

# Computer-Aided Diagnosis Systems in the Classification of Neuroblastoma Histological Images

A Thesis Submitted for the Degree of  
Doctor of Philosophy

By

*Soheila Gheisari*

in

School of Computer Science  
UNIVERSITY OF TECHNOLOGY SYDNEY  
AUSTRALIA

2019

© Copyright by Soheila Gheisari, 2019

UNIVERSITY OF TECHNOLOGY SYDNEY

SCHOOL OF COMPUTER SCIENCE

The undersigned hereby certify that they have read this thesis entitled “**Computer-Aided Diagnosis Systems in the Classification of Neuroblastoma Histological Images**” by **Soheila Gheisari** and that in their opinions it is fully adequate, in scope and in quality, as a thesis for the degree of **Doctor of Philosophy**.

Dated: December 2019

Principal Supervisor:      Prof. Paul J. Kennedy

## CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Soheila Gheisari declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the *school of Computer Science, faculty of Engineering and Information Technology* at the University of Technology Sydney. This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution. This research is supported by the Australian Government research Training Program.

Production Note:

Signature removed prior to publication.

---

Signature of Author

# Acknowledgements

I am greatly indebted to my supervisor, Prof. Paul J. Kennedy for all the efforts, advice, help, and patient guidance that he has provided during my PhD. He is such a nice, generous, helpful and kindhearted person. I feel really happy, comfortable and unconstrained with him during my PhD study. I owe my research achievements to his experienced supervision. Many thanks are also due to my co-supervisor A/P Daniel R. Catchpoole from the Tumour Bank at the Children's Hospital for his support, advice and numerous conversation with him. I would like to express grateful acknowledgement to my co-supervisor Dr. Amanda Charlton from department of histopathology in Auckland City Hospital. She supported me in collecting the dataset of neuroblastoma. She provided the project with necessary clinical resources. I appreciate her attempts in providing invaluable advice and necessary clinical resources. I would also like to thank Zsombor Melegh and Elise Gradhand from department of pathology in Southmead Hospital in Bristol. They supported me in providing neuroblastoma images during this project.

I would like to thank my parents that I could never have done this without their faith, support, and constant prayers. I also thank my dear husband Farshid, love of my life, for his support, patience and managing the life. I also want to thank my five years old daughter Shayna, who has been the light of my life. She has given me the extra motivation to get things done. You will be my greatest achievement.

Last but not least, the financial assistance of an Australian Postgraduate Award scholarship is gratefully appreciated.



*To my father and my mother, who supported me emotionally. I always knew that you believed in me and wanted the best for me. To my husband Farshid, that his understanding, assistance and company during my study enabled the hours of research for me. To my little daughter Shayna you will always be my blessing from God. This thesis could not have been completed without your supports. I love you all.*

# Table of Contents

<b>Table of Contents</b>	<b>vii</b>
<b>List of Tables</b>	<b>viii</b>
<b>List of Figures</b>	<b>xii</b>
<b>Abstract</b>	<b>1</b>
<b>Table of Symbols</b>	<b>5</b>
<b>1 Introduction</b>	<b>8</b>
1.1 Computer-Aided Diagnosis Systems in Pathology . . . . .	9
1.2 Research Significance . . . . .	10
1.3 Research Challenges . . . . .	11
1.4 Research Hypotheses . . . . .	14
1.5 Research Objectives and Contributions . . . . .	15
1.6 Research Structure . . . . .	19
<b>2 Literature Review</b>	<b>21</b>
2.1 Biological Domain . . . . .	21
2.1.1 World Health Organisation Peripheral Neuroblastic Tumour Classification . . . . .	24
2.2 Computer-based Approaches . . . . .	29
2.2.1 Segmentation Based Methods for Classification of Malignant Tumours . . . . .	32
2.2.2 Low-level Feature Extraction . . . . .	35
2.2.3 High-level Feature Extraction Methods . . . . .	42
2.3 Classification of Neuroblastoma Histological Images . . . . .	54
2.4 Classifiers Used in This Thesis . . . . .	55

2.4.1	Support Vector Machine . . . . .	56
2.4.2	k-Nearest Neighbour . . . . .	57
2.4.3	Bag of words model in computer vision . . . . .	57
2.5	Clustering Approaches Used in This Thesis . . . . .	60
2.5.1	k-means clustering . . . . .	60
2.6	Metrics Used in This Thesis . . . . .	61
2.7	Cross-Validation . . . . .	63
2.8	Parameter Tuning Methods Used in This Thesis . . . . .	63
2.8.1	Greedy Method . . . . .	63
2.8.2	Wrapper Approach . . . . .	64
2.9	Research Gaps . . . . .	65
<b>3</b>	<b>Models of Computer-Aided Diagnosis Systems Based on Low-level Features</b>	<b>67</b>
3.1	Data Collection and Cropping . . . . .	68
3.2	Patched Completed Local Binary Pattern . . . . .	72
3.2.1	Methodology . . . . .	73
3.2.2	Experiment Results . . . . .	81
3.3	Scale Invariant Feature Transform . . . . .	86
3.3.1	Methodology . . . . .	87
3.3.2	Experiment Results . . . . .	91
3.3.3	Sensitivity to Dataset . . . . .	97
3.4	Discussion, Contribution and Conclusion . . . . .	103
<b>4</b>	<b>Models of Computer-Aided Diagnosis Systems Based on High-level Features</b>	<b>107</b>
4.1	High-level Feature Extraction . . . . .	108
4.1.1	Convolutional Deep Belief Networks . . . . .	110
4.1.2	Convolutional Neural Network . . . . .	123
4.1.3	Sensitivity of Test Samples . . . . .	129
4.1.4	Discussion, Contribution and Conclusion . . . . .	132
<b>5</b>	<b>Conclusion</b>	<b>134</b>
5.1	Discussion . . . . .	139
5.2	Future Research Directions . . . . .	141

# List of Tables

2.1	Histopathological definitions in neuroblastic tumours . . . . .	27
2.2	Different subtypes of neuroblastic tumours (Shimada et al. 1999) . . .	28
2.3	Specification of favourable and unfavourable histology groups (Robson 2001) based on the Shimada classification (Shimada et al. 1999) . . .	30
2.4	Popular kernel functions . . . . .	56
2.5	Confusion Matrix. Abbreviation N (Negative) and P (positive) . . . .	61
3.1	Number of different categories of neuroblastic tumour cropped images and number of patients . . . . .	73
3.2	Average classification accuracy of SVM over neuroblastic tumour dataset using different C values . . . . .	84
3.3	Average classification accuracy of SVM over neuroblastic tumour dataset using different kernel functions . . . . .	85
3.4	Weighted average precision, recall, and F-measure obtained by the pro- posed method (PCLBP) and Spanhol’s method (CLBP). . . . .	87
3.5	Average classification accuracy of the SIFT over ten experiments on the neuroblastic tumour dataset using different values for $\sigma$ , $C_c = 0.03$ , $C_E = 10$ . . . . .	92

3.6	Average classification accuracy of SIFT over the neuroblastic tumour dataset using different values for the contrast threshold ( $C_C$ ), $\sigma = 1.7$ , $C_E = 10$ . . . . .	92
3.7	Average classification accuracy of SIFT over the neuroblastic tumour dataset using different values for the edge threshold ( $C_E$ ), $\sigma = 1.7$ , $C_c = 0.04$ . . . . .	93
3.8	Average classification accuracy of the SIFT over the neuroblastic tumour dataset using bag of features with different values of the codebook size. . . . .	93
3.9	Weighted average precision, recall and F-measure of the SIFT method and other benchmarks. The bold value indicates the best F-measure achieved by combination of SIFT with bag of features and SVM (histogram intersection) . . . . .	95
3.10	A representative confusion matrix of applying SIFT over the neuroblastoma images. Abbreviations: UD (undifferentiated neuroblastoma), PD (poorly-differentiated neuroblastoma), D (differentiating neuroblastoma), GNB (ganglioneuroblastoma) and GN (ganglioneuroma) . .	97
3.11	The actual and predicted classification for dataset from University of Bristol. Abbreviations: UD (undifferentiated neuroblastoma), PD (poorly-differentiated neuroblastoma), D (differentiating neuroblastoma), GNB (ganglioneuroblastoma) and GN (ganglioneuroma) . . . . .	98
3.12	Image distribution of BreakHis dataset by magnification factor and type	102
3.13	Best average recognition rates (%) of the classifiers trained with different descriptors reported in Spanhol et al. (2016) . . . . .	103

3.14	Average recognition rate (%) of SIFT + bag of features and PFTAS over the BreakHis breast cancer dataset in different magnifications .	104
4.1	Average classification accuracy of the CDBN over the constructed dataset using a different number of hidden layers. Number of groups in the first, second, third layers and number of mini batches are 24, 20, 40 and 4, respectively. . . . .	115
4.2	Average classification accuracy of the CDBN over neuroblastic tumour dataset using a different number of groups in the hidden layers. Number of hidden layers and number of mini batches are 3 and 4, respectively.	116
4.3	Average classification accuracy of the CDBN over the constructed dataset for a different number of mini-batches. The number of hidden layers and number of hidden groups in the first, second and third layer are 3, 24, 20 and 40, respectively. . . . .	117
4.4	Average classification accuracy of the CDBN over the neuroblastic tumour dataset with different values of the codebook size. . . . .	118
4.5	Information of different layers of CDBN . . . . .	118
4.6	Weighted average precision, recall, and F-measure of the CDBN and the benchmarks over the constructed dataset consisting of 1043 neuroblastoma images. Bold values mean the best precision, recall and F-measure. . . . .	120
4.7	T-test for a comparison of CLBP and PCLBP with the combination of CDBN with the bag of features and the histogram intersection kernel SVM, with a significance level $\alpha = 0.05$ . . . . .	122

4.8	A representative confusion matrix. Abbreviations: UD (undifferentiated neuroblastoma), PD (poorly-differentiated neuroblastoma), D (differentiating neuroblastoma), GNB (ganglioneuroblastoma) and GN (ganglioneuroma) . . . . .	123
4.9	Average weighted average precision, recall, and F-measure obtained by convolutional neural network and the benchmarks. Bold values mean the best precision, recall and F-measure. . . . .	128
4.10	T-test for the comparison of the combination of CDBN with the bag of features and SVM histogram intersection kernel with VGG-16, VGG-19, and AlexNet, significance level $\alpha = 0.05$ . . . . .	129
4.11	Weighted average precision, recall, and F-measure of the convolutional deep belief network over randomly selected subimages in 138 TMA cores	132

# List of Figures

1.1	An example of high intra-class variation in the shape and intensity of differentiating neuroblast cells, H&E images with scale $80 \times 80\mu\text{m}$ . . .	12
1.2	High intra-class variation in the size of the neuroblast cell, H&E images with scale $80 \times 80\mu\text{m}$ . . . . .	13
1.3	Thesis RQs and contribution structure . . . . .	16
2.1	Sympathetic nervous system shown in yellow (Children’s-Oncology-Group 2016) . . . . .	22
2.2	Differentiation of neural tumour. Differentiation increases from Neuroblastoma (least differentiated) to ganglioneuroma (most differentiated). . . . .	23
2.3	Classification tree diagram of NT based on the Shimada classification (Shimada et al. 1999), where UH indicates unfavourable histology and FH indicates for favourable histology. Histopathologists place the patient’s tumour into the favourable and unfavourable groups. . . . .	25
2.4	Blue regions are cellular regions and pink regions are neuropil regions in a H&E stained histological image . . . . .	31



2.5	The scheme of keypoint detecting. Each point is compared to its eight neighbours in the current image (middle image) and nine neighbours in the scale above (top image) and below (down image). . . . .	39
2.6	Orientation calculation: (a) orientations for 64 neighbour pixels around the keypoint (b) orientation histogram . . . . .	42
2.7	Keypoint descriptor: a) gradient magnitude and orientation in local neighbourhood of keypoint. b) orientation corrections and spatial coordinate transformations. c) SIFT descriptor establishment . . . . .	43
2.8	Overall framework of the Convolutional Neural Network (Lecun et al. 1998). It consists of a convolutional layer, subsampling layer and fully-connected layer. . . . .	45
2.9	An example RBM with three visible units and two hidden units . . .	51
2.10	Construction of a Convolutional Restricted Boltzmann Machine (CRBM) with probabilistic max-pooling, based on Lee et al. (2011) . . . . .	53
2.11	Pyramid bag of features. At level 0, decomposition consists of just a single cell and the representation is equivalent to a standard bag of features. The image is subdivided into four and 16 quadrants, extracting four and 16 feature histograms in level 1 and 2, respectively (Lazebnik et al. 2006) . . . . .	58
3.1	A sample of tissue microarray (TMA) slide . . . . .	69
3.2	A sample of core from a single tumour . . . . .	70

3.3	Representative images in neuroblastic tumour categories: (a) undifferentiated neuroblastoma, (b) poorly-differentiated neuroblastoma, (c) differentiating neuroblastoma, (d) ganglioneuroblastoma, and (e) ganglioneuroma. . . . .	71
3.4	Quantitative relative and actual size of tissue cores and cropped images	72
3.5	The scheme of the proposed method which consists of four stages: 1) patched images, 2) sign and magnitude binary patterns, 3) histogram of sign and magnitude binary patterns and 4) classification. . . . .	75
3.6	Coordinates in the patched image. Here, this research assumes $W = N/2$ for better visualisation. . . . .	77
3.7	An 8-neighbourhood around $(X_0^{pq}, Y_0^{pq})$ . . . . .	78
3.8	An example of the computed Magnitude Binary Pattern (MBP) and Sign Binary Pattern (SBP): (a) original image, (b) MBP, and (c) SBP. We assume $W = 150$ pixels in this image for better visualization. However, the real $W$ is 60 pixels. . . . .	79
3.9	Concatenation of patches' histograms: (a) patched image, (b) histogram of patches, and (c) concatenation of patches' histograms. . . .	81
3.10	Division of dataset with 1043 neuroblastoma images . . . . .	82
3.11	Accuracy of $k$ -NN classifier versus patch width ( $W$ ) and $k$ in parameter tuning of $k$ -NN classifier . . . . .	83
3.12	Comparison between the proposed method (PCLBP) and Spanhol's method (CLBP) . . . . .	86
3.13	The scheme of the combination of SIFT with feature encoding . . . .	88

3.14	Scheme of the feature encoding block: a) selected keypoints by SIFT b) 128-D extracted features from keypoints c) clustering the 128-D extracted features d) construction of the codebook using clustering e) histogram of codewords in codebook . . . . .	90
3.15	Comparison between SIFT + the bag of feature approach with CLBP and PCLBP . . . . .	96
3.16	a) Whole tissue section 4906 with actual classification PD b) Pre- dicted classes for ten randomly cropped images from the whole tis- sue section. Abbreviations: D = differentiating neuroblastoma, PD = poorly-differentiated neuroblastoma . . . . .	100
3.17	a) Whole tissue section 4909 with actual classification GNB b) Pre- dicted classes for ten randomly cropped images from the whole tissue section. Abbreviations: D = differentiating neuroblastoma, GNB = ganglioneuroblastoma . . . . .	101
4.1	a) Schematic of deep learning network b) High-level feature extrac- tion by deep learning networks. A deep learning network deconstructs the input image into pixels and combine them to reproduce high-level features (Djuric et al. 2017) . . . . .	109
4.2	The overall framework based on the CDBN which consists of whitening, partitioning, CDBN, feature encoding, and classification. The yellow squares show the mini-batches. . . . .	111

4.3	The scheme of the feature encoding block: (1) extracted high-level features by the CDBN algorithm from the test images, where each feature belongs to one test image, (2) clustering the extracted features using the k-means cluster, (3) finding the centroid of each cluster to construct a codebook consisting of codewords (4) visualisation of the constructed codebook consisting of codewords (5) Histogram of codewords which shows the number of each codeword in the codebook . . . . .	113
4.4	The visualization of the weights of CDBN (a) weights between the input layer and first hidden layer (b) weights between the first and second hidden layer (c) weights between the second and third hidden layer . . . . .	119
4.5	Comparison between the proposed algorithm (CDBN + feature encoding) and the benchmarks (CLBP, PCLBP) . . . . .	121
4.6	Examples of actual and predicted images: a) examples of the differentiating type which are identified by the computer as the poorly-differentiated type b) actual poorly-differentiated type . . . . .	124
4.7	Overall framework of the Convolutional Neural Network (CNN). It consists of the convolutional layer, subsampling layer, fully-connected layer and SVM classifier . . . . .	125
4.8	Comparison between CDBN + feature encoding and CLBP, PCLBP, CNN (VGG-16), CNN (VGG-19), and CNN (AlexNet) . . . . .	130
4.9	A sample of a TMA core with randomly selected sub-images. They are likely to include the common artefacts seen in TMA cores. . . . .	131

# Abstract

Neuroblastoma is the most common extracranial solid malignancy in early childhood. Optimal management of neuroblastoma depends on many factors, including histopathological classification. Although histopathological classification by a human histopathologist is considered the gold standard, computers can help to extract many more features, some of which may not be recognisable by the human eye. Neuroblastoma histological images have a complex texture with complicated features which are different from appearance-based features. Computer-aided diagnosis (CAD) systems facilitate the analysis and classification of neuroblastoma histological images which are non-trivial tasks due to the differences in staining, intensity, and instrumentation. This motivates the thesis to work on the classification of neuroblastoma histological images.

In the past, a small number of methods were proposed by previous studies for the classification of neuroblastoma histological images. These methods are based on the geometry and appearance of the different cells. However, there is a high intra-class variation of intensity and size of the neuroblast cells within the same classification group. Therefore, these methods are not applicable to neuroblastoma histological images. This research proposes a solution based on traditional machine learning approaches and deep learning approaches to extract non-appearance-based

features in small regions. This thesis will investigate two research areas of feature extraction: low-level feature extraction and high-level feature extraction. Low-level features are minor details of the image such as lines, curves and edges. However, high-level features are on top of the low-level features to detect object and larger shape in the image. Feature extraction is aggregated with the classifier in this research to classify neuroblastoma histological images into five categories.

This thesis makes four contributions. Contribution 1 is the construction of a dataset comprising neuroblastoma histological images which are labeled by an expert histopathologist. Contribution 2 is the proposal of a local feature extraction method which can extract local features which are robust to high intra-class variations of intensity. Contribution 3 is the extraction of discriminative features which are robust to high intra-class variation of scale of the neuroblast cells within the same class. Contribution 4 is the proposal of deep networks to extract high-level features which are difficult for the human eye to recognise. The performance of all the proposed methods in this research is evaluated on a dataset collected from The Children's Hospital at Westmead, Sydney, Australia. As there was no publicly available dataset in this field, the proposed algorithms were evaluated on the second dataset of neuroblastoma provided by the University of Bristol and the public breast cancer dataset. All the results are compared with state-of-the-art methods. The results indicate the effectiveness of the proposed algorithms.

This is the first time that neuroblastoma histological images have been classified into five subtypes using low-level and high-level features. However, there are limitations in this research. The specificity is not 100% compared with the gold standard.

Moreover, the proposed algorithms are confused in the distinction between poorly-differentiated and differentiating neuroblastoma, a distinction that human pathologists also find difficult in limited fields of view.

## **Publications**

Following is the list of publications resulting from the work in this thesis:

1. Gheisari, S., Catchpoole, D.R., Charlton, A., Kennedy, P.J. (2017), 'Patched Completed Local Binary Pattern is an Effective Method for Neuroblastoma Histological Image Classification' Communications in Computer and Information Science (Springer), pp. 57-71.
2. Gheisari, S., Catchpoole, D.R., Charlton, A., Kennedy, P.J. (2018) 'Computer Aided Classification of Neuroblastoma Histological Images using Scale Invariant Feature Transform with Feature Encoding', Journal of Diagnostics, 8(3), 1-18; doi:10.3390/diagnostics8030056.
3. Gheisari, S., Catchpoole, D.R., Charlton, A., Melegh, Z., Gradhand, E., Kennedy, P.J. (2018) 'Convolutional Deep Belief Network with Feature Encoding for Classification of Neuroblastoma Histological Images', Journal of Pathology Informatics, Vol. 9, No. 1, pp. 1-17, 2018.
4. Gheisari, S., Charlton, A., Catchpoole, D.R., Kennedy, P.J. (2017) 'Computers can Classify Neuroblastic Tumours from Histopathological Images using Machine Learning', Pathology, No. 49, pp. 72-73, 2017.



# Table of Symbols

Symbol	Description
CAD	Computer aided diagnosis
PCLBP	Patched completed local binary pattern
SIFT	Scale-invariant feature transform
CDBN	Convolutional deep belief network
CNN	Convolutional neural network
CT	Computed tomography
CLBP	Completed local binary pattern
SBP	Sign binary pattern
MBP	Magnitude binary pattern
DSBP	Decimal sign binary pattern
DMBP	Decimal magnitude binary pattern
HSBP	Histogram of sign binary patterns in image
HMBP	Histogram of magnitude binary patterns in image
LH	Local histogram
k-NN	k nearest neighbour
SVM	Support vector machine
RBF	Radial basis function

$C_C$	Contrast threshold
$C_E$	Edge threshold
UD	Undifferentiated neuroblastoma
PD	Poorly-differentiated neuroblastoma
D	Differentiating neuroblastoma
GN	Ganglioneuroma
GNB	Ganglioneuroblastoma
QDA	Quadratic Discriminant analysis
GLCM	Gray level co-occurrence matrices
LPQ	Local phase quantization
ORB	Oriented fast rotated brief
PFTAS	Parameter-free threshold adjacency statistics
CDBN	Convolutional deep belief network
CRBM	Convolutional restricted botzmann machine
H	Hidden layer
V	Visible layer
TMA	Tissue microarray
MKI	Mitosis karyorrhexis Index
H & E	Hematoxylin and eosin
$G(x, y)$	Gaussian function
$L(x, y, \sigma)$	Scale space of an image
x and y	Image coordinates
$\sigma$	Width of Gaussian function
DoG	Difference of gaussian

$F_l$	Feature maps of convolutional layer $l$
$y_{i,j}^{\phi,l}$	Output of neuron $(i, j, \phi, l)$ at position $i, j$
$w_{\Delta i, \Delta j}^{\phi, \phi', l}$	Weight of each connection between the neurons in convolutional layer $(i, j, \phi, l)$ and all the neurons in the previous subsampling layer $\phi'$
$b_j$	Hidden unit biases for neuron $j$
$c_i$	Visible unit biases for neuron $i$

# Chapter 1

## Introduction

Neuroblastoma represents 8% of all malignancies in infants and is the most common extracranial solid malignancy in childhood (Pizzo & Poplack 2015). The optimal management of neuroblastoma depends on many factors, one of which is histopathological classification which is performed by medical doctors called pathologists. Pathologists classify neuroblastic tumours by examining thin slices of tissue on a glass slide using an optical microscope, or view digitised whole slide images on a computer screen. Neuroblastic tumours are a heterogeneous group with complex features. Some of the features are complicated and may not be recognisable by the human eye (Yu et al. 2016). A pathologist examines all the tissue with the microscope at low power magnification, and then examines specific regions of the tissue at high power magnification. The classification of the neuroblastoma is based on the integration of the morphological features observed at different magnifications. However, it may be misleading for heterogeneous tumours (Sertel et al. 2009).

A computer-aided diagnosis (CAD) system is a technology that combines elements of artificial intelligence and computer vision with medical image processing. CAD systems have the potential to augment and be complementary to human capability and

may find new relationships which humans cannot detect. Although histopathology study is considered the gold standard for the classification of neuroblastoma histological images, computers could potentially extract features which the human eye may not recognise. The aim of this thesis is to develop novel algorithms for a system to detect the childhood cancer, neuroblastoma. The proposed CAD system is complementary to and may augment human capability to determine an accurate classification for neuroblastoma. The proposed algorithms in this thesis automatically extract many features, some of which may not be recognised by the human eye.

In this thesis, a series of algorithms is proposed for the automated classification of neuroblastoma histological images. As there is no publicly available dataset for the analysis of neuroblastic tumours, a dataset of neuroblastoma histological images is constructed. The patched completed local binary pattern (PCLBP), scale invariant feature transform (SIFT), convolutional deep belief networks (CDBN) and convolutional neural networks (CNN) classify histological images of neuroblastoma into five different classes: undifferentiated neuroblastoma, poorly-differentiated neuroblastoma, differentiating neuroblastoma, ganglioneuroblastoma, and ganglioneuroma.

## 1.1 Computer-Aided Diagnosis Systems in Pathology

Recently, many researchers have become interested in computer-aided diagnosis (CAD) systems due to the improvements in bio-imaging facilities. There are several CAD systems to improve the clinical diagnosis of radiology images such as mammography and computed tomography (CT) (Topol 2019).

Nowadays, because of the huge progress in image processing methods, CAD systems have the potential to play an important role in the analysis and classification of cancer tissue based on histopathology images. The procedure used in CAD systems is to capture digital images from a histology section from glass slides and the subsequent application of machine vision techniques to classify and segregate different tissue types. The histological glass slides can be scanned and converted to digital images by a slide scanner. Generally, digital images consist of millions of colour pixels in numeric form.

Neuroblastoma histological elements have different shapes within the same classification group. Research efforts on the classification of neuroblastoma histological images are divided into two groups: segmentation methods and feature extraction methods. Segmentation methods rely on morphological features such as the size and shape of different cells. Feature extraction methods can identify diagnostic clues that might not be recognised by the human eye. Extracted features are labeled using different classifiers.

## 1.2 Research Significance

Neuroblastoma is the most common extracranial solid malignancy in infants (Pizzo & Poplack 2015). It has the highest proportion of spontaneous and complete regression of all human cancers (Engl 2010). More than 15% of paediatric cancer deaths are the result of neuroblastoma (Park et al. 2008). Pathologists commonly use the World Health Organisation classification, based on the Shimada system (Shimada

et al. 1999) to classify tissue samples as undifferentiated neuroblastoma, differentiating neuroblastoma, poorly-differentiated neuroblastoma, ganglioneuroblastoma, and ganglioneuroma. According to the Shimada system, neuroblastic tumours are classified into five categories helping pathologists to determine the aggressiveness of the tumour.

A difficulty in the classification of neuroblastic tumours is because of their complex texture. There are high intra-class variations in the intensity and size of the neuroblast cells in images which belong to the same class. These characteristics may not be recognisable by the human eye. Therefore, defining features which are not based on the visual features of neuroblastic tumours and gathering a dataset of neuroblastoma histological images allows this research to design and develop algorithms to classify them based on the extracted features.

This thesis proposes CAD systems to assist pathologists with classification decisions. CAD systems are promising tools to facilitate the high throughput classification of neuroblastoma histological images. So, there is an interest in developing approaches for the automated classification of neuroblastoma histological images.

### 1.3 Research Challenges

This thesis addresses four main challenges compared to existing methods:

**Challenge 1** Overall, paediatric tumours are rare and there is not enough data in neuroblastic tumours. As the performance of such CAD systems is strongly dependent on the training set, building accurate CAD systems is a challenging task.

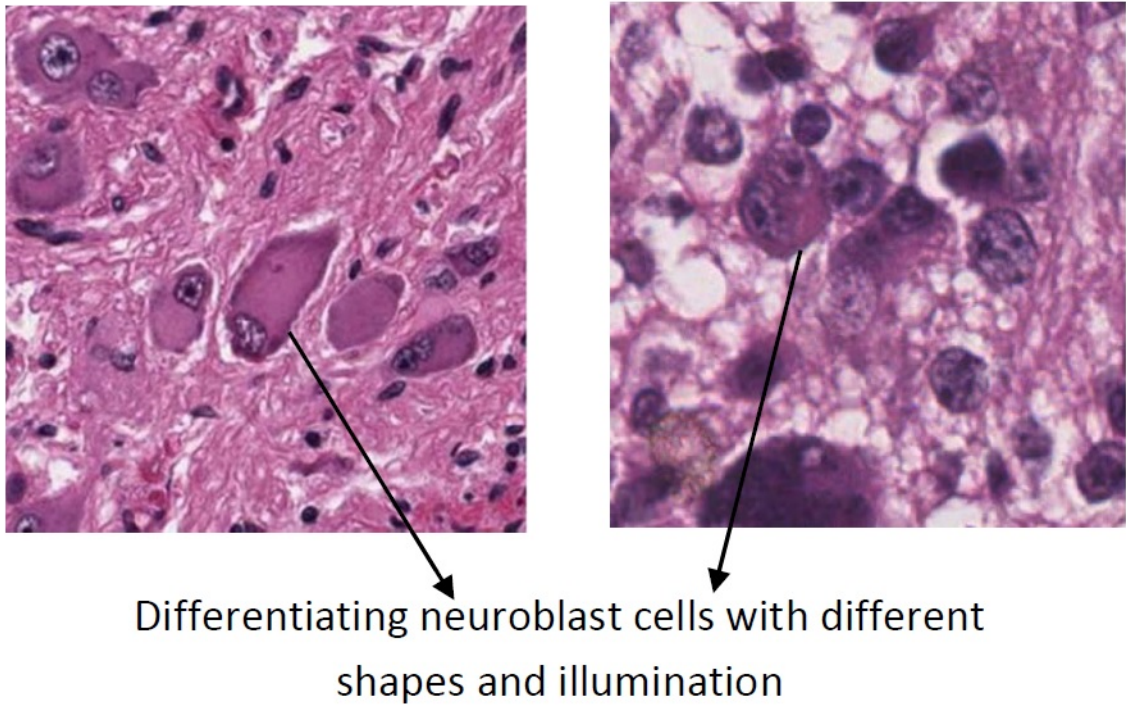
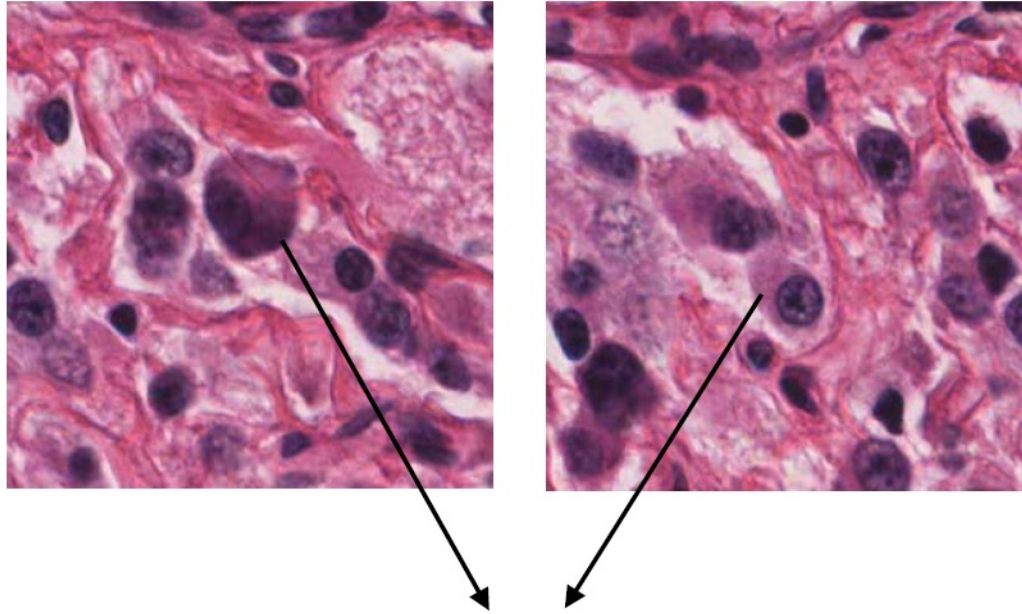


Figure 1.1: An example of high intra-class variation in the shape and intensity of differentiating neuroblast cells, H&E images with scale  $80 \times 80\mu\text{m}$ .

**Challenge 2** There is a high intra-class variation of intensity in neuroblastoma histological images. This means that cells in the same class may have different intensity. Therefore, existing classification methods using segmentation which is based on visual features such as the intensity of the cells struggle to function in this setting. An example of the high intra-class variation of differentiating neuroblastoma is shown in Figure 1.1. Both images are taken in the same magnification and represent the differentiating type while the appearance and intensity of differentiating neuroblast cells are different.





Neuroblast cells with different sizes

Figure 1.2: High intra-class variation in the size of the neuroblast cell, H&E images with scale  $80 \times 80\mu\text{m}$ .

**Challenge 3** There is a high intra-class variation in the size of the nuclei and neuroblast cells in neuroblastoma histological images. Therefore, existing classification methods using segmentation which are based on the size and geometry of the cells are not applicable in this field. A sample of high intra-class variation in the size of the neuroblast cells is shown in Figure 1.2. As can be seen, neuroblast cells which belong to the same class are of different sizes.

**Challenge 4** Histological images of neuroblastoma have a complex texture with complicated features which not only existing CAD systems but also human eye may not recognise. Existing CAD systems extract low-level features using colour, geometry and texture to simulate the visual perception of a human in interpreting the tissue sections. These methods extract the low-level features of different cells in the

neuroblastoma histological images for tissue classification. However, a huge variation in the shape and texture between neuroblast cells from images which belong to the same category reduces the accuracy of the CAD systems using low-level features.

Based on these research issues, the research questions posed in this thesis are as follows:

**RQ1:** How to acquire a dataset of neuroblastic tumours as there is no publicly available dataset in this field?

**RQ2:** How to develop an appropriate low-level feature extraction method to classify neuroblastic tumours with high intra-class variations in the intensity of elements within the same class?

**RQ3:** How to acquire the desired classification method to overcome the high intra-class variations in the size of the nuclei and neuroblast cells in images within the same class?

**RQ4:** How to extract non-appearance-based features from the neuroblastoma histological images?

## 1.4 Research Hypotheses

There are two hypotheses in this research. The first one is that the CAD systems are able to classify neuroblastoma histological images into five subcategories. The proposed CAD systems are able to overcome the issues with the high intra-class variations in the size and intensity of the cells within images in the same class. Neuroblastoma histological images have a complex texture with complicated features which

are different from visual features. Human pathologists look at the image and make the final decision based on visual features such as the colour, shape and size of the cells. However, the proposed CAD systems can assist pathologists in making the final classification decision by extracting high-level features.

The second hypothesis in this thesis is that the classification of neuroblastic tumours based on high-level features is more accurate than low-level ones. The features extracted from histological images are divided into low-level features and high-level features. Low-level features are minor details of the image such as lines, curves and edges. However, high-level features are on top of the low-level features to detect objects and larger shapes in the image.

The experiment results show that the developed CAD system based on high-level features works better than a CAD system based on low-level features. Moreover, CAD systems can assist pathologists by extracting complicated features that may not be recognisable by the human eye in small regions.

## 1.5 Research Objectives and Contributions

By analyzing the main challenges, this thesis will make four main contributions based on the four research questions as shown in Figure 1.3.

