</td>
<td>-0.023</td>
<td><strong>0.026</strong></td>
<td>peritumoral retraction spaces</td>
</tr>
<tr>
<td>No. 107</td>
<td>-0.043</td>
<td><strong>0.022</strong></td>
<td><strong>0.046</strong></td>
<td>solid nests</td>
</tr>
</tbody>
</table>

Fig. 56. The CAM Analysis for the Category of IDC.
The feature analysis has explained the mechanism of how a trained DCNN model recognizes different types of lesions, which follows the clinical rules in diagnosis. In lesion classification, the proposed CAD system not only focuses on the characteristics of cells’ alignment but also learns the co-occurrence properties, such as the wavy collagen fibers in Non-Carcinoma, intraluminal necrosis in DCIS, and the peritumoral retraction spaces in IDC. The mechanism of the proposed system using deep learning approach is recognized by pathologists; therefore, the reliability and validity of the system can be established.
CHAPTER 6
CONCLUSION AND FUTURE WORK

In this dissertation, a CAD system using deep learning approach are proposed, and the validity of the proposed system is verified by its performance and visual interpretability. To fill the clinical requirements, the system performance is evaluated in patch-level, lesion-level, and slide-level through cross-validation rigorously. The multi-view scheme and hierarchical framework were designed in the system to achieve better results on lesion-wise and slide-wise analysis. Moreover, the mechanism of the deep learning approach in both lesion detection and lesion classification is provided with the explanations from the pathological perspective. The results of visual interpretability are accordant with the clinical insights and recognized by experienced pathologists, which enhances the reliability and validity of the proposed system.

For future work, the false positive reduction can be done by decreasing the interference from the cautery artifacts in diagnosis. In addition, as it is shown in Fig. 57, the proposed CAD system will be more versatile for the diagnosis of proliferative lesions if there are (1) further detection of microinvasive carcinoma for cases diagnosed as DCIS and (2) the category of ADH in lesion classification to avoid missed inspections for those high-risk precursor lesions and low-grade forms of DCIS.
Fig. 57. The Well-Developed Framework for the Diagnosis of Proliferative Lesions.
REFERENCES


VITA

Wei-Wen Hsu

Electrical and Computer Engineering Department
Old Dominion University
Norfolk, VA 23529

Education

- **Old Dominion University** Norfolk, VA 08/2012 ~ 08/2019
  *Ph.D. in Electrical and Computer Engineering*
  Dissertation: Using Feature Extraction from Deep Convolutional Neural Networks for Pathological Image Analysis and Its Visual Interpretability
  Advisor: Dr. Chung-Hao Chen

- **National Taiwan University** Taipei, Taiwan 09/2009 ~ 06/2011
  *Master of Science in Computer Science and Information Engineering*
  Thesis: Tumor Detection for Automated Whole Breast Ultrasound Image
  Advisor: Dr. Ruey-Feng Chang

- **National Central University** Jhongli, Taiwan 09/2003 ~ 06/2008
  *Bachelor of Science in Electrical Engineering*

Publications