

McGILL UNIVERSITY

On the Automated Classification and Progression of Early Breast Cancers

Joseph Szymborski

Faculty of Medicine
Division of Experimental Medicine
McGill University, Montréal
September, 2017

Supervised by:
Prof. Luke M. McCAFFREY

A thesis submitted to McGill University in partial fulfilment
of the requirements of the Master's degree.

Abstract

Résumé

Acknowledgements

Habemus Thesim!

Eminentissimum ac reverendissimum Documentum, et cetra...

(Scribimus latinum sine intellectum)

Abbreviations

Preface

Overview

The harbinger of a far more lethal and difficult-to-treat disease, early breast cancer is a key, but often overlooked, point of study. Early breast cancers are often difficult to detect and classify, and the mode of progression between pre-invasive breast cancer lesions is not entirely understood.

Through the development of computational models for the regional detection and classification of early breast cancers, and by characterising the role of tight-junctions in their progression through traditional molecular biology; the projects herein offer novel insight into the fundamental nature of pre-invasive lesions of the mammary gland.

Contributions

Funding

Contents

Abstract	i
Résumé	ii
Acknowledgements	iii
Abbreviations	iv
Preface	v
Overview	v
Contributions	v
Funding	v
1 Introduction	2
1.1 An Anatomy of Early Breast Cancers	2
1.1.1 Stages of Early Breast Cancer Progression	2
1.1.2 Intrinsic Molecular Subtypes of Cancer	5
1.2 Current Practices for the Diagnosis of Early Breast Cancers	6
1.3 Machine Learning and the Medical Context	7
1.3.1 Instance-Based Algorithms & their Medical Applications	8
1.3.2 Perceptron-Based Algorithms & their Medical Applications	9
Preface	10
2 The Role of Tight-Junctions in the Early Progression of Breast Cancer	11
2.1 Introduction	11
2.1.1 Apical Polarity Proteins Complexes in Breast Cancers	11
2.2 Materials & Methods	14
2.3 Results	14
2.4 Discussion	14
Preface	15
3 PPReCOGG: A Model for the Per-Pixel Classification of Early Breast Lesions via Gabor Features	16
3.1 Introduction	16

3.2	Design & Methods	18
3.2.1	Model Design	18
3.2.2	Software Dependencies	18
3.2.3	Datasets	19
3.2.4	Software and Data Availability	19
3.3	Results	22
3.3.1	PPReCOGG Classifies Sub-Regions of Perceptibly Distinct Textures in Synthetic Benchmarks	22
3.3.2	PPReCOGG Classifies Sub-Regions of Different Neoplastic Phenotypes in Synthetic Benchmarks with High Accuracy	25
3.3.3	PPReCOGG Effectively Classifies Sub-Regions of Different Neoplastic Phenotypes in Human Biopsy Samples	25
3.4	Discussion	27
	Preface	29
4	DeepDuct: A Deep-Learning Approach to Regional Breast Cancer Classification using Grad-CAM	30
4.1	Introduction	30
4.1.1	Convolutional Neural Network Models for the Diagnosis of Breast Cancers	31
4.1.2	Localisation-Augmented Visualisation of Convolutional Neural Network Using Grad-CAM	31
4.1.3	Transfer-Learning for Resource Efficient Training of Neural Networks	32
4.2	Design & Methods	33
4.3	Results	35
4.3.1	Accurate Classification of Breast Neoplasmas in the BreakHis Dataset	35
4.3.2	Activation Mapping of BreakHis ConvNet using Grad-CAM . . .	36
4.4	Discussion	40
4.4.1	Explanations for Neural Network Predictions are Essential for Clinical Use of CADe/x Models	40
4.4.2	Dataset Considerations	41
4.4.3	Future Directions & Improvements	42
5	General Discussion	43
	Appendix A: BreakHis Class Codes and Abbreviations	55

1. Introduction

1.1 An Anatomy of Early Breast Cancers

Breast cancer is both a common and lethal disease, having earned the dubious distinction of being both the most common and second most fatal cancer amongst females in Canada and around the world (Canadian Cancer Society's Advisory Committee on Cancer Statistics, 2015). Breast cancer most commonly arises in the epithelium of the mammary gland's many lactiferous ducts, which form a network that delivers to the nipple the milk that is secreted by the lobules of the mammary gland; which is another origin of breast carcinomas (*Figure 1.1*). The epithelium of the lactiferous duct is highly organised, with well-defined tissue and cell polarity that is integral to the structure and function of the duct. The tube-like lactiferous duct is a bilayered structure comprised of the outer myoepithelial and inner epithelial monolayers, both surrounding a hollow lumen at the duct's core. This epithelial inner-layer is surrounded by an outer layer of myoepithelial cells which express smooth-muscle actin (SMA) whose muscle-like contractile properties biomechanically deliver milk along the duct in response to hormonal signal (Hamperl, 1970).

1.1.1 Stages of Early Breast Cancer Progression

When diagnosing a suspected early breast cancer, pathologists analyse needle-core biopsies with the aim of identifying and classifying any lesions that may be present. Classification of lesions allow medical professionals to better understand the nature of the particular disease, what treatment is most appropriate, and what statistical outcomes are associated with the lesion.

The early stages of breast cancer manifest as pre-invasive, hyper-proliferative lesions that exhibit progressive and gradual deterioration of this epithelial organisation. Of these

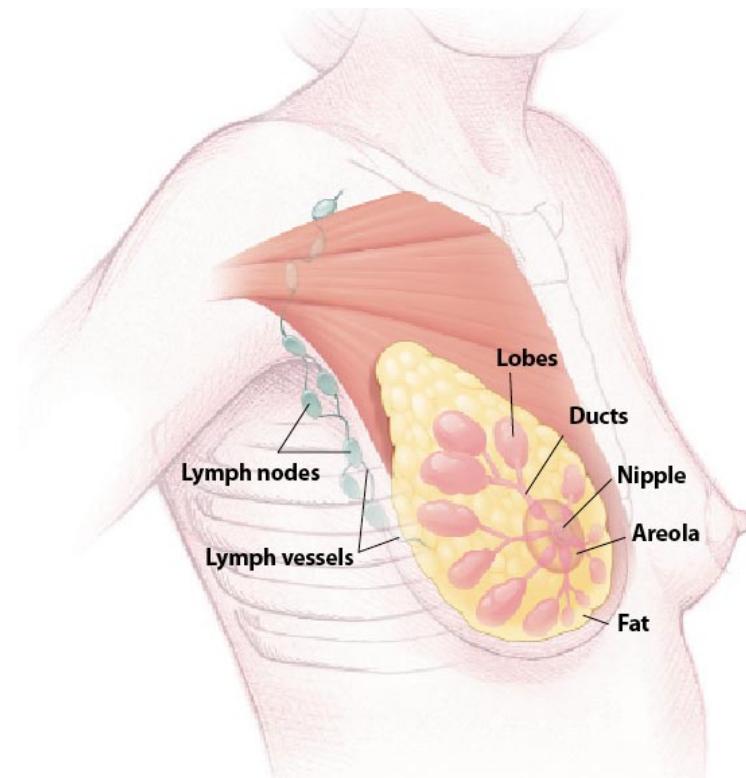


Figure 1.1: Diagram of the human mammary gland and its ducts and lobules. (National Institutes of Health, 2010)

lesions, there are four histologically distinct classes: Usual Hyperplasia (UH), Flat Epithelial Atypia (FEA), Atypical Ductal Hyperplasia (ADH) (or Atypical Lobular Hyperplasia [ALH] when referring to the less common lobular lesion), and Ductal Carcinoma *In Situ* (DCIS).

Ductal or lobular hyperplasias that do not present with abnormal tissue architecture or dysplasia are classified as Usual Hyperplasia (UH), or alternatively Proliferative Disease without Atypia (PDWA). These lesions confer a relative risk of later developing breast cancer as high as 1.9, although this increase in risk is not considered sufficient to warrant any prophylactic measures, including increased follow-up (Mommers *et al.*, 2001). While UH is traditionally believed to progress serially through ADH, DCIS and ultimately IDC due to early Loss of Homozygosity (LOH) analysis, more recent cytokeratin immunophenotype and genetic hybridisation analysis has contested the evolutionary relationship between UH and other proliferative breast lesions (O'Connell *et al.*, 1994; Boecker *et al.*, 2002).

ADH lesions are neoplasias of the lactiferous duct that exhibit subtle dysplasia (as

evidenced by nuclear hyperchromaticity), and can form micropapillary or cribriform patterns (Page *et al.*, 1959; Dion *et al.*, 2016). Of the estimated one million instances of benign breast cancer detected in the USA each year, 10% are classified as ADH (Simpson, 2009). While these lesions have been long-known and extensively proven to impart a low relative risk (approximately 4), recent long-term follow-up studies have shown that one in eight individuals will develop more advanced (local or invasive) breast cancers ten years after their diagnosis. This proportion increases to 46% in individuals with more than one atypical foci twenty-five years after diagnosis (Hartmann *et al.*, 2015).

Arising in the terminal duct-lobule unit of the breast, FEA lesions are a purported precursor to early low-grade ductal carcinomas, and in this regard are similar to ADH. Unlike ADH however, FEA lesions are far-more uncommon, never present with complex architectural patterns (thus the indication “flat”), and are characterised by multi-layered dilated ascini often made-up of columnar cells (Pinder, 2017). While ADH is suspected to arise from FEA lesions due their frequent coincidence, FEA is not independently associated with a long-term increased risk of breast cancer, leaving the matter unclear (Bombonati & Sgroi, 2011; Lerwill, 2008; Acott & Mancino, 2016).

Benign early lesions go on to progress into localised malignant disease, which in the lactiferous duct is termed ductal carcinoma *in situ* (DCIS). DCIS is classified as a Stage 0 cancer and accounts for 20% of all diagnosed breast cancers in the USA in 2003; representing a 500% increase in occurrence over 20 years (Bleicher, 2013; Kerlikowske, 2010).

While DCIS has a relatively low average standardised mortality ratio (SMR) of 1.8, an estimated 30-50% of cases reoccur as invasive breast cancers (Narod *et al.*, 2015; Page *et al.*, 1982; Betsill *et al.*, 1978). When further stratified by how well the lesion is differentiated, individuals with lesions classified as poorly differentiated (using the European Pathologists Working Group guidelines) have recurrence rates above 60% (Badve *et al.*, 1998). At this early stage of cancer progression, the apical domain of the luminal epithelium has begun to shrink, resulting in abnormally small lumen (a phenotype referred to as “luminal collapse” herein). Our understanding of the processes by which transformed

1.1. AN ANATOMY OF EARLY BREAST CANCERS

mammary duct epithelium undergoes luminal collapse is still developing, but recent studies have described a mechanism by which luminal tension is lost as myosin II and RhoA activity is greatly decreased at the luminal membrane of DCIS lesions (Halaoui *et al.*, 2017).

The lesion becomes an invasive ductal carcinoma (IDC, or ILC in the lobular instances) when epithelial cells breach the surrounding myoepithelial layer of the duct and infiltrate into extra-cellular matrix (ECM). By this stage, cellular polarity is entirely disrupted and the apical membrane domain has completely disappeared.

1.1.2 Intrinsic Molecular Subtypes of Cancer

While the stage of a breast lesion may indicate its progression, even lesions of the same stage are heterogenous in their gene expression and their response to therapy. To address this heterogeneity, microarray studies of recent years have identified five “intrinsic” molecular subtypes through unsupervised classification of the gene expression data (Perou *et al.*, 2000; Prat *et al.*, 2015).

Understanding the molecular subtype is an important tool for medical practitioners, as breast cancer lesions of different subtypes are associated with different patient outcomes and respond differently to treatment (Engstrøm *et al.*, 2013; Rouzier *et al.*, 2005). An alternative to microarray studies, molecular subtypes can be inferred by testing biopsy tissue for the co-occurrence of a number of molecular markers, as determined by routine immunohistochemical(IHC) studies. These markers range from hormone receptors (estrogen/progesteron receptors, and human epidermal growth factor receptor 2) to proliferation markers (Ki67) (See *Figure 1.2*).

While molecular subtypes of pre-invasive DCIS lesions have been shown to be detectable in a similar fashion to their invasive counter-parts, there are distinctions to note between the two (Tamimi *et al.*, 2008; Clark *et al.*, 2011). Firstly, the

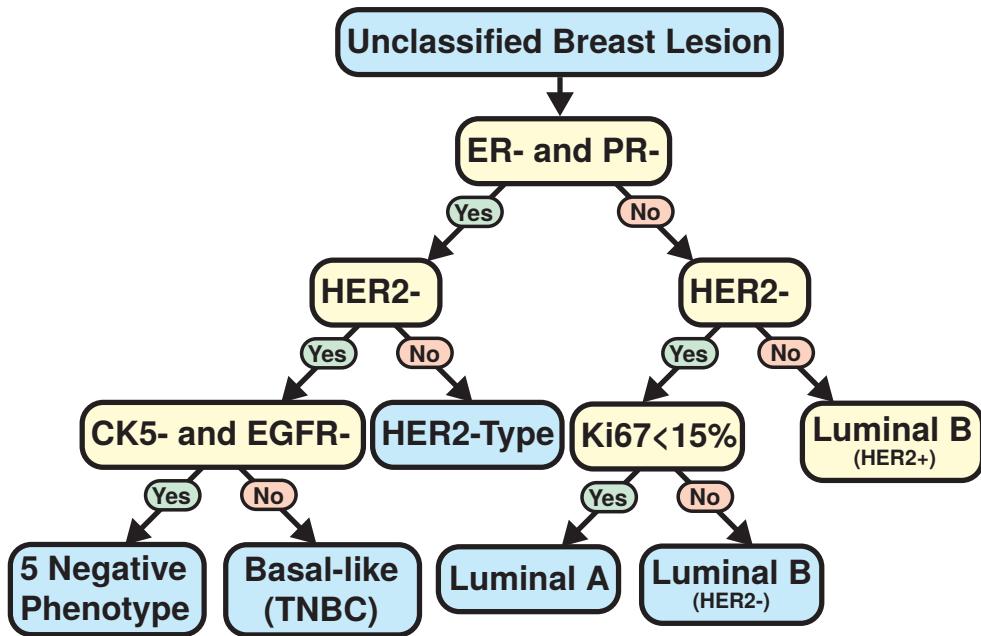


Figure 1.2: Flowchart for the determination of the molecular subtype of breast lesions. Molecular subtype can be determined by staining for the estrogen, progesteron, epithelial growth factor receptors, and human epidermal growth factor receptor 2 (ER, PR, EGFR, and HER2, respectively), as well as the basal/myoepithelial marker cytokeratin 5 (CK5) and proliferation marker Ki67. Adapted from Engstrøm *et al.* (2013).