1. Constructing a dataset consisting of images from neuroblastic tumours.
2. Extracting local features that are robust to intra-class variations of cells of different intensity in the same class.
3. Extracting discriminative features which are robust to scale variations of cells

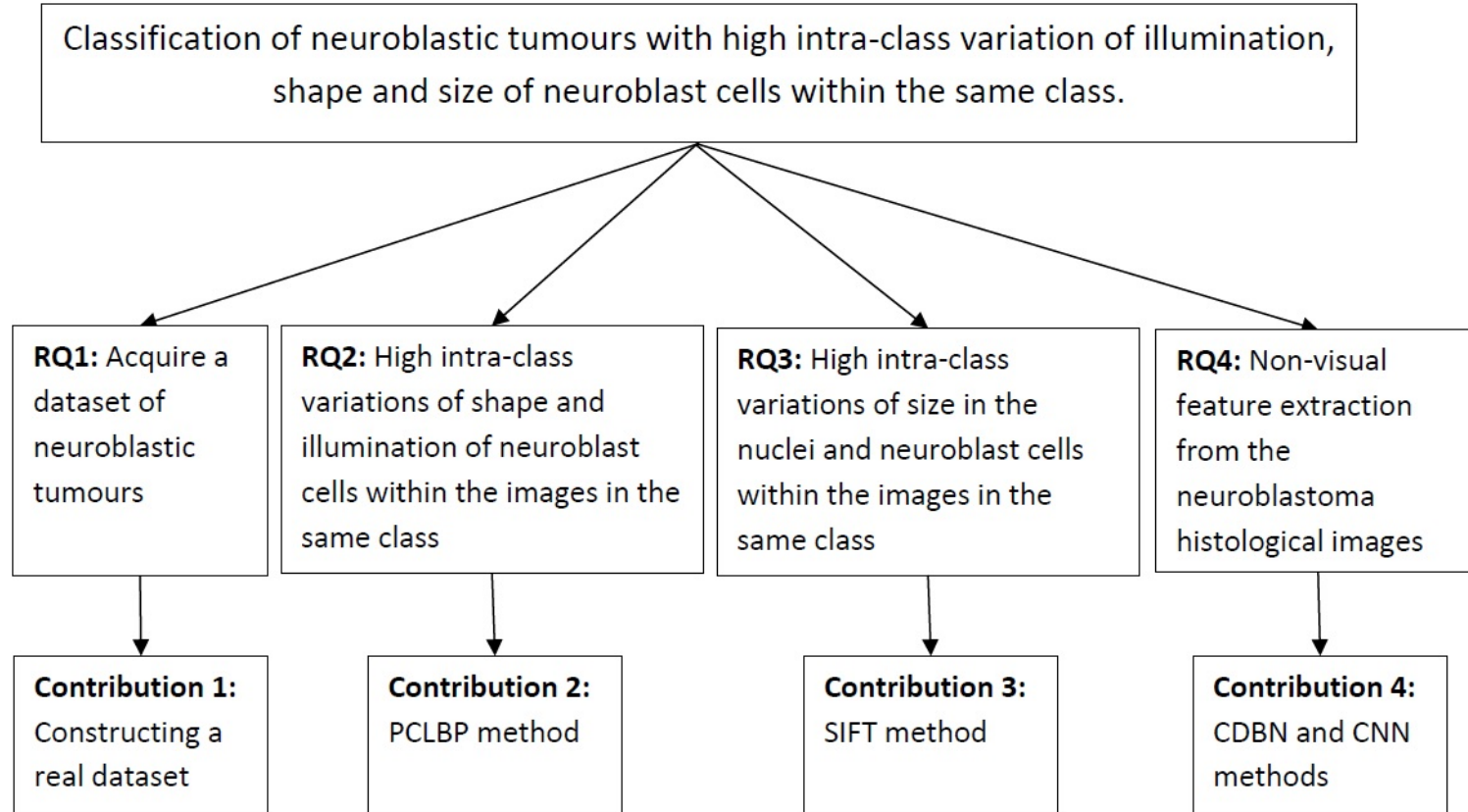


Figure 1.3: Thesis RQs and contribution structure

in the same class.

4. Extracting high-level features which are different from visual features and may be difficult for the human eye to recognise.

These four contributions to knowledge are elaborated as follows:

**Contribution 1: Constructing a dataset consisting of real images from neuroblastic tumours.** This research gathered data from The Tumour Bank of the Kid’s Research Institute at The Children’s Hospital at Westmead, Sydney, Australia. This dataset represents the breadth of the disease and reflects the diagnostic issues faced by pathologists. It is a real representation of patients who present at hospital with this disease which must be determined based on the Shimada classification. All the images in the constructed dataset are labeled by an expert histopathologist [Amanda Charlton]. In Chapter 3, the details of the constructed dataset are explained.

**Contribution 2: Extracting local features that are robust to intra-class variations of cells of different intensity in the same class.** Usually, histological images of neuroblastoma consist of different cells. The cells which belong to the same class have high intra-class variations in intensity. To extract features which are robust to intra-class variations, this research uses the local features as variations in local regions are less than global regions. Local feature extraction is developed by 1) partitioning the input image into equal-sized square patches, 2) extracting features within each patch, 3) concatenating extracted features within patches, and 4) classifying the input images using the extracted features within the whole image.

Chapter 3 details the algorithm. The performance of the proposed approach, patched completed local binary pattern (PCLBP), which is based on the completed local binary pattern (CLBP), is compared with the state-of-the-art method using global features. The results indicate that the proposed CAD system in this thesis outperforms the state-of-the-art method (CLBP) in the classification of neuroblastoma histological images.

**Contribution 3: Extracting discriminative features which are robust to scale variations of cells in the same class.** Neuroblast cells in neuroblastoma histological images have scale variations. This research developed a three-stage algorithm based on the scale-invariant feature transform to classify neuroblastoma based on the features which are robust to scale variations. The proposed algorithm consists of the following stages: 1) extracting scale invariant features, 2) identifying discriminative features and 3) classifying the input images using the extracted features. Chapter 3 describes the algorithm for extracting scale invariant features. The performance of the algorithm is compared with state-of-the-art methods. The proposed algorithm is also evaluated by applying it on a public breast cancer dataset and a second small dataset of neuroblastoma. The results show that the proposed algorithm outperforms state-of-the-art methods such as CLBP and PCLBP on both the neuroblastoma dataset and a benchmark breast cancer dataset.

**Contribution 4: Extracting high-level features which are different from visual features and may be difficult for the human eye to recognise.** Neuroblastoma histological images have complicated features which may not be recognisable by the human eye. Chapter 4 describes the deep learning algorithms, Convolutional

Deep Belief Network (CDBN) and Convolutional Neural Network (CNN), for extracting high-level features in detail. The performance of the algorithm in terms of high-level feature extraction is compared with other state-of-the-art methods such as CLBP and PCLBP. The results of the statistical tests show the proposed CAD systems significantly outperform other CAD systems in the classification of neuroblastic tumours.

## 1.6 Research Structure

This thesis comprises five chapters as follows:

**Chapter 1** provides the research background, research significance, research challenges, research hypotheses and research objectives and contributions.

**Chapter 2** surveys the research relevant to this study. It includes a review of the biological domain of neuroblastic tumours, computer-aided approaches and the classification of neuroblastoma histological images. It also highlights the research gaps in the existing work.

**Chapter 3** describes the dataset constructed of neuroblastoma histological images. This chapter introduces a series of algorithms for classification of the constructed dataset using low-level features which are robust to intensity and scale variations. To validate the performance of the proposed algorithms, the results are evaluated by comparing them with state-of-the-art methods.

**Chapter 4** classifies neuroblastoma histological images by developing a series of algorithms based on deep learning to extract high-level features. Empirical results

indicate the superiority of the proposed approach when compared to state-of-the-art algorithms.

**Chapter 5** completes the thesis by presenting the conclusion. It provides a discussion on the results, lists the benefits and limitations of the contributions, and suggests some future research directions.



# Chapter 2

## Literature Review

This chapter presents the literature related to this project. The biological domain of neuroblastoma histological images, peripheral neuroblastic tumour classification and existing clinical approaches are reviewed in Section 2.1. This chapter reviews the algorithms and CAD systems developed for the automated analysis of images in Section 2.2. The algorithms developed to classify neuroblastoma histological images automatically are described in Section 2.3. Finally, Section 2.9 highlights the research gaps in relation to the development of a CAD system for neuroblastic tumours.

### 2.1 Biological Domain

A tumour is a swelling of a part of the body, generally without inflammation, which is caused by an abnormal growth of tissue. One type of tumour is neuroblastoma which is the most common extracranial solid tumour in children less than five years

old (Pizzo & Poplack 2015). It is a malignancy of peripheral nerve cells of the sympathetic nervous system. The sympathetic nervous system is arranged in groups of nerve cell bodies or a chain of nerve ganglia. These are located parallel to the vertebral column, coloured yellow in Fig. 2.1. Around 40 children are diagnosed with neuroblastoma each year in Australia (*Neuroblastoma Australia* 2018) while this number exceeds to 650 cases in North America (Howlader et al. 2012), (Gurney et al. 1997). There is a 1 in 100,000 chance of a child suffering neuroblastoma (Rare-Cancer-Australia 2018). Ninety percent are younger than five years at diagnosis, with a median age at diagnosis of 19 months (London et al. 2005).

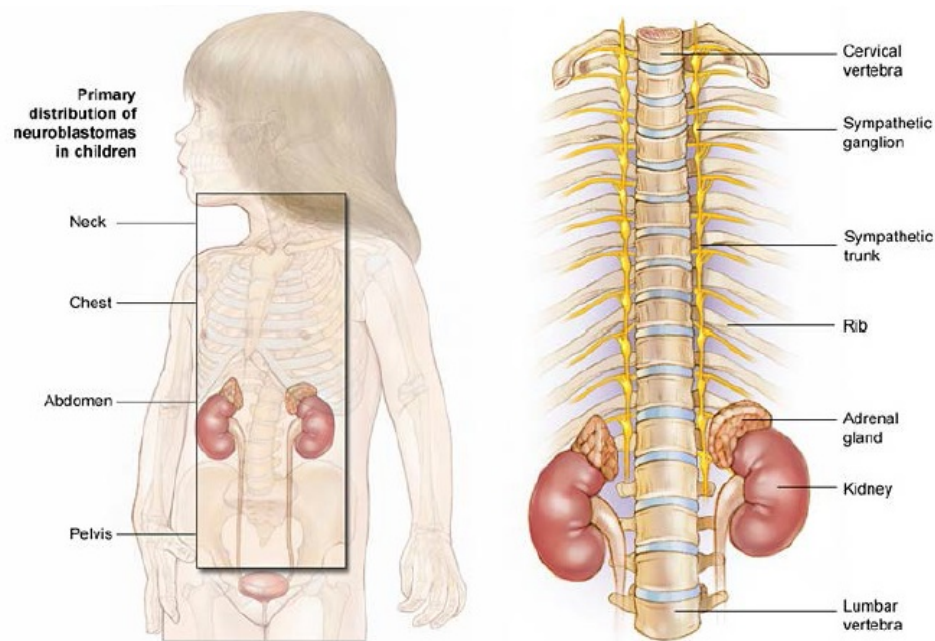


Figure 2.1: Sympathetic nervous system shown in yellow (Children's-Oncology-Group 2016)

Tumours of the nerve cells of the sympathetic nervous system are called neural

tumours and range in differentiation (maturation) from neuroblastoma (least differentiated) to ganglioneuroblastoma (elements of neuroblastoma and ganglioneuroma) to ganglioneuroma (benign, mature) as shown in Figure 2.2.

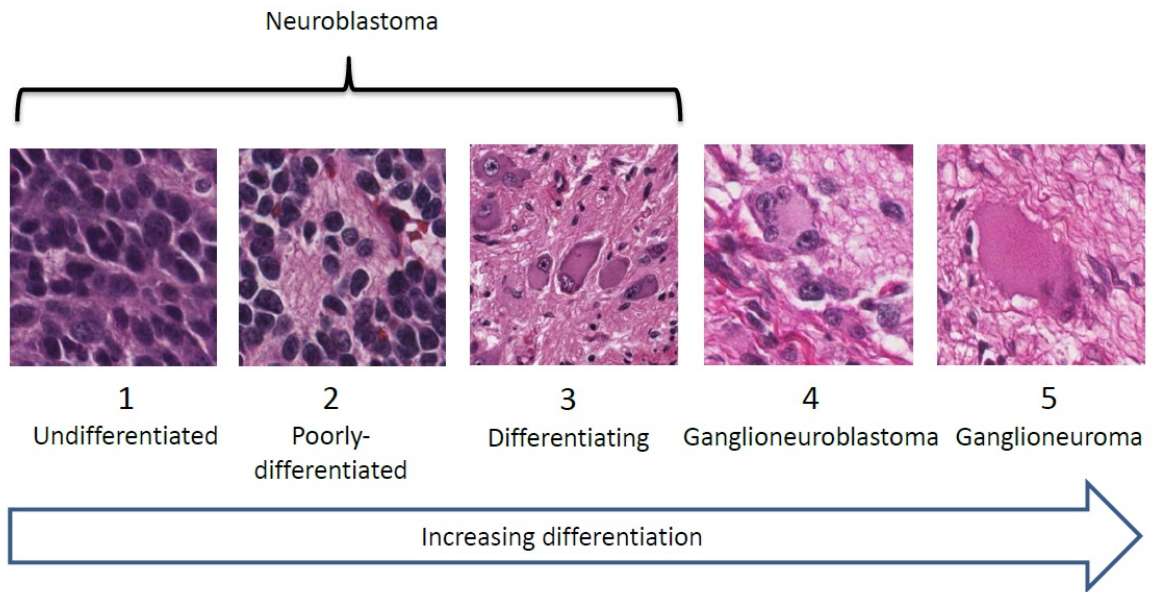


Figure 2.2: Differentiation of neural tumour. Differentiation increases from Neuroblastoma (least differentiated) to ganglioneuroma (most differentiated).

Neuroblasts are primitive, undifferentiated nerve cells present in the developing embryo. During normal development, neuroblasts differentiate (mature) into ganglion cells. Neuroblastoma occurs when the undifferentiated neuroblasts fail to mature into ganglion cells. Neuroblastoma can spontaneously mature from a malignant undifferentiated neuroblastoma to a benign ganglioneuroma, which is a likely explanation of the so-called miracle cures. To predict behaviour and manage treatment, neuroblastoma is divided into risk categories. Variables influencing the natural history of the tumour include age at diagnosis, stage (degree of spread around the body), histopathological classification and status of N-myc amplification. N-myc is a protein

that plays an essential role in brain development and tumours with a high-level of N-myc are undifferentiated or poorly-differentiated (Joshi et al. 1993). Histopathologists determine histopathological classification using tumour subtype, mitotic karyorhectic index (see Table 2.1) and patient age, and the patient's tumour is placed into one of two groups, namely, favourable or unfavourable histology. The classification tree diagram of neuroblastic tumours is shown in Figure 2.3. Distinguishing between favourable histology and unfavourable histology also requires a mitotic karyorhectic index and patient age, neither of which are shown in this diagram. In this research, different subtypes of the neural tumour will be detected. It should be noted that there is no publicly available dataset for neuroblastoma cancer research, hence a new dataset of neuroblastoma tumour tissues had to be collected. This thesis focusses on designing CAD systems for the classification of neuroblastoma histological images. For this purpose, the biological definitions are reviewed in Table 2.1 (Robson 2001). Histopathologists use these definitions to classify neuroblastic tumours.

### **2.1.1 World Health Organisation Peripheral Neuroblastic Tumour Classification**

Histopathological classification is variable with age and the mitosis karyorrhexis index (MKI), the percentage of neuroblastic cells showing mitosis or karyorrhexis. The process of dividing the nucleus of a cell into two nuclei to generate a new cell is called mitosis.

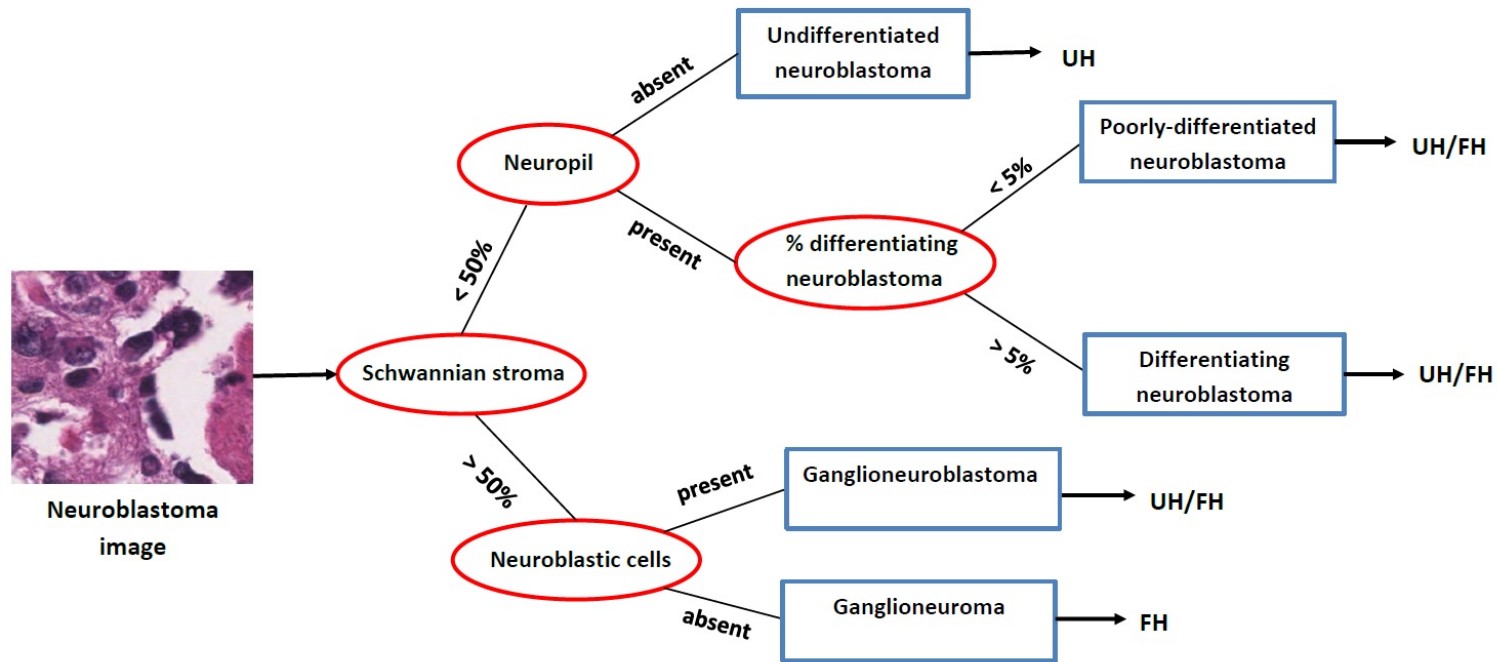


Figure 2.3: Classification tree diagram of NT based on the Shimada classification (Shimada et al. 1999), where UH indicates unfavourable histology and FH indicates for favourable histology. Histopathologists place the patient's tumour into the favourable and unfavourable groups.

The karyorrhexis cells are those which are undergoing cell death and have fragmented nuclei. The MKI is used to assign the patient into either the favourable histology or an unfavourable histology prognostic group. Different subtypes of neural tumours are shown in Table 2.2.

According to the Shimada classification system, pathologists use these definitions to classify neuroblastic tumours into five subcategories:

1. The undifferentiated subtype contains neuroblast cells without identifiable neuropil in the background.
2. The poorly-differentiated subtype is defined as a tumour with a background of recognisable neuropil. 5% or less of the tumour is composed of differentiating neuroblasts.
3. The differentiating subtype is a tumour with usually abundant neuropil in the background. 5% or more of the tumour is composed of differentiating neuroblasts.
4. The ganglioneuroblastoma subtype is composed of differentiating neuroblasts and neuropil. More than 50% of the background is composed of Schwannian stroma.
5. The ganglioneuroma subtype is composed of differentiating neuroblast and fully mature ganglion cells. There is more than 50% Schwannian stroma in the background.

Morphologic histopathological classification is required to place tumours into one of two prognostic groups: favourable or unfavourable histology groups. The specification of favourable and unfavourable histology groups (Robson 2001) is shown

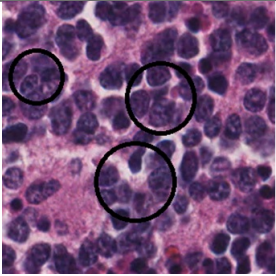
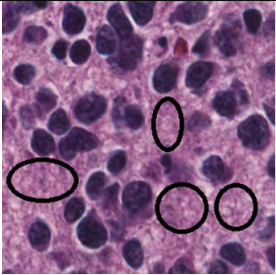
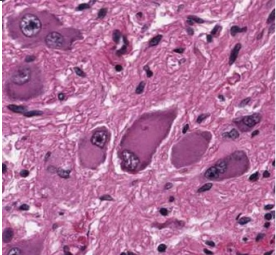
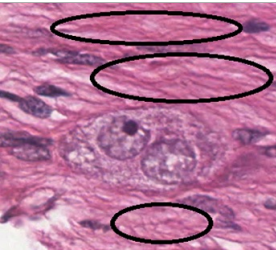
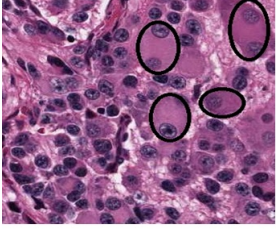
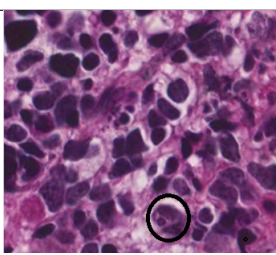
Image	Structure	Definition
	Neuroblast	Cell with a high nuclear-to-cytoplasmic ratio, round nucleus, dark nucleus, inconspicuous nucleolus
	Neuropil	Fine spiderweb-like filamentous cytoplasmic structures that are produced by the neuroblasts
	Differentiating neuroblast	Neuroblast with more cytoplasm, intermediate appearance between neuroblasts and ganglion cell (with a diameter twice or more of the nucleus)
	Schwannian stroma	Wire-like collagenous stroma produced by Schwann cells
	Ganglion cell	Large cell with a large amount of pink cytoplasm, large nucleus with a single nucleolus. The size of the cell is more than twice the size of its nucleus
	MKI (Mitosis Karyorrhexis Index)	Percentage of neuroblastic cells showing mitosis or karyorrhexis. <i>Low</i> $\leq 2\%$ <i>Intermediate</i> = 2-4% <i>High</i> $\geq 4\%$

Table 2.1: Histopathological definitions in neuroblastic tumours



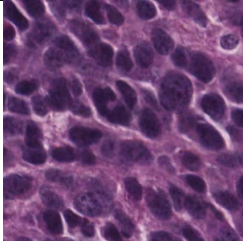
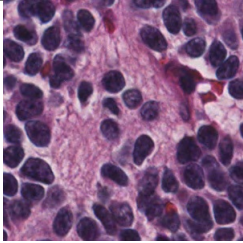
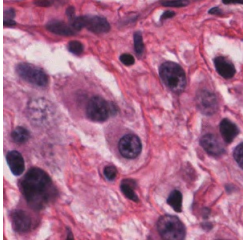
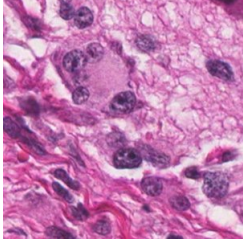
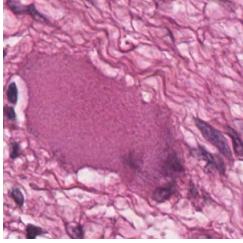
Image	Category	Definition
	Undifferentiated neuroblastoma	Neuroblasts without identifiable neuropil
	Poorly-differentiated neuroblastoma	Neuroblasts with neuropil in the background. 5% or less of the tumour is composed of differentiating neuroblasts
	Differentiating neuroblastoma	Abundant neuropil in the background, $\geq 5\%$ differentiating neuroblasts
	Ganglioneuroblastoma (GNB)	Differentiating neuroblasts, neuropil, $> 50\%$ Schwannian stroma
	Ganglioneuroma (GN)	Differentiating neuroblasts, fully mature ganglion cells, $> 50\%$ Schwannian stroma

Table 2.2: Different subtypes of neuroblastic tumours (Shimada et al. 1999)



in Table 2.3. International Neuroblastoma Pathology Classification distinguishes favourable and unfavourable histology based on age and MKI index (Sano et al. 2006). The combination of tumour subtype, MKI and age put the patient in either the favourable or the unfavourable group.

## 2.2 Computer-based Approaches

Each tumour has a specific classification scheme which allows pathologists to determine the degree of differentiation and classification of the tumour and the chance of survival for patients. Pathologists determine the classification by a combination of histological regions and histological structures observed through a microscope. The important histological regions are cellular regions and neuropil regions. In an H&E stained tissue slide, neuropil regions are stained pink and cellular regions are stained blue (as shown in Figure 2.4)

The histological structures within the histological regions such as neuroblast cells and ganglion cells are described in Table 2.1 (Robson 2001). Pathologists use an optical microscope and classify neuroblastic tumours by examining thin slices of tissue on a glass slide using the Shimada classification system (Shimada et al. 1999). Based on the Shimada classification system the tissue samples are classified as undifferentiated neuroblastoma, poorly-differentiated neuroblastoma, differentiating neuroblastoma, ganglioneuroblastoma and ganglioneuroma. Neuroblastic tumours are an heterogeneous group with complex features. A pathologist examines all the tissue with the microscope at low power magnification, and only examines specific regions of the tissue at high power magnification. The classification of the neuroblastoma is based on

Item	Favourable histology	Unfavourable histology
1	Age < 1.5 years old: neuroblastoma (Schwannian stroma poor), poorly-differentiated subtype with low (< 2% or < 100/5000 cells) or intermediate (2-4% or (100-200)/5000 cells) mitosis karyorrhexis index (MKI)	Any age: neuroblastoma (Schwannian stroma-poor), undifferentiated subtype
2	Age between 1.5 and 5 years old: neuroblastoma (Schwannian stroma-poor), differentiating subtype, with low MKI	Age between 1.5 and 5 years old: neuroblastoma (Schwannian stroma-poor), poorly-differentiated subtype
3	Ganglioneuroblastoma intermixed (Schwannian stroma rich), usually seen in older children	Any age: neuroblastoma (Schwannian stroma-poor) with high (> 4% or >200/5000 cells) MKI
4	Ganglioneuroma, maturing and mature subtype (Schwannian stroma-dominant, usually seen in older children. These tumours are within a framework of age appropriate, maturation sequence from neuroblastoma, poorly-differentiated, and differentiating subtype to ganglioneuroblastoma intermixed to ganglioneuroma. Cut-off points for MKI also changed according to age of the patients.	Age between 1.5 and 5 years old: neuroblastoma (Schwannian stroma-poor) with intermediate MKI
5	...	Age $\geq$ 5 years old: all neuroblastoma (Schwannian stroma-poor) subtypes
6	...	Ganglioneuroblastoma, nodular (composite Schwannian stroma-rich/stroma-dominant and stroma-poor)

Table 2.3: Specification of favourable and unfavourable histology groups (Robson 2001) based on the Shimada classification (Shimada et al. 1999)

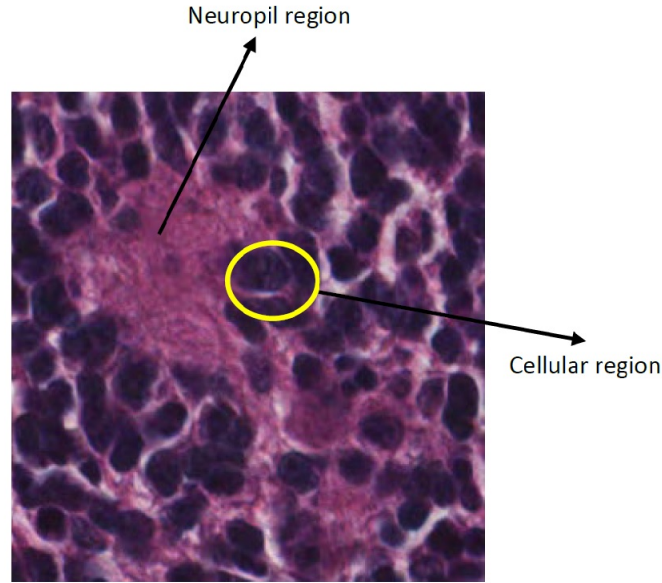


Figure 2.4: Blue regions are cellular regions and pink regions are neuropil regions in a H&E stained histological image

these representative regions which may be misleading for heterogeneous tumours (Sertel et al. 2009). Histopathological classification is currently considered as the gold standard.

Computer-aided diagnosis (CAD) systems have the potential to augment and complement human capability and may find new features which humans cannot detect (Yu et al. 2016). Moreover, CAD systems are a promising tool to facilitate the classification of histological images of neuroblastic tumours with high throughput. Therefore, CAD systems may be able to help pathologists in the classification of neuroblastic tumours. Several CAD systems have been proposed for the classification of histological images of cancerous tumours. These are based on segmentation methods, low-level feature extraction methods and high-level feature extraction methods. In this section, the proposed approaches for the automated analysis of cancerous tumours and their drawbacks are reviewed.

### 2.2.1 Segmentation Based Methods for Classification of Malignant Tumours

In this section, various techniques and approaches for the classification of cancerous tumours are reviewed. Segmentation is an important step in image processing. It identifies regions of interest from the image. In histopathology images, the regions of interest are tissue components such as the nucleus, neuroblast cells and ganglion cells. However, a majority of the works in histopathology images have focussed on the segmentation of nuclei. Different algorithms for the segmentation of regions of interest are summarized in the following paragraphs:

1. Thresholding techniques for segmenting the regions of interest from the images using the threshold value are reviewed.
2. Algorithms based on morphological analysis and geometrical models are discussed.

#### Thresholding Techniques

Thresholding is a widely used algorithm for segmenting the regions of interest from the background using a threshold value. In this method, a binary image is created using an intensity threshold. The optimal thresholding value is achieved by Otsu's method (Otsu 1979). This method determines a new range of intensity values using the gray-level histogram of the low intensity variation images. Otsu's method separates objects such that the inter-class variation is high and the intra-class variation is low (Sridhar 2011).

Zorman et al. (2007) developed an algorithm based on thresholding to segment the follicular regions in the histological images of lymphoma. The threshold value is

calculated as the mean of the brightness value of all pixels in the image. The pixel values which are less than the threshold value are set to 0 and the others are set to 1. They evaluate the proposed approach against the other thresholding methods such as median brightness value and Otsu's method and found similar results. The algorithm identifies the follicular regions where the density of the white pixels is higher than the black ones. Sertel et al. (2010) designed a computer-aided diagnosis approach to detect large malignant cells called centroblasts from histological images of follicular lymphoma. They segment red blood cells and background regions using thresholding. The limitation of the above approaches is their assumption of a uniform intensity in images. Therefore, Otsu's method does not perform well under high variations of intensity.

Michail et al. (2014) implemented histogram equalization as a pre-processing method to enhance the differences between nuclear regions and the background in histopathology images. This pre-processing method improves the performance of the thresholding in histological images with a wide variety of intensity. They integrate thresholding with the Otsu method to remove red blood cells in histological images. The results show that applying the pre-processing algorithm improves the performance of the thresholding. Korzynska et al. (2013) proposed an adaptive threshold method, a novel hybrid merging algorithm, to improve the performance of thresholding in images with high fluctuations of contrast between regions of interest and the background. The algorithm was applied to histological images of follicular lymphoma to classify them into two sub-categories. The experiment results show the significant improvement achieved by of the adaptive thresholding technique compared to the traditional thresholding method. However, thresholding methods do not perform well

under high variations of pixel intensities (Gonzalez & Wintz 2017).

### **Morphological Analysis and Geometrical Models**

Morphological analysis and geometrical models are based on the shape and geometry of the region of interest. Because of variability in the staining process and thickness of the section, the appearance of the tissue in histological images is highly variable. This characteristic makes the use of morphological and geometrical based methods for histological images challenging. The main challenge is the irregular visual aspects of different cells in histological images.

Nedzved et al. (2000) proposed an algorithm to segment cells in histology images. They compute a morphological gradient for object border extraction and thinned the obtained borders to get a line of one-pixel thickness. For this purpose, they propose a fast gray-scale thinning algorithm using the analysis of binary image layers. Then, the obtained one-pixel lines are used to extract cells and measure their features. In addition, they use a split-and-merge segmentation algorithm to split the overlapped cells. The main drawback of this method is the extraction of false boundaries that do not belong to the cells.

Zampirolli et al. (2010) introduced an approach to automatically segment and classify histological images of epithelial tissue and mammary gland adipose tissue. First, they convert the input image to the grayscale using the thresholding method. Then, they define regions of interest with a radius of five pixels. After segmentation, a specialist removes the incorrectly marked regions. Then, the watershed method (Vincent & Soille 1991) segments the regions of interest in an image. Finally, they create a graph using the morphological operations. The vertices of the graph are defined by

the centroids of each cell and the neighbourhoods of each vertic (edges) are defined by the neighbourhoods between the cells. Extracted graphs are classified using an SVM classifier. Over-segmentation is the main drawback of this system. Moreover, the proposed system is not robust to intra-class variations of size as it segments regions of interest based on the Euclidean circle with a radius of five pixels.

Lu et al. (2013) proposed a three-step algorithm to detect nuclei in histological images. First, their algorithm extracts cellular regions using maximally stable extremal regions algorithm (Matas et al. 2004). Second, the proposed method detects the nuclei structure using the integration of a level set function (Li et al. 2005) with an energy function. Finally, the algorithm splits the overlapping nuclei. The main drawback of this method is the low performance after splitting the nuclei.

In Chapter 3 of this thesis, several low-level feature extraction methods are proposed to overcome issues associated with variations in the size and intensity in histological images. Low-level features are robust to variations of size and intensity of cells within the images.

### 2.2.2 Low-level Feature Extraction

Feature extraction within the cells and tissues is used in the classification of histological images. The goal of feature extraction methods is to extract useful information for classification tasks. Feature extraction methods extract appearance-based features, morphology-based features and edge-based features. Appearance-based features including colour and texture are the most important attributes in medical images. Sertel et al. (2009) used nonlinear colour quantization (Jain et al. 2000). They developed a CAD system using intensity-based features including colour and texture to identify

the most aggressive type of follicular lymphoma in histological images. The main drawback of the algorithm is that the extracted features are based on the colour of the pixels. Variations in the staining of the histological images and environmental intensity conditions may affect the results.

Al-Kadi (2010) used fractal features to classify meningioma tumour. He uses a third order Gaussian Markov random field model (Petrou & GaciaSevilla 2006) to extract texture features. Moreover, he extracts non-Euclidean features using fractal functions. The extracted features are combined with each other to extract more useful information. The disadvantage of this method is the low accuracy in non-homogeneous images.

