Deep Learning Approaches to Breast Histopathology Image Classification: A Focus on Dataset Optimisation

by Soyeb Golam, MSc

Submitted to The University of Nottingham
in September 2023
in partial fulfilment of the conditions for the award of the degree of
Master of Science in Computer Science with AI

I declare that this dissertation is all my own work, except as indicated in the text

Content

Abstract

This research delves into the intricate domain of breast histopathology image classification, harnessing the capabilities of deep learning models, with a particular emphasis on the VGG16 model. The study's primary objective was to address the prevalent challenge in the medical imaging field: achieving robust model performance with limited datasets. A comprehensive review of existing literature showed that while deep learning techniques have shown promise in medical imaging, there remains a significant gap in optimising models for smaller datasets, a common scenario in real-world clinical settings.

To address this, the research employed innovative methodologies, including introducing hybrid models such as CNN-SVM and CNN-RF. These models showcased remarkable data efficiency, maintaining consistent performance even when trained on a dataset halved in size.

However, the research also presented some challenges encountered during the evaluations, such as overfitting in VGG16 models and possible solutions to tackle them. Despite these challenges, the research stands as a testament to the potential of AI in revolutionising medical diagnostics. The models developed paralleled the performance of the renowned VGG16 and even surpassed it in certain scenarios. This achievement underscores the research's success and significant contribution to the scientific discourse in breast histopathology image classification.

This research, thus, sheds light on the challenges and opportunities in the field of AI in healthcare industry and sets a robust foundation for future experiments. By building on this groundwork, there is massive potential for further advancements, paving the way for cutting-edge Artificial Intelligence methodologies.

Keywords: "Breast histopathology", "Image classification", "Deep learning", "VGG16", "CNN-SVM", "CNN-RF", "Overfitting", "Kaggle", "Image segmentation", "Medical diagnostics", "Model training", "Diagnostic clarity".

Introduction

1.1 Background

Breast cancer, a predominant health concern, majorly impacts women across the globe. This malignancy arises from abnormal cellular behaviour in the mammary glands, composed of lobules responsible for milk production and ducts transporting milk. Genetic disruptions within these structures can lead to uncontrolled cellular proliferation, potentially resulting in the formation of tumours (Kadota et al., 2010). Depending on their characteristics, these tumours can either be benign, presenting minimal harm, or malignant, which can aggressively spread to surrounding tissues, posing significant health threats (Hollern et al., 2014).

Invasive Ductal Carcinoma (IDC), a primary breast cancer subtype, emerges from the ductal cells. It is known for its invasive nature, enabling it to spread beyond the ducts and permeate nearby tissues. The diagnostic process for IDC is exhaustive. While initial detection might be during routine screenings, advanced imaging techniques, such as mammograms and ultrasounds, provide a more comprehensive view of the breast tissue. However, the definitive diagnosis of IDC relies on biopsy results, where the extracted tissue samples are subjected to detailed histopathological examination. This thorough analysis not only verifies the presence of IDC but also offers insights into its grade and aggressiveness, guiding the subsequent treatment plan (Shen et al., 2017).

Moreover, the advancement of digital pathology has considerably improved the accuracy of histopathological assessments. Cutting-edge imaging technologies and computational algorithms provide a more in-depth and objective evaluation of tissue samples, reducing errors and subjectivity induced by human judgment (Kadota et al., 2010).

1.2 The promise of deep learning

In the past ten years, the field of medical imaging has increasingly adopted deep learning methods, especially convolutional neural networks (CNNs). These sophisticated models have demonstrated remarkable proficiency in identifying and categorising anomalies in images, often matching or surpassing the capabilities of human experts. In the context of breast histopathology image classification, deep learning holds the potential to revolutionise diagnostic methodologies, improving their accuracy and efficiency.

For instance, Huynh et al. (2016) conducted a study that explored the advantages of incorporating transfer learning from deep convolutional neural networks for classifying digital mammographic tumour images. Their results revealed that when CNNs are combined with transfer learning, they can compete with traditional methods relying on manually extracted tumour features, highlighting the effectiveness of deep learning in enhancing diagnostic precision (Huynh et al., 2016).

In a separate study, Bejnordi et al. (2017) introduced an innovative approach utilising context-aware stacked convolutional neural networks for the classification of breast whole-slide images. By integrating cellular-level data with overall tissue structures, their method achieved notable accuracy in distinguishing between benign and malignant slides (Bejnordi et al., 2017).

Collectively, these studies underscore the groundbreaking potential of deep learning in reshaping the landscape of breast histopathology image classification. They lay the foundation for more refined and accurate diagnostic processes, promising to enhance medical imaging significantly.

1.3 Challenges and Motivation

The efficacy of deep learning models is deeply rooted in the calibre and volume of the training data they are exposed to. A comprehensive dataset, rich in diversity and representative of real-world scenarios, is pivotal for these models to generalise well and produce reliable outcomes. Conversely, datasets that are limited in scope or lack the breadth of real-world variations can inadvertently introduce biases, leading to suboptimal model performance. Such constraints can result in overfitting, where the model performs exceptionally well on the training data but falters when exposed to unseen data.