1.2 Current Practices for the Diagnosis of Early Breast Cancers

Patients which exhibit symptoms of mammary neoplasmas, or are otherwise suspected to be at elevated risk due to factors such as age or family history, are subjected to routine screening by mammography. The sensitivity with which mammography is able to detect cancerous lesions, especially in young patients or patients whose breast tissue¹ (Ayyaci *et al.*, 2014). In such cases, screening via ultrasound can be a viable alternative, as it offers greater sensitivity (Nothacker *et al.*, 2009).

Suspected lesions detected during the process of screening are diagnosed by histopathological analysis of tissue biopsies (National Comprehensive Cancer Network, 2017). Tissue obtained through biopsy (core-needle, surgical, or otherwise) are sectioned onto glass slides, fixed and stained with relevant histological stains such as hematoxilin and eosin (H&E).

Despite the standardised and continually refined methods and guidelines clinical

pathologists rely on to identify breast cancer lesions from histological sections, there is a great deal of inconsistency and uncertainty that is becoming increasingly apparent. While some types of breast cancer (such as high-grade DCIS and LCIS) are more consistently and reliably identified than others, inter-observer agreement between clinical pathologists is mixed. In a retrospective study, agreement between pathologists for ADH, FEA, and low-grade DCIS regions was only moderate (0.44, 0.47, and 0.47 Cohen's κ statistic, respectively) (Gomes *et al.*, 2014).

1.3 Machine Learning and the Medical Context

Machine learning is the study of algorithms that can generalise solutions to problem-spaces without having been explicitly programmed them, but instead by “learning” from experiences, employing techniques from the fields of computational statistics, artificial intelligence, and mathematical optimisation.

One of its very first applications, Arthur Samuels coined the term “machine learning” to describe an automated Checkers-playing computer programme he had devised (Samuel, 1959). The Samuels Checkers programme demonstrated that it was possible to have machines solve problems by implementing learning algorithms as opposed to programming solutions in “minute detail” in situations where doing so may be unreasonably onerous or even entirely infeasible.

Medical applications of machine learning can be found early in its history, with Earl Hunt’s application of his Concept Learning System (CLS) for the purpose of medical diagnosis and prognosis as early as 1966 (Hunt *et al.*, 1966). Hunt recognised and stated that machine-learning techniques such as his CLS approach are particularly well-suited for analysing the often large amounts of data collected by medical tests, obviating time-consuming and expensive specialised investigations. The data generated by medical imaging is, in particular, both very rich and difficult to analyse in an efficient, reproducible manner.

Machine-learning systems for medical applications are preferably of high accuracy and

transparent to physicians in its methods, such that unexpected decisions are offered with an explanation that a physician can choose to agree or disagree with (Kononenko, 2001).

The field of machine learning has given rise to a multitude of algorithms, and many of them have been applied to various clinically relevant models of detection and prognosis. Two families of machine learning that have been particularly important to the clinical context will here be surveyed, namely instance-based and perceptron-based machine-learning algorithms.

1.3.1 Instance-Based Algorithms & their Medical Applications

Instance-based learning (IBL) algorithms are machine-learning algorithms that compare features from previous examples to unknown inputs to determine solutions. This is in contrast to other machine-learning algorithm families that generate internal generalised models of a problem-space.

Among IBL algorithms, the k -nearest neighbors (k -NN) and support vector machines (SVMs) are notable for having been widely used for a wide gamut of medical applications.

The k -Nearest Neighbors Algorithm

First described by Fix & Hodges while at the US Air Force as a technical report in 1951, and later formalised by Cover & Hart, the k -NN algorithm is one of the early and fundamental machine learning algorithms, and is used in countless applications today.

The k -NN classification algorithm begins with its training step, whereby an n_F -dimensional feature space is created, where n_F is the number of features per trained data point, all while keeping note of what class each data point in the training feature space belongs to. Subsequent classification steps involves obtaining the features for the unknown data and searching the feature-space generated in the training step for its k nearest features in terms of Euclidean distance of the features, where k is an odd integer. The unknown data is then classified as belonging to the same category as the majority of the k nearest features.

Considered a “lazy” machine learning algorithm, k -NN requires defers heavy computation from the training stage, where no additional feature processing is required, to the classification stage, which make use of memory-complex search algorithms. Data structures such as k -dimensional (k -d) trees, however, can reduce the memory and computational complexity of these searches (Otair, 2013). This, in turn, results in a reduction of time-complexity; in the case of k -d trees, search is performed in $O(\log n)$ time on average.

The k -NN algorithm has long been used for the purpose of clinical diagnostics, with early studies being applied to microcalcification detection systems for mammography, classification of aggressivity of brain tumours, and diagnosis of pigmented skin lesions (Dhawan *et al.*, 1996; Decaestecker *et al.*, 1997; Dreiseitl *et al.*, 2001). Clinical diagnostic tools based on k -NN classifiers continue to be studied, with innovation focused primarily on algorithm performance and feature engineering.

1.3.2 Perceptron-Based Algorithms & their Medical Applications

Preface

2. The Role of Tight-Junctions in the Early Progression of Breast Cancer

2.1 Introduction

Loss of cellular organisation and polarity is a common feature across epithelial cancers, but unlike some other cancers like those that present in the colon where cell polarity is lost at late stages of the disease, loss of polarity is a hallmark of early breast cancer (Hinck & Nähthke, 2014). The mechanisms by which epithelial cell polarity is lost in carcinomas, however, remains elusive and poorly understood.

When discussing the epithelial polarity of the lactiferous duct, one may be referring to asymmetric distribution at either the intercellular or intracellular level.

At the macro, intercellular scale, ductal epithelia is said to exhibit tissue polarity when cells organise into a monolayer forming a single lumen (Bissell *et al.*, 2003).

On the other hand, establishment of cellular apical-basolateral polarity is achieved by the intracellular asymmetric distribution of macromolecules within the inner epithelial monolayer of the mammary duct.

2.1.1 Apical Polarity Proteins Complexes in Breast Cancers

Two protein complexes, the Crumbs complex and the Par complex, are particularly important determinants of the apical identity and play significant roles in the early progression of breast cancer (Horikoshi *et al.*, 2009; Whiteman *et al.*, 2014).

The Crumbs Complex

Localisation of the Crumbs complex to the plasma membrane both contributes to the establishment and maintenance of its apical identity. The Crumbs complex converges upon the apical transmembrane glycoprotein for which it is named, Crumbs3 (*Crb3*), which serves as a scaffold for the complex. Crumbs3 directly binds two proteins via its carboxy-terminal PDZ domain (ERLI): protein associated with *Lin-7* one (Pals1) and partitioning-defective protein six (Par6) (Lemmers *et al.*, 2004; Roh *et al.*, 2002). The presence of Pals1 also brings to the Crumbs complex the Pals1 associated tight junction (PATJ) protein, which is essential for proper polarisation and contributes to the establishment of tight-junctions in mammalian cells (Shin *et al.*, 2005).

Crumbs3 has also been known to interact with FERM (4.1 protein, ezrin, radixin and moesin) domain proteins through its PDZ domain. Crumbs3 also interacts with the FERM-domain proteins EHM2 (also known as Lulu2) and YMO1, homologues of *Drosophila melanogaster* protein Yurt, which helps to establish apical-basolateral polarity and maintain the size of the apical membrane by regulating Crumbs3 (Laprise *et al.*, 2006). Crumbs3 has also been shown to recruit EHM2 and p114RhoGEF to maintain the actomyosin belt and promote cell-cell adhesion in a cancer cell-lines, requiring both the C-terminal FERM-binding and PDZ-binding motifs of Crumbs3 (Loie *et al.*, 2015).

The crumbs complex has been also shown to regulate important proliferative programmes such as organ growth and mammary gland contact inhibition through the Salvador/Warts/Hippo (hereafter Hippo) signalling pathway. Crumbs3 regulates the Hippo pathway through interactions with, among other proteins, the FERM domain-containing protein 6 (*FRMD6*), a mammalian homologue of the *D. melanogaster* gene *Ex* (Robinson *et al.*, 2010). Crumbs3 also regulates the Hippo pathway through direct interaction with WW-domain proteins. One such instance is the direct interaction between the Hippo pathway co-effectors yes-associated protein 1 (YAP1), Tafazzin (TAZ), and Crumbs3; this occurs in response to changes in cell density, which require changes to the cells proliferative program (Varelas *et al.*, 2010; Szymaniak *et al.*, 2015). In a similar, cell-density-sensing manner, Crumbs3 interacts directly with Kibra's WW-domain to sta-

2.1. INTRODUCTION

bilise it, preventing its degradation and promoting Hippo-pathway-mediated proliferation (Moleirinho *et al.*, 2013; Mao *et al.*, 2017).

Cell density is also coupled with transforming growth factor- β (TGF- β)-induced epithelial-mesenchymal transition (EMT) through the Crumbs3-mediated inhibition of SMAD; effectively reducing downstream activation Snail (Varelas *et al.*, 2010). In addition to being an important to the EMT transcriptional programme, the zinc-finger protein Snail (*SNAI1*) is a potent transcriptional inhibitor of Crumbs3 and to a lesser extent, PATJ and PALS1; resulting in mislocalisation of the Crumbs and Par complexes and disruption of tight-junction and polarity formation (Wang *et al.*, 2013; Whiteman *et al.*, 2014).

The Par Complex

The Par complex is named after the eponymous family of proteins first discovered in the 1980s as part of screen to identify maternal effect lethal mutations in the model nematode *Caenorhabditis elegans* (Kemphues *et al.*, 1988; Goldstein & Macara, 2007). Of the six, the par proteins most relevant to apical membrane specification are Par3 and Par6; both PDZ-domain scaffolding proteins and core members of the Par complex (Yu *et al.*, 2014; Hung & Kemphues, 1999). Other members of the Par complex include atypical protein kinase C (aPKC) and the cell division control protein 42 homologue (Cdc42) GTPase, binding to Par3 through Par6 which here acts as an adaptor (Joberty *et al.*, 2000). Par3 is also capable of binding aPKC directly; an interaction that is essential to establishing proper cell polarity, normal ductal architecture, and mammary gland morphogenesis (Nagai-Tamai *et al.*, 2002; McCaffrey & Macara, 2009).

The formation of the Par complex is cued by the establishment of cell-cell contacts. This comes as a result of the complex being anchored to the tight-junctions of the cell by Par3, which is tethered through its binding of phosphotidyl inositols and the junctional adhesion molecule (JAM) through the second and first of Par3's three PDZ domains, respectively (Wu *et al.*, 2007; Ebnet *et al.*, 2001).

At their basal levels of expression, the members of the Par complex are at a regulatory

equilibrium that is often disrupted in neoplasias, resulting in aberrant signal integration and epithelial disorganisation. Human breast cancers often express dramatically reduced levels of Par3, freeing aPKC to inappropriately activate signalling pathways that lead to increased invasive and metastatic potential; namely the human epidermal growth factor receptor 2 (HER2) and janus kinase/two Signal Transducer and Activator of Transcription (JAK/STAT) pathways (Xue *et al.*, 2013; McCaffrey *et al.*, 2012).

aPKC is but one member of the mammalian Protein Kinase C (PKC) super family, of which there are three additional members: the classical/conventional PKCs (cPKCs), the novel PKCs (nPKCs), and the later-discovered PKC-related kinases (PRKs). These families are distinguished by their dependence/independence on Ca^{2+} , and whether they are activated by diacylglycerols (DAGs). aPKC is both Ca^{2+} -independent and DAG-insensitive, and in this respect are identical to PRKs. The two families are primarily differentiated by PRKs association with RhoA, which is unique among the PKCs (Mellor & Parker, 1998).

Each PKC sub-family contains multiple isoforms, which each confer their own unique function. The aPKC exists in two isoforms, aPKC λ/ι and aPKC ζ .

2.2 Materials & Methods

2.3 Results

2.4 Discussion

Preface

3. PPReCOGG: A Model for the Per-Pixel Classification of Early Breast Lesions via Gabor Features

3.1 Introduction

When diagnosing breast cancer lesions, trained clinical pathologists rely on visual interpretation of patient biopsies. While interobserver variability between trained pathologists is a known problem, it is particularly pronounced when pathologists are called to make distinction between early breast cancer lesion stages (Gomes *et al.*, 2014).

Computer-aided detection (CADe) is sometimes used to enhance interpretation of imaging from screening mammographies and can help manage the ambiguous and subjective nature of medical imaging; however computational models are not currently used in most clinical settings to assist the diagnosis of breast cancers.

Despite many being described in the literature, a major limitation of most breast cancer classification models is that they fail to model intra-tumour heterogeneity as they commonly adopt whole-field classification modalities (Pareja *et al.*, 2017; Weigelt *et al.*, 2010). Whole-field classification models similarly do not offer insight into so-called “borderline”, which are lesions containing regions exhibiting features of multiple early lesions (Masood & Rosa, 2011).