Kuse et al. (2010) proposed a method to classify cells in histological images into two classes: lymphocytes and non-lymphocytes. The main stages for the classification are: 1) extracellular matrix segmentation, 2) morphological pre-processing, 3) contour-based overlap resolution, 4) feature extraction and 5) classification. The masks obtained from the segmentation steps are assigned to lymphocyte and non-lymphocyte areas by Kuse and colleagues. The extracted masks were multiplied with the histogram equalized grayscale image of the RGB image to make them robust to intensity variation. Texture features are extracted from the gray level co-occurrence matrix for every detected cell region. Then, the extracted features are classified with an SVM classifier. The solution using the intensity of pixels in classification becomes more complex when the image is blurred or low contrast.

Spanhol et al. (2016) classified histological images of breast cancer as either benign or malignant based on the extracted features in four magnifications. They classify histological images into benign and malignant categories using the combination of six



feature extractors with four classifiers. They apply their algorithm to the BreakHis dataset which consists of 7909 images belonging to 82 patients and reported the recognition rate at the patient level and not at the image level. In section 3.3.3 the performance of the proposed method is compared with that of Spanhol and colleagues because their dataset is publicly available and can be used as a benchmark.

Kothari et al. (2013) introduced an approach for the classification of histological images of renal tumours. They use a three-step algorithm to build a diagnostic model from a set of histological images: 1) shape-based feature extraction, 2) feature selection and 3) classification. The shape-based features are extracted using Fourier shape descriptors. Then, features that are sensitive to noise are rejected in the feature selection part. Finally, they use a directed acyclic graph (DAG) structure for classification. The histological renal images are classified as either tumour or non-tumour. Although the Fourier shape descriptors are robust to high-frequency noise, they are not successful in histological images. This is because histological images have multiple highly variable shapes in the images that belong to the same category.

Irshad et al. (2013) extracted Scale Invariant Feature Transform (SIFT) features from  $80 \times 80$  pixel patches in the histopathology images of breast cancer and classified them into mitosis or non-mitosis categories. They achieve a low accuracy because of the huge number of high-dimensional extracted features which results in overfitting. This thesis combines SIFT with the bag of features detailed in Chapter 3 to overcome this problem. The SIFT algorithm is described in detail below.

## Scale-Invariant Feature Transform

Scale-Invariant Feature Transform was originally proposed by Lowe (2004). In this method, distinctive image features are extracted from the scale-invariant keypoints. Features extracted by SIFT are invariant to rotation and robust to scale and intensity variations. The SIFT algorithm consists of four stages: detection of scale-space extrema, keypoint localization, orientation assignment, and descriptor representation. These stages are described in detail as follows.

### 1. Detection of Scale-space Extrema

The first step involves searching over multiple scales and image locations to identify keypoints that are invariant to differences in scale. These keypoints are detected using a scale-space function (Witkin 1983).

Lindeberg (1993) showed that the only possible scale-space kernel is the Gaussian function,  $G(x, y)$ . The scale space of an image,  $L(x, y, \sigma)$ , is calculated by the convolution of a variable-scale Gaussian function,  $G(x, y, \sigma)$ , with an input image,  $I(x, y)$  defined as

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (2.2.1)$$

and

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (2.2.2)$$

where  $x$  and  $y$  are the image coordinates and  $\sigma$  is the width of the Gaussian function. Stable keypoint locations are defined by the Difference of Gaussian (DoG) function in the scale-space. The scale space extrema are located in

$D(x, y, \sigma)$ , which is calculated as the difference between two images, one with scale  $k$  and the other one with scale  $k\sigma$ , where  $k$  is a multiplicative factor, is computed as

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (2.2.3)$$

Points in the DoG function which are maxima or minima among the 26-neighbouring pixels (eight neighbours in the current image and nine neighbours in the above and nine neighbours in the below scales) are considered as extrema points which are called keypoints. The detection of keypoints is shown in Figure 2.5.

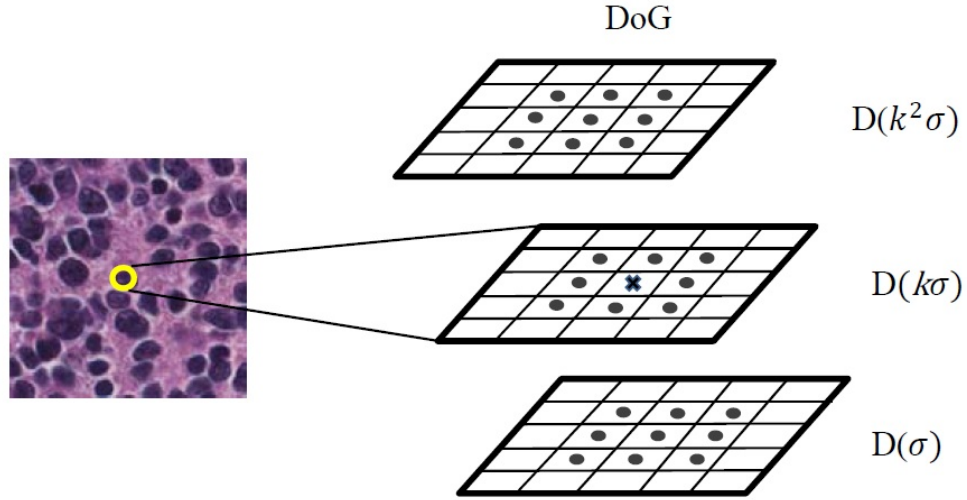


Figure 2.5: The scheme of keypoint detecting. Each point is compared to its eight neighbours in the current image (middle image) and nine neighbours in the scale above (top image) and below (down image).

## 2. Keypoint Localization

Some of the extracted keypoints that have low contrast or are poorly localized on the edges of the image are not stable. These keypoints are eliminated in this stage. The location of the keypoint,  $z$ , is achieved by calculating the Laplacian according to

$$z = - \left( \frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} + \frac{\partial^2 D^{-1}}{\partial y^2} \frac{\partial D}{\partial y} \right) \quad (2.2.4)$$

If the value of the Laplacian function is less than a threshold value, called the contrast threshold ( $C_C$ ), the point is rejected. Therefore, the extracted keypoints with low contrast are eliminated. Some of the extracted keypoints have a strong response along the edges. Therefore, these points need to be eliminated to increase stability. Lowe (2004) used a  $2 \times 2$  Hessian matrix to compute two eigenvalues in edges. A Hessian matrix is a square matrix of second-order partial derivatives of a scalar-valued function. It describes the local curvature of a function of variables. If the ratio between the larger and smaller eigenvalue is greater than the edge threshold ( $C_E$ ), this keypoint is discarded. Therefore, there are more stable keypoints after eliminating the low contrast and edge keypoints. The values of  $C_C$  and  $C_E$  are experimentally set by the authors. This thesis uses these methods to experimentally set the thresholds in Chapter 3.

### 3. Orientation Assignment

Before the extraction of a descriptor for the keypoint, the keypoint is assigned an orientation to make it invariant to rotation. The orientation of the keypoint is calculated from the orientation histogram of local gradients from the smoothed

image  $L(x,y,\sigma)$  which means there is a potential keypoint at  $(x,y)$  with scale  $\sigma$ . The gradient magnitude  $m$  and the orientation  $\theta$  are calculated as

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \quad (2.2.5)$$

$$\theta(x,y) = \tan^{-1} \frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)} \quad (2.2.6)$$

An orientation histogram is created from the gradient orientations of sample points within a region around the keypoint. It consists of 36 bins covering the 360 degrees of orientation (Lowe 2004). Peaks in the orientation histogram are representatives for dominant directions of local gradients which are considered as the orientation for the keypoint. Lowe (2004) stated that any other peaks above 80% of the highest peak are converted into a new keypoint. This new keypoint has the same location and scale, but different orientations. The orientation calculation is illustrated in Figure 2.6.

## Descriptor Representation

The local neighbourhood is rotated with the orientation value that was calculated in the orientation assignment section. In this stage, a SIFT descriptor is calculated for each keypoint as the sum of the gradient magnitudes at each direction (as shown in Figure 2.7). According to Lowe (2004), Smith & Harvey (2011) and Nam et al. (2009), it is common that a  $16 \times 16$  neighbourhood region around each keypoint is divided into

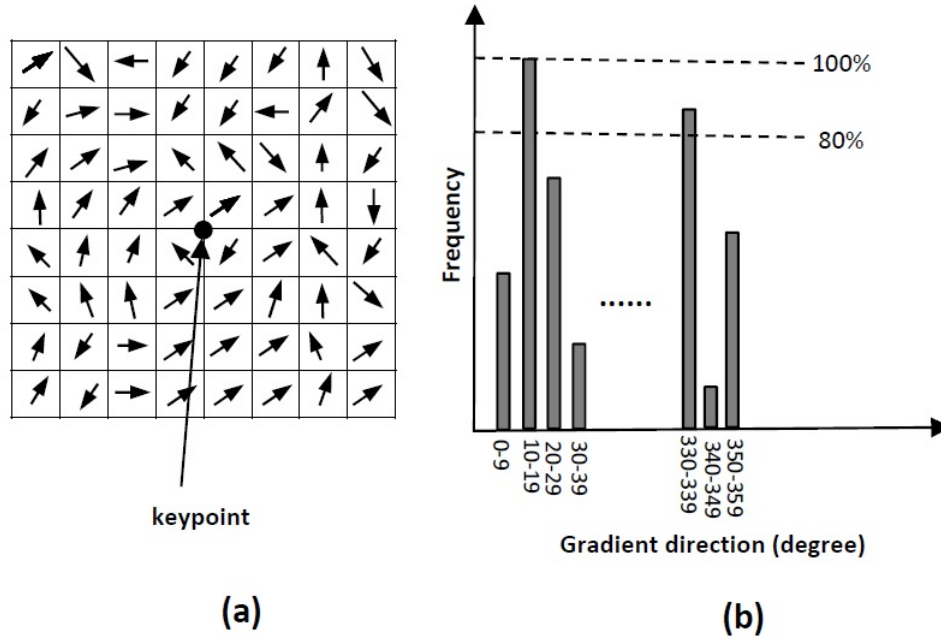


Figure 2.6: Orientation calculation: (a) orientations for 64 neighbour pixels around the keypoint (b) orientation histogram

$4 \times 4$  sub-regions. The gradient vectors are aggregated into 8-bin histogram over the  $4 \times 4$  matrix of subregions. Then, SIFT extracts  $4 \times 4 \times 8 = 128$  element feature vector for each keypoint.

SIFT extracts high dimensional features which can result in overfitting in the state-of-the-art and decreases the accuracy of the algorithm. This thesis combines SIFT with the bag of feature algorithm in Chapter 3 to reduce the dimensionality of the feature space by selecting more discriminative features.

### 2.2.3 High-level Feature Extraction Methods

Deep learning networks which extract high-level features have exciting potential in tissue classification including histopathology images (Janowczyk & Madabhushi 2016).

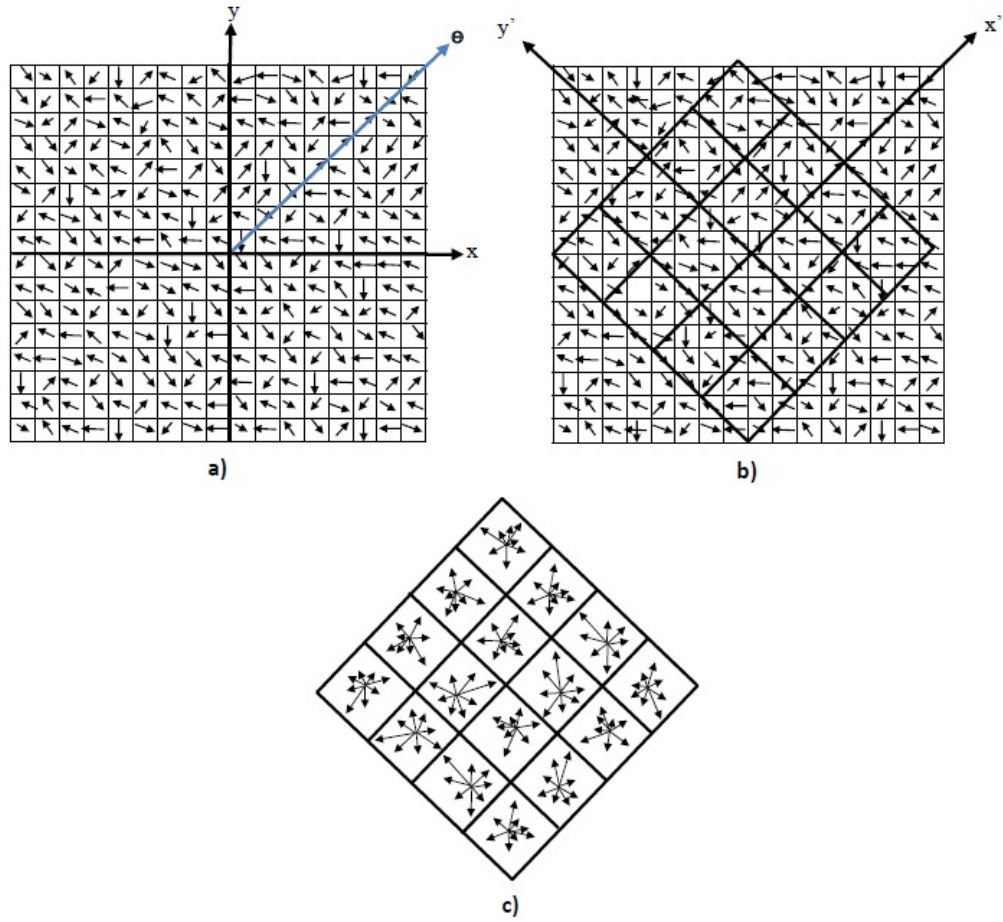


Figure 2.7: Keypoint descriptor: a) gradient magnitude and orientation in local neighbourhood of keypoint. b) orientation corrections and spatial coordinate transformations. c) SIFT descriptor establishment

Deep learning architectures are the evolution of multilayer neural networks with a different design to make them competitive. These approaches extract the high-level features of the data by combining multiple linear and non-linear transformations (Montavon 2009). High-level feature extraction is part of the learning process in deep learning approaches. These features are extracted from a large amount of training data to discriminate between different classes of images. However, low-level feature extraction methods which are handcrafted approaches require considerable effort and

several algorithmic iterations to extract similar features. Low-level features are image features that are mostly application independent and refer to the minor details of the image such as lines, curves and edges. This is in contrast to the high-level features that are built on top of the low-level features to detect objects and larger shapes in the image. Convolutional Neural Networks (CNNs) and Convolutional Deep Belief Networks (CDBNs) are two types of deep networks that are used frequently in image analysis tasks. Deep networks combined low-level features that learned in lower layers to extract high-level features in the later layers. In the following paragraphs, these networks are reviewed.

### **Convolutional Neural Network**

The Convolutional Neural Network (CNN) was proposed by Fukushima (1980) and is composed of feed-forward networks which are stacked together to recognise features of a 2-D input image. As shown in Figure 2.8, CNN consisting of two parts; feature extraction and classification. The feature extraction part consists of convolutional layers and subsampling layers. The classification is performed using the fully connected multi layer Perceptron based on the extracted features from the previous layers. The main part of the CNN is the convolutional layer which is a mathematical operation to merge two sets of information. The convolution is applied on the input data using a convolution filter to produce a feature map. The filter slides over the input image to do the convolution operation. After sliding the filter over all the locations and doing element-wise matrix multiplication, there is an array of numbers which is called feature map.



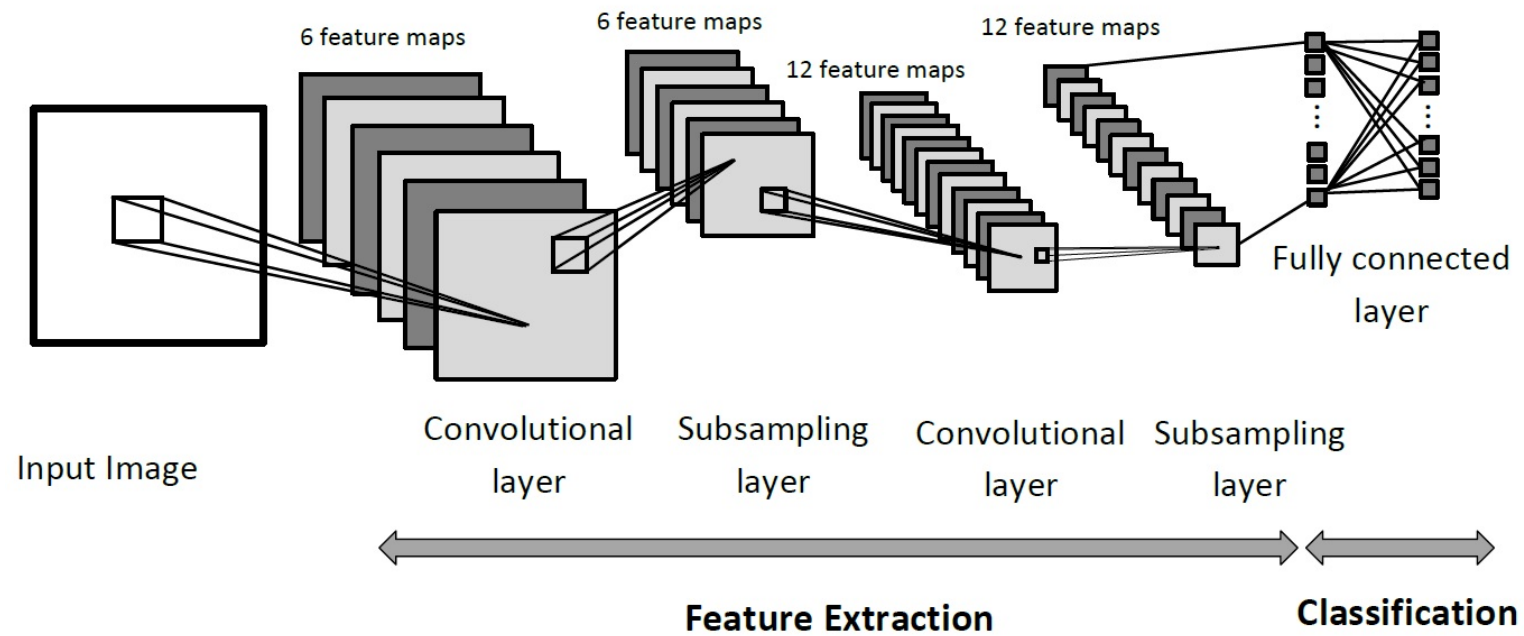


Figure 2.8: Overall framework of the Convolutional Neural Network (Lecun et al. 1998). It consists of a convolutional layer, subsampling layer and fully-connected layer.

A convolutional layer  $l$  takes a set of feature maps  $F_l$  which comes after a subsampling layer with feature maps  $F_{l-1}$ . The size of the feature maps in layer  $l$  is defined as  $m_l \times n_l$ . The output of neuron  $(i, j, \phi, l)$  at position  $(i, j)$  in feature map  $\phi$  of layer  $l$  is defined as  $y_{i,j}^{\phi,l}$  while  $i$  and  $j$  are as follow

$$0 \leq i < m_l, \quad 0 \leq j < n_l \quad (2.2.7)$$

The size of the input image is  $m_0 \times n_0$  pixels. The input layer is followed by convolutional and subsampling layers. The input of the neurons of feature map  $\phi$  in convolutional layer  $l$  comes from feature maps  $F'_{\phi,l}$ . The subsampling layer which is connected to the convolutional layer reduces the size of the feature maps. Finally, the input image is classified using a fully-connected multilayer perceptron network according to the extracted features. Convolutional layers which consist of multiple feature maps detect the features in the input images. Feature extraction is done by passing the sub-images through a filter. For simplicity, a feature map in the convolutional layer is referred to as a convolutional map and a feature map in the subsampling layer as a subsampling map.

As shown in Figure 2.8, each neuron in the convolutional map takes its input as  $r_c^l \times r_c^l$  from neurons in the previous subsampling map. The dimension of convolutional maps is calculated as follows for a given subsampling map of size  $m_{l-1}$  by  $n_{l-1}$ .

$$(m_l, n_l) = \left( \left\lceil \frac{m_{l-1} - r_c^l + 1}{s_c^l} \right\rceil, \left\lceil \frac{n_{l-1} - r_c^l + 1}{s_c^l} \right\rceil \right) \quad (2.2.8)$$

Each neuron in the convolutional layer which is defined as  $(i, j, \phi, l)$  is connected to all the neurons  $(is_c^l + \Delta i, js_c^l + \Delta j, \phi', l - 1)$  in the previous subsampling layer where

$\phi' \in F_{\phi,l}$ ,  $0 \leq \Delta i$  and  $\Delta j < r_c^l$ . The weight of each connection within the neuron  $(i, j, \phi, l)$  is  $w_{\Delta i, \Delta j}^{\phi, \phi', l}$  which is shared for all the neurons within the same feature map. The number of required weights by convolutional layer  $l$  is calculated as

$$|W_l| = |F_l| \cdot (r_c^l)^2 \quad (2.2.9)$$

The output,  $y_{i,j}^{\phi,l}$ , of the neuron  $(i,j,\phi,l)$  is calculated as

$$y_{i,j}^{\phi,l} = \sum_{\phi' \in F_{\phi,l-1}} \sum_{\Delta i=0}^{r_c^l-1} \sum_{\Delta j=0}^{r_c^l-1} y_{is_c^l+\Delta i, js_c^l+\Delta j}^{\phi',l-1} w_{\Delta i, \Delta j}^{\phi, \phi', l} \quad (2.2.10)$$

The algorithm calculates the sum of the weights for each neuron. There are  $|F_l| \cdot m_l \cdot n_l$  neurons in layer  $l$  that sums  $|F_{l-1}| \cdot (r_c^l)^2$  weights.

In CNNs, subsampling layers come after the convolutional layers. They reduce the size of the feature maps by retaining the most important information. Each convolutional map in layer  $l-1$  is connected only to the subsampling map in layer  $l$ . The sizes of the feature maps are calculated as follows (Fukushima 1980)

$$m_l = \left\lceil \frac{m_{l-1}}{r_x^l} \right\rceil, \quad n_l = \left\lceil \frac{n_{l-1}}{r_y^l} \right\rceil \quad (2.2.11)$$

where  $r_x^l$  and  $r_y^l$  are the non-overlapping receptive fields of the subsampling neurons. Each neuron from the convolutional layer is connected to just one neuron in the subsampling layer and each convolutional map is connected to just one subsampling map, keeping the design of the network simple. The input of the neurons in the subsampling maps comes from their receptive fields. Multiple inputs  $r_x^l \times r_y^l$  are combined to make a single value which is defined as the neuron potential. The inputs

are combined through averaging or finding the maximum value as follows

$$\xi_{i,j}^{\phi,l} = \frac{1}{r_x^l r_y^l} \sum_{\Delta i=0}^{r_x^l-1} \sum_{\Delta j=0}^{r_y^l-1} y_{ir_x^l+\Delta i, jr_y^l+\Delta j}^{\phi,l-1} \quad (2.2.12)$$

$$\xi_{i,j}^{\phi,l} = \max_{\substack{\Delta i \in (0, r_x^l-1) \\ \Delta j \in (0, r_y^l-1)}} (y_{ir_x^l+\Delta i, jr_y^l+\Delta j}^{\phi,l-1}) \quad (2.2.13)$$

The output of the neuron is calculated as follows

$$y_{i,j}^{\phi,l} = f(a^{\phi,l} \xi_{i,j}^{\phi,l} + b^{\phi,l}) \quad (2.2.14)$$

where  $a^{\phi,l}$  is a trainable coefficient and  $b^{\phi,l}$  is a trainable bias.

## Convolutional Neural Networks in Histopathology Images

Convolutional neural networks are valuable tools for the generation of complex multi-parametric decision algorithms for processing histological images. These networks extract features that represent distinct diagnostic patterns.

Coudray et al. (2018) classified non-small cell lung cancer histopathology images into normal and tumour tissues and distinguished the type of the lung cancer using CNN. CNN is applied on whole slide images obtained from The Cancer Genome Atlas data. The results are evaluated by calculating the area under the curve (AUC). The experimental results show that the proposed algorithm classified histopathology images into normal and tumour tissues with 0.99 AUC and distinguished lung cancer type with 0.97 AUC. To compare the results with the pathologists, they asked three pathologists to classify the images. The experiment results show half of all the slide

images were misclassified by the CNN. These slides are misclassified by at least one pathologist. The main drawback of this algorithm is that it does not work on less common lung cancers.

Bejnordi et al. (2017) used CNN to detect metastases in H&E stained tissue sections of the lymph nodes of breast cancer and compared the results with pathologists' diagnoses. They use 129 whole slide images to assess the method. The 11 pathologists registered their diagnostic confidence as definitely normal, probably normal, equivocal, probably tumour, or definitely tumour. The experiments were conducted in two stages. In the first experiment, the pathologists specified the level of confidence in their interpretation for each slide in high magnification without any time constraints. In the second experiment, the pathologists determined their level of confidence within two hours. The experiment results show that the deep learning algorithm is comparable with the experiments without the time constrained. This algorithm discriminates between normal and cancerous tissue in the histological images of breast cancer. The main drawback of this algorithm is that it was unable to classify cancerous tissue into subcategories.

Kieffer et al. (2017) classified the Kimia Path24 dataset using CNN-VGG16 which is publicly available. This dataset consists of 28,380 histopathology images in 24 tissue texture classes. They used 27,055 images for training and 1325 images for testing. The results they achieved were low in accuracy because the selected architecture was not well suited to the problem because of the high intra-class variability between the images within the same class.

Zheng et al. (2017) used the CNN to classify histological images of breast cancer into benign and malignant categories. They applied CNN in three levels: the patch

level, block level and image level to extract features from the cell nuclei. The network was pre-trained with abundant unlabeled data and then fine-tuned using the labeled images. First, nucleus components were extracted using the colour deconvolution algorithm (Ruifrok & Johnston 2001). Second, the cell nuclei was highlighted using a Gaussian filter. Finally, features were extracted from the different levels of images: patch level, block level and image level. The input images were classified using an SVM classifier based on the extracted features from the nuclei. However, the background regions consist of important information that is useful in the classification of histological images especially when they are classified into more than two classes.

Litjens et al. (2016) detected cancerous areas on prostate histological images by implementing a deep CNN. Although their model achieved 0.9 AUC, they did not address the challenge of distinguishing high-grade versus low-grade cancer.

While much work has been done in relation to the classification of histological images using CNN, it is new in relation to neuroblastoma histological images. In Chapter 4, CNN is used to classify the histological images of neuroblastoma. Moreover, this research combines CNN with the SVM classifier. Most of the examples in the literature output a simple binary classification of histologic images into benign or malignant. This work further refines classification of a tumour into the multiple subtypes required for real-world tumour classification.

### **Convolutional Deep Belief Network**

Lee et al. (2011) proposed Convolutional Deep Belief Network (CDBN) which is an hierarchical generative model composed of stacked Convolutional Restricted Boltzmann Machines (CRBMs)(Desjardins & Bengio 2008). The CRBM is a modified version of a Restricted Boltzmann Machine (RBM). Therefore, before going through

the CRBM, this thesis reviews the RBM. As can be seen in Figure 2.9, structure of the RBM has two hidden units  $\mathbf{h}$  and three visible units  $\mathbf{v}$ . These two layers are connected by a symmetric connection weight matrix  $W$ .

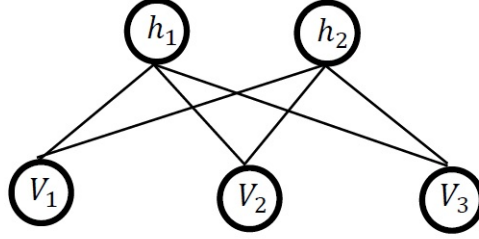


Figure 2.9: An example RBM with three visible units and two hidden units

The probabilistic semantics for an RBM are defined by its energy function as follows (Lee et al. 2011):

$$P(v, h) = \frac{1}{Z} \exp(-E(v, h)) \quad (2.2.15)$$

where  $Z$  is a normalization constant. If the visible units have a binary value, the energy function is defined according to equation (2.2.16):

$$E(v, h) = - \sum_{i,j} v_i W_{i,j} h_j - \sum_j b_j h_j - \sum_i c_i v_i \quad (2.2.16)$$

where  $b_j$  and  $c_i$  are hidden unit biases and visible unit biases, respectively. If the visible units are real-valued, the energy function is defined according to the following equation (Lee et al. 2011):

$$E(v, h) = \frac{1}{2} \sum_i v_i^2 - \sum_{i,j} v_i W_{i,j} h_j - \sum_j b_j h_j - \sum_i c_i v_i \quad (2.2.17)$$

According to the energy function, hidden units give the visible layer and vice versa. These units are conditionally independent of each other.

The RBM is limited when it is used on its own. It is more powerful when RBMs are stacked to form a Deep Belief Network (DBN) (Lee et al. 2011). In a DBN each layer consists of binary units. Two layers are fully connected to each other, but units in the same layer are not connected to each other. DBNs are trained from lowest to highest using greedy training of each layer. The inputs of the DBNs are activations of the previous layer.

Both RBMs and DBNs ignore the 2-D structure of images. Therefore, weights that detect a given feature must be learned separately for every location. Lee et al. (2011) proposed a Convolutional DBN where weights are shared among all locations in an image. Each CRBM consists of a visible layer (V), a hidden layer (H), and a pooling layer (P) (as shown in Figure 2.10).

The visible layer is a  $N_V \times N_V$  array of binary units. The hidden layer consists of  $k$  groups, where each group is a  $N_H \times N_H$  array of binary units resulting in  $k \times N_H^2$  hidden units. For each group, there is a dedicated  $N_W \times N_W$  convolutional window in the visible layer where  $[N_W = N_V - N_H + 1]$ . The symmetric connections between hidden and visible units are represented by a weight matrix  $W$ . Each weight matrix is interpreted as a filter. One hidden group's weight is called "basis".

### **Probabilistic Max-pooling**

According to the algorithm proposed by Lee et al. (2011), CRBMs are stacked to build DBNs. This architecture is based on the probabilistic max-pooling. As shown in Figure 2.10, a CRBM consists of three layers: visible layer V, hidden layer H and pooling layer P. The hidden and pooling layers have  $K$  groups of units, and each



group of the pooling layer has  $N_P \times N_P$  binary units. The pooling layer shrinks the representation of the hidden layer by a pre-defined constant factor  $C$ , a small integer such as 2 or 3. The hidden layer is partitioned into blocks of size  $C \times C$ . Each block in the hidden layer is connected to one binary unit  $p_\alpha^k$  in the pooling layer. For this purpose, the pooling layer selects the maximum values in the  $C \times C$  windows of the hidden layer. Applying the max-pooling not only allows the higher layer outputs to be invariant to slight changes in the input, it also reduces the overall computational cost.

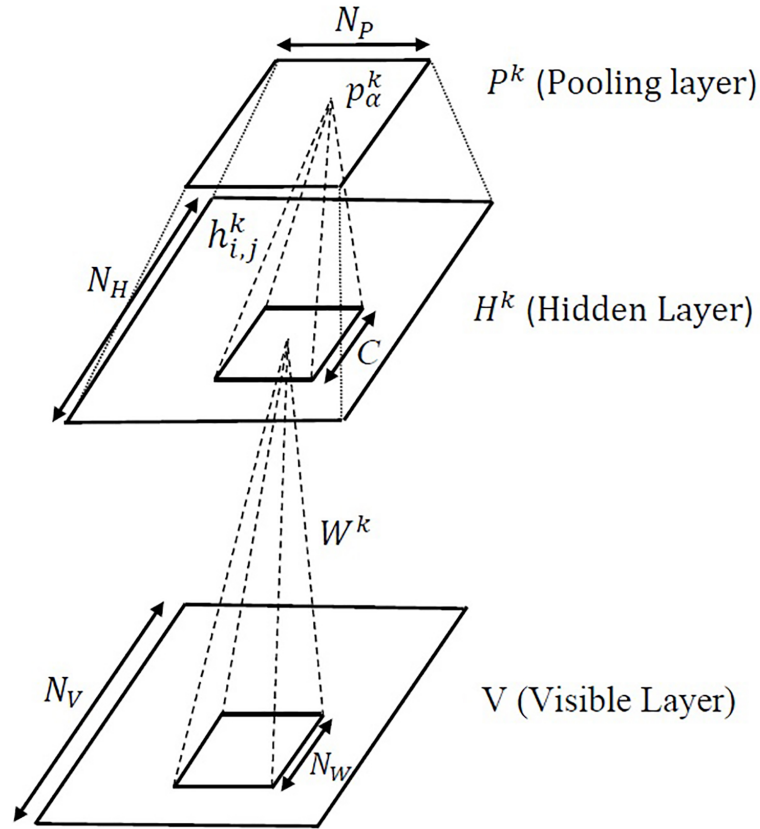


Figure 2.10: Construction of a Convolutional Restricted Boltzmann Machine (CRBM) with probabilistic max-pooling, based on Lee et al. (2011)

A CDBN consists of several stacked max-pooling-CRBMs. The weights of a trained layer are frozen and the output of the layer is used as the input for the next layer. An efficient approach for training DBNs was proposed by Hinton et al. (2006). In this approach, DBN layers are trained by a greedy algorithm in which the output of the lower RBM layers is used as the input of the upper RBM layers.

While much work has been done in the classification of histological images using CNN, the CDBN has not been used in this field. In Chapter 4, CDBN is used to classify the histological images of neuroblastoma. Moreover, this research combines CDBN with a feature selection algorithm.

## 2.3 Classification of Neuroblastoma Histological Images

Several research studies have been conducted for the automated classification of neuroblastoma histological images. Kong et al. (2009) classified neuroblastoma into three established categories: undifferentiated, poorly-differentiated, and differentiating, using image segmentation techniques. Images at each resolution level were segmented into cellular, neuropil and background elements and were classified by integrating different classifiers such as Linear Discriminative Analysis (LDA), Support Vector Machine (SVM) and k-Nearest Neighbour (k-NN). Tafavogh et al. (2014) classified neuroblastoma tumour images into undifferentiated and poorly-differentiated categories using Otsu segmentation (Otsu 1979). Images at each resolution level were segmented into cellular, neuropil, and background elements and were classified based on the Shimada classification (Shimada et al. 1999). They used structural features

of nucleus, ganglion cells, and neuroblast cells. The main drawback of the aforementioned algorithms is that they use visual features which are based on the colour, geometry and appearance of the region of interest in histological images. However, a high variation of shape and texture between neuroblast cells from images which belong to the same category make the classification based on the visual features more challenging. Sertel et al. (2009) designed a CAD system to determine the degree of Schwannian stromal in whole-slide histological images of neuroblastoma. They classified neuroblastoma into favourable and unfavourable categories based on the stroma-rich and stroma-poor regions using co-occurrence statistics and the Local Binary Pattern. They addressed the problem of neuroblastoma stroma-rich versus stroma-poor. The step prior to favourable or unfavourable histology requires typing the neuroblastoma tumours into subtypes. This thesis develops different methods to predict five different categories of neuroblastoma histological images required for real-world tumour classification.