Moreover, the intricacies of medical imaging, especially in fields like breast histopathology, further amplify the challenges. The subtle variations between benign and malignant samples, the presence of rare conditions, and the inter-patient variability necessitate a robust and comprehensive dataset for training. In scarce data sets, techniques like data augmentation, transfer learning, and synthetic data generation become crucial. These methods aim to artificially enhance the dataset's size and diversity, ensuring the model is exposed to a broader range of scenarios during training.

This research endeavours to delve deep into these challenges, exploring the intricacies of training deep learning and traditional machine learning models on limited datasets. By understanding the pitfalls and potential solutions, the goal is to pave the way for more robust diagnostic tools, even in the face of data constraints.

1.4 Research Objectives

The primary objective of this research is to delve into the intricacies of breast histopathology image classification, especially when confronted with data constraints. By embarking on a comprehensive exploration of data preprocessing, model fine-tuning, and evaluation metrics, this study aims to:

1. Illuminate the best breast histopathology image classification practices, ensuring that the developed models are robust and reliable.

2. Investigate the challenges posed by limited datasets, which might not encapsulate the full spectrum of variability in real-world samples.

3. Explore strategies such as data augmentation, transfer learning, and synthetic data generation to counteract the limitations of smaller datasets.

4. Evaluate the performance of various deep learning architectures in breast histopathology image classification, identifying the most promising models in terms of accuracy and computational efficiency.

5. Contribute valuable insights to the burgeoning domain of medical image classification, emphasising the importance of data quality and quantity in the success of deep learning models.

By addressing these objectives, the research aspires to develop more robust diagnostic tools, ensuring that they remain effective even in the face of data constraints.

Literature review

This literature review delves into the multifaceted landscape of breast histopathology image classification, exploring the nuances of traditional machine learning models, the prowess of deep learning architectures, and the innovative synergy of hybrid approaches. From the challenges posed by limited datasets to the untapped potential of novel model combinations, this review will offer a comprehensive insight into the current state of the art, highlighting research gaps and illuminating avenues for future exploration.

1.5 Breast cancer overview

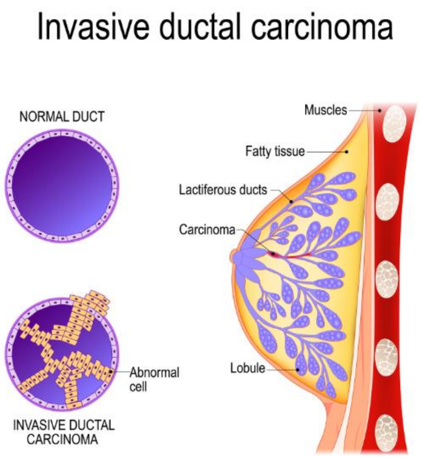


Figure 1-Breast cancer visualisation

Breast cancer, one of the predominant malignancies in women globally, is marked by anomalous cellular activity within the mammary glands, resulting in tumour formation. The mammary glands are an intricate network of lobules (which produce milk) and ducts (tubular structures facilitating milk's transportation towards the nipple). Genetic aberrations in the ductal cells can instigate unrestrained cell division, leading to ductal hyperproliferation (Perou et al., 2000). Over time, these excessive cell aggregations can evolve into tumours, which might be benign (non-hazardous) or malignant (hazardous).

When malignancy is involved, these tumours can infiltrate neighbouring tissues and eventually disseminate to other remote body parts, known as metastasis. Such metastatic episodes can jeopardise vital organ functions and drastically influence the prognosis.

Invasive Ductal Carcinoma (IDC) is notably the most widespread within the broader spectrum of breast cancer types. Originating from the ductal linings, IDC is characterised by its invasive nature, allowing it to breach the confines of the ducts and potentially spread to adjacent tissues. If left unchecked, IDC might progress to metastatic stages, affecting lymph nodes and distant organs.

The diagnostic trajectory for IDC is multi-pronged. Initial detection often occurs during routine physical examinations when anomalies or masses in the breast are identified. After this, imaging modalities such as mammography come into play, offering a radiographic perspective of the breast tissue. In scenarios demanding a more granular view, ultrasounds can furnish in-depth imaging of the breast's architecture. Nonetheless, the gold standard for IDC confirmation remains the biopsy, wherein a tissue specimen is procured and subjected to microscopic histopathological scrutiny. This examination confirms IDC's presence and offers insights into its grade and aggressiveness, thereby guiding therapeutic interventions.

Histopathology, particularly within breast cancer diagnostics, entails meticulously evaluating breast cell tissue samples obtained via biopsy. This examination is executed under a light microscope by a proficient pathologist to discern between benign and malignant cell growth and classify the specific subtype of breast cancer. While this method has been a mainstay in diagnostics, it bears inherent challenges. While incredibly informed, the manual analysis overseen by a pathologist remains intrinsically subjective, potentially leading to discrepancies in diagnosis. Moreover, this traditional approach is labour-intensive and may incur considerable time and monetary costs. Nevertheless, with the technological surge of digital image scanning systems, particularly the advent of cost-effective whole slide image scanners, the landscape of histopathology is undergoing a digital transformation (Pantanowitz et al., 2011). This digital shift is ushering in the era of automated image analysis, offering potential solutions to the limitations of conventional manual examination methods.