To account for borderline lesions and the heterogeneous nature of cancer, a practical model for the classification of early lesions would be required to classify sub-regions of whole sections. As the characteristic “cobblestone” epithelial phenotype is altered distinctly in the early progression of breast cancers, such a model would benefit from leveraging features which are based on differences in cell patterning.

The observation that early lesions exhibit distinct cell patterning unique to themselves

3.1. INTRODUCTION

are reported in the first descriptions of hyperplasia and carcinoma *in situ* of both the mammary duct and lobules by Page *et al.* (1982). The descriptions provided of the early lesion are strongly based on cellular architecture and patterning; distinguishing, for example, ductal and lobular carcinomas *in situ* (DCIS and LCIS) from atypical ductal and lobular hyperplasias (ADH and ALH) by “...round, regular spacing” in the former and their absence in the latter, sometimes exhibiting “...swirls or streaming”.

With the aim of developing a model that classifies regions of early neoplasms from whole sections of breast biopsy according to aberrations in their cell patterning, the authors here present a model for the per-pixel recognition of cancers using oriented Gabor filters on the GPU (henceforth referred to by the acronym PPReCOGG). Analogous to the manner in which simple cortical cells perceive patterns and texture, the Gabor filter and has long been used to programmatically discern textures from one another (Fogel & Sagi, 1989; Mar  elja, 1980). Its purpose in this model is as a texture-dependent feature that differs between early cancer lesions according to their distinct epithelial patterning.

The PPReCOGG model achieves a high rate of accuracy with an average of $\approx 94.3\%$ of pixels being correctly classified on synthetic validation classification tasks, and also is demonstrated to effectively identify sub-regions exhibiting characteristic neoplastic cell patterning in images of human early breast lesions.

3.2 Design & Methods

3.2.1 Model Design

The PPreCOGG model is a k -nearest-neighbours (k -NN) trained on Gabor features extracted from human breast tumour tissue stained immunofluorescently for E-cadherin.

Gabor features were extracted in a similar fashion as Melendez *et al.* (2008) and is described by *Figure 3.1*. Namely, for each pixel in an image, six windows of increasing size (3×3 , 5×5 , 9×9 , 17×17 , 33×33 , 65×65) centred on the pixel are defined. Each window is then filtered through four Gabor kernels with quarter-turn orientations (*i.e.*: $\theta = \left\{ \frac{1}{2}\pi, \pi, \frac{3}{2}\pi, 2\pi \right\}$). Each Gabor kernel also has a sinusoidal wavelength of 0.25 pixels ($\lambda = 0.25$), which has been previously described as providing good discrimination in general-purpose texture classification (Manjunath & Ma, 1996).

The mean and the standard deviation of the resulting Gabor energies are then added to a vector for the relevant pixel. This results in a total of 48 features per pixel (6 windows $\times 4$ orientations $\times [1 \text{ mean} + 1 \text{ standard deviation}] = 48$ feature per pixel).

Gabor energy vectors extracted from the unknown image (i_u) are classified according to a known set of classes $C = \{\text{ADH}, \text{DCIS}\}$ extracted from images that are representative of these known classes ($I_C = \{I_{\text{ADH}}, I_{\text{DCIS}}\}$), as illustrated in *Figure 3.2*. Classification was performed using the k -NN algorithm.

3.2.2 Software Dependencies

The implementation of the k -NN algorithm by the Scikit-Learn (sklearn) python library was used for the purpose of classifying pixels (Pedregosa *et al.*, 2011). Multidimensional Scaling was also implemented by sklearn. Gabor kernels were generated by the OpenCV library via python language-bindings (Bradski, 2000) and convolved against the extracted windows on the GPU using the Theano library (Al-Rfou *et al.*, 2016). Training and classification tasks were performed on consumer hardware with an nVIDIA GTX 1050M GPU and Intel i7-7700HQ CPU with stock 2.8 GHz clockspeed.

Three-dimensional plots depicted in *Figure 3.3* were generated using the Plot.ly python

bindings from Plotly Technologies Inc. (2015). Accuracy and classification plots were created with the matplotlib python library (Hunter, 2007).

3.2.3 Datasets

Brodatz textures were obtained from the University of South California's Signal and Image Processing Institute (USC-SIPI) image database, and used within the terms under which they are distributed by Dover Publications, Incorporated (Brodatz, 1999).

Human sample tissue was provided by **clinician** in accordance with the guidelines of **ethics-council** and immunofluorescently labelled by **lab-member** as described in (Halaoui *et al.*, 2017).

3.2.4 Software and Data Availability

The software and documentation for the PPReCOGG model is available at:

<https://github.com/jszym/pprecogg/>

Computed gabor features can be downloaded at:

<https://jszym.com/software/pprecogg/> **404**

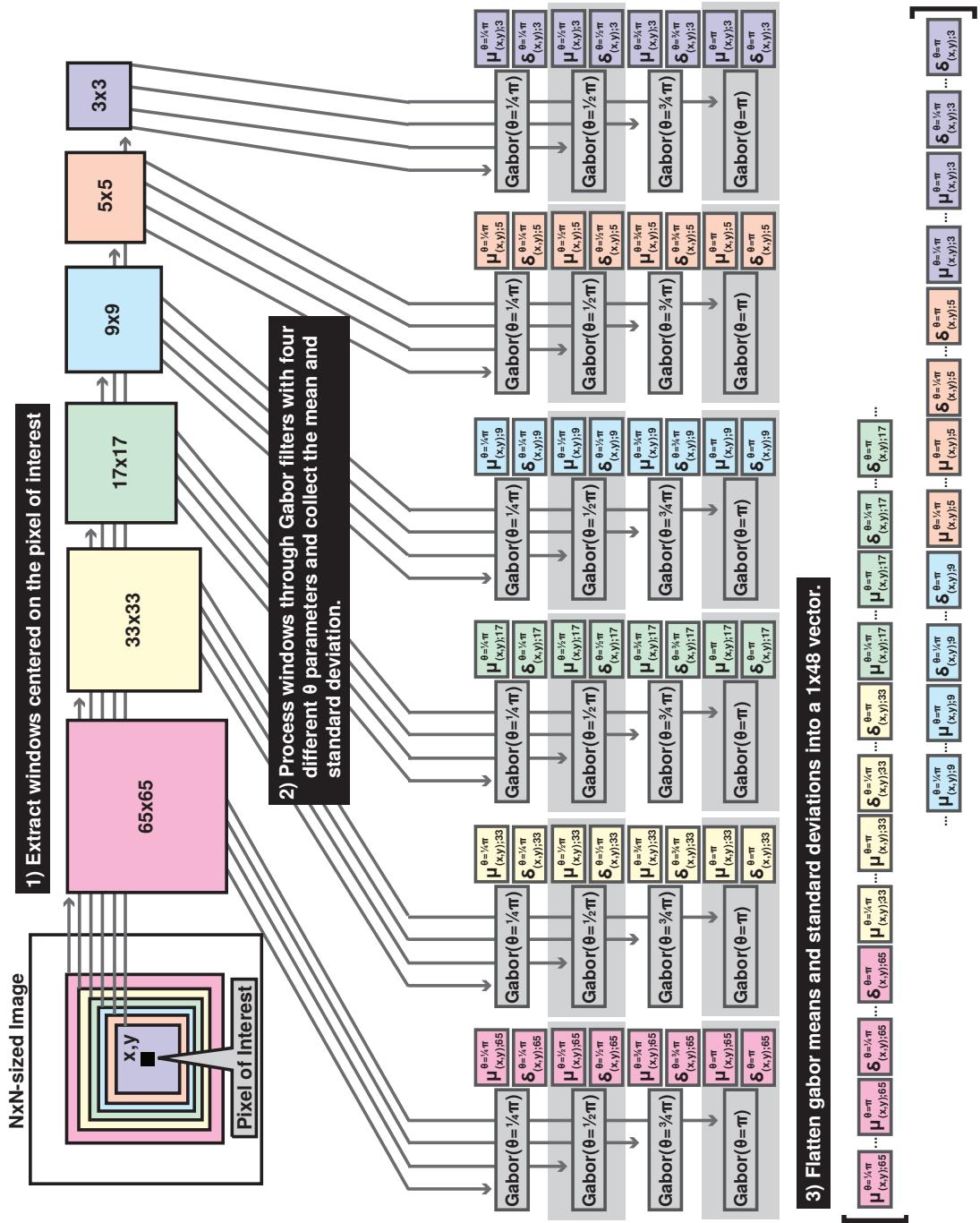


Figure 3.1: Diagram illustrating the method by which Gabor features are extracted from images in the PPreCOGG model.

3.2. DESIGN & METHODS

To reduce computational complexity, images are resized to a resolution of 256×256 pixels. Computation time of training is further reduced by extracting the features of a random sampling of one-quarter of the population of pixels (*i.e.*: For an $N \times N$ image where $N = 256$, $\lfloor N^2/4 \rfloor = 16,384$ pixels) from the total population .

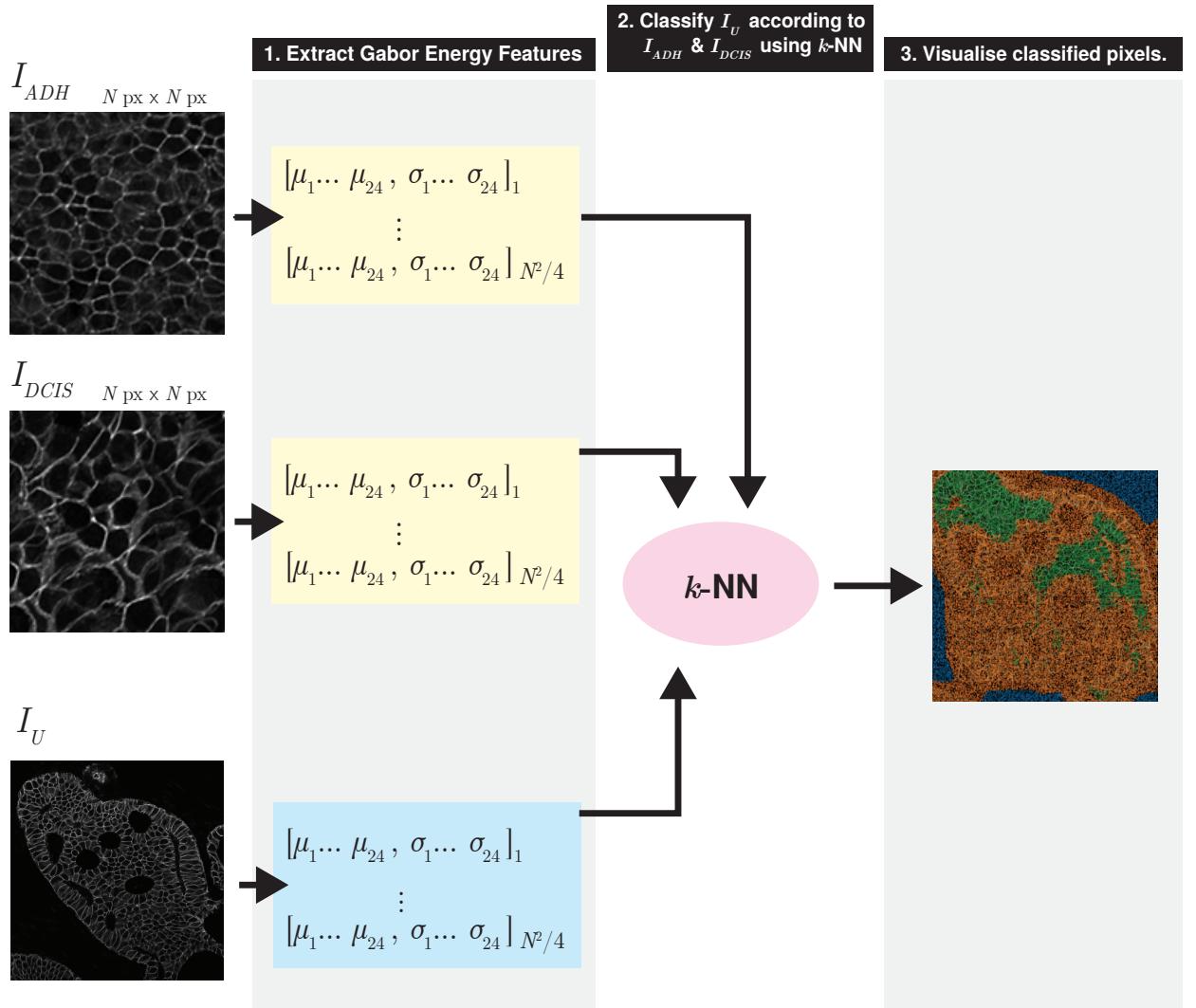


Figure 3.2: Schematic overview of the PPrECOGG model, wherein Gabor energy features are extracted from images of known classes (I_{ADH} , I_{DCIS}), as well as an image with regions belonging to different classes (I_U). Pixels are classified according to the class of their Gabor energy features, which are classified using the k -NN algorithm.