## 2.4 Classifiers Used in This Thesis

This thesis uses three popular classifiers: support vector machine (SVM), k-nearest neighbour (k-NN) and bag of features to evaluate the performance of the proposed feature extraction methods. These algorithms are summarized in the following paragraphs.

Kernel type	Kernel function
Linear	$k(x_i, x_p) = x_i^T x_p$
Polynomial	$k(x_i, x_p) = (\eta x_i^T x_p + c)^d$
Gaussian	$k(x_i, x_p) = \exp()$
Sigmoidal	$k(x_i, x_p) = \tanh(\eta x_i^T x_p + c)$

Table 2.4: Popular kernel functions

### 2.4.1 Support Vector Machine

Support Vector Machine (SVM) was proposed by Vapnik (Vapnik & Media. 2000) and is one of the best algorithms that used for the classification (Wu et al. 2008). The SVM separates a set of points using an optimal hyperplane to maximize the distance between the hyperplane and the nearest points from each class, which are known as support vectors (Chang & Lin 2011). The SVM is a supervised learning algorithm which builds a model based on a set of samples with known labels or categories, and the model predicts the label or category of new observations. Support vector machine classifiers are divided into binary and multi-class versions. This thesis uses the multi-class SVM which deals with non-linearly separable data and consequently needing a nonlinear classifier to separate the classes.

A linear SVM is transformed to the nonlinear case by using different kernel functions. Kernel methods transform data from the original data space to a feature space. There are several kernel functions that are shown in Table 2.4 (Kavzoglu & Colkesen 2009), where  $x_i$ ,  $x_p$  are two samples,  $\eta$  is the slope,  $c$  is the constant term, and  $d$  is the polynomial degree.

### 2.4.2 k-Nearest Neighbour

The k-nearest neighbour classifier (k-NN) (Cover & Hart 1967) classifies a new point based on the distance or similarity function to its neighbours. The algorithm selects the closest  $k$  points and classifies the new point based on the majority class of the  $k$  neighbours. For example, if  $k$  is three, the algorithm will select the closest three points and classify the new point based on the majority class of the three neighbours. The K-NN generally uses the Euclidean distance between points, although other distance functions may also be used. It is a simple machine learning algorithm that is used to solve classification and regression problems. However, it has the following disadvantages (a. Cunningham & Delany 2007).

1. It is sensitive to noise
2. It decides based on the Euclidean distance which may not work in the case of high dimensions and a low number of samples.
3. The weights of all the features are equal in this classifier, including noise and irrelevant data

### 2.4.3 Bag of words model in computer vision

The bag of words model can be applied to image classification by considering image features as words (Csurka et al. 2004). In computer vision, bag of visual words is a histogram showing occurrence counts of the local image features. Three steps should be considered for defining features in images: feature detection, feature description and codebook generation.

Keypoints are the “stand out” points in an image and descriptors are the description of the keypoints. Keypoints and descriptors construct the vocabularies in the images. Each image is represented by histograms showing the frequency of features in the image. Bag of features can only compare entire images to one another and does not allow for partial matching. However, Pyramid matching is an efficient method proposed by Lazebnik et al. (2006) to map sets of features to multi-resolution histograms and then compare the histograms with a weighted histogram intersection to approximate the similarity between the feature set. The input sets are converted to multi-resolution histograms in the pyramid match kernel (see Figure 2.11).

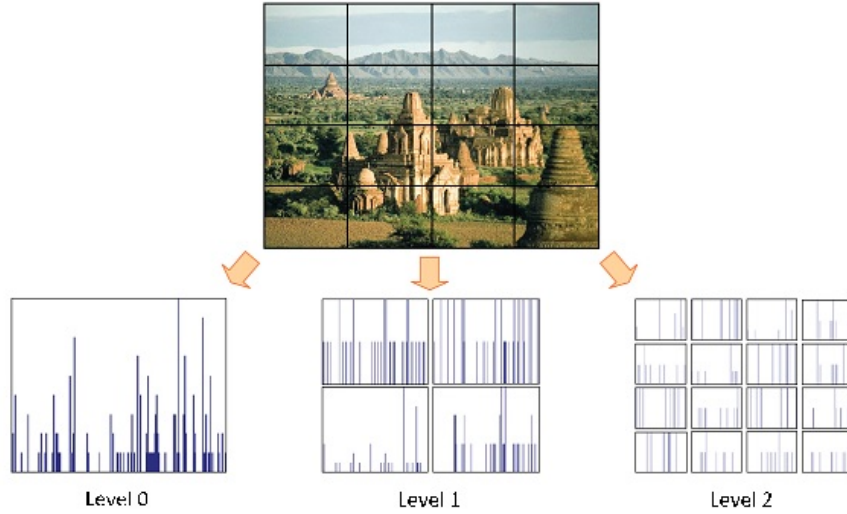


Figure 2.11: Pyramid bag of features. At level 0, decomposition consists of just a single cell and the representation is equivalent to a standard bag of features. The image is subdivided into four and 16 quadrants, extracting four and 16 feature histograms in level 1 and 2, respectively (Lazebnik et al. 2006)

The feature extraction function  $\phi$  is defined as

$$\phi = [H_{-1}(X), H_0(X), \dots, H_L(X)] \quad (2.4.1)$$

where  $L = \lceil \log_2 D \rceil$ ,  $x \in X$  and  $H_i(X)$  is a histogram vector calculated over data  $x$  using  $d$ -dimensional bins of side length  $2^i$  and  $H_i(X)$  has a dimension  $r_i = (\frac{D}{2^i \sqrt{D}})^d$  where  $\phi$  is a vector of concatenated histograms and  $H_i(X)$  is a histogram vector formed over data  $x$ . The bins in level  $H_{-1}$  are small and each data point from  $X$  data falls into its own bin and increases the bin size until all data points from  $X$  fall into a single bin at level  $L$ .

The pyramid match kernel is shown with  $K_\Delta$  that measures the similarity between point sets. However, similarity between two input sets is defined as the weighted sum of the number of feature matching found at each level of the pyramid formed by  $\phi$  and is calculated as

$$K_\Delta(\Phi(y), \Phi(z)) = \sum_{i=0}^L w_i N_i \quad (2.4.2)$$

where  $N_i$  signifies the number of pairs that has been matched newly at level  $i$ . A new match is defined as a pair of features that were not in correspondence at any resolution level. If two points are matched and they fall into the same histogram bin, the kernel implicitly finds correspondences between point sets. The matching is an hierarchical process and vectors that did not find a matched pair at higher resolution can be matched at lower resolutions. The output of  $K_\Delta$  shows the similarity of the matching. Each new matched pair at level  $i$  contributes a value  $w_i$  that shows the similarity between two points matching at level  $i$ .

The histogram intersection  $I$  counts the number of points in two sets which are fell into the same bin which is calculated as

$$I(A, B) = \sum_{j=1}^r \min(A^{(j)}, B^{(j)}) \quad (2.4.3)$$

where  $A$  and  $B$  are histograms with  $r$  bins, and  $A^{(j)}$ ,  $B^{(j)}$  show the count of the  $j^{th}$  bin of  $A$ . The number of newly matched pairs at level  $i$  is calculated as

$$N_i = I(H_i(y), H_i(z)) - I(H_{i-1}(y), H_{i-1}(z)) \quad (2.4.4)$$

weights  $w_i = \frac{1}{2^i}$  of the pyramid level  $i$  are inversely proportional to the bin size.

## 2.5 Clustering Approaches Used in This Thesis

### 2.5.1 k-means clustering

The k-means clustering algorithm is an iterative algorithm that partitions the dataset into  $k$  distinct non-overlapping subgroups (clusters) where each data point belongs to only one cluster with the nearest mean (MacQueen 1967). A cluster is applied on a collection of datapoints aggregated together because of certain similarities. The k-means algorithm identifies  $k$  number of centroids. Every datapoint is allocated to the nearest cluster. The position of the centroids in k-means algorithm is randomly selected and then optimized. If  $X = \{X_1, X_2, \dots, X_n\}$  be a set of data points and  $V = \{V_1, V_2, \dots, V_n\}$  be a set of centroids. The steps for k-means clustering are as follow.

1. The ‘C’ cluster centers are selected randomly.
2. The distance between each data point and centers of the cluster is calculated.
3. Assign the data point to each cluster center, the distance from the cluster center is minimum between all the cluster centers.



Table 2.5: Confusion Matrix. Abbreviation N (Negative) and P (positive)

Actual \ Predicted	N	P
	True Negative (TN)	False Positive (FP)
N		
P	False Negative (FN)	True Positive (TP)

4. The new cluster center is recalculated as follow:

$$V_i = (1/C_i) \sum_{j=1}^{C_i} X_i \quad (2.5.1)$$

where,  $C_i$  represents the number of data points in  $i^{th}$  cluster.

5. Recalculate the distance between each data point and new obtained cluster centers.
6. If no data point was reassigned then stop, otherwise repeat from step 3). K-means clustering is fast and understandable and works well with distinct or well separated data.

## 2.6 Metrics Used in This Thesis

This research evaluates the performance of the model using precision, recall and F-measure. Evaluating the proposed model is the most important task in designing the algorithms which indicates how good the predictions are. In binary classification, classification results can be shown in a confusion matrix as showing in Table 3.10. The two classes in this case are labelled Positive and Negative.

Precision shows the proportion of the predicted positives that are truly positive,

total number of predicted positives is calculated by summation of the true positives plus the false positives ( $TP + FP$ ). Precision is calculated by

$$Precision = \frac{TP}{TP + FP} \quad (2.6.1)$$

Another measure is recall, which indicates the proportion of actual positives, total number of actual positives is calculated by summation of false negative and true positive ( $FN + TP$ ). Recall is calculated by

$$Recall = \frac{TP}{TP + FN} \quad (2.6.2)$$

F-measure is an overall measure of the accuracy that combines precision and recall by their geometric mean

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (2.6.3)$$

However, this research considers multi-class classification and needs to categorize each sample into 1 of  $N$  different classes. The precision and recall for multi-class classification are calculated as follow.

$$Precision = \frac{TP_1 + TP_2 + \dots + TP_k}{TP_1 + \dots + TP_k + FP_1 + \dots + FP_k} \quad (2.6.4)$$

$$Recall = \frac{TP_1 + TP_2 + \dots + TP_k}{TP_1 + \dots + TP_k + FN_1 + \dots + FN_k} \quad (2.6.5)$$

## 2.7 Cross-Validation

Cross-validation (Geisser 1993) is a technique for evaluating machine learning methods by training them on subsets of the input data and evaluating on the complementary subset of the data. It is mainly used in prediction to estimate how accurately a model will predict in practice. In  $k$ -fold cross-validation, the input data is split into  $k$  subsets of data (known as folds). The model is trained on  $k-1$  subsets of the data and tested on a subset that has not been seen in the training set. The goal of cross-validation is to evaluate the ability of the model in prediction of the new data (test subsets). In most methods multiple rounds of cross-validation are performed to reduce variability. They use different partitions, and the test results are the average of the rounds to give an estimate of the model's predictive performance. In summary, averages of the cross-validation measure the most accurate estimation of model prediction performance.

## 2.8 Parameter Tuning Methods Used in This Thesis

This research uses greedy and wrapper methods for parameter tuning of the network. These methods are described in the following paragraphs.

### 2.8.1 Greedy Method

A greedy algorithm (Jungnickel 1999) makes a locally optimal choice in the hope that this choice will lead to an optimal solution. In some problems, a greedy algorithm does

not usually produce an optimal solution but approximate a globally optimal solution in a reasonable amount of time. The optimal solution for the problem contains optimal solutions to the sub-problems. In a greedy algorithm the parameters of the network are tuned one by one while holding the other parameters constant. In general, greedy algorithms have five steps:

1. A candidate set, from which a solution is created
2. A selection function, which chooses the best candidate to be added to the solution
3. A feasibility function, that is used to determine if a candidate can be used to contribute to a solution
4. An objective function, which assigns a value to a solution, or a partial solution, and
5. A solution function, which will indicate when does a complete solution is discovered

This algorithm has some advantages; the first one is that working with a greedy algorithm is quite easy and straightforward. The second advantage is the short run time that is because at each level of recursion the size of the variables gets smaller. However, the disadvantage of this method is that it is hard to prove its correctness.

### **2.8.2 Wrapper Approach**

Feature selection identifies the subset of features in the dataset that are most useful or most relevant for solving the problem. Feature selection is an automatic selection of

features in the dataset that are most relevant features for use in model construction. The simplest approach is to test each possible subset of features to find the subset of features that minimizes the error rate. However, this quickly becomes computational untenable as the number of features increases. Feature selection methods include or exclude extracted features from data without changing them. They identify and remove irrelevant features from the dataset. These methods choose the features to achieve high accuracy while using less data and reduce the complexity of the model to make it more understandable. One method used in feature selection is the wrapper method (Guyon & Elisseeff 2003). This method is based on a greedy search and evaluates the extracted features from the data and select the features that produce the best results. The main issue with this method is deciding when to stop the algorithm. This problem is solved with cross-validation.

## 2.9 Research Gaps

In summary, to develop a CAD system to assist pathologists in the classification of neuroblastoma histological images, four main steps must be taken: 1) construct a dataset of neuroblastoma, 2) extract features that are more robust to variations in intensity, 3) extract features that are more robust to variations in the size of the different cells in neuroblastoma histological images that belong to the same class and 4) extract high-level features which may not be recognisable by human eyes. As reported in the literature review, the common issues among the developed CAD systems are that they cannot deal with images with 1) a wide range of intensity variations in the histological images, 2) high intra-class variations in the size of different cells within

the same class in the histological images and 3) non-linear features which may be difficult to spot by the human eye.

In the case of neuroblastoma, a small number of methods such as the proposed approaches by Tafavogh et al. (2014) and Kong et al. (2009), are based on the geometry and appearance of the different cells. However, there is a wide variety of intensity in neuroblastoma histological images. Moreover, neuroblast cells have different shapes within the same classification group which demonstrates the high intra-class variation of neuroblastoma. As mentioned in section 2.3, Sertel et al. (2009) proposed the only CAD system to classify neuroblastoma histological images based on the handcrafted features. However, the huge variation of shape and texture between neuroblast cells from images which belong to the same category make the classification based on handcrafted features more challenging.

In this thesis, different CAD systems are proposed to address the aforementioned issues. For more evaluations, the proposed algorithms are applied to different datasets.

## Chapter 3

# Models of Computer-Aided Diagnosis Systems Based on Low-level Features

This chapter addresses **Contribution 1, 2 and 3** listed in Section 1.5 by 1) collecting a set of histological images of neuroblastic tumours, 2) developing a local feature extraction approach to extract features which are robust to intra-class variations of cells of different intensity, and 3) proposing a feature extraction algorithm for extracting discriminative features which are robust to scale variations.

**Contribution 1** is due to the lack of large and publicly available image datasets for the analysis of neuroblastic tumours, which significantly hinders development and validation of methods. Therefore, to undertake this research, a dataset of images from neuroblastic tumours was collected.

**Contribution 2** is motivated by the fact that cells in neuroblastoma histological images have high intra-class variations in intensity because of variations in staining, fixation, tissue thickness and inherent variability. This issue is addressed by the proposal of the Patched Completed Local Binary Pattern (PCLBP) algorithm. This algorithm extracts local features which are robust to high intra-class variations.

**Contribution 3** arises from the fact that neuroblast cells have different sizes for tumours of the same class. This research combines the Scale Invariant Feature Transform (SIFT) algorithm with bag of features to extract specific features which are robust to high intra-class variations in size.

This chapter contains three main sections. First, Section 3.1 describes the details of the constructed dataset of neuroblastoma histological images. Second, Section 3.2 introduces patched completed local binary pattern in local regions to extract low-level features which are robust to intensity variations. Finally, Section 3.3 explains the scale-invariant feature transform approach to extract features which are robust to size variations.

### 3.1 Data Collection and Cropping

This section addresses Contribution 1 of this thesis, constructing a dataset which consists of images of neuroblastic tumours. The images that are used in this thesis were collected by the Tumour Bank of the Kid's Research Institute at the Children's Hospital at Westmead, Sydney, Australia. The tumour bank is compliant with local policy, national legislation, and ethical mandates in relation to the use of human tissue in research. All patient-specific details were removed and a de-identified dataset was used for this research. The ethics approval for this study was obtained from the children's hospital at Westmead on Human Research Ethics Committee (HREC) reference number LNR/17/SCHN/195.

The initial dataset was generated from cancer tissue biopsy slides and consisted of images of haematoxylin and eosin (H&E) stained tissue microarrays (TMAs) of



neuroblastic tumours that were in .svs format with resolution  $0.2\mu\text{m}$ . Staining with H&E allows the identification of different histological structures to enable discrimination between types of cells. This research used six TMA slides and seven whole sections representing 125 patients. Each TMA slide contains from 20 to 40 cores of different neuroblastic tumours. The TMA images were generated by the Aperio ScanScope system scanned to  $20\times$  objective lens; images were viewed and extracted using ImageScope software (*ImageScope, pathology slide viewing software* 2016).

Samples of the TMA slides containing cores of different neuroblastic tumours are shown in Figures 3.1 and 3.2, respectively. The diameter of each TMA core is 1.2mm, they are cut at  $3\mu\text{m}$  thickness and then stained with H&E method.

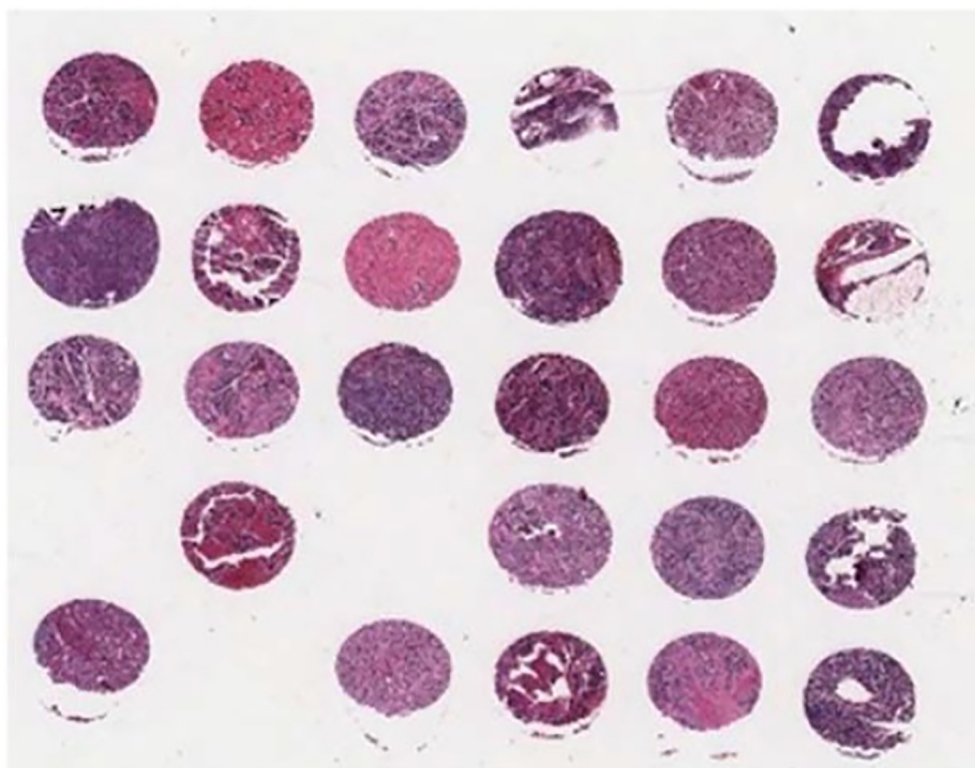


Figure 3.1: A sample of tissue microarray (TMA) slide

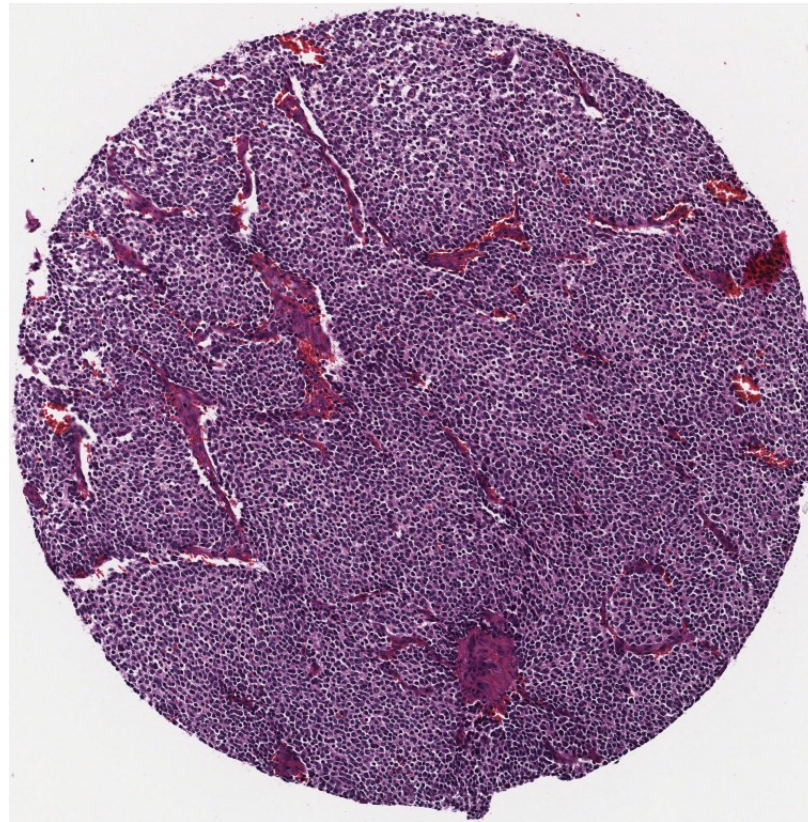


Figure 3.2: A sample of core from a single tumour

Although most cores originate from different patients, some of them are duplicates of patients. All the collected tissue samples were classified before construction of the TMA as undifferentiated neuroblastoma, poorly-differentiated neuroblastoma, differentiating neuroblastoma, ganglioneuroblastoma, or ganglioneuroma by pathologists. Representative images in each of the categories are shown in Figure 3.3. As can be seen, Figure 3.3 (a) represents undifferentiated neuroblastoma with the blue colour in the background, there is no identifiable neuropil in this subtype. The poorly-differentiated neuroblastoma is defined in Figure 3.3 (b) with a recognisable neuropil (pink area) in the background. The differentiating neuroblastoma which contains

neuroblast cells is shown in Figure 3.3 (c). The ganglioneuroblastoma representative is indicated in Figure 3.3 (d) with three specifications; ganglion cells, schwannian stroma and neuropil in the background. The ganglioneuroma is shown in Figure 3.3(e) consisting of fully mature ganglion cells and schwannian stroma in the background.

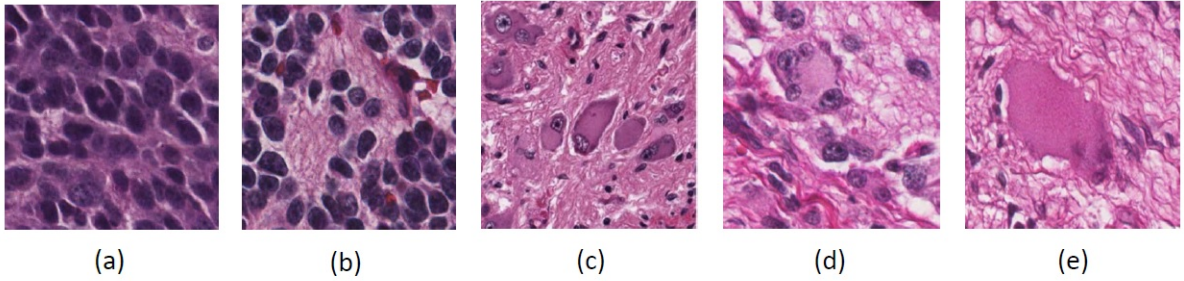


Figure 3.3: Representative images in neuroblastic tumour categories: (a) undifferentiated neuroblastoma, (b) poorly-differentiated neuroblastoma, (c) differentiating neuroblastoma, (d) ganglioneuroblastoma, and (e) ganglioneuroma.

Due to the size of the scanned neuroblastic tumour images, up to a gigabyte each, the computational cost is high. Therefore, they were cropped to equal-sized square regions ( $300 \times 300$  pixels) by an expert histopathologist in which all the specifications of a particular subtype of a neuroblastic tumour are present in the cropped image. At  $40\times$  viewing magnification, the cropped image size was  $300 \times 300$  pixels with real size  $80 \times 80 \mu\text{m}$  which is approximately forty percent of the area of an optical microscope high power field of view. The quantitative actual size of tumour cores relative to the cropped images is shown in Figure 3.4. Information on the images and patients in the dataset is given in Table 3.1. As can be seen, the highest number of images belongs to the poorly-differentiated neuroblastoma. The number of differentiating neuroblastoma and undifferentiated neuroblastoma decrease to 187 and 155, respectively. This

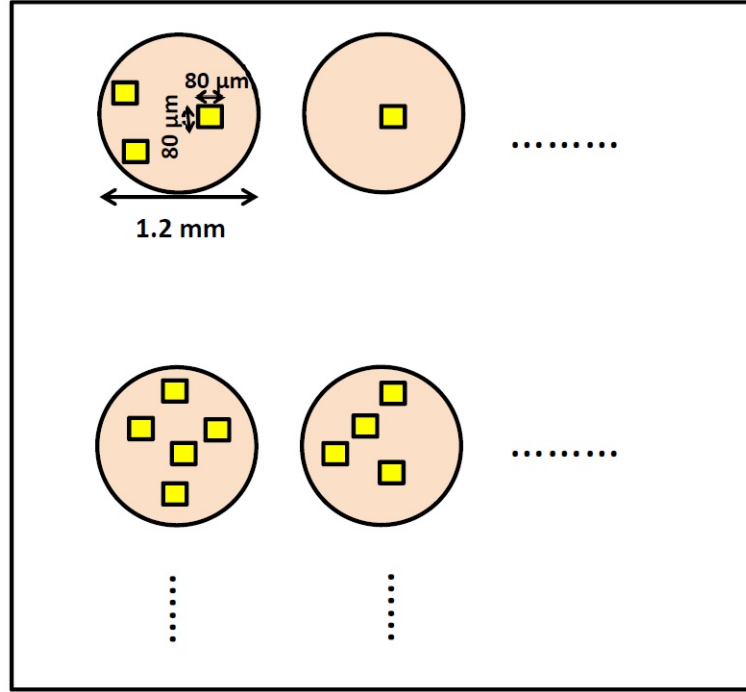


Figure 3.4: Quantitative relative and actual size of tissue cores and cropped images

number drops to 84 images in ganglioneuroma. The lowest number of images belongs to the ganglioneuroblastoma which is 46. In this research the collected images representing 125 patients are classified using different methods.

## 3.2 Patched Completed Local Binary Pattern

Neuroblastoma histological images have different variations in intensity which demonstrate the high intra-class variation observed in neuroblastoma. This characteristic makes their classification challenging for CAD systems. As a result, this research develops an algorithm to extract local features which are robust to the intra-class variations of intensity.

Table 3.1: Number of different categories of neuroblastic tumour cropped images and number of patients

Category of Neuroblastic Tumour	Number of Cropped Images	Number of Patients
Poorly-differentiated	571	77
Differentiating	187	12
Undifferentiated	155	10
Ganglioneuroma	84	18
Ganglioneuroblastoma	46	8
<b>Total</b>	<b>1043</b>	<b>125</b>

This section addresses Contribution 2 of this thesis, namely extracting local features that are robust to intra-class variations of cells of different intensity in the same class. The proposed aches which are more robust to variations in the intensity of different cells within the same class in neuroblastoma histological images.

### 3.2.1 Methodology

The proposed Patched Completed Local Binary Pattern (PCLBP) algorithm is based on Completed Local Binary Pattern (CLBP) (Guo et al. 2010), one of the recent variants of Local Binary Pattern (LBP) (Ojala et al. 2002). The LBP operator computes the distribution of binary patterns in the circular neighbourhood characterized by a radius  $R$  and a number of neighbours  $P$ . The idea is to threshold neighbouring pixels, compared to the central pixel to the  $P$  neighbours. If the intensity of a neighbour pixel is greater than or equal to that of the central pixel, the value 1 is assigned,

otherwise 0. Therefore, a binary pattern is obtained from the neighbourhood. The LBP function at pixel  $p$  is (Ojala et al. 2002)

$$LBP(f(X, Y)) = \sum_{i=0}^{P-1} 2^i \cdot u(f(X_i, Y_i) - f(X, Y)) \quad (3.2.1)$$

where  $f(X_i, Y_i)$  and  $f(X, Y)$  are gray levels of pixels  $(X_i, Y_i)$  and central pixel. The  $u(\cdot)$  is the unit step function defined as

$$u(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (3.2.2)$$

The CLBP is the completed modeling of LBP which is based on three components extracted from the local region: the center pixel, sign, and magnitude. The center pixel is coded by a binary code after thresholding, with the threshold set as the average gray level of the whole image. In order to compute the sign and magnitude, a neighbourhood of radius  $R$  and number of neighbours  $P$  is considered. Signs and magnitudes are computed and coded into a binary format so that they can be combined to form the final CLBP histograms (Guo et al. 2010).

Patched Completed Local Binary Pattern (PCLBP), the proposed approach, extends Completed Local Binary Pattern. The overall framework consists of four stages: 1) patched images, 2) sign and magnitude binary patterns, 3) histogram of sign and magnitude binary patterns and 4) classification, as shown in Figure 3.5. First, the images are partitioned into 25 equal-sized square patches with size  $60 \times 60$  pixels. Second, Sign Binary Patterns (SBPs) and Magnitude Binary Patterns (MBPs) are computed within patches. Third, histograms of SBPs and MBPs are computed and



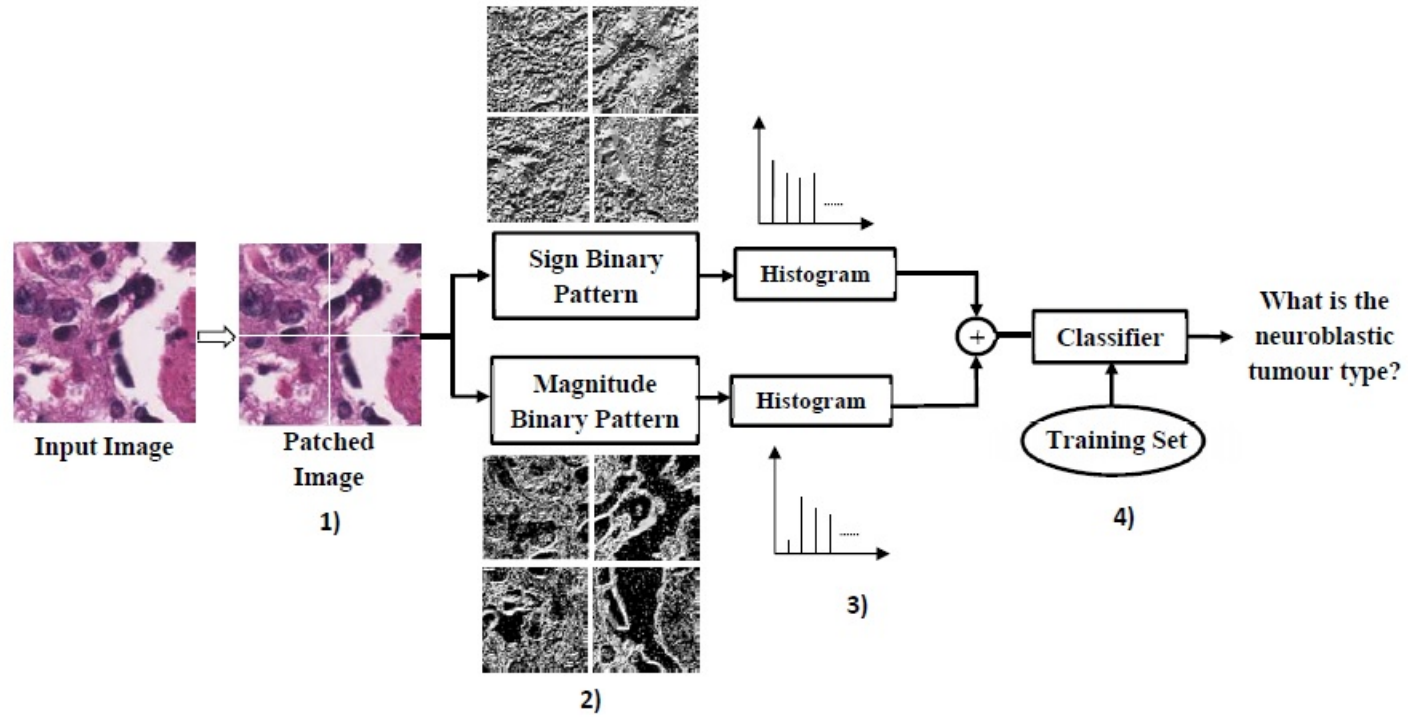


Figure 3.5: The scheme of the proposed method which consists of four stages: 1) patched images, 2) sign and magnitude binary patterns, 3) histogram of sign and magnitude binary patterns and 4) classification.

concatenated to build a feature vector for each patch. A feature vector for the whole image is created by concatenating the feature vectors of all patches. Finally, the input image is classified by comparing the related feature vector with the feature vectors of all images in the training set. The algorithm is described in detail as follows.

An  $N \times N$  pixel input image is partitioned into  $W \times W$  pixel non-overlapping patches. All patches are indicated with  $p$  and  $q$  indices, ranging from 1 to  $N/W$ , as

$$p = \left\lfloor \frac{X}{W} \right\rfloor + 1, \quad q = \left\lfloor \frac{Y}{W} \right\rfloor + 1 \quad (3.2.3)$$

where  $0 \leq X \leq N$  and  $0 \leq Y \leq N$  are the coordinates of the input neuroblastoma image and  $\lfloor \cdot \rfloor$  is the floor function. The  $(p,q)$ th patch in the input neuroblastoma image (see Figure 3.6) is defined as

$$f_{p,q}(X^{pq}, Y^{pq}) = f(W(p-1) + X^{pq}, W(q-1) + Y^{pq}) \quad (3.2.4)$$

where  $f(X, Y)$  denotes the original image and  $X^{pq}, Y^{pq}$  denote the coordinates of the  $(p,q)$ th patch.