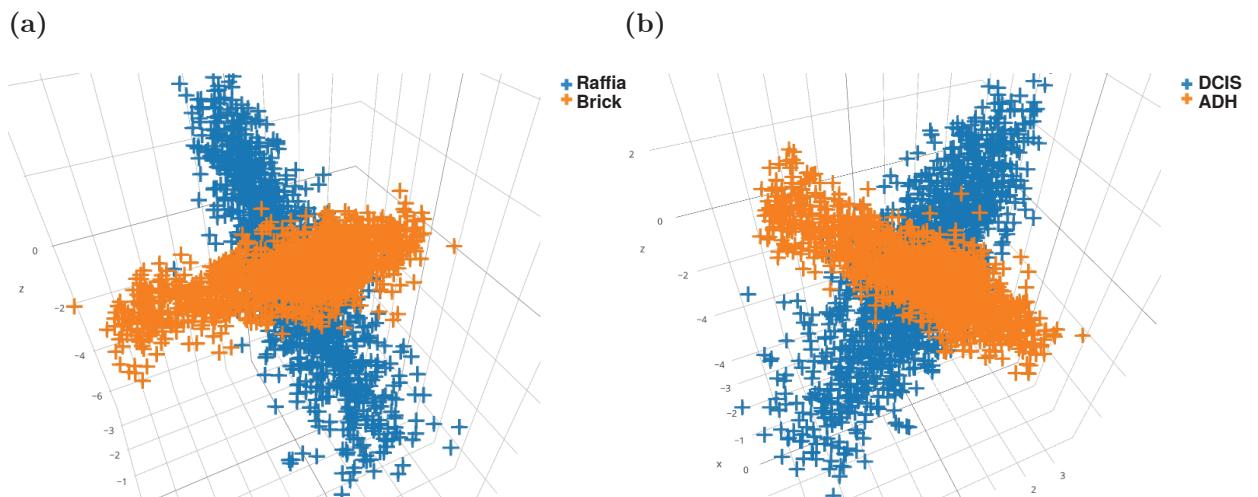
3.3 Results

3.3.1 PPreCOGG Classifies Sub-Regions of Perceptibly Distinct Textures in Synthetic Benchmarks

To assess the baseline ability of the PPreCOGG model to distinguish natural patterns from one-another, the PPreCOGG model was trained and benchmarked against the Brodatz textures (namely the “Raffia” and “Brick” textures) compiled in the USC-SIPI image dataset, which is a common image dataset used in the evaluation of image processing and texture recognition community (Weber, 1997). The brick and raffia textures were selected as they are visually distinct, and test images comprised of sub-regions of the two are perceptibly distinct upon visual inspection (*Table 3.1a*).

Three-dimensional multidimensional scaling (MDS) embeddings of the forty-eight-dimensional Gabor energy features calculated from the Brick and Raffia Brodatz textures reveal that while they intersect, the feature embeddings describe two distinct planes, each belonging to one of the two classes (*Figure 3.3a*).

Figure 3.3: MDS Embedding of Gabor Energy Features from Brodatz and Early Human Breast Lesion Datasets



Legend — Three-dimensional multidimensional scaling (MDS) embedding of Gabor energy features from known classes of the (a) Brodatz dataset and (b) the early human breast lesion dataset.

3.3. RESULTS

Table 3.1: Accuracy of the PPReCOGG Model Trained on Brodatz Textures

	Test Image One	Test Image Two	Test Image Three
(a) Original			
(b) Ground Truth			
(c) Classified Image			
(d) Raffia Accuracy Map			
(e) Brick Accuracy Map			
(f) Accuracy	90.94%	85.29%	94.00%

Legend — (a) Test images comprised of composites of the Brodatz textures entitled “Raffia” (pg. D84) and “Brick Wall” (pg. D94). (b) The ideal classification (or “ground truth”) of the original test image, where black pixels code for the raffia texture and white codes for the brick texture. (c) A random sampling of pixels classified by the PPReCOGG model. Pixels classified as Raffia are coded by blue points, and pixels classified as Brick are coded by orange pixels. (d) and (e) Classified pixels are here compared to and overlaid on their ground truth. Pixels which are correctly classified are coded in green, while false-positives are coded in red. (f) Accuracy of the model, as calculated by the quotient of the number of correctly classified pixels and the total number of classified pixels.

Table 3.2: Accuracy of the PPReCOGG Model Trained on Textures Derived from E-cadherin Staining of Human Lesions

	Test Image One	Test Image Two	Test Image Three
(a) Original			
(b) Ground Truth			
(c) Classified Image			
(d) Hyperplasia Accuracy Map			
(e) Carcinoma Accuracy Map			
(f) Accuracy	93.17%	93.60%	96.00%

Legend — (a) Test images comprised of composites of textures derived from E-cadherin staining of human lesions exhibiting characteristic hyperplastic or carcinomic cell patterning. (b) The ideal classification (or “ground truth”) of the original test image, where black pixels code for the hyperplastic texture and white codes for the carcinomic texture. (c) A random sampling of pixels classified by the PPReCOGG model. Pixels classified as hyperplasia are coded by blue points, and pixels classified as carcinoma are coded by orange pixels. (d) and (e) Classified pixels are here compared to and overlaid on their ground truth. Pixels which are correctly classified are coded in green, while false-positives are coded in red. (f) Accuracy of the model, as calculated by the quotient of the number of correctly classified pixels and the total number of classified pixels.

3.3. RESULTS

The PPreCOGG model achieved high average accuracy rates of 90% in benchmarks consisting of the Raffia and Brick Wall Brodaz textures (*Table 3.1*).

3.3.2 PPreCOGG Classifies Sub-Regions of Different Neoplastic Phenotypes in Synthetic Benchmarks with High Accuracy

Similar benchmarks were performed on test patterns composed of images of early lesions (ADH and DCIS) from human patient samples which had been immunofluorescently labelled for E-cadherin. These synthetic benchmarks are meant to simulate and quantitatively measure the efficiency of the PPreCOGG model in the task of classifying whole fields into sub-regions which exhibit cell patterning characteristic to certain early lesions.

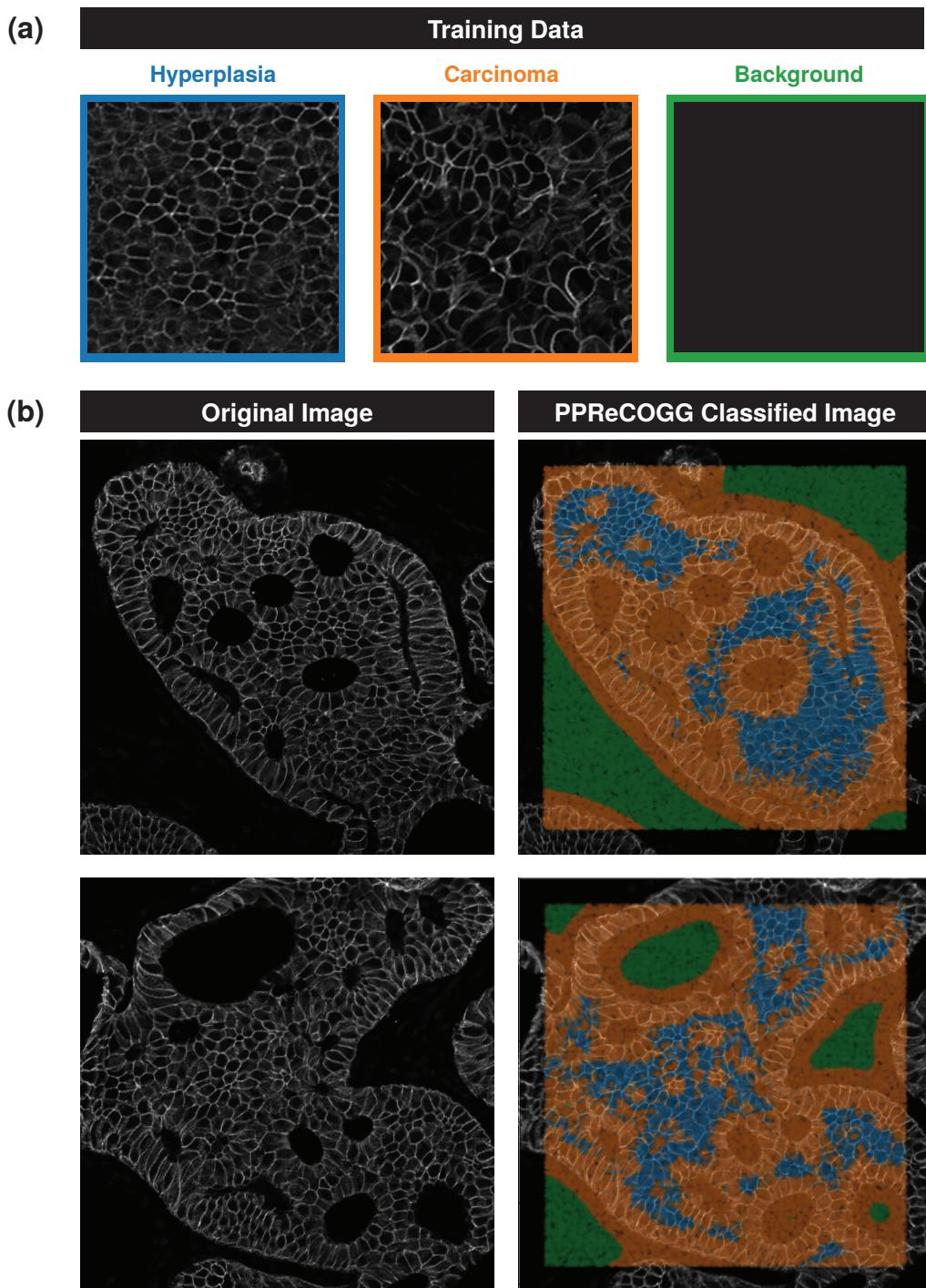
MDS embeddings of the 48-dimensional Gabor energy features for the human samples reveal results very similar to the embeddings of the features of the Brodaz textures; two distinct but intersecting planes (*Figure 3.3b*).

The PPreCOGG model achieved high accuracy rates on the human lesion benchmarks, achieving an accuracies ranging from 93.17% to 96.00% across all test images (*Table ??*).

3.3.3 PPreCOGG Effectively Classifies Sub-Regions of Different Neoplastic Phenotypes in Human Biopsy Samples

PPreCOGG model trained on E-cadhering patterning found in early breast lesion phenotypes (same features as those used in §3.3.2, see *Figure 3.4a*). This model was subsequently used to classify the pixels in images of early human breast lesions in with E-cadherin had been immunofluorescently labelled (*Figure 3.4b*). PPreCOGG effectively classifies sub-regions within the image that exhibit different cell-patterning characteristically found in early lesions.

Figure 3.4: Human Lesions as Classified by the PPReCOGG Model



Legend — Sections of human breast biopsies were immunofluorescently stained for E-cadherin and classified using the PPReCOGG model trained on two different early transformed phenotypes, and one background control. **(a)** Three representative fields of the 512 pixel by 512 pixel images used to train the PPReCOGG model. **(b)** Fields of breast E-cadherin labelled human breast lesion (left column) were classified by the trained PPReCOGG model (right column). Blue and orange pixels are classified as belonging to a hyperplastic carcinomic regions, respectively. Green pixels are classified as background signal.

3.4 Discussion

The PPreCOGG model described has here been shown to effectively recognise sub-regions of cell patterning in immunofluorescent confocal imagery, provided a training set that exemplifies said patterns. The PPreCOGG model has here been employed the task of identifying, within images of human breast lesions, cell patterns that are characteristic of early breast lesions. PPreCOGG classification in this task is effective and, thanks to its GPU accelerated implementation, performed in a practical time-frame.

The performance of the PPreCOGG model on the Brodatz texture synthetic benchmark attest to the models proficiency at texture recognition tasks between textures that are easily distinguished by human observers upon casual visual inspection (*Table 3.1*). However, PPreCOGGs true utility is most clearly evinced in the human breast lesion synthetic benchmarks, whereby texture recognition was performed on the test images (*Table 3.2a*) after being trained on a dataset of human lesions. In these visually challenging tasks, PPreCOGG’s accuracy is equal to or greater than those observed in the visually distinct Brodatz texture benchmarks.

A partial explanation for PPreCOGGs efficiency in both visually distinct and visually challenging texture recognition tasks is provided by the MDS embeddings of the underlying Gabor features of the training set for both the Brodatz and Human Breast Lesion datasets). The MDS embedding of both datasets are remarkably similar, with both classes in each case forming distinct but intersecting planes when scaled to three-dimensional space (*Figure 3.3*). The distance between the feature-spaces of each class defines the degree to which it is possible for PPreCOGG to distinguish between them. This is largely due to PPreCOGGs reliance on the k -nearest neighbour algorithm for the classification of features.

While the PPreCOGG model readily recognises textural sub-regions within clinical samples, any clinical utility of the PPreCOGG model is dependent on and currently precluded by an immature and incomplete training set. In order for the PPreCOGG model to offer meaningful interpretation and classification of early lesions, a rich dataset is required to capture the many textural manifestations of early lesions. ADH and DCIS

lesions are not homogenously or universally comprised of single textures, and so a sufficiently large and comprehensive dataset is required before the PPreCOGG model can be used to identify the many faces of early breast lesions.

In addition to a complete training dataset, it is possible to extend the current model to recognise sub-regions according to the identity of neighbouring sub-regions; such that some sub-region identified as belonging to some texture class A would only be reported as belonging to sub-type X if neighbouring regions belong to some texture class B but not C . Rule-sets for these contextual classifications can be learned through random-forest models trained on annotated images, manually according to existing pathology guidelines, or some combination of the two.

The underlying conventional machine-learning algorithm that is the basis of the PPreCOGG model does shape the nature of the conclusions that can be drawn from its output. Namely, the PPreCOGG model is a manifestation of our current understanding of the histopathology of early breast lesions. While this approach results in highly desirable and much needed quantitative and reproducible interpretation of the pathology of biopsy tissue, PPreCOGG as a consequence does not implement feature learning. This is in contrast to models based on neural network algorithms, which forego feature engineering for hidden layers which discover them independently through optimisation. Careful inspection of the hidden layers of the neural network can potentially lead to understanding of early lesion pathology interpretation previously overlooked or otherwise unknown, however such interpretation is nuanced and often provide incomplete “snapshots” of the internal state of the network (Erhan *et al.*, 2010; Zeiler & Fergus, 2013).

Preface

4. DeepDuct: A Deep-Learning Approach to Regional Breast Cancer Classification using Grad-CAM

4.1 Introduction

Histopathology is the study and practice of examining biological tissue at a microscopic level with the aim of detecting and/or analysing any disease that may happen to be manifesting therein. It is through the clinical histopathological analysis of tissue resulting from human mammary biopsies that all breast cancers are diagnosed.

Early breast lesions are associated with increased risk of invasive recurrence, and present important challenges for diagnosis by histopathology. Notably, in a consultation with clinical pathologists, a majority had cited distinguishing atypical ductal hyperplasias (ADH) from usual epithelial hyperplasias (UEH) and ductal carcinoma *in situ* (DCIS) as the most common challenge among their breast biopsy consultations (Putti *et al.*, 2005).

Challenges like these can be mitigated in part by computationally assisted detection and diagnosis (CADe/CADx) software, which analyse medical images in a reproducible and quantitative manner with the aim of making the interpretation of these data by clinicians a less complex and subjective practice. While CADe software is sometimes used to aid in the screening of breast mammographies, challenges such as dimensional complexity has historically prevented the use of CADe/x to help interpret histological data (Rangayyan *et al.*, 2007; Madabhushi, 2009).

4.1.1 Convolutional Neural Network Models for the Diagnosis of Breast Cancers

At the heart of an increasing amount of modern CADe/x solutions are the use of convolutional neural networks (CNNs or ConvNets) (?Cheng *et al.*, 2016). ConvNets are deep machine learning algorithms that use multiple weighted hidden layers of convolutional filters to make decisions about a given input. Recent developments in general-purpose computing on graphics processing units (GPGPU) have made ConvNets, whose use had historically been regarded as “unrealistic”, practical many years after their first conception (Crick, 1989). As a result, ConvNets have since been shown to be particularly well suited for the task of classifying and analysing images (Ciresan *et al.*, 2011, 2012).

ConvNets have been successfully used to create very accurate models for the classification of breast cancer lesions. Binary models for classifying benign and malignant lesions, as well as multi-class models for distinguishing between multiple subtypes of breast lesions from H&E stained biopsy slides have established with very high ($> 90\%$) accuracy (Wei *et al.*, 2017; Han *et al.*, 2017). These models, however, are severely limited in that they classify whole imaging fields as belonging to a single class. These approaches entirely ignore the heterogeneous nature of breast lesions and are entirely “black-boxes” for clinicians, offering no added dimensions of information and little understanding as to why the model has interpreted a lesion the way it has. To mitigate this limitation, a classifier would be required to identify, classify and annotate sub-regions that exhibit characteristics of early lesions in medical images of breast biopsies; transparently offering insights into the classifications being made. One such method to do so is to generate localisation annotations with the Gradient-weighted Class Activation Mapping algorithm.

4.1.2 Localisation-Augmented Visualisation of Convolutional Neural Network Using Grad-CAM

Recent work by Selvaraju *et al.* has made the interpretation of ConvNets much more clear by visually annotating inputs with general localisations of objects identified by the model. This technique, named Gradient-weighted Class Activation Mapping (Grad-CAM), can

produce heatmaps of areas within an input image that, according to a given ConvNet model, are likely to belong to a given class. Grad-CAM is generalisable to most ConvNet architectures, and does not require to be trained on example localisation annotations.

Grad-CAM has been since used for the dual purpose of localising classified regions and better understanding differences between classes in practical applications ranging from plant stress phenotyping to classifying colorectal polyps (Ghosal *et al.*, 2017; Korbar *et al.*, 2017).

4.1.3 Transfer-Learning for Resource Efficient Training of Neural Networks

Two common limitations of adapting convolutional networks to domain-specific tasks such as classifying medical imagery for computer-aided detection are the large dataset and computational power requirements. These two limitations can be largely addressed by the process of transfer learning, which uses an existing convolutional architecture that has been previously trained (“pre-trained”) on a sufficiently generalised dataset appropriate for the target task (Pan & Yang, 2010). The ImageNet ILSVRC2014 dataset is an example of a widely-adopted, readily-available, and comprehensive general-purpose dataset that is commonly the basis of pre-trained models used for transfer-learning image classification tasks (Russakovsky *et al.*, 2015).

Transfer learning has in-fact been used in a number of computer-aided detection, ranging from thoraco-abdominal lymph node detection and interstitial lung disease classification from chest X-ray and CT scan imaging to classification of skin cancers from dermatoscope imagery (Shin *et al.*, 2016; Esteva *et al.*, 2017).

Transfer learning uses the weights of the many hidden layers of the pre-trained network. The last fully-connected layer of the network is removed from the architecture and a new linear classifier for the network is trained on the new dataset using the pre-trained hidden layers as features.

A limitation of transfer learning is that the later, more specialised, hidden layers of the pre-trained network can lead to reduced accuracy of the model if the original

dataset the pre-trained layer was trained against is extremely different from the new dataset. This challenge is usually met by an additional process known as fine-tuning, which continues to train the hidden layers of the pre-trained network against the new dataset using backpropagation (Yosinski *et al.*, 2014).

In this work, we describe the DeepDuct model, a pre-trained ConvNet model (namely, VGG16) fine-tuned on a dataset comprised of histological images of breast biopsies classified across eight different lesion types (the BreakHis dataset), and combines it with the Grad-CAM algorithm to provide general localisation of the various lesions identified, while providing an opportunity to better understand the model.

VGG16 is a sixteen weight-layer ConvNet architecture (*Figure 4.2*) that has been engineered by (and named for) the Visual Geometry Group at the University of Oxford (Simonyan & Zisserman, 2014). The VGG16 model has been shown to generalise very well to a number of different datasets, and is particularly well suited for localisation tasks, having been awarded first and second place in the classification & localisation task of the ImageNet ILSVRC2014 contest (Russakovsky *et al.*, 2015). This makes the use of the VGG16 architecture well-suited for the function of localisation and classification of medical imagery for the purpose of computational detection and diagnosis.

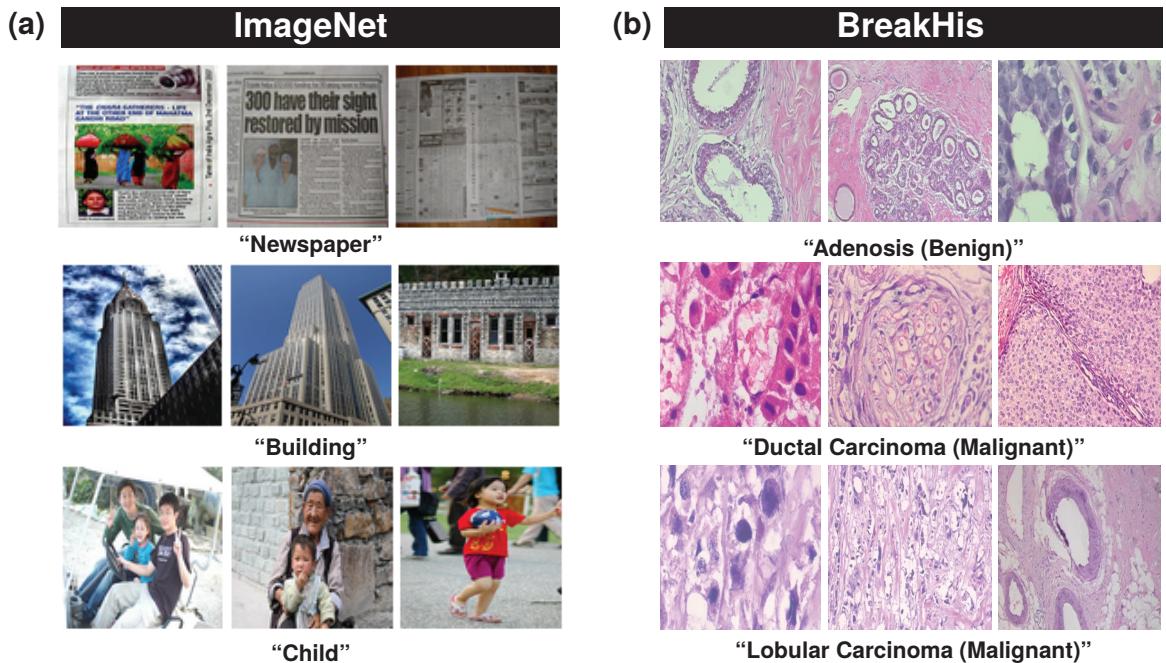
4.2 Design & Methods

The DeepDuct model begins with a so-called “off-the-shelf” ConvNet architecture; namely the VGG16 model pre-trained on the general-purpose ImageNet dataset (Simonyan & Zisserman, 2014; Deng *et al.*, 2009; Russakovsky *et al.*, 2015). The pre-trained VGG16 model was fine-tuned on the BreakHis dataset, a dataset comprised of approximately 8,000 images of H&E stained human mammary biopsy sections classified according to World Health Organisation (WHO) guidelines (Spanhol *et al.*, 2016; Sunil R. Lakhani, 2012). Abbreviations for the class names present in the BreakHis dataset are used throughout this manuscript, and within the model itself. Refer to Appendix A for a legend of these class codes.

The VGG16 model was implemented with Keras using TensorFlow as the back-end (Chollet *et al.*, 2015; Abadi *et al.*, 2016). A Keras/TensorFlow implementation of Grad-CAM implemented in the Keras-Vis library was used to generate attention maps from the BreakHis-trained VGG16 model (Kotikalapudi & contributors, 2017).

Plots created with the matplotlib and searborn libraries (Hunter, 2007; Waskom *et al.*, 2014).

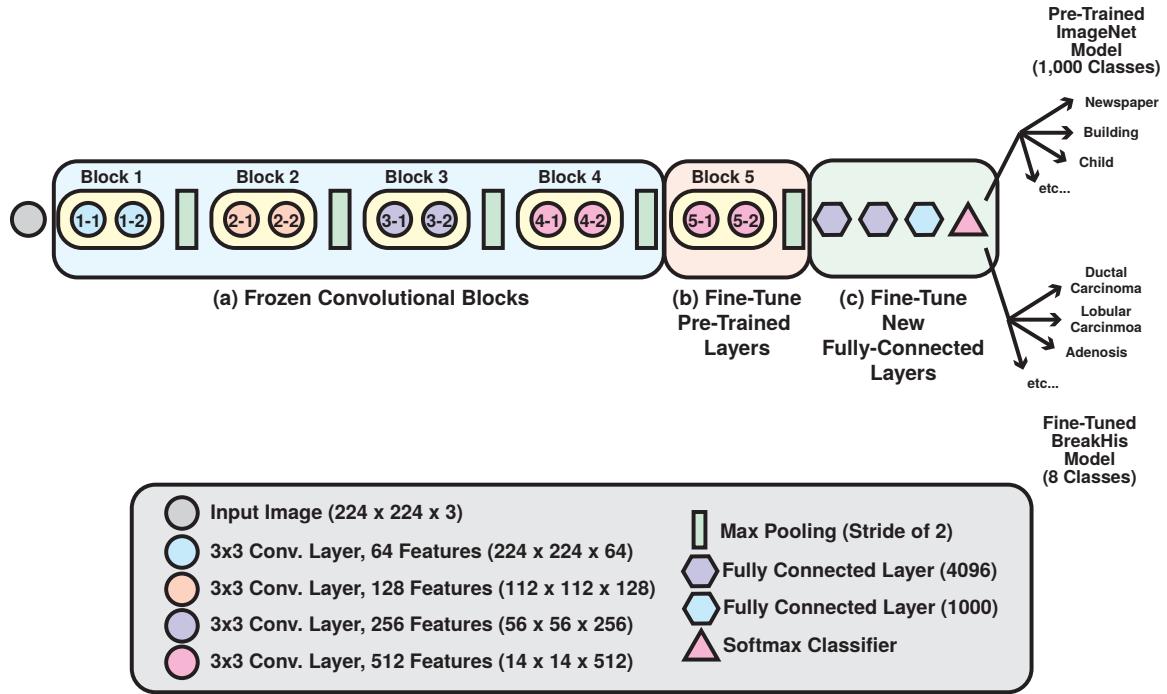
Figure 4.1: Examples from the ImageNet and BreakHis Datasets



Legend — Thumbnails from some selected classes of the (a) ImageNet and (b) BreakHis datasets. The DeepDuct model makes use of a ConvNet model pre-trained on the ImageNet dataset and repurposed for classifying images according to the BreakHis dataset via transfer learning. The ImageNet ILSVRC2014 dataset is comprised of $\approx 150,000$ images belonging to 1,000 classes, and the BreakHis dataset is comprised of 8,000 images belonging to 8 classes.

4.3. RESULTS

Figure 4.2: Schematic of the VGG16 ConvNet Architecture and its Fine-Tuning



Legend — The VGG16 model is comprised of 16 weight layers, making up five convolutional blocks and a fully-connected classifier. Fine-tuning a pre-trained VGG16 model involves freezing the first four blocks (a), continuing to train the fifth block (b) against the new dataset (in this case, BreakHis) through backpropagation, and finally training a new fully-connected classifier against the new dataset (c).

4.3 Results

4.3.1 Accurate Classification of Breast Neoplasms in the BreakHis Dataset

Transfer-learning the VGG16 on the unmodified BreakHis dataset results in acceptable overall classification accuracy (70%), however closer inspection reveals bias towards one of the classes (ductal carcinoma) due to imbalances in the number of examples between classes (*Figure 4.3a*). Oversampling the dataset such that all classes have an equal number of training examples resulted in a small improvement in overall classification accuracy (72%), and shows a demonstrable reduction in bias (*Figure 4.3b*).