The local differences of the 8-neighbourhood around  $(X_0^{pq}, Y_0^{pq})$ , as shown in Figure 3.7, are computed as

$$f'_{p,q,k}(X_0^{pq}, Y_0^{pq}) = f_{p,q}(X_k^{pq}, Y_k^{pq}) - f_{p,q}(X_0^{pq}, Y_0^{pq}) \text{ for } k = 1, \dots, 8 \quad (3.2.5)$$

The Sign Binary Pattern of the  $(p,q)$ th patch,  $SBP_{p,q}(f_{p,q}(X_0^{pq}, Y_0^{pq}))$ , is defined



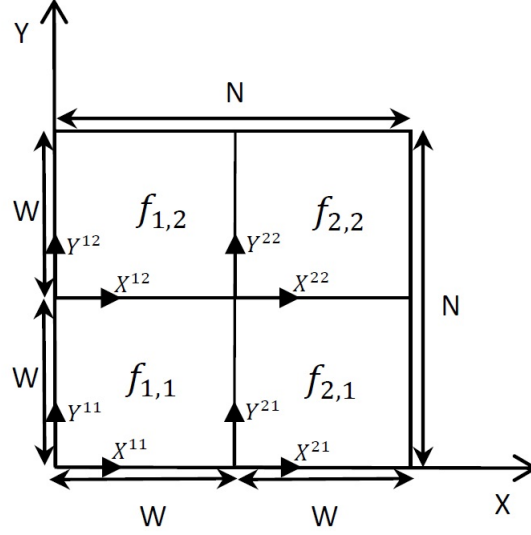


Figure 3.6: Coordinates in the patched image. Here, this research assumes  $W = N/2$  for better visualisation.

as the concatenation of 8 bits as

$$SBP_{p,q}(f_{p,q}(X_0^{pq}, Y_0^{pq})) = \left\{ u(f'_{p,q,1}(X_0^{pq}, Y_0^{pq})), \dots, u(f'_{p,q,8}(X_0^{pq}, Y_0^{pq})) \right\} \quad (3.2.6)$$

where  $u(x)$  is defined as

$$u(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (3.2.7)$$

and the Magnitude Binary Pattern of the  $(p,q)$ th patch,  $MBP_{p,q}(f_{p,q}(X_0^{pq}, Y_0^{pq}))$ , is defined as

$$MBP_{p,q}(f_{p,q}(X_0^{pq}, Y_0^{pq})) = \left\{ u(f'_{p,q,1}(X_0^{pq}, Y_0^{pq}) - m), \dots, u(f'_{p,q,8}(X_0^{pq}, Y_0^{pq}) - m) \right\} \quad (3.2.8)$$

$(X_1^{pq}, Y_1^{pq})$	$(X_2^{pq}, Y_2^{pq})$	$(X_3^{pq}, Y_3^{pq})$
$(X_8^{pq}, Y_8^{pq})$	$(X_0^{pq}, Y_0^{pq})$	$(X_4^{pq}, Y_4^{pq})$
$(X_7^{pq}, Y_7^{pq})$	$(X_6^{pq}, Y_6^{pq})$	$(X_5^{pq}, Y_5^{pq})$

Figure 3.7: An 8-neighbourhood around  $(X_0^{pq}, Y_0^{pq})$ 

where  $m$  is a threshold to be set as the mean of the absolute values of all  $f'_{p,q,k}(X_0^{pq}, Y_0^{pq})$  from the whole image.

After computing the SBP and the MBP, they are converted into decimal values as

$$DSBP_{p,q}(f_{p,q}(X_0^{pq}, Y_0^{pq})) = \sum_{l=1}^8 2^{l-1} SBP_{p,q}^l(f_{p,q}(X_0^{pq}, Y_0^{pq})) \quad (3.2.9)$$

$$DMBP_{p,q}(f_{p,q}(X_0^{pq}, Y_0^{pq})) = \sum_{l=1}^8 2^{l-1} MBP_{p,q}^l(f_{p,q}(X_0^{pq}, Y_0^{pq})) \quad (3.2.10)$$

where  $SBP_{p,q}^l(f_{p,q}(X_0^{pq}, Y_0^{pq}))$  and  $MBP_{p,q}^l(f_{p,q}(X_0^{pq}, Y_0^{pq}))$  denote the  $l$ -th bit of the SBP and the MBP, respectively. The Decimal Sign Binary Pattern (DSBP) and Decimal Magnitude Binary Pattern (DMBP) for each pixel in the neuroblastoma image are computed according to equations (3.2.9) and (3.2.10), respectively. Figure 3.8

shows an example of DSBP and DMBP for an arbitrary representative neuroblastoma image. As can be seen, DSBP and DMBP operators both can capture the special pattern and contrast of local image texture such as edges and corners.

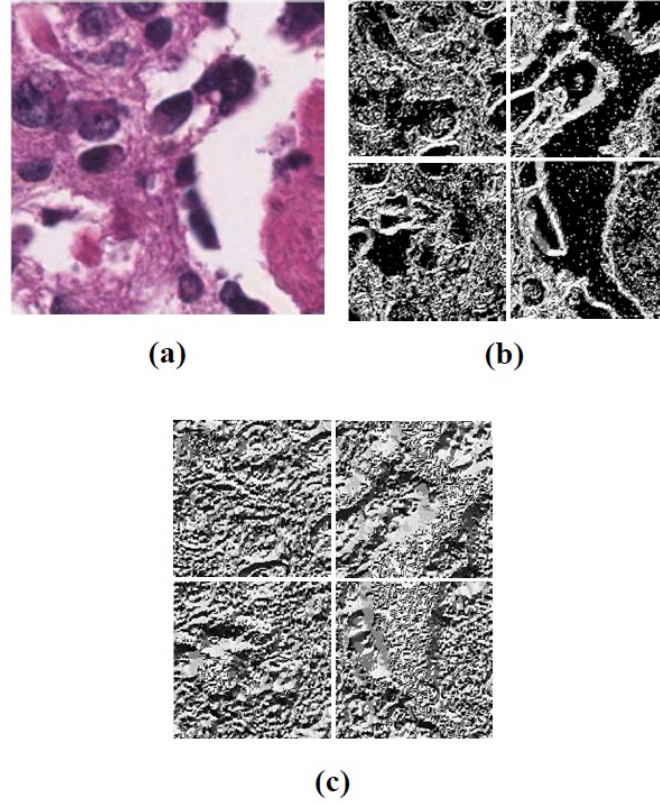


Figure 3.8: An example of the computed Magnitude Binary Pattern (MBP) and Sign Binary Pattern (SBP): (a) original image, (b) MBP, and (c) SBP. We assume  $W = 150$  pixels in this image for better visualization. However, the real  $W$  is 60 pixels.

For each patch, this research models the distribution of DSBP and DMBP using the histogram operator. For an 8-bit gray-scale image there are 256 different intensities in the range  $[0, 255]$ . Therefore, the histogram displays 256 numbers showing the distribution of pixels among different gray-scale values.

$$HSBP_{p,q}(f_{p,q}(X, Y)) = H\left\{DSBP_{p,q}(f_{p,q}(X, Y))\right\} \quad (3.2.11)$$

$$HMBP_{p,q}(f_{p,q}(X, Y)) = H\left\{DMBP_{p,q}(f_{p,q}(X, Y))\right\} \quad (3.2.12)$$

This section concatenates the histograms of the SBP and the MBP for each patch to build a local histogram (LH) within each patch

$$LH_{p,q}(f_{p,q}(x, y)) = \left\{HSBP_{p,q}(f_{p,q}(x, y)), HMBP_{p,q}(f_{p,q}(x, y))\right\} \quad (3.2.13)$$

Then, the histograms of all patches are concatenated to build the PCLBP for the neuroblastoma image (as shown in Figure 3.9)

$$PCLBP = \left\{LH_{p,q}(f_{p,q}(X, Y)) \mid p, q = 1, \dots, N/W\right\} \quad (3.2.14)$$

Finally, two classifiers are used to classify the extracted PCLBPs: k-nearest neighbour ( $k$ -NN) (Cover & Hart 1967) and Support Vector Machine (SVM) (Cortes & Vapnik 1995). This research chooses k-NN because of its low cost learning process. Also, SVMs can efficiently perform a non-linear classification using different kernels to map the inputs into the high-dimensional feature spaces. Although this research uses k-NN and SVM, the PCLBP features can be classified using other existing classifiers without customization.

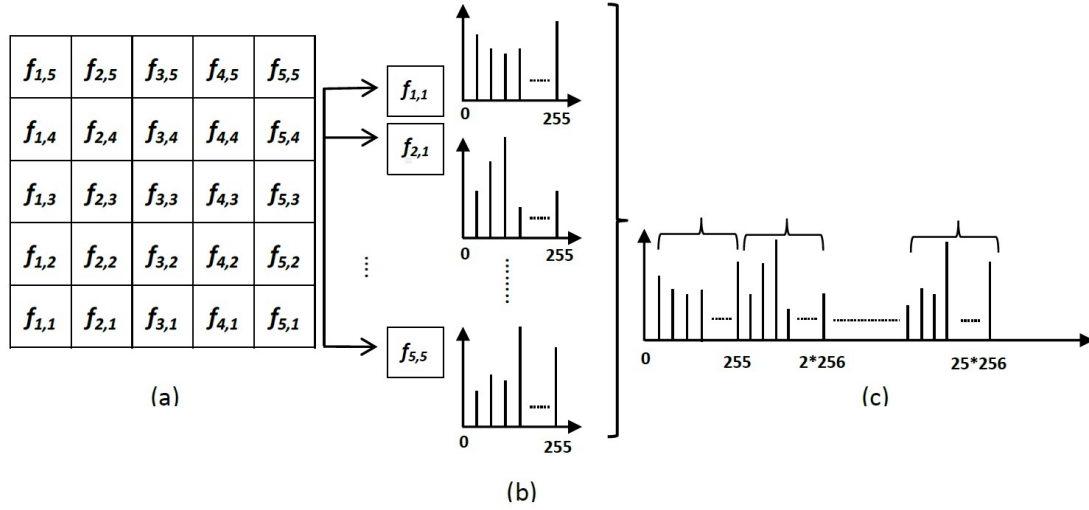


Figure 3.9: Concatenation of patches' histograms: (a) patched image, (b) histogram of patches, and (c) concatenation of patches' histograms.

### 3.2.2 Experiment Results

In this section, the performance of the proposed method for the classification of neuroblastoma histological images is evaluated. Experiments are conducted on the collected neuroblastic tumour database. As shown in Figure 3.10, the dataset is divided randomly into two subsets: parameter-tuning (211 images) and validation (832 images). This section selects the optimum values for free parameters using the parameter-tuning dataset and fixes them for the validation. Then, the method is evaluated using the validation dataset and selected parameters. To make a more accurate comparison between the different methods, this research uses the same division with the same images throughout the thesis.

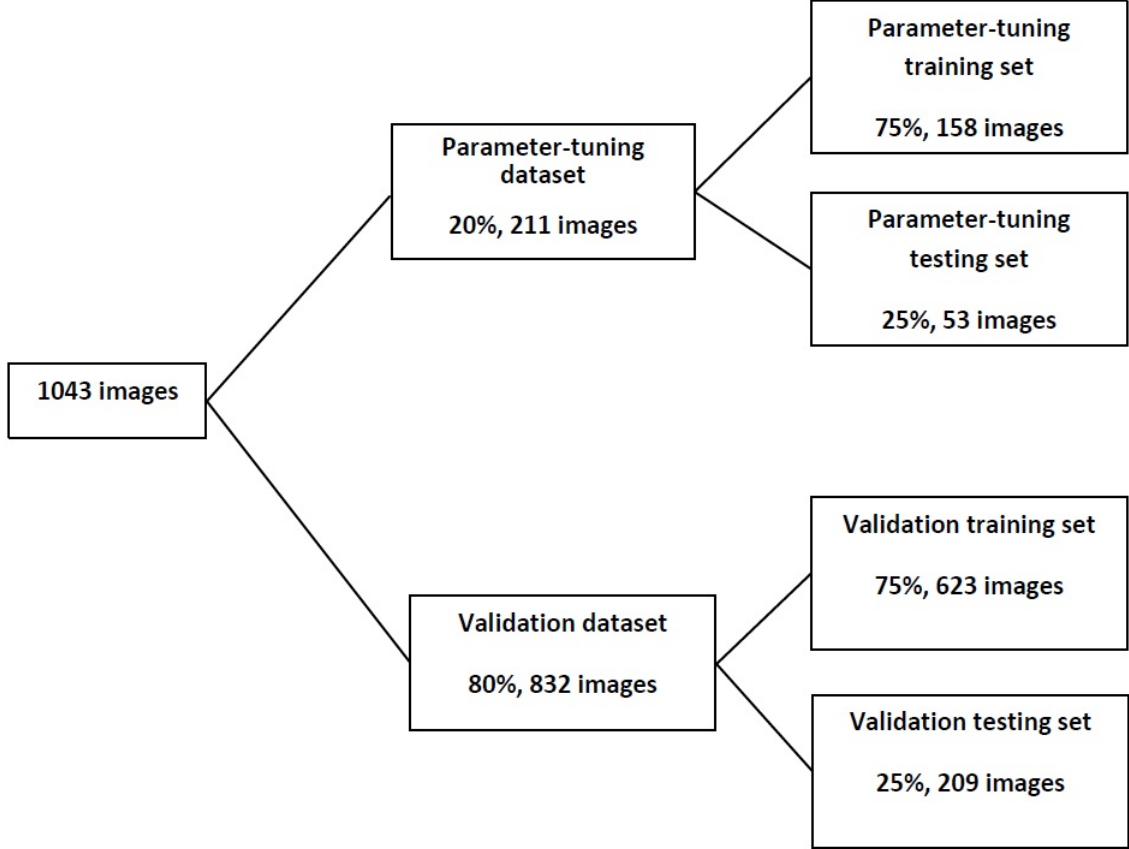


Figure 3.10: Division of dataset with 1043 neuroblastoma images

### Parameter Tuning

This research uses the wrapper technique (Kohavi & John 1998) to find the parameter values that maximize classification accuracy on the parameter tuning dataset. The parameter-tuning dataset is randomly divided into training (75%, 158 images) and testing (25%, 53 images) subsets. The algorithm is trained using the parameter-tuning training set and computes the accuracy of the technique using the parameter-tuning testing set to tune the free parameters. To make a better estimation of the accuracy, the above procedure is repeated multiple (10) times and the average is computed over

all the experiments.

For  $k$ -NN, the free parameters are the width of the patch ( $W$ ) and the  $k$  numbers of neighbours in the feature space. Accuracy was computed for  $k$  ranging from 1 to 10 and patch width  $W \in \{10, 15, 30, 50, 60, 75, 100, 150, 300\}$  pixels and the results are shown in Figure 3.11. The highest average classification accuracy of 72.7% was found with  $W=60$  pixels and  $k=5$ . So, these values are set in the next experiments.

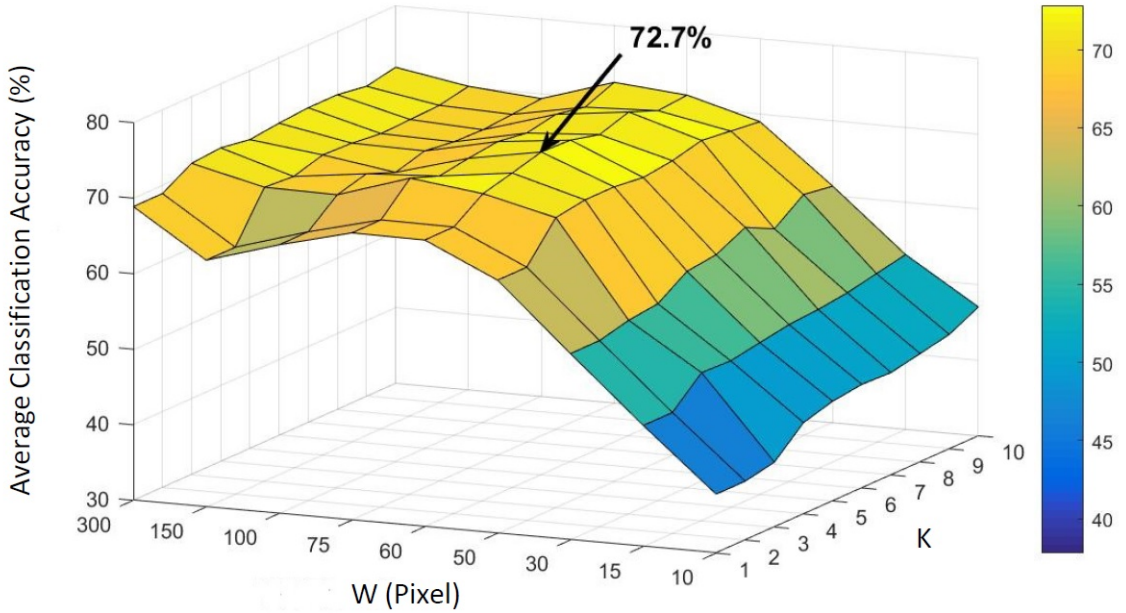


Figure 3.11: Accuracy of  $k$ -NN classifier versus patch width ( $W$ ) and  $k$  in parameter tuning of  $k$ -NN classifier

This section used the C-SVC type (Boser et al. 1992) of SVM using the LIBSVM tool (Chang & Lin 2011). The  $C$  is a regularization parameter that controls the trade off between a smooth decision boundary and classifying training points correctly. The classification accuracy was calculated for different values of  $C$ , the average accuracy are shown in Table 3.2. The highest accuracy was achieved when  $C$  was 10. This research tested different kernels: linear, polynomial, Radial Basis Function (RBF),

Table 3.2: Average classification accuracy of SVM over neuroblastic tumour dataset using different C values

C values	Classification Accuracy (%)
0.01	$64.15 \pm 1.17$
0.1	$68.54 \pm 2.14$
1	$71.35 \pm 1.35$
10	$75.14 \pm 3.18$

and sigmoid. The RBF and polynomial parameters are  $\gamma = 1/2\sigma^2$  and  $d$ , respectively. These parameters were empirically defined through the experiments with the best value taking  $1/256$  for  $\gamma$  and 3 for  $d$  (256 is the number of different intensities in the images and  $d$  is the degree of polynomial kernel function). In this section, different kernel functions are tested and the average classification accuracy is given in Table 3.3. The best result is achieved using the RBF kernel, so it is used in the next experiments. However, the results for RBF and Polynomial kernels are overlapping with considering the variances.

### System Validation

To evaluate the proposed method, this research uses the validation dataset which is not used in the parameter tuning section. As shown in Figure 3.10, the validation dataset is randomly divided into validation training (75%, 623 images) and validation testing subsets (25%, 209 images). The final model is trained using the validation



Table 3.3: Average classification accuracy of SVM over neuroblastic tumour dataset using different kernel functions

Kernel	Classification Accuracy (%)
Linear	$62.22 \pm 3.14$
Polynomial	$71.82 \pm 1.78$
Radial Basis Function	$72.33 \pm 1.15$
Sigmoid	$49.45 \pm 2.76$

training set and computes the accuracy of the method using the validation testing set. This procedure is repeated multiple (10) times and the average F-measure is reported.

The algorithm performance for  $k$ -NN and SVM is reported with the average F-measure, recall, and precision (Powers 2011) weighted by the number of examples in each of the five classes. The distribution of the computed F-measures for two classifiers and feature extraction approaches over the ten trials is presented in Figure 3.12 which shows that SVM is superior to  $k$ -NN.

The  $t$ -test with  $P$  value = 0.03 and  $\alpha=0.05$  (significance level) shows that the combination of PCLBP with an SVM classifier significantly improves the accuracy of classification in comparison with the  $k$ -NN classifier. As shown in Table 3.4, the proposed algorithm obtains approximately 5% higher accuracy when it is combined with the SVM classifier compared to the  $k$ -NN classifier.

This section also tests the Completed Local Binary Pattern (CLBP) algorithm (Guo

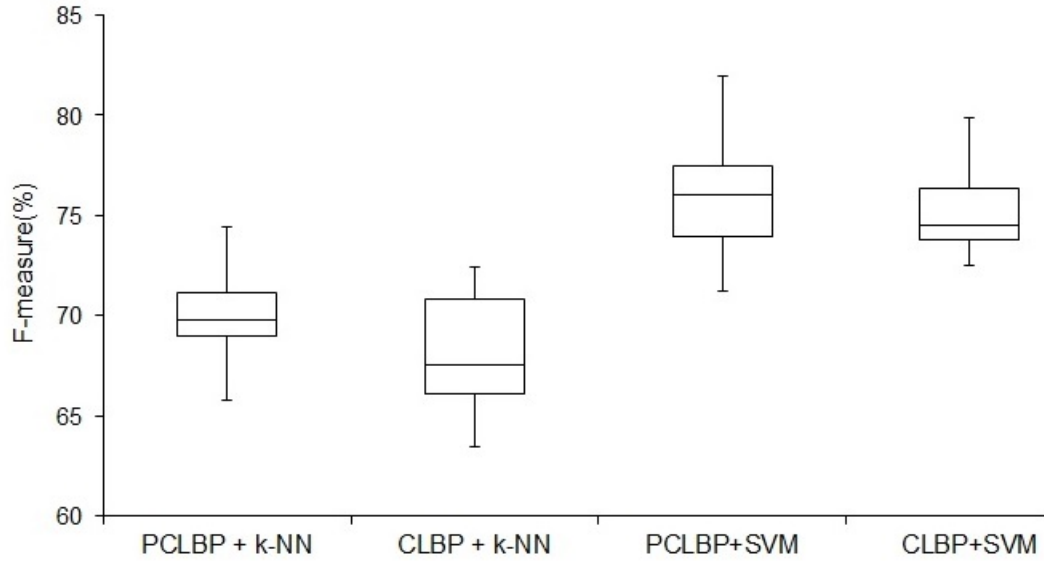


Figure 3.12: Comparison between the proposed method (PCLBP) and Spanhol's method (CLBP)

et al. 2010) on the test images as a benchmark, again reporting the weighted average of precision, recall, and F-measure. The weighted average precision, recall, and F-measure of PCLBP and CLBP are reported in Table 3.4. The weighted average precision, recall, and F-measure of PCLBP are better than CLBP, although not statistically significantly so.

### 3.3 Scale Invariant Feature Transform

One of the challenges with the histological images of neuroblastoma is the high intra-class variation in the size of the nuclei and neuroblast cells. This section continues to address contribution 3 by extracting the features which are invariant to variations in the size of the different cells in the neuroblastoma histological images.

Table 3.4: Weighted average precision, recall, and F-measure obtained by the proposed method (PCLBP) and Spanhol's method (CLBP).

	PCLBP (%)	CLBP (%)
Precision kNN	70.49±3.37	68.22±3.23
Precision SVM	75.59±3.15	74.1±2.35
Recall kNN	71.02±2.87	69.53±2.81
Recall SVM	76.35±3.41	76.25±2.23
F-measure kNN	70.75±3.09	68.86±6.04
F-measure SVM	75.96±3.27	75.15±2.28

### 3.3.1 Methodology

The overall framework of the proposed approach is shown in Figure 3.13. The Scale Invariant Feature Transform (SIFT) proposed by Lowe (2004) extracts features from the test images and feeds them into the feature encoding block to summarize them into the new discriminative features. This research aggregates the bag of feature algorithm (Lazebnik et al. 2006) as the feature selector with SIFT to extract more discriminative features. Here, the SVM is used to classify the test images by comparing their extracted features with the training set images. The SIFT descriptor extracts a feature vector with 128 elements for each keypoint. Using all of the extracted features in the classification of neuroblastoma histological images is time-consuming and has a high computational complexity. So, this research combines SIFT with the bag of features algorithm that was proposed by Lazebnik et al. (2006) as a feature encoding method to compute more discriminative representations from neuroblastoma histological images. The scheme of the bag of features is shown in Figure 3.14. It starts by

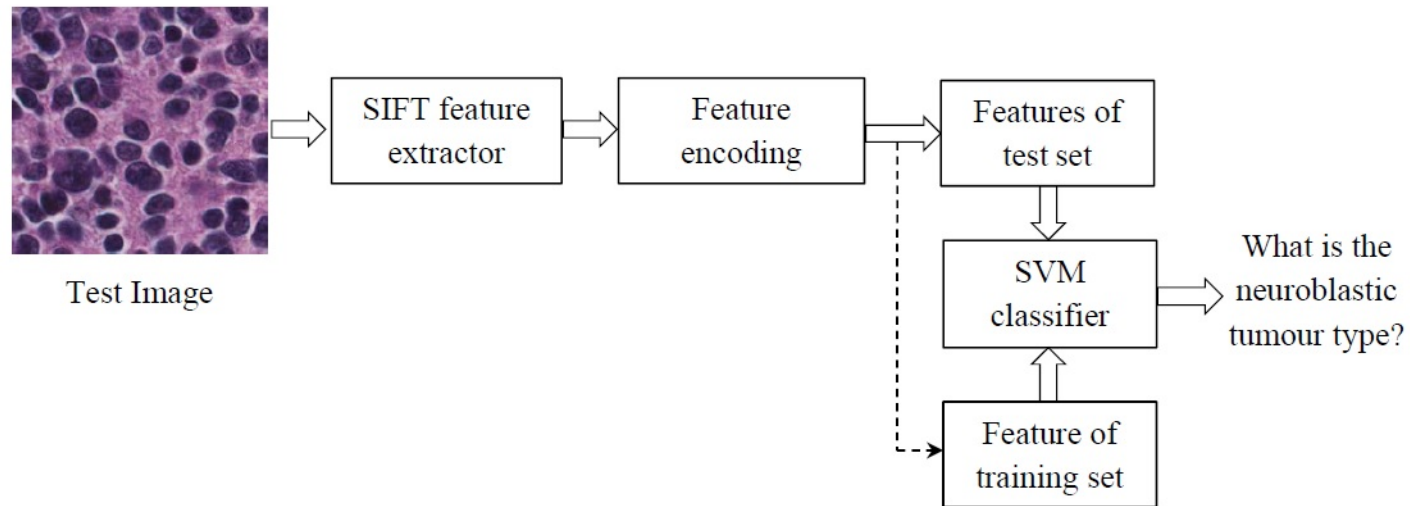


Figure 3.13: The scheme of the combination of SIFT with feature encoding

selecting keypoints (step a) and then describes the keypoints from the input image (step b) by extracting the feature vectors with 128 elements for each keypoint. In the next step (c), the extracted features are clustered and the codebook which consists of the codewords is constructed in step (d). Codewords are the extracted features from the keypoints. Then, a bag of features histogram counts the frequency of each feature in step e.

SIFT has three free parameters which must be set. The first one is  $\sigma$ , the width of the Gaussian. Increasing  $\sigma$  reduces the number of features generated from an image (Tang et al. 2008). The second free parameter is the contrast threshold ( $C_C$ ) which eliminates keypoints with low contrast. The third free parameter is the edge threshold ( $C_E$ ) which eliminates the unstable keypoints that are near the edges. These parameters are set in Section 3.3.2.

This research uses the k-means clustering algorithm proposed by Frogy (1965) to construct a codebook in the bag of feature algorithm. All the extracted features from the keypoints are clustered to find a set of centroids. An important factor in the construction of the codebook is the selection of the number of codewords. This research evaluates different codebook sizes and selects the best one with the highest classification accuracy; the experiment results are shown in Section 3.3.2. The input image is represented by a histogram of codewords using the extracted features from keypoints. In the final stage, histological images are classified by training a classifier which classifies the histograms of the feature encoding block. This section uses an SVM classifier which can efficiently perform a non-linear classification using an optimal hyperplane in a high-dimensional space.

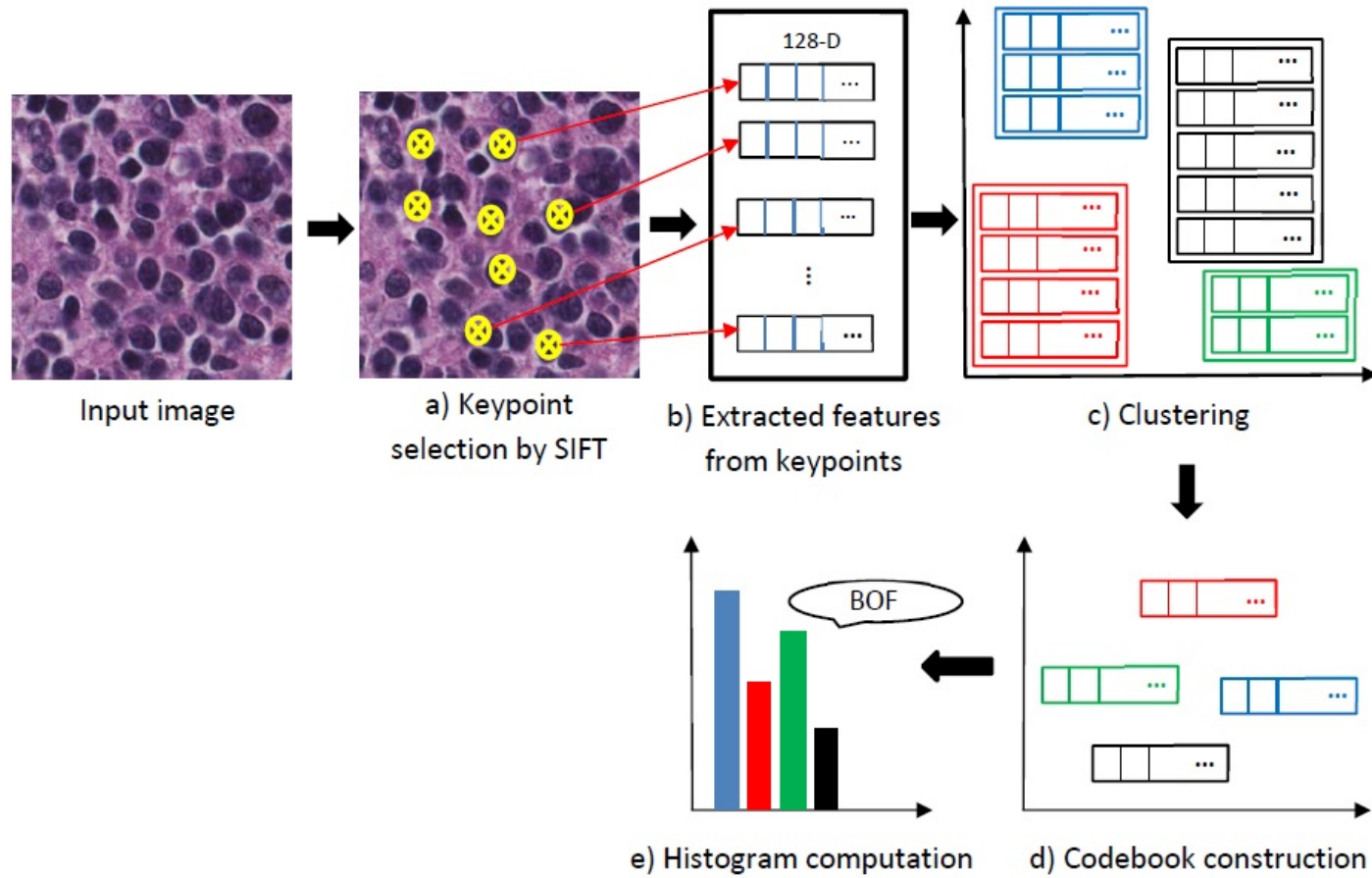


Figure 3.14: Scheme of the feature encoding block: a) selected keypoints by SIFT b) 128-D extracted features from keypoints c) clustering the 128-D extracted features d) construction of the codebook using clustering e) histogram of codewords in codebook

### 3.3.2 Experiment Results

The performance of the proposed approach is evaluated on the constructed dataset of the histological images of neuroblastoma. This section uses the split of the data that was described in Section 3.2.2 (see Figure 3.10). The optimum values for the free parameters of the proposed approach are selected using the parameter-tuning set (211 images) and the approach is evaluated using the validation set (832 images).

#### Parameter Tuning

SIFT has three parameters to be selected: Gaussian  $\sigma$ , contrast threshold ( $C_C$ ), and edge threshold ( $C_E$ ). This section trains the algorithm using the parameter-tuning training set (158 images) and computes the accuracy of the algorithm using the parameter-tuning testing set (53 images) to tune the free parameters of the algorithm. The dataset is randomly split multiple (10) times and the average accuracy computed over all the experiments to obtain an accurate result. SIFT has three parameters,  $\sigma$ ,  $C_C$ ,  $C_E$  that should be tuned. This research uses the greedy method to tune these parameters one by one while holding the other parameters constant. SIFT is initially tested with  $\sigma = 1.6$ , as  $C_C$  and  $C_E$  are 0.03 and 10, respectively; these values are selected based on Lowe's work (2004). This research evaluates  $\sigma$  in 1.4 to 1.8 in step of 0.1. According to Table 3.5, the best accuracy is obtained when  $\sigma$  is 1.7. So,  $\sigma$  is set to 1.7 in the next experiments. The mentioned accuracies are the average over ten experiments. The parameters of the SVM are selected according to Section 3.2.2.

This section described the results of another experiment to find the optimum value for  $C_C$  while the  $C_E$  is 10. According to Table 3.6, the best accuracy was found when  $C_C$  is 0.04. Also, this research identified the optimum value for  $C_E$ .  $C_E$  was evaluated

Table 3.5: Average classification accuracy of the SIFT over ten experiments on the neuroblastic tumour dataset using different values for  $\sigma$ ,  $C_c = 0.03$ ,  $C_E = 10$

$\sigma$	Classification Accuracy (%)
1.4	$73.25 \pm 2.4$
1.5	$75.24 \pm 2.76$
1.6	$76.52 \pm 1.1$
1.7	<b><math>78.96 \pm 1.72</math></b>
1.8	$76.24 \pm 3.54$

for five different values and, as shown in Table 3.7, the best accuracy was obtained when  $C_E$  is 11. Furthermore, this section describes the proposed technique for five different codebook sizes, starting with 300, followed by 400, 500, 600 and 700. The average classification accuracy versus the size of the codebook is shown in Table 3.8. The best accuracy was obtained when the codebook size was 500. This value was set for the next experiments.