Notably, lobular carcinomas were mistaken for ductal carcinomas in just over two-thirds of the validation set, and correctly identified a quarter of lobular carcinoma ex-

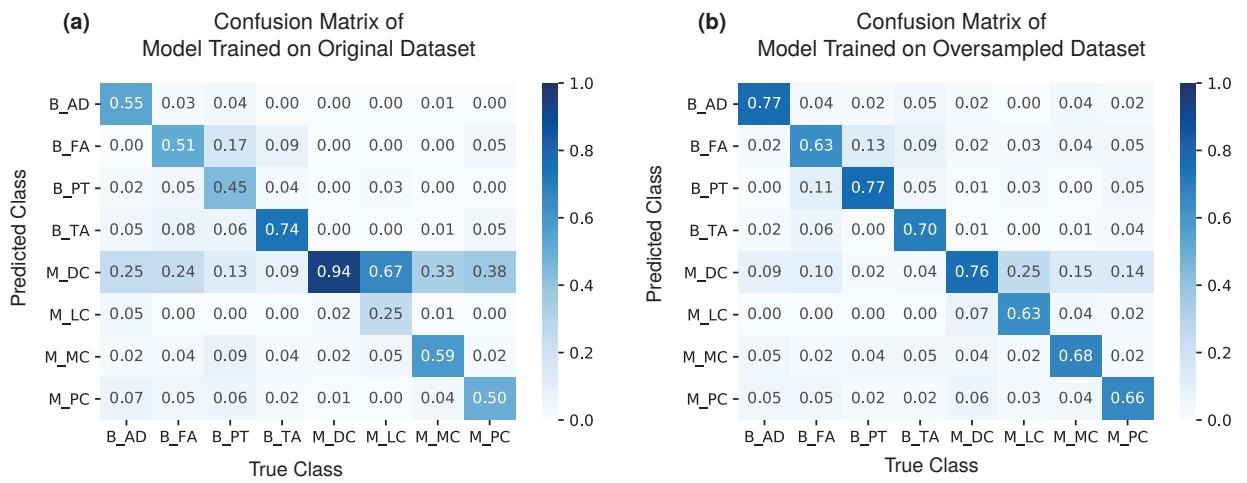
amples. After oversampling the BreakHis dataset, the accuracy and ductal carcinoma false-positive rates have nearly replaced one another, with lobular carcinomas being correctly classified in 63% of cases and false ductal carcinoma classifications in a quarter of cases. Ductal carcinoma false-positives were in fact halved, on average, across nearly all classes after oversampling.

4.3.2 Activation Mapping of BreakHis ConvNet using Grad-CAM

The Grad-CAM algorithm was used to compute class activation maps from the BreakHis fine-tuned VGG16 ConvNet model (*Figure 4.4*). ...

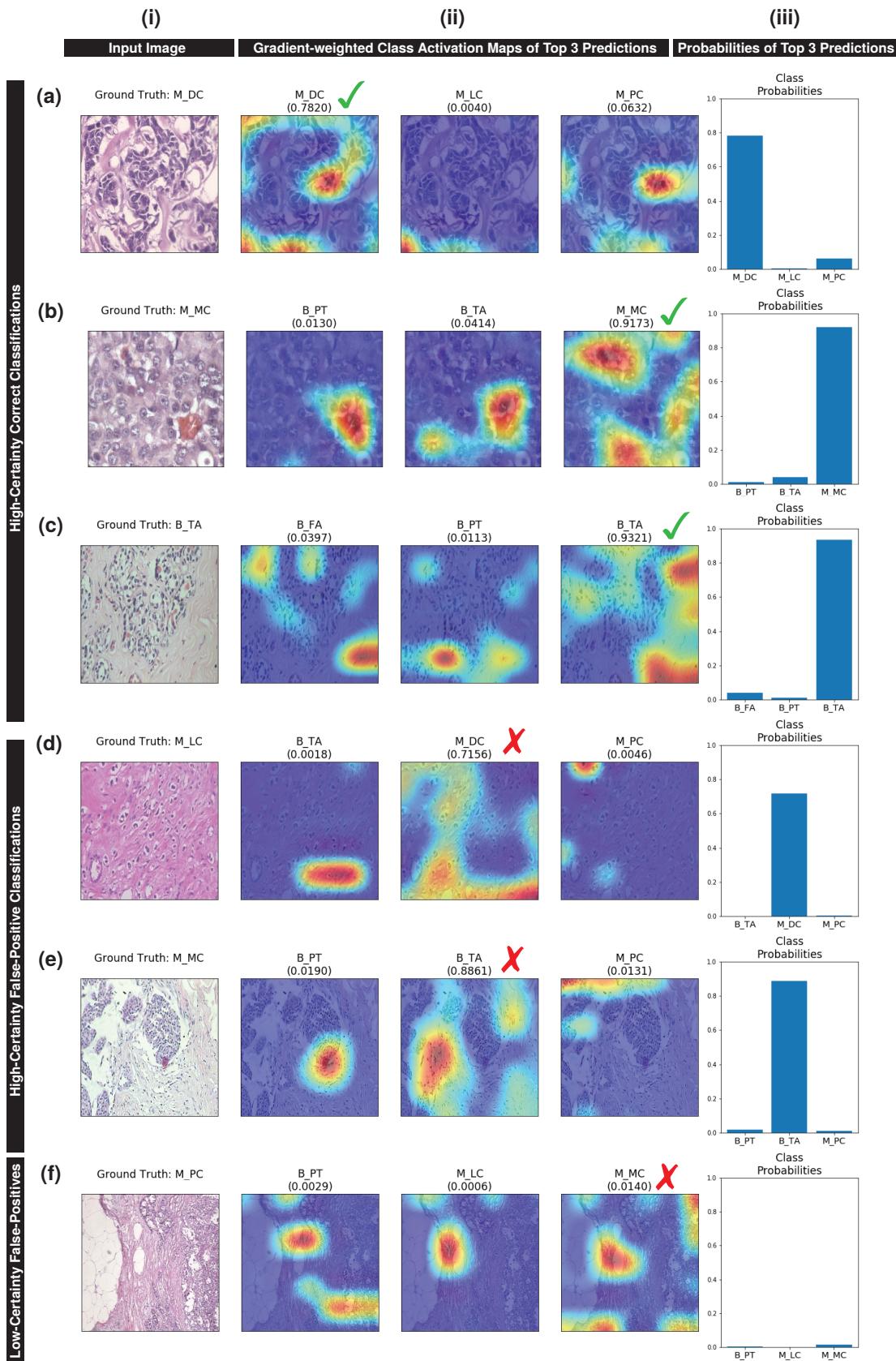
4.3. RESULTS

Figure 4.3: Confusion Matrix of VGG16 Model Trained on the Original and Oversampled BreakHis Datasets



Legend — Confusion matrices of the VGG16 model either trained on (a) the original BreakHis dataset as provided by its creators, or (b) a modified version of the BreakHis dataset oversampled such that each class has the same number of training examples. The model trained on the original dataset achieved an average accuracy of 70% on the validation set across all classes, while the oversampled model saw a modest increase for a final average accuracy of 72%. See Appendix A for class code legend used in this figure.

Figure 4.4: Activation Maps of the BreakHis Fine-Tuned VGG16 ConvNet



Legend — Selected examples of activation maps of the BreakHis fine-tuned VGG16 model described herein, as calculated by the Grad-CAM algorithm. Column (i) displays the original input image and its ground truth classification. Column (ii) shows the top three predictions of the ConvNet and their activation maps overlayed on the input image; as well as the probability of the classification as reported by the model’s softmax layer (shown in brackets). Column (iii) reports the probabilities of the top three predictions in a bar chart. Top probabilities are indicated with a checkmark (\checkmark) when they match the ground truth (correct classification), and an “x” mark (\times) when they do not (false-positives). Rows (a–c) are examples of correct classifications made by the ConvNet model where the top prediction is of high probability. Rows (d–e) are examples of incorrect classification made by the ConvNet model where the top prediction is of high probability (false-positives). Row (f) is an example of an incorrect classification where none of the classes are reported to be of high-probability.

4.4 Discussion

The DeepDuct model described herein provides a proof-of-concept framework for the localisation of breast lesions from H&E staining that does not rely on manual feature selection, transparently reports explanations for its predictions via class activation maps and allows for the potential discovery of new features that could inform future manual diagnosis.

Neural networks have the advantage of dynamically “learning” and optimising features, as opposed to relying on a manual process of feature engineering that is often driven by limited powers of intuition and conventional knowledge. The automated process of feature engineering allows for the potential discovery of new underlying concepts previously not described in the literature that fundamentally define a class from others, and also prevents assumptions and misconceptions from biasing features and the resulting inaccuracies.

4.4.1 Explanations for Neural Network Predictions are Essential for Clinical Use of CADe/x Models

Whole-image ConvNet classification models fine-tuned on the BreakHis dataset have been previously described reporting high-accuracy, as well as patch-based whole-slide classifiers which apply whole-image classification to small patches of an imaged slide resulting, in a form of tumour localisation (Han *et al.*, 2017; Wang *et al.*, 2016). None of these models, however, offer the same extent of transparency and resolution offered by the activation maps provided by the Grad-CAM algorithm used in the DeepDuct model. Existing models of breast lesion classification and localisation remain “black-box” solutions to end-users, particularly those without in-depth knowledge of deep learning algorithms.

Unique among deep-learning based breast lesion classifiers, DeepDuct reports which regions of the input image have lead the model to classify the image as it had. This simultaneously allows for general localisation of classified objects and a glimpse into the internal state of the model, informing and not prescribing a diagnosis. This transparency is essential for a model to see use in contexts such as clinical settings where acting

on predictions in blind faith is not an option due to the high-risk associated with the decisions being made. Models that implement “explanations” for their predictions have indeed been shown to increase end-user trust in model predictions, as well as help identify false-positive predictions made by a given model (Ribeiro *et al.*, 2016). To the author’s understanding, DeepDuct is the first application of such explanatory algorithms to the classification and localisation of breast lesions from medical imaging.

4.4.2 Dataset Considerations

subsection-needs-better-name

The BreakHis dataset, while covering a number of relevant lesion types with a significantly large number of examples for each type, presents some important challenges. Firstly, despite the multiple magnifications provided, images in the BreakHis dataset are not of high-resolution, taken with a digital camera with pixel size of $6.5\mu\text{m}$ and resolution of 480 TV lines. Secondly, the number of examples is extremely imbalanced between classes, with as much of a 7.5-fold difference between the least represented class and the most represented class.

While low-resolution images can be useful for learning so-called “global features”, they’ve proven to be problematic when distinguishing differences between objects with similar high-level features, as is the case between two H&E images exhibiting different lesion subtypes. This problem is illustrated well in the description of Baidu’s Deep Image model, whereby similar objects (such as insects of the same species) can only be distinguished from one another when higher-resolution images are considered in the model (Wu *et al.*, 2015). Training the DeepDuct model on higher resolution datasets would address this concern. High-resolution datasets of breast lesions do exist, but many offer too few examples (INESCTC) or do not offer histological type information outside grade (CAMELYON16) [citation-needed](#).

As described early, imbalances in the number of examples provided per class in the BreakHis dataset had lead to a strong bias towards over-represented classes (*Figure 4.3a*). This bias was addressed by oversampling all under-represented classes by duplicating

examples until all classes in the training set contained the same number of examples (*Figure 4.3b*).

4.4.3 Future Directions & Improvements

Implementing the DeepDuct model on smartphones would afford clinicians low-cost, mobile tools for the annotated classification of breast histology slides through use of commercial or 3D-printable smartphone-microscope adapters (Roy *et al.*, 2014). A less computationally-complex mobile DeepDuct implementation would be required to account for the limited resources available on the platform. This is typically achieved either by a networked server-client model supported by computation in the cloud, or by replacing the deep, resource-heavy VGG16 model with a more shallow mobile-oriented model such as SqueezeNet (Iandola *et al.*, 2016). While the former is limited by patient-privacy compliance and network connectivity, the latter requires retraining the network on a shallower ConvNet architecture with potential losses in accuracy.

Regional convolutional neural networks (R-CNNs), such as Facebook’s Mask R-CNN, have been developed to provide pixel-resolution object detection in complex scenes (He *et al.*, 2017). Using a Mask R-CNN model trained on a breast lesion dataset can provide higher resolution lesion detection than existing patch-based breast lesion models (Wang *et al.*, 2016).

5. General Discussion

Bibliography

Abadi, Martín, Agarwal, Ashish, Barham, Paul, Brevdo, Eugene, Chen, Zhifeng, Citro, Craig, Corrado, Gregory S., Davis, Andy, Dean, Jeffrey, Devin, Matthieu, Ghemawat, Sanjay, Goodfellow, Ian J., Harp, Andrew, Irving, Geoffrey, Isard, Michael, Jia, Yangqing, Józefowicz, Rafal, Kaiser, Lukasz, Kudlur, Manjunath, Levenberg, Josh, Mané, Dan, Monga, Rajat, Moore, Sherry, Murray, Derek Gordon, Olah, Chris, Schuster, Mike, Shlens, Jonathon, Steiner, Benoit, Sutskever, Ilya, Talwar, Kunal, Tucker, Paul A., Vanhoucke, Vincent, Vasudevan, Vijay, Viégas, Fernanda B., Vinyals, Oriol, Warden, Pete, Wattenberg, Martin, Wicke, Martin, Yu, Yuan, & Zheng, Xiaoqiang. 2016. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *CoRR*, **abs/1603.04467**.

Acott, A. A., & Mancino, A. T. 2016. Flat epithelial atypia on core needle biopsy, must we surgically excise? *The American Journal of Surgery*, **212**(6), 1211–1213.

Al-Rfou, Rami, Alain, Guillaume, Almahairi, Amjad, Angermueller, Christof, Bahdanau, Dzmitry, Ballas, Nicolas, Bastien, Frédéric, Bayer, Justin, Belikov, Anatoly, Belopolsky, Alexander, & et al. 2016. Theano: A Python framework for fast computation of mathematical expressions. *arXiv e-prints*, **abs/1605.02688**(May).

Ayyaci, Mehmet US, Alagoz, Oguzhan, Chhatwal, Jagpreet, Munoz del Rio, Alejandro, Sickles, Edward A., Nassif, Houssam, Kerlikowske, Karla, & Burnside, Elizabeth S. 2014. Predicting invasive breast cancer versus DCIS in different age groups. *BMC Cancer*, **14**(Aug), 584.

Badve, Sunil, A'hern, Roger P, Ward, Ann M, Millis, Rosemary R, Pinder, Sarah E, Ellis, Ian O, Gusterson, Barry A, & Sloane, John P. 1998. Prediction of local recurrence of ductal carcinoma in situ of the breast using five histological classifications: A comparative study with long follow-up. *Human Pathology*, **29**(9), 915–923.

Betsill, William L., Rosen, Paul Peter, Lieberman, Philip H., & Robbins, Guy F. 1978. Intraductal Carcinoma: Long-Term Follow-up After Treatment by Biopsy Alone. *JAMA*, **239**(18), 1863–1867.

BIBLIOGRAPHY

- Bissell, Mina J, Rizki, Aylin, & Mian, I Saira. 2003. Tissue architecture: the ultimate regulator of breast epithelial function. *Current Opinion in Cell Biology*, **15**(6), 753–762.
- Bleicher, Richard J. 2013. Ductal carcinoma in situ. *Surgical Clinics of North America*, **93**(2), 393–410.
- Boecker, Werner, Moll, Roland, Dervan, Peter, Buerger, Horst, Poremba, Christopher, Diallo, Raihanatou Ina, Herbst, Hermann, Schmidt, Ansgar, Lerch, Markus M, & Buchwalow, Igor B. 2002. Usual ductal hyperplasia of the breast is a committed stem (progenitor) cell lesion distinct from atypical ductal hyperplasia and ductal carcinoma in situ. *The Journal of Pathology*, **198**(4), 458–467.
- Bombonati, Alessandro, & Sgroi, Dennis C. 2011. The molecular pathology of breast cancer progression. *The Journal of Pathology*, **223**(2), 308–318.
- Bradski, G. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.
- Brodatz, Phil. 1999. *Textures: A Photographic Album for Artists and Designers*. Dover Publications.
- Canadian Cancer Society's Advisory Committee on Cancer Statistics. 2015. *Canadian Cancer Statistics 2015*.
- Cheng, Jie-Zhi, Ni, Dong, Chou, Yi-Hong, Qin, Jing, Tiu, Chui-Mei, Chang, Yeun-Chung, Huang, Chiun-Sheng, Shen, Dinggang, & Chen, Chung-Ming. 2016. Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans. *Scientific Reports*, **6**(Apr), srep24454.
- Chollet, François, et al. 2015. *Keras*. <https://github.com/fchollet/keras>.
- Cireşan, Dan, Meier, Ueli, & Schmidhuber, Juergen. 2012. Multi-column Deep Neural Networks for Image Classification. *arXiv:1202.2745 [cs]*, Feb. arXiv: 1202.2745.
- Cireşan, Dan C., Meier, Ueli, Masci, Jonathan, Gambardella, Luca M., & Schmidhuber, Jürgen. 2011. Flexible, High Performance Convolutional Neural Networks for Image Classification. *Pages 1237–1242 of: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two*. IJCAI'11. AAAI Press.
- Clark, S E, Warwick, J, Carpenter, R, Bowen, R L, Duffy, S W, & Jones, J L. 2011. Molecular subtyping of DCIS: heterogeneity of breast cancer reflected in pre-invasive disease. *British Journal of Cancer*, **104**(1), 120–127.
- Cover, T., & Hart, P. 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, **13**(1), 21–27.

- Crick, Francis. 1989. The recent excitement about neural networks. *Nature*, **337**(6203), 129–132.
- Decaestecker, Christine, Salmon, Isabelle, Dewitte, Olivier, Camby, Isabelle, Van Ham, Philippe, Pasteels, Jean-Lambert, Brotchi, Jacques, & Kiss, Robert. 1997. Nearest-neighbor classification for identification of aggressive versus nonaggressive low-grade astrocytic tumors by means of image cytometry-generated variables. *Journal of Neurosurgery*, **86**(3), 532–537.
- Deng, J., Dong, W., Socher, R., Li, L. J., Li, Kai, & Fei-Fei, Li. 2009 (June). ImageNet: A large-scale hierarchical image database. *Pages 248–255 of: 2009 IEEE Conference on Computer Vision and Pattern Recognition*.
- Dhawan, A. P., Chitre, Y., & Kaiser-Bonasso, C. 1996. Analysis of mammographic microcalcifications using gray-level image structure features. *IEEE Transactions on Medical Imaging*, **15**(3), 246–259.
- Dion, L., Racine, A., Brousse, S., Beltjens, F., Cauchois, A., Levêque, J., Coutant, C., & Lavoue, V. 2016. Atypical epithelial hyperplasia of the breast: state of the art. *Expert Review of Anticancer Therapy*, **16**(9), 943–953.
- Dreiseitl, Stephan, Ohno-Machado, Lucila, Kittler, Harald, Vinterbo, Staal, Billhardt, Holger, & Binder, Michael. 2001. A Comparison of Machine Learning Methods for the Diagnosis of Pigmented Skin Lesions. *Journal of Biomedical Informatics*, **34**(1), 28–36.
- Ebnat, Klaus, Suzuki, Atsushi, Horikoshi, Yosuke, Hirose, Tomonori, Brickwedde, Maria-Katharina Meyer zu, Ohno, Shigeo, & Vestweber, Dietmar. 2001. The cell polarity protein ASIP/PAR-3 directly associates with junctional adhesion molecule (JAM). *The EMBO Journal*, **20**(14), 3738–3748.
- Engstrøm, M. J., Opdahl, S., Hagen, A. I., Romundstad, P. R., Akslen, L. A., Haugen, O. A., Vatten, L. J., & Bofin, A. M. 2013. Molecular subtypes, histopathological grade and survival in a historic cohort of breast cancer patients. *Breast Cancer Research and Treatment*, **140**(3), 463–473.
- Erhan, Dumitru, Courville, Aaron, & Bengio, Yoshua. 2010. Understanding representations learned in deep architectures.
- Esteva, Andre, Kuprel, Brett, Novoa, Roberto A., Ko, Justin, Swetter, Susan M., Blau, Helen M., & Thrun, Sebastian. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, **542**(7639), nature21056.
- Fix, Evelyn, & Hodges, Jr. 1951. *Discriminatory Analysis - Nonparametric Discrimination: Consistency Properties*.

BIBLIOGRAPHY

- Fogel, I., & Sagi, D. 1989. Gabor filters as texture discriminator. *Biological Cybernetics*, **61**(2), 103–113.
- Ghosal, Sambuddha, Blystone, David, Singh, Asheesh K., Ganapathysubramanian, Baskar, Singh, Arti, & Sarkar, Soumik. 2017. Interpretable Deep Learning applied to Plant Stress Phenotyping. *arXiv:1710.08619 [cs, stat]*, Oct. arXiv: 1710.08619.
- Goldstein, Bob, & Macara, Ian G. 2007. The PAR Proteins: Fundamental Players in Animal Cell Polarization. *Developmental Cell*, **13**(5), 609–622.
- Gomes, Douglas S., Porto, Simone S., Balabram, Débora, & Gobbi, Helenice. 2014. Inter-observer variability between general pathologists and a specialist in breast pathology in the diagnosis of lobular neoplasia, columnar cell lesions, atypical ductal hyperplasia and ductal carcinoma in situ of the breast. *Diagnostic Pathology*, **9**(1), 121.
- Halaoui, Ruba, Rejon, Carlis, Chatterjee, Sudipa June, Szymborski, Joseph, Meterissian, Sarkis, Muller, William J., Omeroglu, Atilla, & McCaffrey, Luke. 2017. Progressive polarity loss and luminal collapse disrupt tissue organization in carcinoma. *Genes & Development*, **31**(15), 1573–1587.
- Hamperl, H. 1970. *The Myothelia (Myoepithelial Cells)*. Springer, Berlin, Heidelberg. Page 161–220.
- Han, Zhongyi, Wei, Benzheng, Zheng, Yuanjie, Yin, Yilong, Li, Kejian, & Li, Shuo. 2017. Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model. *Scientific Reports*, **7**(1), 4172.
- Hartmann, Lynn C., Degnim, Amy C., Santen, Richard J., Dupont, William D., & Ghosh, Karthik. 2015. Atypical Hyperplasia of the Breast — Risk Assessment and Management Options. *The New England Journal of Medicine*, **372**(1), 78–89.
- He, Kaiming, Gkioxari, Georgia, Dollár, Piotr, & Girshick, Ross. 2017. Mask R-CNN. *arXiv:1703.06870 [cs]*, Mar. arXiv: 1703.06870.
- Hinck, Lindsay, & Näthke, Inke. 2014. Changes in cell and tissue organization in cancer of the breast and colon. *Current opinion in cell biology*, **0**(Feb), 87–95.
- Horikoshi, Yosuke, Suzuki, Atsushi, Yamanaka, Tomoyuki, Sasaki, Kazunori, Mizuno, Keiko, Sawada, Hajime, Yonemura, Shigenobu, & Ohno, Shigeo. 2009. Interaction between PAR-3 and the aPKC–PAR-6 complex is indispensable for apical domain development of epithelial cells. *Journal of Cell Science*, **122**(10), 1595–1606.
- Hung, T. J., & Kemphues, K. J. 1999. PAR-6 is a conserved PDZ domain-containing protein that colocalizes with PAR-3 in *Caenorhabditis elegans* embryos. *Development*, **126**(1), 127–135.

- Hunt, Earl B., Marin, Janet, & Stone, Philip J. 1966. *Experiments in Induction*. Academic Press.
- Hunter, J. D. 2007. Matplotlib: A 2D graphics environment. *Computing In Science & Engineering*, **9**(3), 90–95.
- Iandola, Forrest N., Han, Song, Moskewicz, Matthew W., Ashraf, Khalid, Dally, William J., & Keutzer, Kurt. 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and \approx0.5MB model size. *arXiv:1602.07360 [cs]*, Feb. arXiv: 1602.07360.
- Joberty, Gérard, Petersen, Clark, Gao, Lin, & Macara, Ian G. 2000. The cell-polarity protein Par6 links Par3 and atypical protein kinase C to Cdc42. *Nature Cell Biology*, **2**(8), 531–539.
- Kemphues, Kenneth J., Priess, James R., Morton, Diane G., & Cheng, Niansheng. 1988. Identification of genes required for cytoplasmic localization in early *C. elegans* embryos. *Cell*, **52**(3), 311–320.
- Kerlikowske, Karla. 2010. Epidemiology of Ductal Carcinoma In Situ. *Journal of the National Cancer Institute. Monographs*, **2010**(41), 139–141.
- Kononenko, Igor. 2001. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, **23**(1), 89–109.
- Korbar, Bruno, Olfson, Andrea M, Miraflor, Allen P, Nicka, Catherine M, Suriawinata, Matthew A, Torresani, Lorenzo, Suriawinata, Arief A, & Hassanpour, Saeed. 2017. Looking Under the Hood: Deep Neural Network Visualization to Interpret Whole-Slide Image Analysis Outcomes for Colorectal Polyps. *Pages 821–827 of: Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*. IEEE.
- Kotikalapudi, Raghavendra, & contributors. 2017. *keras-vis*. <https://github.com/raghakot/keras-vis>.
- Laprise, Patrick, Beronja, Slobodan, Silva-Gagliardi, Nancy F., Pellikka, Milena, Jensen, Abbie M., McGlade, C. Jane, & Tepass, Ulrich. 2006. The FERM Protein Yurt Is a Negative Regulatory Component of the Crumbs Complex that Controls Epithelial Polarity and Apical Membrane Size. *Developmental Cell*, **11**(3), 363–374.
- Lemmers, Céline, Michel, Didier, Lane-Guermonprez, Lydie, Delgrossi, Marie-Hélène, Médina, Emmanuelle, Arsanto, Jean-Pierre, & Bivic, André Le. 2004. CRB3 Binds Directly to Par6 and Regulates the Morphogenesis of the Tight Junctions in Mammalian Epithelial Cells. *Molecular Biology of the Cell*, **15**(3), 1324–1333.

BIBLIOGRAPHY

- Lerwill, MF. 2008. Flat epithelial atypia of the breast. *Archives of Pathology & Laboratory Medicine*, **132**(4), 615–621.
- Loie, Elise, Charrier, Lucie E., Sollier, Kévin, Masson, Jean-Yves, & Laprise, Patrick. 2015. CRB3A Controls the Morphology and Cohesion of Cancer Cells through Ehm2/p114RhoGEF-Dependent Signaling. *Molecular and Cellular Biology*, **35**(19), 3423–3435.
- Madabhushi, Anant. 2009. Digital pathology image analysis: opportunities and challenges.
- Manjunath, B. S., & Ma, W. Y. 1996. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(8), 837–842.
- Mao, Xiaona, Li, Pingping, Wang, Yaochun, Liang, Zheyong, Liu, Jie, Li, Juan, Jiang, Yina, Bao, Gang, Li, Lei, Zhu, Bofeng, & et al. 2017. CRB3 regulates contact inhibition by activating the Hippo pathway in mammary epithelial cells. *Cell Death & Disease*, **8**(1), e2546.
- Marćelja, S. 1980. Mathematical description of the responses of simple cortical cells*. *JOSA*, **70**(11), 1297–1300.
- Masood, Shahla, & Rosa, Marilin. 2011. Borderline Breast Lesions: Diagnostic Challenges and Clinical Implications. *Advances in Anatomic Pathology*, **18**(3), 190–198.
- McCaffrey, Luke Martin, & Macara, Ian G. 2009. The Par3/aPKC interaction is essential for end bud remodeling and progenitor differentiation during mammary gland morphogenesis. *Genes & Development*, **23**(12), 1450–1460.
- McCaffrey, Luke Martin, Montalbano, JoAnne, Mihai, Constantina, & Macara, Ian G. 2012. Loss of the Par3 Polarity Protein Promotes Breast Tumorigenesis and Metastasis. *Cancer Cell*, **22**(5), 601–614.
- Melendez, Jaime, Garcia, Miguel Angel, & Puig, Domenec. 2008. Efficient distance-based per-pixel texture classification with Gabor wavelet filters. *Pattern Analysis and Applications*, **11**(3–4), 365–372.
- Mellor, Harry, & Parker, Peter J. 1998. The extended protein kinase C superfamily. *Biochemical Journal*, **332**(2), 281–292.
- Moleirinho, S., Chang, N., Sims, A. H., Tilston-Lünel, A. M., Angus, L., Steele, A., Boswell, V., Barnett, S. C., Ormandy, C., Faratian, D., & et al. 2013. KIBRA exhibits MST-independent functional regulation of the Hippo signaling pathway in mammals. *Oncogene*, **32**(14), 1821–1830.