Table 3.6: Average classification accuracy of SIFT over the neuroblastic tumour dataset using different values for the contrast threshold ( $C_C$ ),  $\sigma = 1.7$ ,  $C_E = 10$

Contrast Threshold	Classification Accuracy (%)
0.02	$73.19 \pm 2.7$
0.03	$76.21 \pm 0.94$
0.04	<b><math>76.58 \pm 1.5</math></b>
0.05	$75.68 \pm 1.12$
0.06	$74.25 \pm 1.88$



Table 3.7: Average classification accuracy of SIFT over the neuroblastic tumour dataset using different values for the edge threshold ( $C_E$ ),  $\sigma = 1.7$ ,  $C_c = 0.04$

Edge Threshold	Classification Accuracy (%)
10	$78.25 \pm 2.63$
11	<b><math>81.76 \pm 1.6</math></b>
12	$77.24 \pm 3.7$
13	$76.15 \pm 1.95$
14	$74.18 \pm 1.75$

Table 3.8: Average classification accuracy of the SIFT over the neuroblastic tumour dataset using bag of features with different values of the codebook size.

Codebook Size	Classification Accuracy (%)
300	$78.2 \pm 2.04$
400	$80.31 \pm 2.41$
500	<b><math>82.25 \pm 0.81</math></b>
600	$79.77 \pm 2.72$
700	$81.62 \pm 1.16$

### System Validation

This section evaluates the proposed method using the validation dataset. The final model is trained using the validation training set and tested using the validation testing set. The dataset is randomly split multiple (10) times and the average accuracy

is reported. This section uses the standard bag of features and the pyramid bag of features algorithms (Lazebnik et al. 2006) to evaluate the proposed method. In each case, the RBF and histogram intersection kernels (Maji et al. 2008) for SVM are used. In the histogram intersection SVM, each image is split into blocks and each block is represented by the extracted features from the SIFT keypoints. The histogram of each image is constructed by counting the extracted features from the SIFT keypoints. The histogram Intersection kernel is built based on these histograms.

This section measures the performance of the algorithm by the average F-measure, recall and precision, weighted by the number of samples in each class. The results are shown in Table 3.9. This research compared the combination of SIFT and bag of features with PCLBP and CLBP as benchmarks. The distribution of the computed F-measures for the different approaches over the ten trials is presented in Figure 3.15, showing that the combination of SIFT with the bag of feature is superior to CLBP and PCLBP. Also, this section described the results of t-test (with significance level of 5%) giving p-values of  $1.5 \times 10^{-6}$  and  $1.2 \times 10^{-4}$  for CLBP and PCLBP, respectively. If the p-value is less than 0.05, the null hypothesis that there is no difference between the mean of the f-measures is rejected. Therefore, the combination of SIFT with the bag of features and SVM with the histogram intersection kernel statistically significantly improved the classification accuracy in comparison with CLBP and PCLBP with the RBF kernel SVM. The F-measure accuracy of the combination of SIFT and the bag of features in the best case is 10.04% and 9.23% higher than CLBP and PCLBP, respectively. The weighted average of precision, recall, and F-measure of the proposed algorithm is better than PCLBP and CLBP. A representative confusion matrix is shown in Table 3.10. Interestingly, the poorest computer performance was

in discriminating between poorly-differentiated and differentiating neuroblastoma, a distinction that human pathologists also find difficult in limited fields of view.

Table 3.9: Weighted average precision, recall and F-measure of the SIFT method and other benchmarks. The bold value indicates the best F-measure achieved by combination of SIFT with bag of features and SVM (histogram intersection)

Method	Precision (%)	Recall (%)	F-measure (%)
* CLBP + SVM (RBF)	74.1±2.35	76.25±2.23	75.15±2.28
* PCLBP + SVM (RBF)	75.59±3.15	76.35±3.41	75.96±3.27
SIFT + bag of features + SVM (RBF)	81.62±3.72	84.54±1.66	83.03±2.63
SIFT + pyramid bag of features + SVM (RBF)	80.63±3.55	83.57±1.81	82.08±2.65
SIFT + bag of features + SVM (histogram intersection)	83.81±3.33	86.61±1.87	<b>85.19±2.42</b>
SIFT + pyramid bag of features + SVM (histogram intersection)	83.5±3.64	86.41±1.22	84.93±2.35

---

\* Same data is presented in Table 3.4.

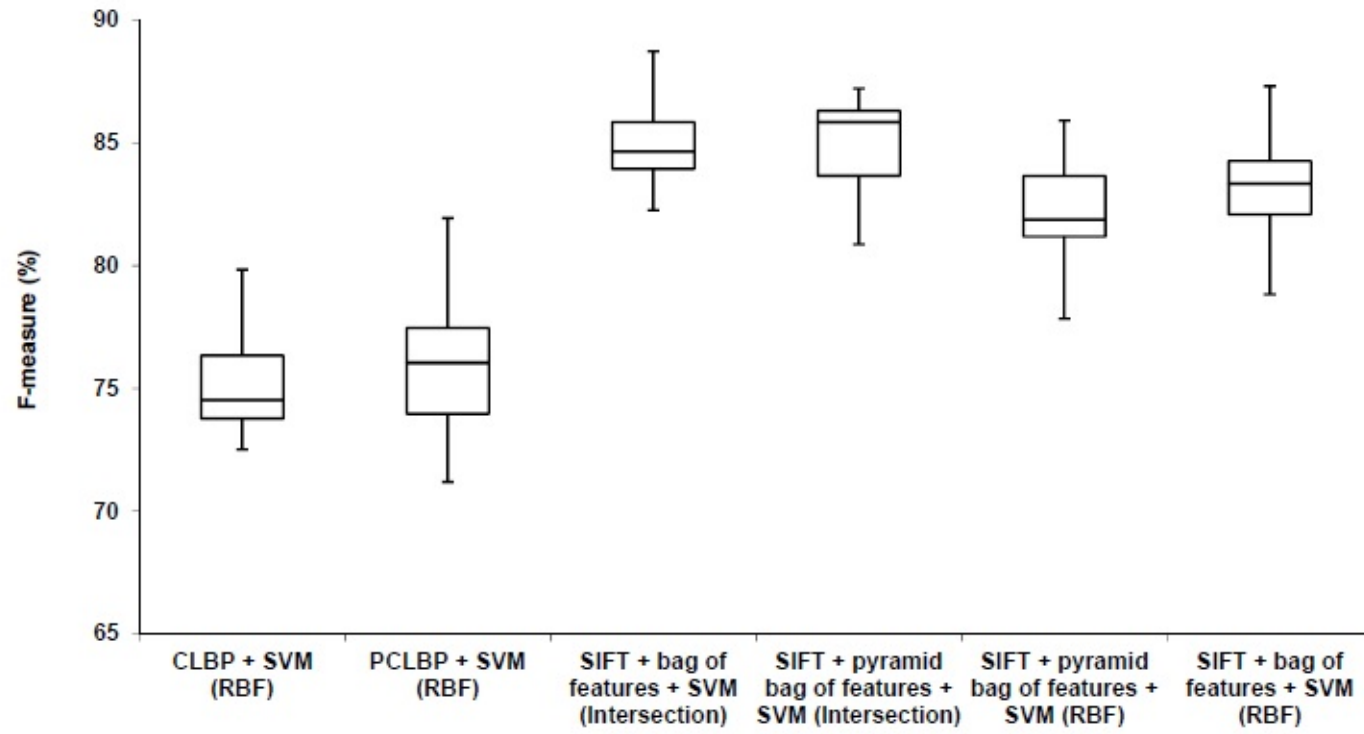


Figure 3.15: Comparison between SIFT + the bag of feature approach with CLBP and PCLBP

Table 3.10: A representative confusion matrix of applying SIFT over the neuroblastoma images. Abbreviations: UD (undifferentiated neuroblastoma), PD (poorly-differentiated neuroblastoma), D (differentiating neuroblastoma), GNB (ganglioneuroblastoma) and GN (ganglioneuroma)

Predicted \ Actual	UD	PD	D	GNB	GN
UD	1	2	0	0	0
PD	0	119	8	0	1
D	0	12	6	4	1
GNB	0	0	1	34	0
GN	0	0	0	5	14

### 3.3.3 Sensitivity to Dataset

This research tested the algorithm on the collected dataset which consisted of 1043 (300×300 pixel) neuroblastoma histological images. As there is no public available dataset of neuroblastoma histological images, for more evaluation, the proposed algorithm was evaluated on two more datasets: 1) second neuroblastoma dataset provided by the University of Bristol and 2) BreakHis breast cancer dataset. The details are described in the following paragraphs.

#### Second Neuroblastoma dataset

The neuroblastoma dataset provided by the University of Bristol consists of five whole tissue sections, consisting of one ganglioneuroblastoma, three poorly-differentiated neuroblastoma and one ganglioneuroma which belong to five different patients. To evaluate the proposed method, this research uses a validation training set, 623 cropped images that are selected from the Children’s Hospital at Westmead dataset as a training set and five whole tissue sections from the University of Bristol as the testing

set, training images that are used in this section are same as the training images that mentioned in Section 3.2.2. Here, ten sub-images ( $300 \times 300$  pixels) were selected randomly from each whole tissue section. First, SIFT + bag of feature + SVM histogram intersection assigns a label to each sub-image. Then, each tissue section is classified using the majority vote among ten labels corresponding to the ten sub-images. The whole tissue section predicted labels and the result of the majority vote are shown in Table 3.11.

Table 3.11: The actual and predicted classification for dataset from University of Bristol. Abbreviations: UD (undifferentiated neuroblastoma), PD (poorly-differentiated neuroblastoma), D (differentiating neuroblastoma), GNB (ganglioneuroblastoma) and GN (ganglioneuroma)

Tissue Section Number	Actual Classification	Predicted Classification	Majority Vote for actual class	Majority Vote for Predicted Class	Majority Vote for other classes
4905	GN	GN	10 out of 10	10 out of 10	0
4906	PD	D	4 out of 10	6 out of 10	0
4907	PD	PD	8 out of 10	8 out of 10	2 out of 10 D
4908	PD	PD	9 out of 10	9 out of 10	1 out of 10 D
4909	GNB	D	2 out of 10	6 out of 10 D	2 out of 10 PD

As can be seen, the proposed algorithm correctly classified three out of five whole tissue sections and the other two were misclassified as differentiating type while their actual types were poorly-differentiated and ganglioneuroblastoma. As mentioned before, distinction between poorly-differentiated and differentiating subtypes also is difficult for human in limited field of view. Moreover, having three specifications of ganglioneuroblastoma in  $300 \times 300$  pixel images which are selected randomly are rare. These specifications are differentiating neuroblasts, neuropil and schwannian stroma (see Table 2.2). The sub-images for whole tissue section 4906 and 4909 are shown in Figures 3.16 and 3.17. The actual label of whole tissue section 4906 is poorly-differentiated. The four out of ten sub-images extracted from 4906 were predicted as poorly-differentiated neuroblastoma and the other six images were predicted as differentiating neuroblastoma. The proposed algorithm assigns label differentiating to the whole tissue section 4906 based on the majority vote. Moreover, two out of the ten extracted sub-images from 4909 with an actual label ganglioneuroblastoma were predicted as ganglioneuroblastoma and the others were predicted as differentiating. Therefore, according to the majority vote, the whole tissue section 4909 was predicted as differentiating.

### **Breast Cancer Histological Images**

This research evaluated the combination of SIFT with bag of features on the BreakHis breast cancer dataset (Spanhol et al. 2016). It consists of 7909 breast cancer histopathology images divided into benign and malignant tumours. The images are in four different levels of magnification with dimension of  $700 \times 460$  pixels. The image distribution is shown in Table 3.12. As can be seen, there are 2480 benign images and 5429

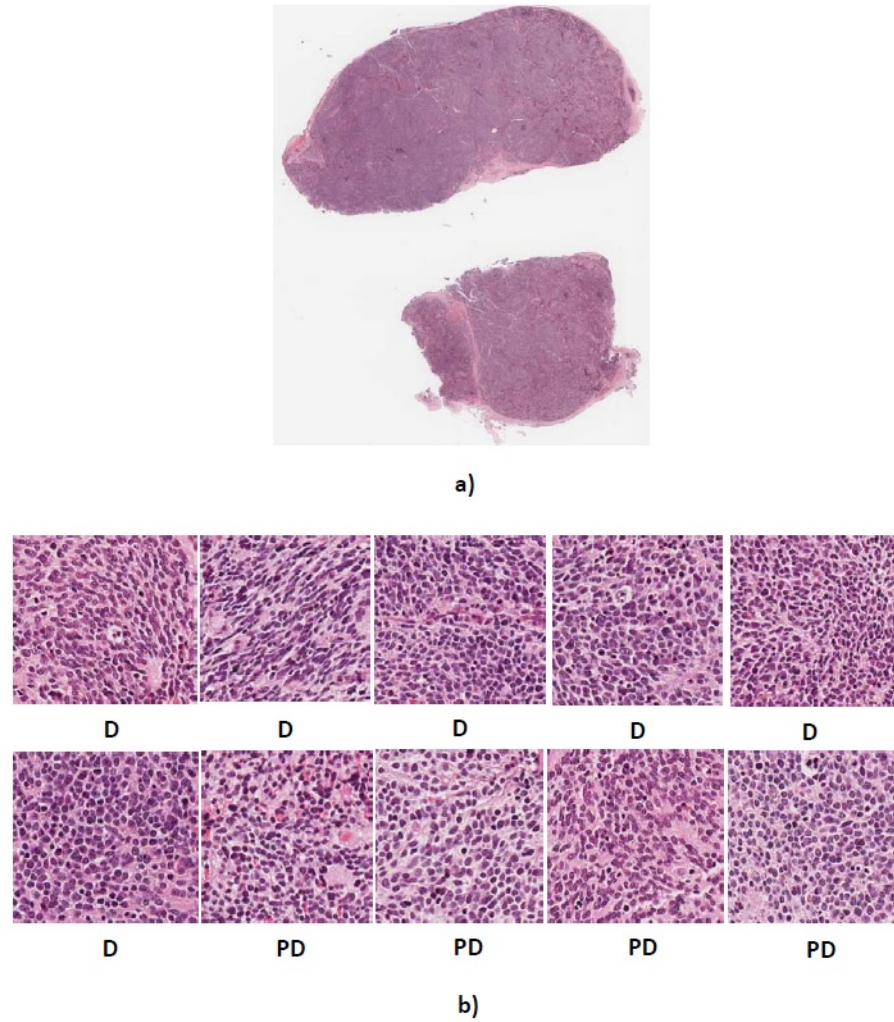


Figure 3.16: a) Whole tissue section 4906 with actual classification PD b) Predicted classes for ten randomly cropped images from the whole tissue section. Abbreviations: D = differentiating neuroblastoma, PD = poorly-differentiated neuroblastoma



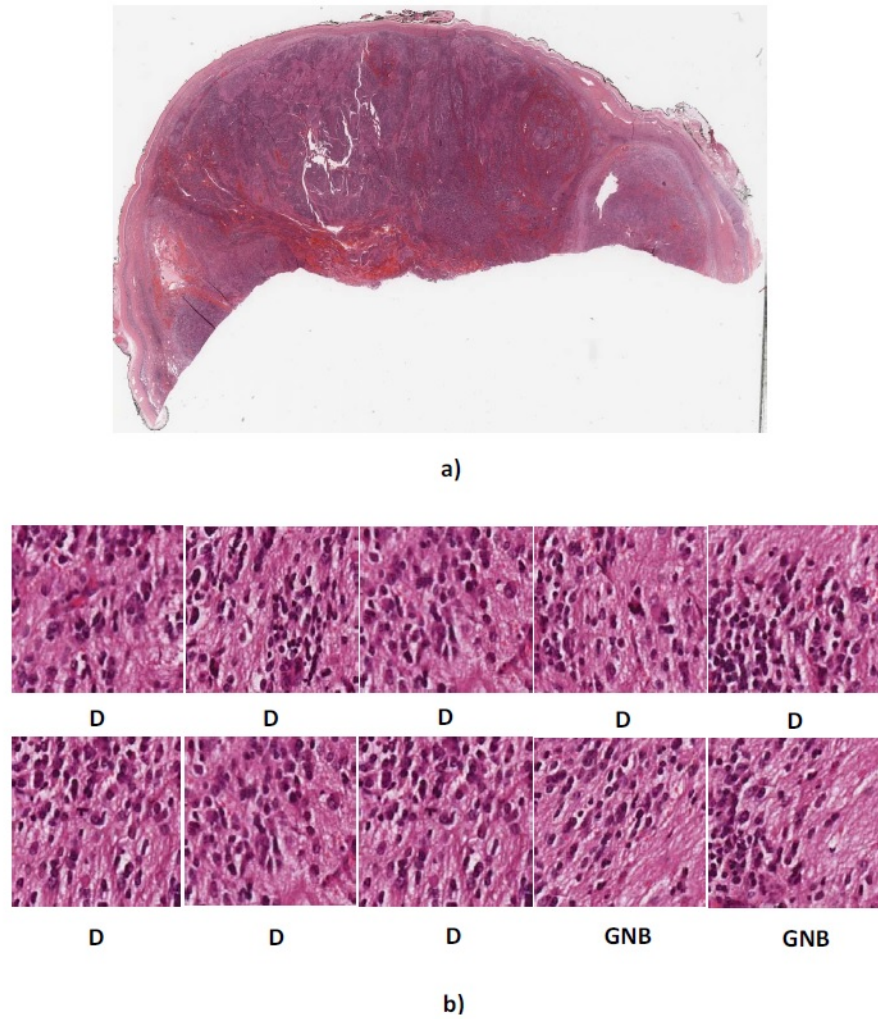


Figure 3.17: a) Whole tissue section 4909 with actual classification GNB b) Predicted classes for ten randomly cropped images from the whole tissue section. Abbreviations: D = differentiating neuroblastoma, GNB = ganglioneuroblastoma

malignant images which belong to 24 and 58 patients, respectively.

Spanhol et al. (2016) applied a combination of six feature extractors with four classifiers (1-nearest neighbour (1-NN), quadratic linear analysis (QDA), random forest (RF) and SVM classifiers) to classify breast cancer histopathology images. They used CLBP, LBP, gray level co-occurrence matrices (GLCM), Local Phase Quantization (LPQ), Oriented FAST Rotated BRIEF (ORB) and Parameter-free threshold adjacency statistics (PFTAS) as feature extractors. They calculated the average recognition rates of five trials for magnification factors of  $40\times$ ,  $100\times$ ,  $200\times$  and  $400\times$ . They achieved the best average recognition rates using PFTAS as a descriptor. Spanhol et al. (2016) reported the best average recognition rates of the classifiers trained with the PFTAS descriptor as shown in Table 3.13.

Table 3.12: Image distribution of BreaKHis dataset by magnification factor and type

<b>Magnification</b>	<b>Benign</b>	<b>Malignant</b>	<b>Total</b>
$40\times$	625	1370	1995
$100\times$	644	1437	2081
$200\times$	623	1390	2013
$400\times$	588	1232	1820
<b>Total</b>	2480	5429	7909
<b>No of Patients</b>	24	58	82

For comparison, this section follows the same testing protocol as that used in their paper. The BreaKHis dataset is randomly divided into a training set (70%) and a testing set (30%). All of the models showed in Table 3.14 are trained using the training set and evaluated using the testing set and the recognition rates are computed. This procedure is repeated five times and the average recognition rates for all the magnification factors are reported. The results are shown in Table 3.14.

Table 3.13: Best average recognition rates (%) of the classifiers trained with different descriptors reported in Spanhol et al. (2016)

Descriptor	Classifier	40×	100×	200×	400×
PFTAS	1-NN	80.9±2.0	80.7±2.4	81.5±2.7	79.4±3.9
PFTAS	QDA	<b>83.8±4.1</b>	<b>82.1±4.9</b>	84.2±4.1	78.2±5.9
PFTAS	RF	81.8±2	81.3±2.8	83.5±2.3	81.0±3.8
PFTAS	SVM	81.6±3.0	79.9±5.4	<b>85.1±3.1</b>	<b>82.3±3.8</b>

According to Spanhol et al. (2016), the recognition rate is calculated as follows

$$Recognition\ Rate = \frac{\sum Patient\ Score}{Total\ number\ of\ patients} \quad (3.3.1)$$

$$Patient\ Score = \frac{N_{rec}}{N_P} \quad (3.3.2)$$

where  $N_{rec}$  is the number of images for each patient which are correctly classified and  $N_P$  is the number of images for patient  $P$  existing in the dataset. The results show that the combination of SIFT with the bag of features works better than the methods that Spanhol et al. (2016) tested.

### 3.4 Discussion, Contribution and Conclusion

Neuroblast cells have different intensity and size for tumours of the same class. Many of the CAD systems are dependent on the visual features of the histological images and they fail to provide the reliable results on neuroblastoma images. To overcome these issues, this chapter proposes two low-level feature extraction methods, PCLBP and SIFT, which are appearance-based features. These features are robust to variations

Table 3.14: Average recognition rate (%) of SIFT + bag of features and PFTAS over the BreakHis breast cancer dataset in different magnifications

Method	40×	100×	200×	400×
PFTAS (Spanhol et al. 2016)	83.8±4.1	82.1±4.9	85.1±3.1	82.3±3.8
SIFT + bag of features				
+ SVM (RBF)	82.86±1.51	85.11±0.87	81.55±0.78	84.15±0.69
SIFT + pyramid bag of features				
+ SVM (RBF)	80.07±1.56	80.75±3.91	76.31±0.99	79.21±2.85
SIFT + bag of features				
+ SVM (histogram intersection)	<b>85.4±1.39</b>	<b>87.25±2.51</b>	<b>85.6±0.76</b>	<b>86.04±1.56</b>
SIFT + pyramid bag of features				
+ SVM (histogram intersection)	83.06±1.4	85.6±0.43	81.29±1.64	84.65±0.25

of size and intensity of cells within the images.

This chapter addresses **Contributions 1, 2 and 3** of this thesis by constructing a real dataset of neuroblastoma histological images and developing a series of algorithms for classification of these images with high intra-class variations of intensity and size. The first contribution is constructing a dataset consisting of real images from neuroblastoma. As there is no publicly available dataset of neuroblastoma, this research constructs a dataset from the Tumour Bank of the Kid's Research Institute at the Children's Hospital at Westmead to address the first contribution.

The second contribution is extracting local features that are robust to intra-class variations of cells of different intensity in the same class. To address this contribution this thesis proposes a method called Patched Completed Local Binary Pattern (PCLBP) based on the CLBP. The aim of this method is to extract features in local

regions which are robust to intra-class variations in intensity. The results of the experiments show that the proposed algorithm outperforms the state-of-the-art in the constructed dataset of neuroblastoma.

The third contribution, extracting discriminative features which are robust to scale variations of cells in the same class, is addressed by proposing a combination of Scale Invariant Feature Transform (SIFT) with the bag of features. The aim of the proposed method is to extract more discriminative features which are robust to scale variations. The results of the experiments show that the proposed algorithm outperforms the state-of-the-art methods. For further evaluation, the algorithm is evaluated on a public dataset of histological images of breast cancer and a second neuroblastoma dataset provided by the University of Bristol. Experimental results show the effectiveness of the proposed algorithm.

The results of the experiments show that the proposed algorithms outperform the state-of-the-art. Based on the classification results SIFT and PCLBP outperform the CLBP. The F-measure of SIFT, PCLBP and CLBP is 85.19 and 75.96 and 75.15, respectively. The results show that SIFT is statistically significantly better than PCLBP and CLBP.

The PCLBP and SIFT extract low-level features which are based on the intensity and size of the different cells in neuroblastoma images. The main issue with the results of the low-level feature extraction methods is due to the complicated features of neuroblastoma which are different from visual features such as intensity and size. It is not only difficult for low-level feature extraction methods to extract these features, they may also not be recognisable to the human eye. Consequently, in the following chapter high-level feature extraction methods are developed to improve the

performance of the system in the classification of neuroblastic tumours.

## Chapter 4

# Models of Computer-Aided Diagnosis Systems Based on High-level Features

This chapter addresses **Contribution 4** listed in Section 1.5, extracting high-level features which are different from appearance-based features and may be difficult to recognise by human eye. The contribution is addressed by developing deep learning algorithms to extract high-level features which are robust to colour, shape, and texture of the cells. These features may not be recognisable by the human eye. Contribution 4 is motivated by the understanding that neuroblastoma histological images have a complex structure with complicated features which may be difficult for the human eye to recognise.

This chapter comprises two sections: first, extracting high-level features using Convolutional Deep Belief Network (CDBNs), and second, extracting high-level features using Convolutional Neural Network (CNNs).

## 4.1 High-level Feature Extraction

This research predicts that neuroblastoma histological images have complicated features which are different from visual features of the cells and may not be recognisable by the human eye. Human pathologists look at the image and pick up everything from colour, shape and size which are based on the visual features of the images. Pathologists are at risk of inter- and intra- observer variability. To promote consistency and inter- and intra- observer agreement, pathologists follow a simple algorithmic decision tree to classify patients into subgroups (as shown in Figure 2.3 in Chapter 2).

Computers abstract appearance-based features to extract high-level features and distinguish one type of image from another. This section describes the development of Computer Aided Diagnosis System (CAD) based on deep learning networks to extract high-level features for tumour classification to improve diagnostic efficiency and consistency. Recent research shows that deep learning networks have an exciting potential for various applications in medical image pattern recognition and tissue classification including histopathology (Komura & Ishikawa 2018). These networks are valuable tools for the implementation of complex multi-parametric decision algorithms (Djuric et al. 2017). Deep learning networks are applied in pixel level in images to extract features such as lines and edges in lower layers and aggregate them to produce high-level features in the upper layers. (as shown in Figure 4.1).



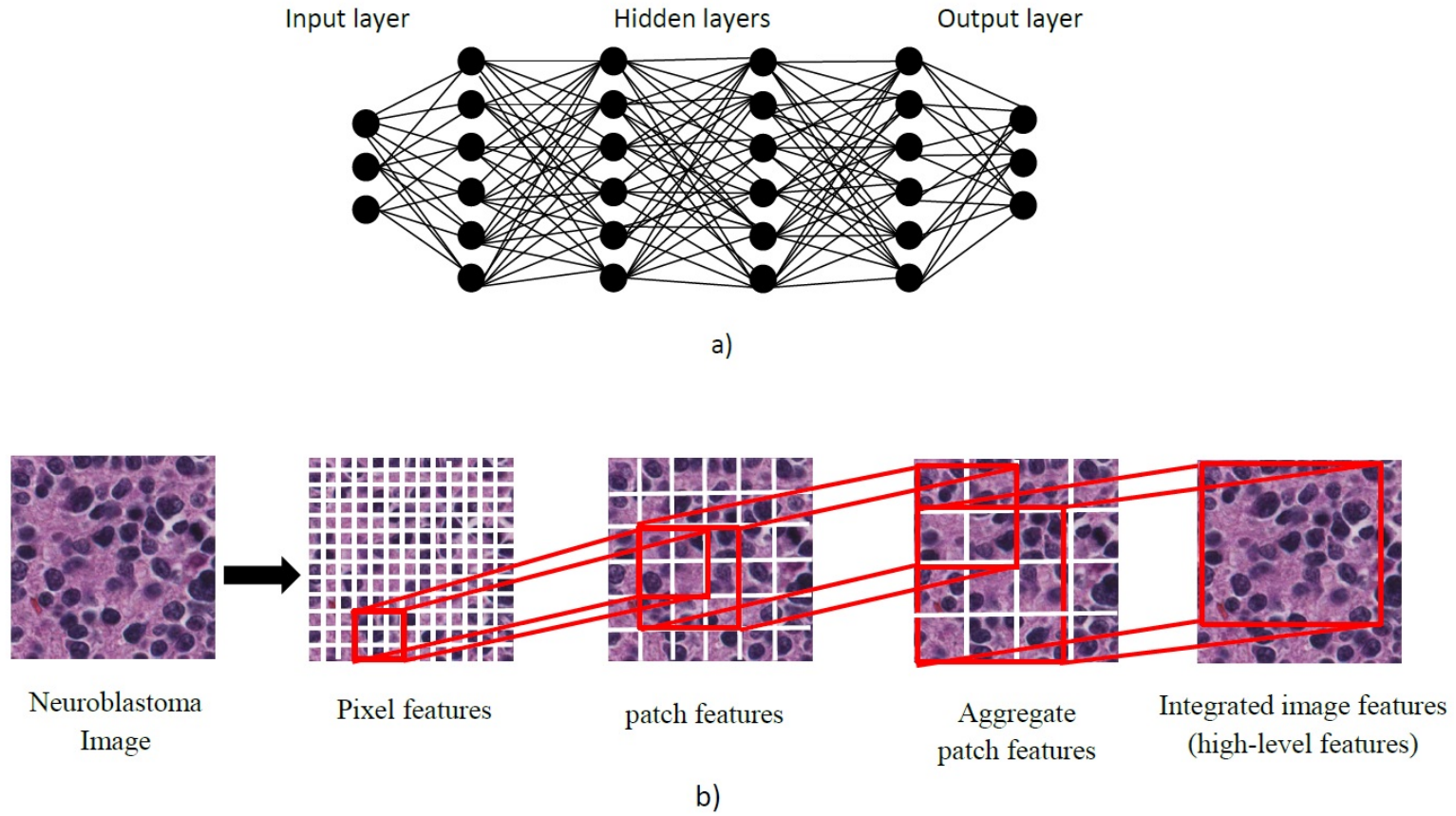
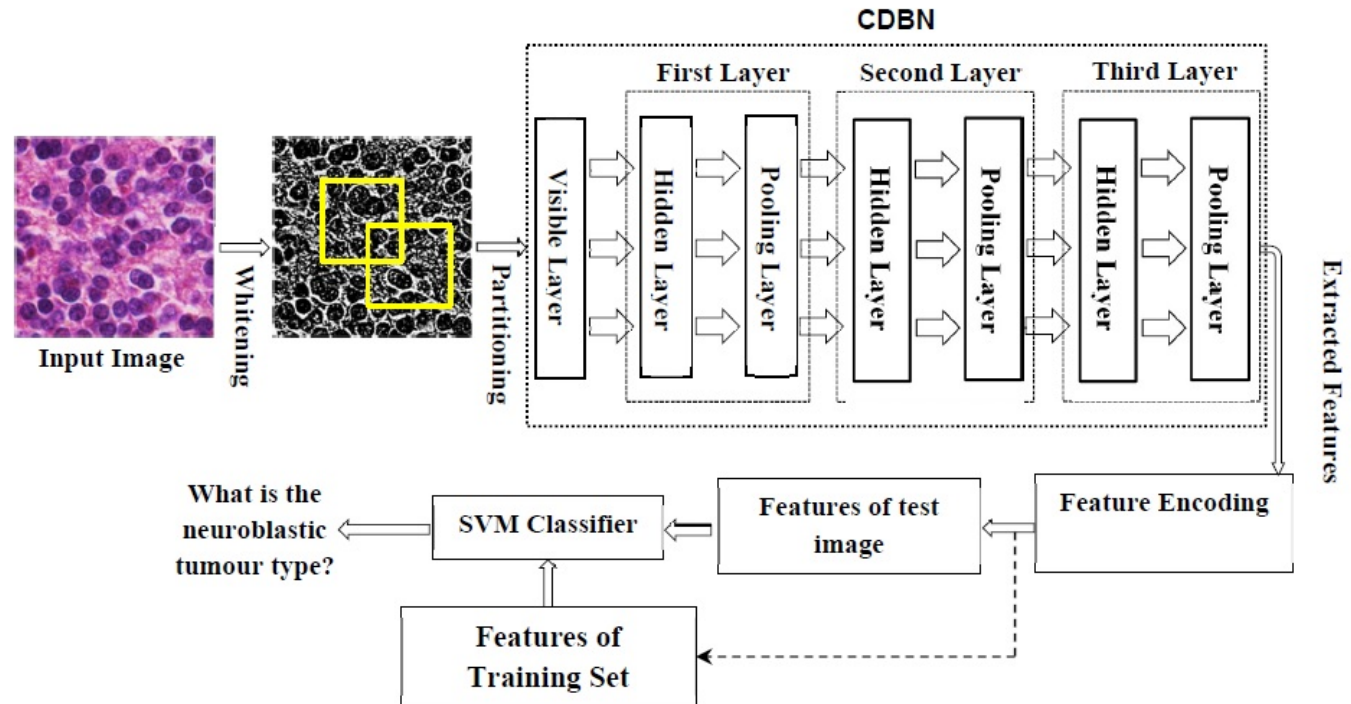


Figure 4.1: a) Schematic of deep learning network b) High-level feature extraction by deep learning networks. A deep learning network deconstructs the input image into pixels and combine them to reproduce high-level features (Djuric et al. 2017)

### 4.1.1 Convolutional Deep Belief Networks

The proposed algorithm is designed to learn additional representations using multi-parametric algorithms compared to handcrafted image descriptors. The overall framework of the proposed approach is shown in Figure 4.2. First, the test (input) image is pre-processed by the whitening method which modifies the image intensity values to be uncorrelated with each other. As a result, the image in this stage has more contrast compared to the input image. Whitening projects the input image into the eigenvectors of the covariance matrix and normalizes the images to have a variance of 1 for all intensity values (Kessy et al. 2018). In the next step, the whitened image is divided randomly into overlapping mini-batches which are uploaded to the visible layer of the CDBN. Mini-batches partition the input images into small chunks and perform the learning for each chunk. They also decrease the noisy information when the cropped images are not a good representation of the whole data.

The CDBN model uses an hierarchical structure to extract high-level features from the low-level ones. Lower layers extract low-level features that input them into the higher layers. Hierarchical generative models are composed of stacked convolutional restricted boltzmann machines (CRBMs) consisting of a visible layer (V), a hidden layer (H), and a pooling layer (P) (see Figure 2.10 in Chapter 2). This research uses a three-layer CDBN to extract features from neuroblastoma histological images. The proposed CDBN has four free parameters which need to be determined: number of hidden layers, number of groups in hidden layers, number of mini-batches and the size of the codebook. The first one is the number of hidden layers. Training a CDBN with fewer hidden layers takes less time, but the performance may be suppressed because the network extracts low-level features from the input images. Increasing the number



schematic.jpg

Figure 4.2: The overall framework based on the CDBN which consists of whitening, partitioning, CDBN, feature encoding, and classification. The yellow squares show the mini-batches.

of hidden layers can improve the performance by learning more complicated features, but it may cause over-fitting and a high computational rate. The second free parameter is the number of groups in the hidden layers. There is no generally accepted theoretical approach for choosing this parameter. Like the number of hidden layers, more or fewer units result in slow learning time or low performance, respectively. The third free parameter is the number of mini-batches. Mini-batches have been shown to improve generalization performance and use significantly smaller memory by decreasing the computational cost of the CDBN network (Masters & Luschi 2018). The parameter tuning subsection describes the approach for setting the free parameters.

### **Feature Encoding**

To achieve higher performance, the CDBN features are uploaded to a feature encoding block which computes more discriminative representations. This thesis uses the bag of features algorithm (O'Hara & Draper 2011) as a feature encoding method. The scheme of the bag of features is shown in Figure 4.3. First, the codebook which is a set of codewords is constructed; codewords are the extracted CDBN features. Then, the input image is represented by a histogram showing the frequency of the codewords in the image.

The codebook is generated using the k-means clustering algorithm (Kanungo et al. 2002). All the extracted CDBN features are clustered to find a set of centroids in feature space. In this work, the k-means algorithm is used for clustering. A key parameter in the construction of the codebook is the selection of the number of codewords, i.e. codebook size. Csurka et al. (2004) showed that a larger codebook size

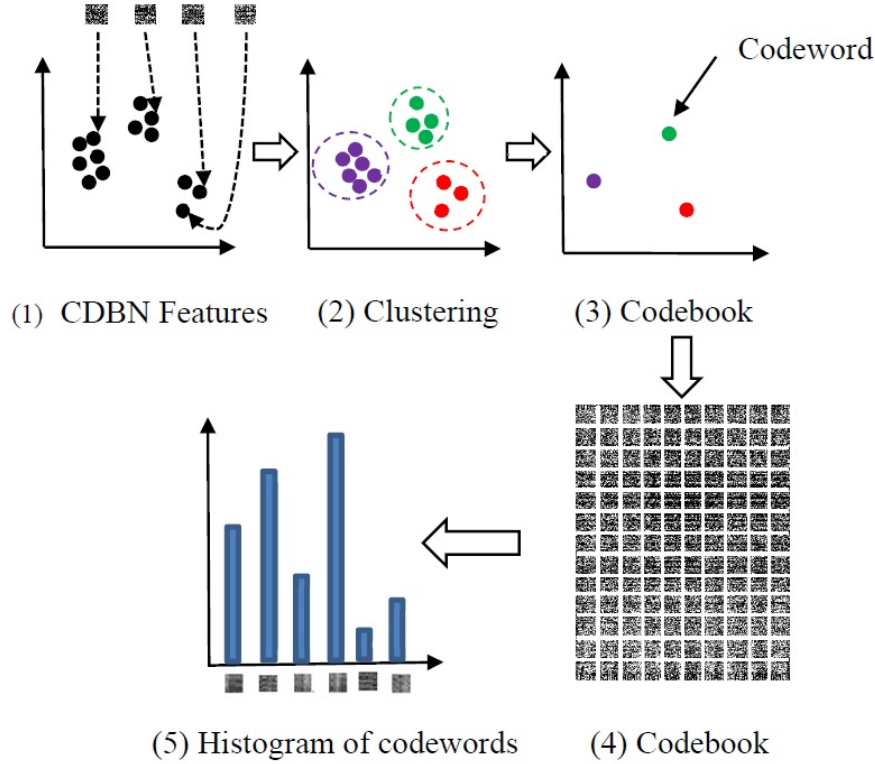


Figure 4.3: The scheme of the feature encoding block: (1) extracted high-level features by the CDBN algorithm from the test images, where each feature belongs to one test image, (2) clustering the extracted features using the k-means cluster, (3) finding the centroid of each cluster to construct a codebook consisting of codewords (4) visualisation of the constructed codebook consisting of codewords (5) Histogram of codewords which shows the number of each codeword in the codebook

works better for natural image classification. However, Tatiana et al. (2007) demonstrated that the size of the codebook does not always noticeably affect the accuracy of the medical image classification. To select the best codebook size, this thesis evaluates a different range of codebook sizes in the classification of neuroblastoma histological images and chooses the best one based on the average classification accuracy of the algorithm. Using the feature encoding block, the input image is represented by a

histogram of codewords. In the final stage, different categories of neuroblastoma histological images are represented by training an SVM classifier which classifies the histograms of the feature encoding block.

### **Experiment Results**

In this section, the performance of the proposed approach for the classification of neuroblastoma histological images is evaluated by conducting experiments on the constructed neuroblastic tumour database. Since the first layer in the CDBN learns common visual information such as edges, this research uses the Kyoto natural image dataset (Lee et al. 2009) to train the first layer of the CDBN; these images contain a large number of low-level features including edges. The second and third layers are trained with the neuroblastic tumour database. This section uses the same parameter-tuning set (211 images) and validation set (832 images) that was reviewed in Section 3.2.2 of Chapter 3. This research selects the optimum values for free parameters using the parameter-tuning dataset and fixes them for the validation. Then, the approach is evaluated using the validation dataset and selected parameters.

### **Parameter Tuning**

Four parameters need to be selected for the proposed algorithm, namely the number of hidden layers, the number of groups in each hidden layer, the number of mini-batches and the size of the codebook. This section uses the greedy method to tune the parameters of the CDBN based on the Kim & Hebbo (2013)'s work. The parameter-tuning dataset is randomly divided into training (75%) and testing (25%) subsets. These subsets are the same as the data division that was reviewed in Section 3.2.2 of Chapter 3. The algorithm is trained using the parameter-tuning training

set and computes the accuracy of the method using the parameter-tuning testing set to tune the free parameters. To better estimate the accuracy, the above procedure is repeated multiple (10) times and the average is computed over all experiments.

This research tested CDBN with one layer and increased it to two and three layers, while the number of mini-batches was held at 10. As shown in Table 4.1, the best accuracy was found for a three-layer network, while the number of groups in the first, second, and third layers was 24, 20, and 40, respectively. Consequently, this research considers a three-layer network in the next experiments.

Table 4.1: Average classification accuracy of the CDBN over the constructed dataset using a different number of hidden layers. Number of groups in the first, second, third layers and number of mini batches are 24, 20, 40 and 4, respectively.

Number of Hidden Layers	Classification Accuracy (%)
1	$50.81 \pm 2.26$
2	$62.28 \pm 1.57$
3	<b><math>79.69 \pm 1.11</math></b>
4	$69.56 \pm 1.45$

To choose the optimal number of groups in the hidden layers, this research computes the accuracy of the CDBN for a range of the number of groups in the first, second, and third layers, while the number of mini-batches was held at 10. As shown in Table 4.2, the best accuracy was found when the number of groups in the first, second, and third layers was 24, 40, and 40, respectively. These values are considered in the next experiments.

Table 4.2: Average classification accuracy of the CDBN over neuroblastic tumour dataset using a different number of groups in the hidden layers. Number of hidden layers and number of mini batches are 3 and 4, respectively.

Number of Hidden Groups in First, Second, and Third Layer	Classification Accuracy(%)
22-40-40	75.64 $\pm$ 3.69
24-40-40	<b>78.69 <math>\pm</math> 2.62</b>
26-40-40	76.25 $\pm$ 2.79
30-40-40	75.03 $\pm$ 1.32
24-20-40	77.25 $\pm$ 2.75
24-30-40	75.47 $\pm$ 2.16
24-40-20	74.42 $\pm$ 2.6
24-40-30	73.81 $\pm$ 1.06

An experiment was conducted to find the optimum number of mini-batches. The accuracy of the classification for a variety of mini-batches is shown in Table 4.3. The best accuracy was found when the number of mini-batches was seven. The size of the mini-batches was  $70 \times 70$  pixels. The impact of batch size is mostly on the computational time and not on the accuracy of the system. It is faster for smaller mini-batches but requires more examples.

Finally, this research explored the optimum value of the codebook size for the bag of features algorithm. Five codebook sizes  $\{100, 200, 300, 400, 500\}$  were tested. The average accuracy versus the size of the codebook is shown in Table 4.4. The best



Table 4.3: Average classification accuracy of the CDBN over the constructed dataset for a different number of mini-batches. The number of hidden layers and number of hidden groups in the first, second and third layer are 3, 24, 20 and 40, respectively.

Number of Mini-Batches	Classification Accuracy(%)
4	$73.2 \pm 2.33$
6	$74.42 \pm 3.15$
7	<b><math>77.47 \pm 1.8</math></b>
8	$76.86 \pm 2.13$
10	$75.64 \pm 1.76$

accuracy was found when the size of the codebook was 300. This value was considered in the next experiments.

### System Validation

Based on the experiments to find the best parameter setting, this research trains a CDBN with three hidden layers. The information for the different layers is shown in Table 4.5. In this thesis, the learned first layer bases (as shown in Figure 4.4 (a)) are oriented edge filters which are trained with natural images. They show the weights between the input layer and the first hidden layer. The learned second layer bases (are shown in Figure 4.4(b)) which are visualized as a weighted linear combination of the first layer bases and many of them respond to contours, corners, and angles.

The learned third layer bases which are the weighted linear combinations of the first and second layer bases are shown in Figure 4.4 (c)). The learned weights are visualized as the intensity values between 0 and 1 with the lowest weight is black

Table 4.4: Average classification accuracy of the CDBN over the neuroblastic tumour dataset with different values of the codebook size.

Codebook Size	Classification Accuracy (%)
100	$82.04 \pm 2.14$
200	$82.69 \pm 1.89$
300	<b><math>86.01 \pm 1.07</math></b>
400	$82.14 \pm 1.96$
500	$82.87 \pm 2.15$

Table 4.5: Information of different layers of CDBN

Layer	Number of bases	Size of filter (pixels)	Pooling ratio
1	24	$10 \times 10$	2
2	40	$10 \times 10$	2
3	40	$10 \times 10$	2

and the highest weight is white. The number of groups in the first, second and third layer is 24, 40 and 40, respectively. Therefore, there are 24, 40 and 40 blocks in Figure 4.4 (a-c) with size  $10 \times 10$  pixels where 10 is the size of the filter. As can be seen, Figures 4.4 (a) and 4.4 (b) show the edges and corners of different objects in the image ,respectively. These features are combined to form the overall shape of the objects in the image (as shown in Figure 4.4 (c)).

To evaluate the proposed method, this research uses the validation dataset which is not detailed in the parameter tuning section. The validation dataset is randomly

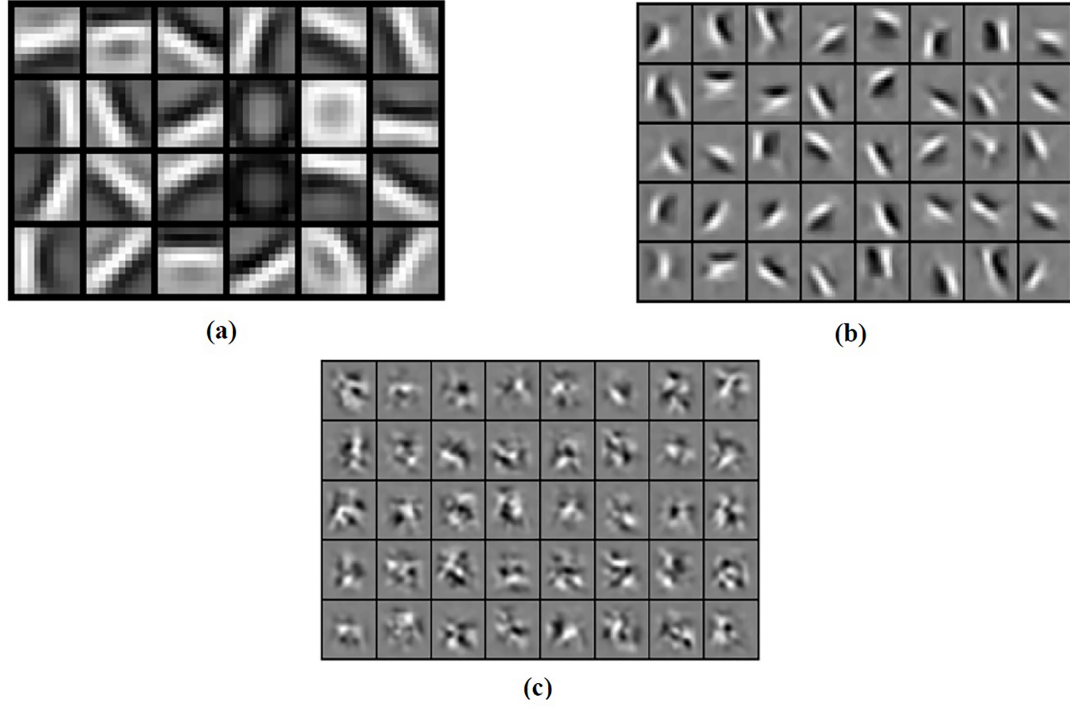


Figure 4.4: The visualization of the weights of CDBN (a) weights between the input layer and first hidden layer (b) weights between the first and second hidden layer (c) weights between the second and third hidden layer

divided into validation training (75%) and validation testing (25%) subsets (these subsets are the same as the subsets that were discussed in Section 3.2.2 of Chapter 3). The final model is trained using the validation training set and the accuracy is assessed for the model using the validation testing set. This procedure is repeated multiple (10) times and the average accuracy (weighted by the number of samples in each class) is reported.

This thesis combines the proposed method ,CDBN, with the standard bag of features and the pyramid bag of features (Lazebnik et al. 2006) algorithms as the feature encoding techniques. In each case, two kernel types are used for the SVM classifier: radial basis function (RBF) and the histogram intersection kernel (Maji et al. 2008).

Table 4.6: Weighted average precision, recall, and F-measure of the CDBN and the benchmarks over the constructed dataset consisting of 1043 neuroblastoma images. Bold values mean the best precision, recall and F-measure.

Method	Precision (%)	Recall (%)	F-measure
CLBP + SVM (RBF)	$74.1 \pm 2.35$	$76.25 \pm 2.23$	$75.15 \pm 2.28$
PCLBP + SVM (RBF)	$75.59 \pm 3.15$	$76.35 \pm 3.41$	$75.96 \pm 3.27$
CDBN + bag of features + SVM (RBF)	$82.54 \pm 4.49$	$85.63 \pm 2.04$	$84.02 \pm 3.03$
CDBN + pyramid bag of features + SVM (RBF)	$79.87 \pm 4.26$	$82.21 \pm 2.14$	$81.02 \pm 3.02$
CDBN + bag of features + SVM (histogram intersection)	<b><math>84.51 \pm 3.57</math></b>	<b><math>87.55 \pm 1.55</math></b>	<b><math>86.01 \pm 2.09</math></b>
CDBN + pyramid bag of features + SVM (histogram intersection)	$83.97 \pm 4.14$	$86.5 \pm 1.91$	$85.21 \pm 2.8$

The performance of the algorithm is measured by the weighted average F-measure, recall, and precision and is reported in Table 4.6. For comparison, the performance of the Patched Completed Local Binary Pattern (PCLBP) and Completed Local Binary Pattern (CLBP) (Guo et al. 2010) are reported as the benchmarks.

As shown in Table 4.6, the proposed method in the best case obtains 10.05% and 10.86% higher F-measure compared to PCLBP and CLBP, respectively. The weighted average of precision, recall, and F-measure of the proposed algorithm is better than PCLBP and CLBP. The distribution of the computed F-measures for different approaches over the ten trials is presented in Figure 4.5, showing that the combination of CDBN with the bag of features works better than CLBP and PCLBP.

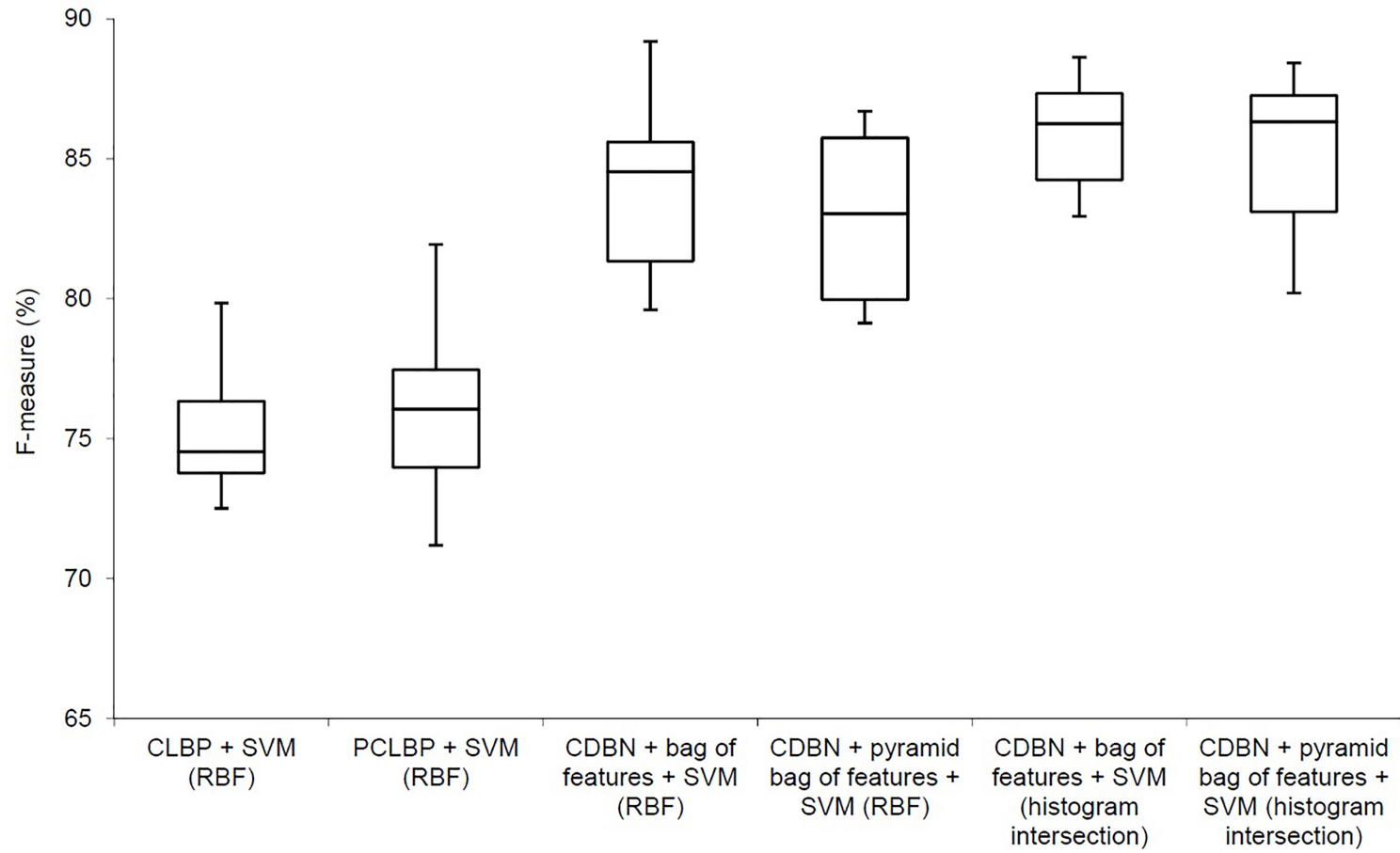


Figure 4.5: Comparison between the proposed algorithm (CDBN + feature encoding) and the benchmarks (CLBP, PCLBP)

To check if the proposed method is statistically significantly better than the benchmarks, a t-test was completed (with significance level of 5%) for CLBP and PCLBP (see Table 4.7).

Table 4.7: T-test for a comparison of CLBP and PCLBP with the combination of CDBN with the bag of features and the histogram intersection kernel SVM, with a significance level  $\alpha = 0.05$

Method	p-Value
CLBP + SVM (RBF)	$2 \times 10^{-5}$
PCLBP + SVM (RBF)	$1 \times 10^{-9}$

The results show that the combination of CDBN with the bag of features and histogram intersection kernel SVM significantly improves the weighted average classification accuracy compared to CLBP and PCLBP. A representative confusion matrix is shown in Table 4.8. As can be seen, the poorest performance is discriminating between poorly-differentiated and differentiating neuroblastoma, a distinction that human pathologists also find difficult in limited fields of view.

Examples of the differentiating type which are misclassified as poorly-differentiated type are shown in Figure 4.6(a). As can be seen, they are very similar to the actual poorly-differentiated ones that are shown in Figure 4.6(b).

Table 4.8: A representative confusion matrix. Abbreviations: UD (undifferentiated neuroblastoma), PD (poorly-differentiated neuroblastoma) , D (differentiating neuroblastoma), GNB (ganglioneuroblastoma) and GN (ganglioneuroma)

Predicted \ Actual	UD	PD	D	GNB	GN
UD	15	2	0	0	0
PD	0	121	8	0	0
D	0	7	17	2	0
GNB	0	1	1	15	0
GN	0	1	0	4	15

### 4.1.2 Convolutional Neural Network

Another type of deep neural network is the Convolutional Neural Network (CNN) (Lecun et al. 1998). It is composed of feed-forward networks which are stacked together to recognise the features of a 2-D input image. The neurons in a CNN consider subregions of the image. Then, features from the subregions are extracted by the feature maps of the network. The extracted features are combined to classify the image. A CNN consists of two parts: feature extraction and classification. The feature extraction part is built of stacked convolutional layers and subsampling layers. The classification part is performed using a fully connected layer. The overall framework of the CNN is shown in Figure 4.7.

The first layer in a CNN is the convolutional layer that filters (convolves) the inputs to get input information. The convolutional layer computes the output of the filters connected to the local regions in the input. Each filter computes a dot product between the values of the filter and the regions that are connected in the input image. The output of the dot product is a single number that is the output of the convolutional operation. This process is repeated for every location in the region

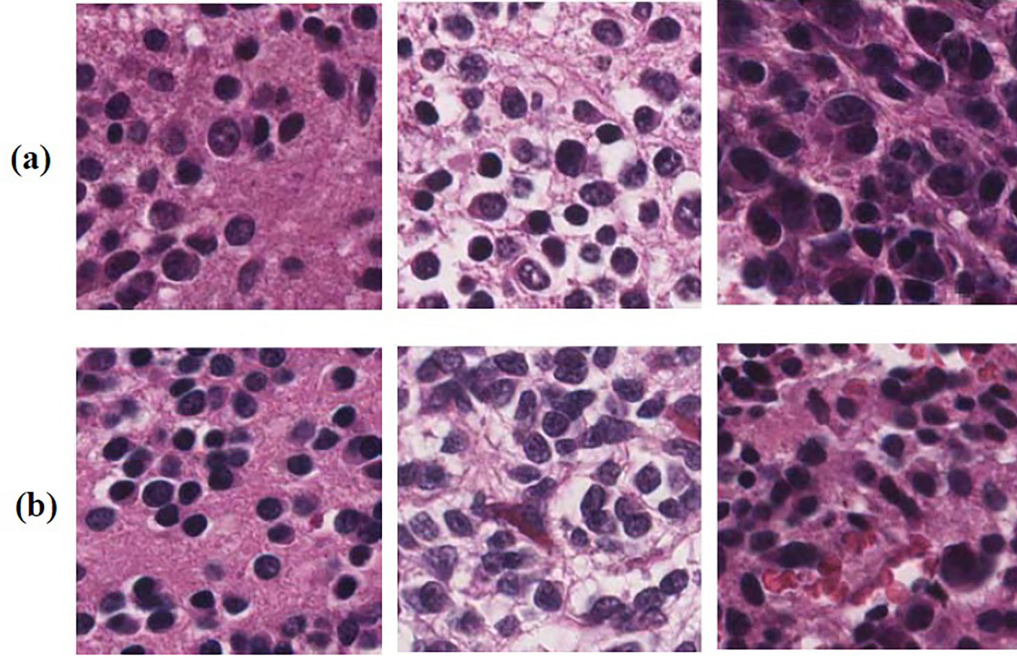


Figure 4.6: Examples of actual and predicted images: a) examples of the differentiating type which are identified by the computer as the poorly-differentiated type b) actual poorly-differentiated type

by convolution, moving the filter to the right by the stride which in this case is one unit. Filters learn frequent patterns that occur in the image by sharing the weights across neurons.

The second layer in a CNN is the subsampling layer which comes after the convolutional layer. This layer not only reduces the size of the representations and computational rates, it also controls overfitting in the network. The final layer in a CNN is a fully-connected layer. Each neuron in a fully-connected layer is connected to all the neurons of the following layer. They are completely independent and do not share any connections. In the last fully-connected layer, each neuron represents a class. Since Huang & Yann (2006) showed that CNNs extract excellent features for



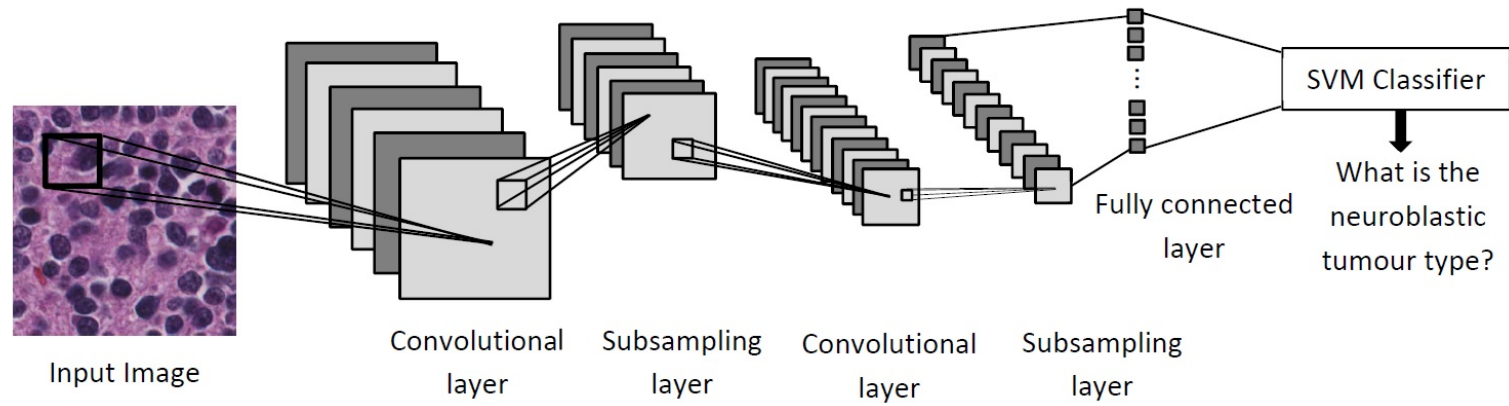


Figure 4.7: Overall framework of the Convolutional Neural Network (CNN). It consists of the convolutional layer, subsampling layer, fully-connected layer and SVM classifier

a SVM classifier, this thesis adds SVM into the last fully-connected layer to classify input images. For this purpose, an SVM is trained with the extracted features from the fully connected layer.

As CNNs have a huge number of parameters, often in the range of millions, they need a large dataset for training. This thesis uses a dataset of neuroblastoma histological images consisting of 1043 images which could be viewed as a small dataset for training the CNN. As training with 1043 images may lead the CNN to overfit, this research uses existing pretrained CNN models from Simonyan & Zisserman (2015) work. They used the ImageNet dataset (Deng et al. 2009) for pretraining.

ImageNet is a large benchmark dataset in object category classification and detection on hundreds of object categories and millions of images. This thesis uses the weights of the standard pretrained approaches which is called transfer learning. The idea of using a pretraining network is to obtain the initial weights of the network. The output of the last layer, the fully-connected layer, is 1000 classes. However, the neuroblastoma dataset has five categories. Therefore, the top layer of the fully-connected layer is connected to an SVM classifier that is trained with the neuroblastoma dataset to classify the neuroblastoma images into five categories.

This research applies three standard CNN models: VGG-16 (Simonyan & Zisserman 2015), VGG-19 (Simonyan & Zisserman 2015), and AlexNet (Krizhevsky et al. 2012). VGG-16 contains twelve convolutional layers and five pooling layers, VGG-19 has fourteen convolutional layers and five pooling layers and AlexNet consists of five convolutional layers and three pooling layers. Different categories of neuroblastoma histological images are modeled by training an SVM classifier using the extracted high-level features by CNN.

### Experiment Results

In this section, the performance of the CNNs is evaluated for the classification of neuroblastoma histological images. The SVM classifier with the RBF kernel which is connected to the fully connected layer of the CNN is trained with the neuroblastic tumour dataset. For this purpose, the algorithm is trained using the validation training set (623 images) and computes the accuracy of the method using the validation testing set (209 images). These subsets are the same as the subsets, as shown in Section 3.2.2 of Chapter 3.

### System Validation

This research compares the CNN method with VGG16, VGG19, AlexNet, Completed Local Binary Pattern (CLBP), Patched Completed Local Binary Pattern (PCLBP) and Convolutional Deep Belief networks (CDBN). The performance of the algorithm is measured by the weighted average (the values are weighted by the number of samples in each class) F-measure, recall and precision. This procedure is repeated 10 times and the average of the weighted averages are reported in Table 4.9.

As can be seen, the CNN-AlexNet obtains 9.65% and 8.84% higher F-measure compared to CLBP and PCLBP, respectively. However, CNN-AlexNet achieved 1.21% lower F-measure compared to the combination of CDBN with the bag of features. The average weighted average of precision, recall, and F-measure of the CDBN algorithm is better than VGG-16, VGG-19 and AlexNet.

To check if the combination of CDBN and the bag of features was statistically significantly better than CNN, this research completed a t-test (with significance level of 5%) for VGG-16, VGG-19 and AlexNet (as shown in Table 4.10). The results show that the combination of CDBN with the bag of features and SVM histogram

Table 4.9: Average weighted average precision, recall, and F-measure obtained by convolutional neural network and the benchmarks. Bold values mean the best precision, recall and F-measure.

Method	Precision (%)	Recall (%)	F-measure
CLBP + SVM (RBF)	$74.1 \pm 2.35$	$76.25 \pm 2.23$	$75.15 \pm 2.28$
PCLBP + SVM (RBF)	$75.59 \pm 3.15$	$76.35 \pm 3.41$	$75.96 \pm 3.27$
CDBN + bag of features + SVM (RBF)	$82.54 \pm 4.49$	$85.63 \pm 2.04$	$84.02 \pm 3.03$
CDBN + pyramid bag of features + SVM (RBF)	$79.87 \pm 4.26$	$82.21 \pm 2.14$	$81.02 \pm 3.02$
CDBN + bag of features + SVM (histogram intersection)	<b><math>84.51 \pm 3.57</math></b>	<b><math>87.55 \pm 1.55</math></b>	<b><math>86.01 \pm 2.09</math></b>
CDBN + pyramid bag of features + SVM (histogram intersection)	$83.97 \pm 4.14$	$86.5 \pm 1.91$	$85.21 \pm 2.8$
CNN (VGG-16) + SVM (RBF)	$81.75 \pm 4.59$	$83.47 \pm 1.93$	$82.6 \pm 2.75$
CNN (VGG-19) + SVM (RBF)	$82.85 \pm 4.45$	$85.21 \pm 2.5$	$84.01 \pm 3.16$
CNN (AlexNet) + SVM (RBF)	$83.67 \pm 3.7$	$85.98 \pm 2.15$	$84.8 \pm 2.41$

intersection kernel significantly improved the classification F-measure compared to VGG-16.

The distribution of the computed F-measures for different approaches over the ten trials is presented in Figure 4.8. It shows that CNNs work better than CLBP and PCLBP. However, the combination of CDBN with the bag of features and histogram intersection kernel SVM works better than CNNs.

Table 4.10: T-test for the comparison of the combination of CDBN with the bag of features and SVM histogram intersection kernel with VGG-16, VGG-19, and AlexNet, significance level  $\alpha = 0.05$

Method	p-Value
CNN (VGG-16)	0.005
CNN (VGG-19)	0.11
CNN (AlexNet)	0.25

All the CDBN and CNN algorithms are developed in MATLAB R2017b, Deep learning toolbox.

### 4.1.3 Sensitivity of Test Samples

This research evaluated the deep learning algorithms with manually selected  $300 \times 300$  pixel sub-images that are the best representative of the neuroblastoma tissues. The research in this subsection evaluates whether the careful selection of regions biases the proposed algorithm. To evaluate the CDBN method, this section describes the use of validation training images as the training set and 138 TMA cores as the testing set. Here, the method selects ten  $300 \times 300$  pixels sub-images randomly from each TMA core. The randomly selected sub-images are likely to include common artefacts seen in TMA cores such as tissue holes, red blood cells, over-staining and under-staining. A sample of the randomly selected sub-regions of a TMA core is shown in Figure 4.9.

The method classifies each randomly selected sub-image using the CDBN algorithm and assigns a class to the TMA core by the majority vote among ten classes

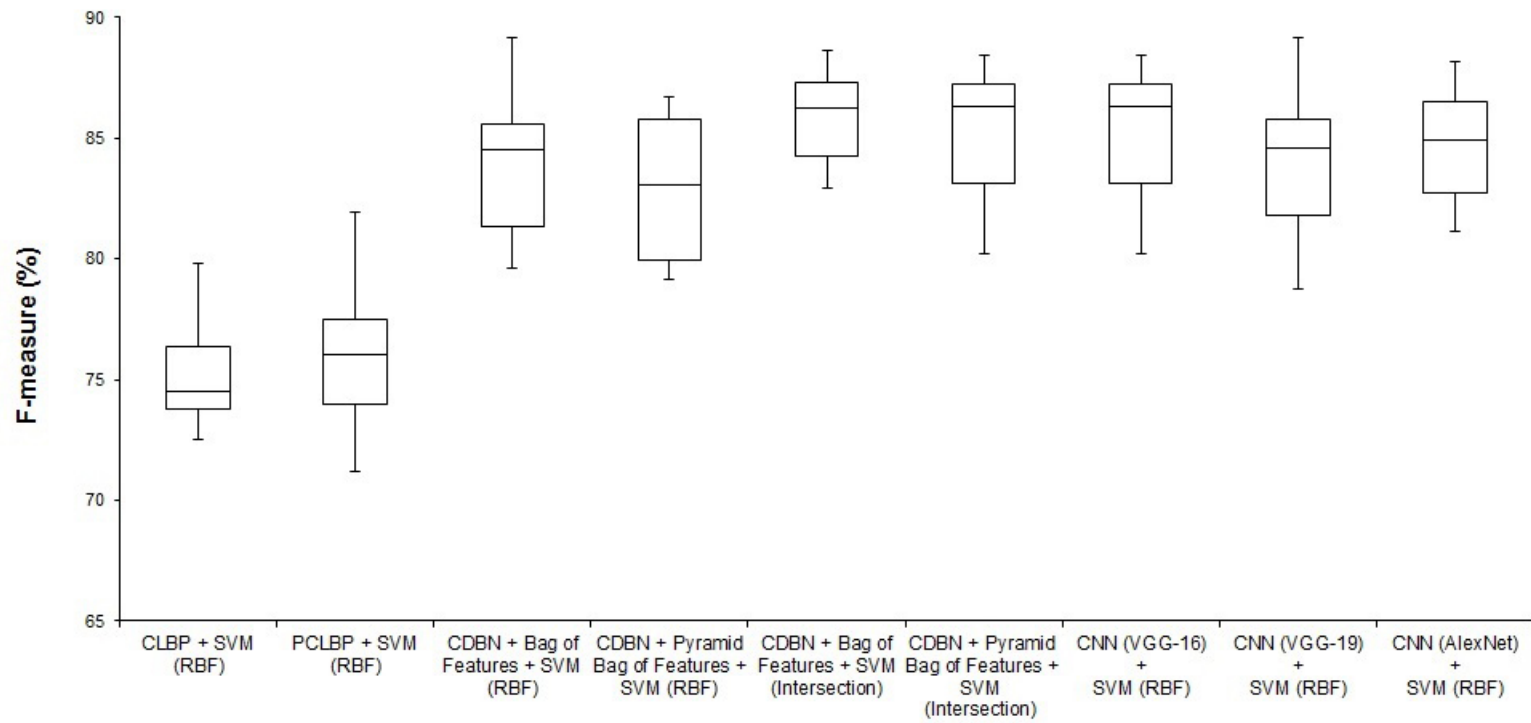


Figure 4.8: Comparison between CDBN + feature encoding and CLBP, PCLBP, CNN (VGG-16), CNN (VGG-19), and CNN (AlexNet)

corresponding to ten sub-images. In cases with two or more equal majority votes, the algorithm assigns the class with the highest prior probability (the class with the highest number of samples in the training set) to the TMA core.