- Mommers, Ellen C.M., Page, David L., Dupont, William D., Schuyler, Peggy, Leonhart, Angelique M., Baak, Jan P.A., Meijer, Chris J.L.M., & Van Diest, Paul J. 2001. Prognostic value of morphometry in patients with normal breast tissue or usual ductal hyperplasia of the breast. *International Journal of Cancer*, **95**(5), 282–285.
- Nagai-Tamai, Yoko, Mizuno, Keiko, Hirose, Tomonori, Suzuki, Atsushi, & Ohno, Shigeo. 2002. Regulated protein–protein interaction between aPKC and PAR-3 plays an essential role in the polarization of epithelial cells. *Genes to Cells*, **7**(11), 1161–1171.
- Narod, Steven A., Iqbal, Javaid, Giannakeas, Vasily, Sopik, Victoria, & Sun, Ping. 2015. Breast Cancer Mortality After a Diagnosis of Ductal Carcinoma In Situ. *JAMA Oncology*, **1**(7), 888–896.
- National Comprehensive Cancer Network. 2017 (June). *Breast Cancer Screening and Diagnosis*. Tech. rept. 1.2017.
- National Institutes of Health. 2010. Breast Cancer Basics and You: Introduction. *NIH MedlinePlus*, **5**(2), 17.
- Nothacker, Monika, Duda, Volker, Hahn, Markus, Warm, Mathias, Degenhardt, Friedrich, Madjar, Helmut, Weinbrenner, Susanne, & Albert, Ute-Susann. 2009. Early detection of breast cancer: benefits and risks of supplemental breast ultrasound in asymptomatic women with mammographically dense breast tissue. A systematic review. *BMC Cancer*, **9**(Sep), 335.
- O'Connell, Peter, Pekkel, Vladimir, Fuqua, Suzanne, Osborne, C. Kent, & Allred, D. Craig. 1994. Molecular genetic studies of early breast cancer evolution. *Breast Cancer Research and Treatment*, **32**(1), 5–12.
- Otair, Dr Mohammed. 2013. Approximate k-nearest neighbour based spatial clustering using k-d tree. *arXiv:1303.1951 [cs]*, Mar. arXiv: 1303.1951.
- Page, David L, Dupont, William D, Rogers, LW, & Rados, MS. 1959. Atypical hyperplastic lesions of the female breast. A long-term follow-up study. *Cancer*.
- Page, David L., Dupont, William D., Rogers, Lowell W., & Landenberger, Marie. 1982. Intraductal carcinoma of the breast: Follow-up after biopsy only. *Cancer*, **49**(4), 751–758.
- Pan, S. J., & Yang, Q. 2010. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, **22**(10), 1345–1359.
- Pareja, Fresia, Marchiò, Caterina, Geyer, Felipe C., Weigelt, Britta, & Reis-Filho, Jorge S. 2017. Breast Cancer Heterogeneity: Roles in Tumorigenesis and Therapeutic Implications. *Current Breast Cancer Reports*, **9**(1), 34–44.

BIBLIOGRAPHY

- Pedregosa, Fabian, Varoquaux, Gaël, Gramfort, Alexandre, Michel, Vincent, Thirion, Bertrand, Grisel, Olivier, Blondel, Mathieu, Prettenhofer, Peter, Weiss, Ron, Dubourg, Vincent, & et al. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, **12**(Oct), 2825–2830.
- Perou, Charles M., Sørlie, Therese, Eisen, Michael B., van de Rijn, Matt, Jeffrey, Stefanie S., Rees, Christian A., Pollack, Jonathan R., Ross, Douglas T., Johnsen, Hilde, Akslen, Lars A., & et al. 2000. Molecular portraits of human breast tumours. *Nature*, **406**(6797), 747–752.
- Pinder, Sarah E. 2017. *Proliferative Breast Lesions*. Springer International Publishing. Page 33–41.
- Plotly Technologies Inc. 2015. *Collaborative data science*.
- Prat, Aleix, Pineda, Estela, Adamo, Barbara, Galván, Patricia, Fernández, Aranzazu, Gaba, Lydia, Díez, Marc, Viladot, Margarita, Arance, Ana, & Muñoz, Montserrat. 2015. Clinical implications of the intrinsic molecular subtypes of breast cancer. *The Breast*, **24**(Nov), S26–S35.
- Putti, T C, Pinder, S E, Elston, C W, Lee, A H S, & Ellis, I O. 2005. Breast pathology practice: most common problems in a consultation service. *Histopathology*, **47**(5), 445–457.
- Rangayyan, Rangaraj M., Ayres, Fábio J., & Leo Desautels, J. E. 2007. A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs. *Journal of the Franklin Institute*, **344**(3), 312–348.
- Ribeiro, Marco Tulio, Singh, Sameer, & Guestrin, Carlos. 2016. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. *arXiv:1602.04938 [cs, stat]*, Feb. arXiv: 1602.04938.
- Robinson, Brian S., Huang, Juang, Hong, Yang, & Moberg, Kenneth H. 2010. Crumbs Regulates Salvador/Warts/Hippo Signaling in Drosophila via the FERM-Domain Protein Expanded. *Current Biology*, **20**(7), 582–590.
- Roh, Michael H., Makarova, Olga, Liu, Chia-Jen, Shin, KunYoo, Lee, Seonok, Laurinec, Stephanie, Goyal, Meera, Wiggins, Roger, & Margolis, Ben. 2002. The Maguk protein, Pals1, functions as an adapter, linking mammalian homologues of Crumbs and Discs Lost. *The Journal of Cell Biology*, **157**(1), 161–172.
- Rouzier, Roman, Perou, Charles M., Symmans, W. Fraser, Ibrahim, Nuhad, Cristofanilli, Massimo, Anderson, Keith, Hess, Kenneth R., Stec, James, Ayers, Mark, Wagner, Pe-

- ter, & et al. 2005. Breast Cancer Molecular Subtypes Respond Differently to Preoperative Chemotherapy. *Clinical Cancer Research*, **11**(16), 5678–5685.
- Roy, Somak, Pantanowitz, Liron, Amin, Milon, Seethala, Raja R., Ishtiaque, Ahmed, Yousem, Samuel A., Parwani, Anil V., Cucoranu, Ioan, & Hartman, Douglas J. 2014. Smartphone adapters for digital photomicrography. *Journal of Pathology Informatics*, **5**(Jul).
- Russakovska, Olga, Deng, Jia, Su, Hao, Krause, Jonathan, Satheesh, Sanjeev, Ma, Sean, Huang, Zhiheng, Karpathy, Andrej, Khosla, Aditya, Bernstein, Michael, & et al. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, **115**(3), 211–252.
- Samuel, A. L. 1959. Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, **3**(3), 210–229.
- Selvaraju, Ramprasaath R., Das, Abhishek, Vedantam, Ramakrishna, Cogswell, Michael, Parikh, Devi, & Batra, Dhruv. 2016. Grad-CAM: Why did you say that? *arXiv:1611.07450 [cs, stat]*, Nov. arXiv: 1611.07450.
- Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. 2016. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, **35**(5), 1285–1298.
- Shin, Kunyoo, Straight, Sam, & Margolis, Ben. 2005. PATJ regulates tight junction formation and polarity in mammalian epithelial cells. *The Journal of Cell Biology*, **168**(5), 705–711.
- Simonyan, K., & Zisserman, A. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR*, **abs/1409.1556**.
- Simpson, Jean F. 2009. Update on atypical epithelial hyperplasia and ductal carcinoma in situ. *Pathology*, **41**(1), 36–39.
- Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. 2016. A Dataset for Breast Cancer Histopathological Image Classification. *IEEE Transactions on Biomedical Engineering*, **63**(7), 1455–1462.
- Sunil R. Lakhani, International Agency for Research on Cancer, Weltgesundheitsorganisation (ed). 2012. *WHO classification of tumours of the breast: views of a working group that convened for a consensus and editorial meeting at the International Agency for Research on Cancer (IARC), Lyon, September 1 - 3, 2011*. 4. ed edn. World Health Organization Classification of tumours. Internat. Agency for Research on Cancer.

BIBLIOGRAPHY

- Szymaniak, Aleksander D., Mahoney, John E., Cardoso, Wellington V., & Varelas, Xaralabos. 2015. Crumbs3-Mediated Polarity Directs Airway Epithelial Cell Fate through the Hippo Pathway Effector Yap. *Developmental Cell*, **34**(3), 283–296.
- Tamimi, Rulla M, Baer, Heather J, Marotti, Jonathan, Galan, Mark, Galaburda, Laurie, Fu, Yineng, Deitz, Anne C, Connolly, James L, Schnitt, Stuart J, Colditz, Graham A, & et al. 2008. Comparison of molecular phenotypes of ductal carcinoma in situ and invasive breast cancer. *Breast Cancer Research*, **10**(4), R67.
- Varelas, Xaralabos, Samavarchi-Tehrani, Payman, Narimatsu, Masahiro, Weiss, Alexander, Cockburn, Katie, Larsen, Brett G., Rossant, Janet, & Wrana, Jeffrey L. 2010. The Crumbs Complex Couples Cell Density Sensing to Hippo-Dependent Control of the TGF-Beta-SMAD Pathway. *Developmental Cell*, **19**(6), 831–844.
- Wang, Dayong, Khosla, Aditya, Gargeya, Rishab, Irshad, Humayun, & Beck, Andrew H. 2016. Deep Learning for Identifying Metastatic Breast Cancer. *arXiv:1606.05718 [cs, q-bio]*, Jun. arXiv: 1606.05718.
- Wang, Yifan, Shi, Jian, Chai, Kequn, Ying, Xuhua, & Zhou, Binhu P. 2013. The Role of Snail in EMT and Tumorigenesis. *Current cancer drug targets*, **13**(9), 963–972.
- Waskom, Michael, Botvinnik, Olga, Hobson, Paul, Cole, John B., Halchenko, Yaroslav, Hoyer, Stephan, Miles, Alistair, Augspurger, Tom, Yarkoni, Tal, Megies, Tobias, & et al. 2014. seaborn: v0.5.0 (November 2014). Nov.
- Weber, Allan G. 1997. The USC-SIPI image database version 5. *USC-SIPI Report*, **315**, 1–24.
- Wei, Benzheng, Han, Zhongyi, He, Xueying, & Yin, Yilong. 2017. Deep learning model based breast cancer histopathological image classification. IEEE. Page 348–353.
- Weigelt, Britta, Geyer, Felipe C., & Reis-Filho, Jorge S. 2010. Histological types of breast cancer: How special are they? *Molecular Oncology*, **4**(3), 192–208.
- Whiteman, Eileen L., Fan, Shuling, Harder, Jennifer L., Walton, Katherine D., Liu, Chia-Jen, Soofi, Abdul, Fogg, Vanessa C., Hershenzon, Marc B., Dressler, Gregory R., Deutsch, Gail H., & et al. 2014. Crumbs3 Is Essential for Proper Epithelial Development and Viability. *Molecular and Cellular Biology*, **34**(1), 43–56.
- Wu, Hao, Feng, Wei, Chen, Jia, Chan, Ling-Nga, Huang, Siyi, & Zhang, Mingjie. 2007. PDZ Domains of Par-3 as Potential Phosphoinositide Signaling Integrators. *Molecular Cell*, **28**(5), 886–898.

- Wu, Ren, Yan, Shengen, Shan, Yi, Dang, Qingqing, & Sun, Gang. 2015. Deep Image: Scaling up Image Recognition. *arXiv:1501.02876 [cs]*, Jan. arXiv: 1501.02876.
- Xue, Bin, Krishnamurthy, Kannan, Allred, D. Craig, & Muthuswamy, Senthil K. 2013. Loss of Par3 promotes breast cancer metastasis by compromising cell–cell cohesion. *Nature Cell Biology*, **15**(2), 189–200.
- Yosinski, Jason, Clune, Jeff, Bengio, Yoshua, & Lipson, Hod. 2014. *How transferable are features in deep neural networks?* Curran Associates, Inc. Page 3320–3328.
- Yu, Cao Guo, Tonikian, Raffi, Felsensteiner, Corinna, Jhingree, Jacquelyn R., Desveaux, Darrell, Sidhu, Sachdev S., & Harris, Tony J. C. 2014. Peptide Binding Properties of the Three PDZ Domains of Bazooka (Drosophila Par-3). *PLOS ONE*, **9**(1), e86412.
- Zeiler, Matthew D., & Fergus, Rob. 2013. Visualizing and Understanding Convolutional Networks. *CoRR*, **abs/1311.2901**.

Appendix A: BreakHis Class Codes and Abbreviations

Abbreviations for the class names present in the BreakHis dataset are used throughout this manuscript, and within the model itself. These class names refer to the breast lesion types outlined in the WHO guidelines for the classification of breast tumours (Sunil R. Lakhani, 2012).

Benign Tumours

B_AD Benign Adenosis

B_FA Benign Fibroadenoma

B_TA Benign Tubular Adenoma

B_PT Benign Phylodes Tumour

Malignant Tumours

M_DC Malignant Ductal Carcinoma

M_LC Malignant Lobular Carcinoma

M_MC Malignant Mucinous Carcinoma

M_PC Malignant Papillary Carcinoma

Notes

¹In the former group, the distinction between invasive and pre-invasive lesions is often unclear, while the dense tissue of the latter group can obscure and mask possible lesions