The performance of the CDBN algorithm is measured by the weighted average of F-measure, precision, and recall over 138 TMA cores from Children's Hospital at Westmead and the 1043 cropped images belong to 138 TMA cores. This procedure is repeated 10 times and the average accuracy of precision, recall, and F-measure are calculated, as shown in Table 4.11.

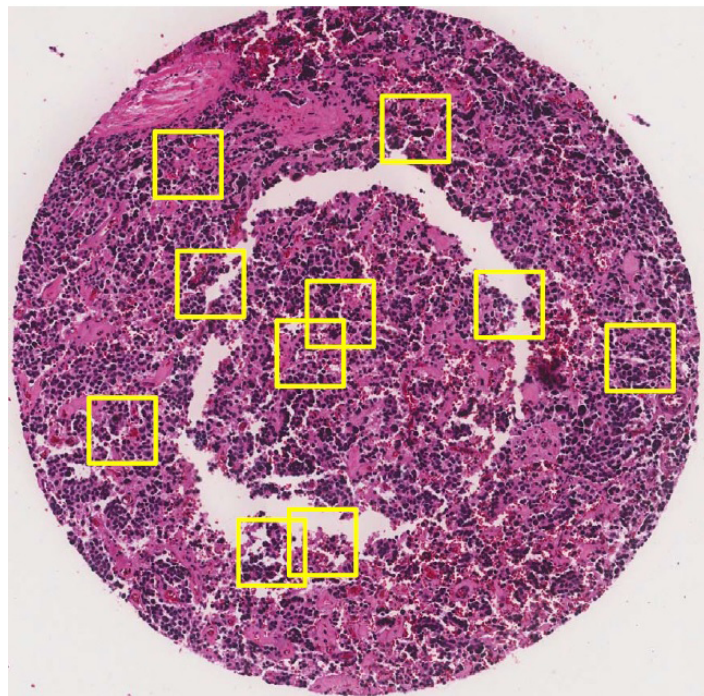


Figure 4.9: A sample of a TMA core with randomly selected sub-images. They are likely to include the common artefacts seen in TMA cores.

As can be seen, the accuracy of the method decreased when it was applied on randomly selected sub-images from 138 TMA cores compared to manually selected

sub-images, but this is not surprising because manual selection intentionally included features to discriminate between the classes. With the random selection of sub-images, some images will include features to discriminate between classes, and some images will not. Therefore, a collection of randomly selected sub-images would, on average, be less likely to contain class discriminating features and would be expected to be classified with reduced accuracy.

Table 4.11: Weighted average precision, recall, and F-measure of the convolutional deep belief network over randomly selected subimages in 138 TMA cores

Method	Precision (%)	Recall (%)	F-measure
CDBN + Bag of Features + SVM (RBF)	66.91±4.7	70.94 ± 1.29	68.86 ± 3.07
CDBN + Pyramid Bag of Features + SVM (RBF)	65.45±4.45	70.21 ± 1.15	67.74 ± 2.81
CDBN + Bag of Features + SVM (Histogram Intersection)	66.96±5.39	75.79±1.84	71.1±3.65
CDBN + Pyramid Bag of Features + SVM (Histogram Intersection)	67.01±4.39	64.42±2.5	65.69±3.26

#### 4.1.4 Discussion, Contribution and Conclusion

Neuroblastoma images have a complex texture with complicated features different to other types of cancers. Pathologists classify the histologic subtype of a neuroblastic tumour by identifying the morphologic features and then matching them against a pre-defined standard, known colloquially as ‘wallpaper matching’. Pathologists can use an



optical microscope and glass slide, or a digital image on a screen. The human workflow is a simple algorithmic approach based on the presence, absence or proportion of a limited range of defined morphologic features, and identifying where the aggregated morphologic features fit into a pre-defined, internationally standardised classification system. Unlike simplified algorithms pathologists use, deep learning networks are applied at the pixel level of images to extract low-level features such as lines and curves and aggregate them to detect high-level features such as objects and larger shapes in the upper levels. Deep learning networks can be a valuable tool in digital pathology due to the ability to learn useful features using multi-parametric algorithms, some of which may not be recognisable by the human eye.

This chapter addresses Contribution 4 of this thesis by developing Convolutional Neural Network (CNN) and Convolutional Deep Belief Network (CDBN) for the classification of neuroblastoma histological images. The aim of the proposed methods is extracting high-level features which are different from low-level features and may be difficult for the human eye to recognise.

The results of the experiments show that the proposed algorithm outperformed the state-of-the-art. Based on the classification results of the approaches above, CDBN and CNN outperform the CLBP and PCLBP. The results of t-test show that CDBN and CNN are statistically significantly better than CLBP and PCLBP.

# Chapter 5

## Conclusion

Neuroblastoma histological images have a complex texture with complicated features compared to other types of cancer that make their classification challenging. This is because cells in the same class may have a different shape, intensity and size. Because of these characteristics, CAD systems which extract mathematically high-level features have become more powerful than those which extract low-level features within the images.

Pathologists perform histological classification of neuroblastic tumours by matching the microscopic images to an international standardised classification scheme. They view all the tissue at low magnification and selected areas at medium and high magnification to identify the presence, absence or quantity of discriminatory diagnostic features, typically using a simple algorithmic approach. Although the gold standard is histopathological classification, CAD systems are able to extract novel features within small patches that are not readily observable by the human eye. These features are mathematically-based features and are different from appearance-based features such as size, colour and shape. Therefore, CAD systems can provide decision support for pathologists in regions that require further assessment (Yu et al. 2016).

Efforts have been made toward the development of CAD systems for tumour classification to mechanise human tasks. But still many CAD systems fail to provide a reliable result as they focused on handcrafted features, and therefore biased towards appearance based features. Especially for neuroblastoma images, a CAD system that extracts non-appearance-based features to classify them into five categories based on the Shimada classification has not been proposed. The system proposed by Sertel et al. (2009) is the only system which classifies histological images of neuroblastoma based on the mathematical features which are independent of the appearance-based features. Their diagnostic algorithm classifies neuroblastoma histological images into favourable and unfavourable categories based on the stroma-rich and stroma-poor regions. However, we aim to subclassify neuroblastoma in more detail, that is, into five different subtypes. The main advantage of deep learning methods is automatic feature extraction. These methods do the feature extraction of a richer high-level feature sets that are useful in automatic classification. However, non-deep learning approaches focused on the handcrafted features that are biased towards the appearance-based features.

The research in this thesis is motivated to collect a dataset of neuroblastoma histological images and design a CAD system to classify them for the following reasons highlighted by the literature review as research gap:

1. A lack of a publicly available large dataset for neuroblastoma histological images.
2. High intra-class variations of intensity of different cells in neuroblastoma histological images which belong to the same class.
3. A wide variety in the size of cells which belong to the same class.

4. Neuroblastoma has a complex texture with a great deal of complicated morphological information compared to other types of cancer such as breast cancer.

In light of these research gaps, this thesis asked the following research questions:

**RQ1:** How to acquire a dataset of neuroblastoma tumours as there is no publicly available dataset in this field?

This question is answered by constructing a dataset from The Children’s Hospital at Westmead, Sydney, Australia, as discussed in Section 3.1 of Chapter 3.

**RQ2:** How to develop an appropriate low-level feature extraction method to classify neuroblastic tumours with high intra-class variations of intensity of elements within the same class?

This question is answered by developing an algorithm which is based on the Completed Local Binary Pattern (CLBP), namely, Patched Completed Local Binary Pattern (PCLBP), as described in Section 3.2 of Chapter 3, to extract local features.

**RQ3:** How to acquire a desired classification method to overcome the high intra-class variations of size in the nuclei and neuroblast cells in the images within the same class?

To answer this question, a combination of Scale Invariant Feature Transform (SIFT) and bag of features was applied on the neuroblastic tumour images, as discussed in Section 3.3 of Chapter 3.

**RQ4:** How to extract high-level features from the neuroblastoma histological images?

This research applied Convolutional Deep Belief Network (CDBN) and Convolutional Neural Network (CNN) in Chapter 4 to classify neuroblastic tumour images based on the high-level features.

This research makes the following contributions to knowledge:

**Contribution 1: “Constructing a dataset consisting of images from neuroblastic tumours”**

This research constructs a dataset of histological images of neuroblastoma in Chapter 3 as there is no publicly available dataset in this field. The constructed dataset consists of 1043 cropped images from TMA slides and seven whole sections representing 125 patients from the Tumour Bank of the Kid’s Research Institute at the Children’s Hospital at Westmead. This research plans to publish the constructed dataset and hopes this could be of use to future researches.

**Contribution 2: “Extracting local features that are robust to intra-class variations of cells of different shapes and intensity in the same class”**

This research develops a low-level feature extraction algorithm in Chapter 3 to extract local features that are more robust to high intra-class variations of intensity of cells that belong to the same class. These characteristics make the performance of classification, based on the appearance-based features, more challenging. In this thesis, the issue is addressed by local feature extraction. The proposed algorithm consists of three stages: 1) dividing the input image into equal-sized square patches, 2) extracting features from the local patches which are more robust to variations compared to the global features by extracting sign binary patterns and magnitude binary patterns from each patch, 3) computing a local feature vector for each patch

by concatenation of the histogram of sign binary patterns and magnitude binary patterns within each patch, 4) creating a global feature vector for the whole image by concatenating the feature vectors of all patches and 5) classifying neuroblastoma histological images using SVM and k-NN classifiers. A global feature vector is created by concatenation of the local features extracted from each patch. The rationale behind this technique is to extract features within local regions which are more robust to high intra-class variations of intensity. The performance of the algorithm was compared with state-of-the-art approaches and the results show the weighted average F-measure of the proposed algorithm is better than the state-of-the-art methods which extract features within the global regions.

**Contribution 3: “Extracting discriminative features which are robust to scale variations of cells in the same class”.**

There is a high intra-class variation in the size of the neuroblast cells and nuclei cells in images which belong to the same class. These characteristics make the progress of the classification based on the size and geometry of the cells more challenging. This research introduces a combination of feature extraction with feature encoding in Chapter 3 to extract more discriminative features which are robust to scale variations. As extracted features are vectors with 128 elements, using all of the extracted features results in overfitting. Therefore, by using feature encoding, more discriminative features are extracted. The performance of the proposed algorithm was compared with state-of-the-art methods, and the results show the statistically significant superiority of the proposed algorithm over the state-of-the-art approaches. For further evaluation, the algorithm was evaluated on a public dataset of histological images of breast cancer and a second neuroblastoma dataset provided by the

University of Bristol. The experiment results show the effectiveness of the proposed algorithm.

**Contribution 4: “Extracting high-level features which are difficult for the human eye to recognise”** This research develops a deep learning network, discussed in Chapter 4, to extract high-level features using several hidden layers to hierarchically learn the high-level features of the image. These features are the non-linear combination of low-level features and may not be recognisable by the human eye. This thesis sends the features extracted by a deep learning network to a bag of features as a feature encoding algorithm to compute more discriminative representations, and to achieve higher performance. The aim of this technique is to extract high-level features which are independent of the appearance of the cells in neuroblastoma histological images. The performance of the algorithm which extracts high-level features was compared with the low-level feature extractors. The experiment results show that the deep learning networks are statistically significantly better than the handcrafted feature extractor which extracts low-level features.

## 5.1 Discussion

Computer Aided Diagnosis (CAD) systems have the potential to delve deeper into images and identify new characteristics that relate to tumour growth and treatment response. CAD systems facilitate the analysis and classification of neuroblastoma histological images which is a non-trivial task due to differences in staining, intensity and instrumentation. They can help pathologists by extracting complicated features

that may not be recognisable by the human eye especially in small regions. These features are mathematical features not appearance-based features. Although the gold standard is the human eye and pathologist assessment, the proposed CAD systems provide further information for pathologist decision making. By directly training CNNs with treatment outcome data it may even be possible to identify biomarkers unknown to histopathologists, potentially providing a method to surpass that of the histopathologist.

In terms of computer science, it is the first time that neuroblastoma histological images have been classified into five different categories using low-level and high-level features. Therefore, the proposed algorithms can be used as a platform for other types of CAD systems to improve the accuracy of the classification.

There are limitations in this research. The specificity is not 100% when compared with the gold standard. Perhaps this is a reflection of the fidelity of the algorithm to find really minor variations, possibly unobservable to the human eye. Also, human pathologists are at risk of bias, inter- and intra- observer variability and fatigue. Moreover, the poorest performance of the proposed algorithms is in discriminating between poorly differentiated and differentiating neuroblastoma, a distinction that human pathologists also find difficult in limited fields of view. Another limitation in proposed method in Chapter 3, Patched Completed Local Binary Pattern (PCLBP), is that this research did not consider the edge effects between patches. Furthermore, n-fold cross validation experiments could be run while repeating the tuning and testing experiments. However, we have run 10 experiments instead of potentially 16 experiments that can be mentioned as another limitation of this thesis.

The last limitation in this research is using the greedy method for parameter



tuning. However, parameter tuning in the 3D space may achieve more consistent results.

## 5.2 Future Research Directions

This thesis focuses on designing CAD systems to classify neuroblastoma histological images into five categories, as discussed in Chapters 3 and 4. In the future, the research will aim to extend the proposed methods and overcome the limitations of the proposed approaches, as summarized in the following tasks:

**Classification of neuroblastoma histological images by different pathologists.** 1043 cropped images will be given to different pathologists to identify the differences in their classification and to compare this with the CAD algorithms. As experience and fatigue may cause significant inter- and intra-observation variations among pathologists, the specificity would not be 100%. Perhaps CAD systems will not be good as the human eye. This distinction has not yet been tested in this thesis and would be the subject of future study.

**Classification of whole slide images in different magnifications.** Pathologists examine all the regions of neuroblastoma histological images on low magnification and choose some representative regions to examine on high magnification. The final decision about the whole slide image is made based on combining the low power features and the representative regions. To mimic the method that pathologists use, it is reasonable to evaluate the proposed algorithms on whole slide images at different magnifications in future work.

**Investigating patient outcome linkage.** The research in this thesis tested the proposed CAD systems on a dataset consisting of images of histological images of neuroblastoma tumour from 125 children. Designing CAD systems to include the metadata of patients, especially patient outcomes, will assist treatment stratification and prognostication.

**Applying deep learning on immunohistochemistry staining images.** The research in this thesis only focuses on images of haematoxylin and eosin (H&E) stained tissue microarrays (TMAs) of neuroblastic tumours. However, immunohistochemistry staining identifies cancer cell immunophenotype, which is information additional to morphology that can contain information to assist diagnosis, prognosis, genetic status and response to treatment. Applying the proposed CAD systems on images stained with immunohistochemistry will be another research direction.

**Applying new methods of deep learning on unbalanced neuroblastoma histological images.** As neuroblastoma is a rare tumour among children, there is not enough data in this field. Therefore, there are a small number of images in two categories of the constructed dataset. The zero-shot learning (Rohrbach et al. 2011), which is an extension of supervised learning, is used to solve the classification problem when there is not enough labeled data for all classes. In future work, zero-shot learning will be applied on the constructed dataset of neuroblastoma and the results will be compared with the CNN and CDBN.

**Using the locally interpretable model-explanations (LIME) algorithm on neuroblastoma histological images.** LIME (Ribeiro et al. 2016) is an algorithm that may explain the prediction of classifier by approximating it with an

interpretable model. LIME algorithm can highlight the parts of the image that gives intuition to why the model would think that class may be present. Therefore, using LIME algorithm is likely to make the CAD system more trustable for the pathologists.

**Identifying another dataset and validate the proposed CAD systems on it** As neuroblastoma is a very rare disease there is no publicly available dataset for that. For more evaluation, this research validated the proposed algorithms on the second neuroblastoma dataset that is a very small one. However, it would be valuable to evaluate the proposed CAD systems on the third neuroblastoma dataset.

The first hypothesis in this thesis was to provide a robust and accurate CAD system based on the low-level features to classify neuroblastoma histological images into five different subtypes. The results demonstrate that proposed CAD systems are able to do the classification, but they just confused with the poorly-differentiated and differentiating subtypes. The second hypothesis in this thesis was the classification of neuroblastic tumours based on the high-level features is more accurate than low-level features. The proposed deep learning systems are not only feasible, but they are statistically significantly better than the handcrafted approaches which extract low-level features.

With the advent of deep learning networks, this is an exciting time for pathologists to work together with CAD to achieve superior results compared to either a human diagnosis or a CAD alone.

# Bibliography

- a. Cunningham, P. & Delany, S. J. (2007), ‘k-nearest neighbour classifiers’.
- Al-Kadi, O. S. (2010), ‘Texture measures combination for improved meningioma classification of histopathological images’, *Pattern Recognition* **43**(6), 2043–2053.
- Bejnordi, B. E., Veta, M., van Diest, P. J. & et al. (2017), ‘Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer’, *Journal of American Medical Association* **318**(22), 2199–2210.
- Boser, B., Guyon, I. & Vapnik, V. (1992), A training algorithm for optimal margin classifiers, in ‘Proceedings of the Fifth Annual Workshop on Computational Learning Theory’, pp. 144–152.
- Chang, C. & Lin, C. (2011), ‘LIBSVM: A Library for Support Vector Machines’, *ACM Transactions on Intelligent Systems and Technology (TIST)* **2**(3), 1–27. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Children’s-Oncology-Group (2016), ‘Neuroblastoma information’. viewed at July 2018 <https://www.childrensoncologygroup.org/index.php/neuroblastom>.
- Cortes, C. & Vapnik, V. (1995), ‘Support-vector networks’, *Machine Learning* **20**(3), 273–297.

- Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., an D. Fenyő, M. S., Moreira, A. L., Razavian, N. & Tsirigos, A. (2018), ‘Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning’, *Nature Medicine* **24**, 1559–1567.
- Cover, T. & Hart, P. (1967), ‘Nearest neighbour pattern classification’, *IEEE Transactions on Information Theory* **13**(1), 21–27.
- Csurka, G., Dance, C., Fan, L., Willamowski, J. & Bray, C. (2004), Visual categorization with bags of keypoints, *in* ‘Workshop on Statistical Learning in Computer Vision, ECCV’, pp. 1–22.
- Deng, J., Dong, W., Socher, R., Li, L., Li, K. & Fei-Fei, L. (2009), Imagenet: a large-scale hierarchical image database, *in* ‘IEEE Conference on Computer Vision and Pattern Recognition’, pp. 248–255.
- Desjardins, G. & Bengio, Y. (2008), *Empirical evaluation of convolutional RBMs for vision.*, Technical report, University of Montreal, Quebec, Canada.
- Djuric, U., Zadeh, G., Aldape, K. & Diamandis, P. (2017), ‘Precision histology: how deep learning is poised to revitalize histomorphology for personalized cancer care’, *NPJ Precision Oncology* **1**(1), 1–22.
- Engl, N. (2010), ‘Recent advances in neuroblastoma’, *Journal of PubMed* **362**(231), 2202–2211.
- Frogy, E. W. (1965), ‘Efficiency versus interpretability of classifications’, *Biometrics* **21**(3), 768–769.

- Fukushima, K. (1980), 'Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position', *Biological Cybernetics* **36**(4), 193–202.
- Geisser, S. (1993), *Predictive inference: an introduction*, Chapman and Hall, New York.
- Gonzalez, R. C. & Wintz, P. (2017), *Digital Image Processing (4th Ed.)*, Addison-Wesley Longman Publishing Co., Inc., Boston, USA.
- Guo, Z., Zhang, L. & Zhang, D. (2010), 'A completed modeling of local binary pattern operator for texture classification', *IEEE Transaction, Image Processing* **19**(6), 1657–1663.
- Gurney, J., Ross, J. & Wall, D. (1997), 'Histology specific incidence and trends', *Infant Cancer* **19**(5), 428–432.
- Guyon, I. & Elisseeff, A. E. (2003), 'An introduction to variable and feature selection', *Journal of Machine Learning* **3**, 1157–1182.
- Hinton, G. E., Osindero, S. & Teh, Y. W. (2006), 'A fast learning algorithm for deep belief nets.', *Neural Computation* **18**, 1527–1554.
- Howlader, N., Noone, A. & Krapcho, M. (2012), *SEER Cancer Statistics Review*, National Cancer Institute.
- Huang, F. & Yann, L. (2006), Large-scale learning with SVM and convolutional nets for generic object categorization, in 'Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition', pp. 284–291.

*ImageScope, pathology slide viewing software* (2016). viewed at June 2018, <http://www.leicabiosystems.com/digital-pathology/digital-pathology-management/imagescope/>.

Irshad, H., Roux, L. & Racocceanu, D. (2013), Multi-channels statistical and morphological features based mitosis detection in breast cancer histopathology, *in* ‘35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)’, pp. 3–7.

Jain, A. K., Duin, R. P. W. & Mao, J. (2000), ‘Statistical pattern recognition: a review’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(1), 4–37.

Janowczyk, A. & Madabhushi, A. (2016), ‘Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases’, *Journal of Pathology Informatics* **7**(1), 7–29.

Joshi, V. V., Cantor, A. B., Brodeur, G. M., Look, A. T., Shuster, J. J., Altshuler, G., Larkin, E. W., Holbrook, C. T., Silverman, J. F., Norris, H. T., Hayes, F. A., Smith, E. I. & Castleberry, R. P. (1993), ‘Correlation between morphologic and other prognostic markers of neuroblastoma’, *Cancer* **71**(10), 3173–3181.

Jungnickel, D. (1999), *The Greedy Algorithm. In: Graphs, Networks and Algorithms.*, Springer, Berlin, Heidelberg.

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R. & Wu,

- A. Y. (2002), ‘An efficient k-means clustering algorithm: analysis and implementation’, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* **24**(7), 881–892.
- Kavzoglu, T. & Colkesen, I. (2009), ‘A kernel functions analysis for support vector machines for land cover classification’, *International Journal of Applied Earth Observation and Geoinformation* **11**(5), 352–359.
- Kessy, A., Lewin, A. & Strimmer, K. (2018), ‘Optimal whitening and decorrelation’, *The American Statistician* **72**(4), 309–314.
- Kieffer, B., Babaie, M., Kalra, S. & Tizhoosh, H. R. (2017), ‘Convolutional neural networks for histopathology image classification: Training vs. using pre-trained networks’, *Computer Vision and Pattern Recognition* **abs/1710.05726**.
- Kim, J. W. & Hebbo, H. (2013), ‘Classification with deep belief networks’.
- Kohavi, R. & John, G. (1998), ‘Feature extraction, construction and selection’, *The Springer International Series in Engineering and Computer Science* **453**.
- Komura, D. & Ishikawa, S. (2018), ‘Machine learning methods for histopathological image analysis’, *Computational and Structural Biotechnology Journal* **16**, 34–42.
- Kong, J., Sertel, O., Shimada, H., Boyer, K., Saltz, J. & Gurcan, M. (2009), ‘computer-aided evaluation of neuroblastoma on whole-slide histology images: classifying grade of neuroblastic differentiation’, *Pattern Recognition* **42**(6), 1080–1092.
- Korzynska, A., Roszkowiak, L., Lopez, C., Bosch, R. & Witkowski, L. (2013), ‘Validation of various adaptive threshold methods of segmentation applied to follicular



- lymphoma digital images stained with 3,3'-diaminobenzidine and haematoxylin', *Diagnostic Pathology* **8**(48), 1–29.
- Kothari, S., Sonal, P., Phan, J., Young, A. & Wang, M. (2013), 'Histological image classification using biologically interpretable shape-based features', *BMC Medical Imaging* **13**(1), 1–16.
- Krizhevsky, A., Sutskever, I. & Hinton, G. (2012), Imagenet classification with deep convolutional neural networks, *in* 'Proceedings of the 25th International Conference on Neural Information Processing Systems.', pp. 1097–105.
- Kuse, M., Sharma, T. & Gupta, S. (2010), A classification scheme for lymphocyte segmentation in H&E stained histology images, *in* 'International Conference on Pattern Recognition', pp. 235–243.
- Lazebnik, S., Schmid, C. & Ponce, J. (2006), Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, *in* 'IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)', pp. 2169–78.
- Lecun, Y., Bottou, L., Bengio, Y. & H., P. (1998), 'Gradient-based learning applied to document recognition', *Proceeding of the IEEE* **86**, 2278–2324.
- Lee, H., Grosse, R., Ranganath, R. & Ng, A. (2011), 'Unsupervised learning of hierarchical representations with convolutional deep belief networks', *Communications of the ACM*. **54**(10), 95–103.
- Lee, H., Pham, P., Largman, Y. & Ng, A. Y. (2009), Unsupervised feature learning for audio classification using convolutional deep belief networks, *in* 'Advances in Neural Information Processing Systems', pp. 1096–1104.

- Li, C., Xu, C., Gui, C. & Fox, M. (2005), ‘Level set evolution without reinitialization: A new variational formulation’, *Computer Vision and Pattern Recognition (CVPR)* **1**, 430–436.
- Lindeberg, T. (1993), ‘Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention’, *International Journal of Computer Vision* **11**(3), 283–318.
- Litjens, G., Sánchez, C., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., van de Kaa, C. H., Bult, P., van Ginneken, B. & van der Laak, J. (2016), ‘Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis’, *Scientific Reports* **6** (262–86).
- London, W., Castleberry, R. & Matthay, K. (2005), ‘Evidence for an age cutoff greater than 365 days for neuroblastoma risk group stratification in the children’s-oncology group’, *Clinical Oncology* **23**(27), 6459–6465.
- Lowe, D. (2004), ‘Distinctive image features from scale-invariant keypoints’, *International Journal of Computer Vision* **60**(2), 91–110.
- Lu, Z., Carneiro, G. & Bradley, A. P. (2013), Automated nucleus and cytoplasm segmentation of overlapping cervical cells, in ‘International Conference on Medical Image Computing and Computer-Assisted Intervention’, pp. 452–460.
- MacQueen, J. (1967), Some methods for classification and analysis of multivariate observations, in ‘Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability’, pp. 281–297.

- Maji, S., Berg, A. C. & Malik, J. (2008), Classification using intersection kernel support vector machines is efficient, *in* ‘Computer Vision and Pattern Recognition (CVPR)’, pp. 1–8.
- Masters, D. & Luschi, C. (2018), ‘Revisiting small batch training for deep neural networks’, *Computing Research Repository (CoRR)* **arXiv:1804.07612**.  
**URL:** <http://arxiv.org/abs/1804.07612>
- Matas, J., Chum, O., Urban, M. & Pajdla, T. (2004), ‘Robust wide-baseline stereo from maximally stable extremal regions’, *Image and Vision Computing* **22**(10), 761–767.
- Michail, E., Kornaropoulos, E. N., Dimitropoulos, K., Grammalidis, N., Koletsa, T. & Kostopoulos, I. (2014), Detection of centroblasts in *H&E* stained images of follicular lymphoma, *in* ‘22nd Signal Processing and Communications Applications Conference’, pp. 2319–2322.
- Montavon, G. (2009), *A machine learning approach to classification of low resolution histological samples*, A Master Thesis in Computer and Communication Sciences in Ecole Polytechnique Federale de Lausanne.
- Nam, J. E. J., Maurer, M. & Mueller, K. (2009), ‘A high-dimensional feature clustering approach to support knowledge-assisted visualization’, *Computers and Graphics* **33**(5), 607–615.
- Nedzved, A., Ablameyko, S. & Pitas, I. (2000), ‘Morphological segmentation of histology cell images’, *15th International Conference on Pattern Recognition. ICPR* **4**, 3–7.

- Neuroblastoma Australia* (2018). viewed at December 2018, <https://neuroblastoma.org.au/about-neuroblastoma/overview/>.
- O'Hara, S. & Draper, B. A. (2011), 'Introduction to the bag of features paradigm for image classification and retrieval', *Computing Research Repository* **arXiv:1101.3354**.
- Ojala, T., Pietikainen, M. & Maenpaa, T. (2002), 'Multiresolution gray-scale and rotation invariant texture classification with local binary patterns', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(7), 971–987.
- Otsu, N. (1979), A threshold selection method from gray-level histograms, in 'IEEE Transactions on Systems, Man, and Cybernetics', pp. 284–291.
- Park, J., Caron, H. & Eggert, A. (2008), 'Neuroblastoma: biology, prognosis, and treatment', *Paediatric Clinics of North America* **55**(1), 97–120.
- Petrou, M. & GaciaSevilla, P. (2006), *Image Processing: Dealing with Texture*, Wiley, New York.
- Pizzo, P. A. & Poplack, D. G. (2015), *Principles and practice of pediatric oncology*, Philadelphia, Lippincott Williams Wilkins.
- Powers, D. (2011), 'Evaluation: from precision, recall and f-measure to roc, informedness, markedness, and correlation', *Journal of Machine Learning Technologies* **2**(1), 37–63.
- Rare-Cancer-Australia (2018), 'Neuroblastoma'. Viewed at July 2018 <https://www.neuroblastoma.org.au/about-neuroblastoma/overview/>.

- Ribeiro, M. T., Singh, S. & Guestrin, C. (2016), "why should I trust you?": explaining the predictions of any classifier, *in* 'Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', pp. 1135–1144.
- Robson, D. (2001), *Pathology and genetics of tumours of the nervous system*, International Agency for Research on Cancer (IARC Press).
- Rohrbach, M., Stark, M. & Schiele, B. (2011), 'Evaluating knowledge transfer and zero-shot learning in a large-scale setting', *CVPR 2011* pp. 1641–1648.
- Ruifrok, A. & Johnston, D. (2001), 'Quantification of histochemical staining by color deconvolution', *Analysis of Quantitative Cytology and Histology* **23**(4), 291–299.
- Sano, H., Bonadio, J., Gerbing, R. B., London, W. B., Matthay, K. K., Lukens, J. N. & Shimada, H. (2006), 'International neuroblastoma pathology classification adds independent prognostic information beyond the prognostic contribution of age', *Journal of Science Direct* **42**, 1113–1119.
- Sertel, O., Kong, J., Shimada, H., Catalyurek, U. V., Saltz, J. & Gurcan, M. (2009), 'Computer-aided prognosis of neuroblastoma on whole-slide images: classification of stromal development', *Pattern Recognition* **42**(6), 1093–1103.
- Sertel, O., Lozanski, G., Shana'ah, A. & Gurcan, M. N. (2010), 'Computer-aided detection of centroblasts for follicular lymphoma grading using adaptive likelihood based cell segmentation', *IEEE Transactions on Biomedical Engineering* **57**(10), 2613–2616.
- Shimada, H., Ambros, I. M., ichi Hata V. V. Joshi, L. P. D. J., Roald, B., Stram, D. O., Gerbing, R. B., Lukens, N., Matthay, K. K. & Castleberry, R. P. (1999),

- ‘The international neuroblastoma pathology classification(the shimada system)’, *American Cancer Society* **86**(2), 364–372.
- Simonyan, K. & Zisserman, A. (2015), ‘Very deep convolutional networks for large-scale image recognition’, *Computer Vision and Pattern Recognition* **arXiv:1409.1556**.
- Smith, D. & Harvey, R. (2011), ‘document retrieval using sift image features’, *Journal of Universal Computer Science* **17**(1), 3–15.
- Spanhol, F., Oliveira, L., Caroline, P. & Laurent, H. (2016), ‘A dataset for breast cancer histopathological image classification’, *IEEE Transactions on Biomedical Engineering* **63**(7), 1455–1462.
- Sridhar, S. (2011), *Digital Image Processing*, Oxford University Press, Oxford.
- Tafavogh, S., Meng, Q., Catchpoole, D. R. & Kennedy, P. J. (2014), Automated quantitative and qualitative analysis of the whole slide images of neuroblastoma tumour for making a prognosis decision, *in* ‘Proceedings of the IASTED 11th International Conference on Biomedical Engineering’, pp. 244–251.
- Tang, C. Y., Wu, Y. L., Hor, M. K. & Wang, W. H. (2008), Modified SIFT descriptor for image matching under interference, *in* ‘International Conference on Machine Learning and Cybernetics’, pp. 3294–3300.
- Tatiana, T., Francesco, O. & Barbara, C. (2007), CLEF2007 image annotation task: an SVM-based cue integration approach, *in* ‘Proceedings of ImageCLEF 2007 LNCS’, pp. 1096–1104.

- Topol, E. (2019), ‘High-performance medicine: the convergence of human and artificial intelligence’, *Nature Medicine* **25**(1), 44–56.
- Vapnik, V. & Media. (2000), *The nature of statistical learning theory*, Springer Science Business media.
- Vincent, L. & Soille, P. (1991), ‘Watersheds in digital spaces : an efficient algorithm based on immersion simulations’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**(6), 583–598.
- Witkin, A. (1983), Scale-space filtering, *in* ‘International joint conference on artificial intelligence’, pp. 1019–1022.
- Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Liu, A. N. B., Yu, P. S., Zhou, Z. H., Steinbach, M., Hand, D. J. & Steinberg, D. (2008), ‘Top 10 algorithms in data mining’, *Knowl Inf Syst* **14**, 1–37.
- Yu, K., Zhang, C., Berry, G., Altman, R., Re, C., Rubin, D. & Snyder, M. (2016), ‘Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features’, *Nature Communications* **7**, 1–10.
- Zampierioli, F., Stransky, B., Lorena, A. & Paulon, F. (2010), Segmentation and classification of histological images - application of graph analysis and machine learning methods, *in* ‘2010 23rd SIBGRAPI Conference on Graphics, Patterns and Images’, pp. 331–338.
- Zheng, Y., Jiang, Z., Xie, F., Zhang, H., Ma, Y., Shi, H. & Zhao, Y. (2017), ‘Feature extraction from histopathological images based on nucleus-guided convolutional neural network for breast lesion classification’, *Pattern Recognition* **71**, 14–25.

- Zorman, M., Kokol, P., Lenic, M., d. l. Rosa, J. L. S., Sigut, J. F. & Alayon., S. (2007), Symbol-based machine learning approach for supervised segmentation of follicular lymphoma images, *in* ‘12st IEEE International Symposium on Computer-Based Medical Systems (CBMS)’, pp. 115–120.