1.6 Neural Network overview in the context of deep learning

Deep learning, a specialised branch of machine learning, has garnered significant attention due to its utilisation of neural networks. Through their intricate architecture, these networks are adept at extracting intricate patterns from raw data across multiple layers. This capability allows them to make informed predictions on previously unseen data. The versatility of deep learning is evident in its wide range of applications, from image classification to natural language processing (Sumathi et al., 2022).

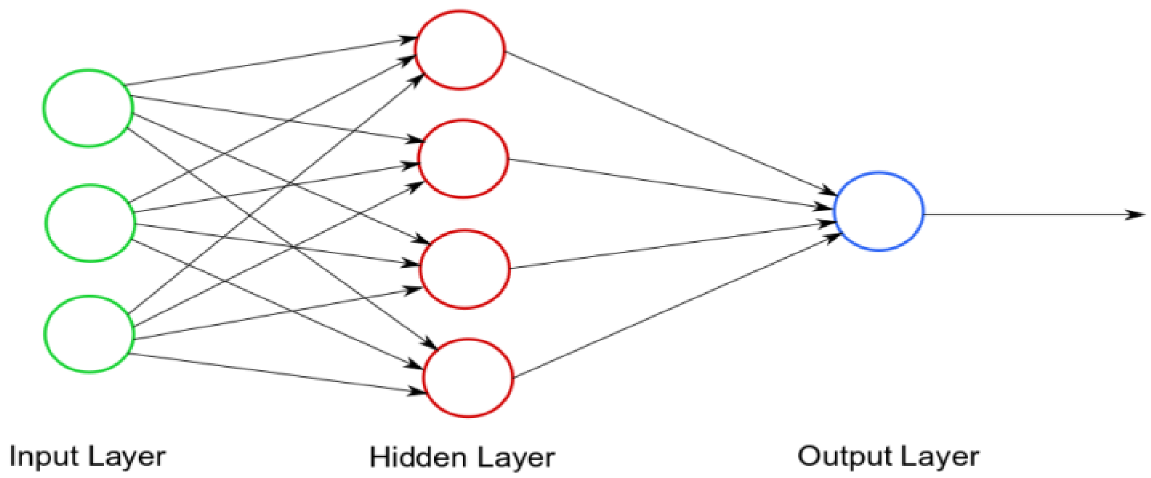


Figure 2-A simple neural network overview

At the heart of deep learning lies the neural network, a sophisticated structure comprising interconnected nodes or neurons. These are systematically organised into three primary layers: input, hidden, and output. The input layer is the initial recipient of data, the hidden layers process this data, and the output layer delivers a prediction based on the processed information. Each connection between neurons is assigned a weight, symbolising its significance in the data processing chain. The intricate process of training a neural network revolves around fine-tuning these weights to minimise prediction discrepancies (Qi et al., 2022).

The training regimen for neural networks is multifaceted. Initially, the input layer is fed data, which undergoes processing by being multiplied with the connection weights, resulting in a summation. A bias value is added to this sum, establishing an activation threshold. This combined value is subsequently passed through an activation function, such as the sigmoid function. This function's role is pivotal as it scales the input value between 0 and 1, determining the neuron's activation level. This output then serves as the input for the subsequent layer. This iterative process continues until the data reaches the output layer. Here, the neuron with the highest activation level dictates the final prediction. The accuracy of this prediction is then gauged against the actual desired output, with discrepancies calculated using a loss function. The backpropagation optimisation algorithm then adjusts the weights to rectify these errors (Nurcahyati et al., 2022).

Backpropagation is a cornerstone of neural network training. It meticulously evaluates the contribution of each connection to the overall error by retracing its steps through the layers. This involves computing the gradient of the loss function with respect to each weight, followed by weight adjustments to minimise the error. This adjustment process leverages the gradient descent algorithm, a pivotal component in neural network training (Arora et al., 2022).

Gradient descent is a methodical approach to refining the weights of a neural network to curtail errors. Its primary objective is to minimise the error by making incremental adjustments to the weights, moving in the direction opposite to the gradient of the loss function. This ensures the network continually improves its predictions, honing its accuracy over time.

In the vast landscape of deep learning, two deep neural network types have emerged as frontrunners: convolutional neural networks (CNNs) and recurrent neural networks (RNNs). CNNs have carved a niche for themselves in computer vision tasks. Their design and functionality make them exceptionally effective in tasks like image classification, recognition, and even intricate processes like image segmentation. Their prowess in processing visual data is unparalleled, making them the go-to choice for many computer vision applications. Conversely, RNNs shine when dealing with sequential data formats like text or audio. Their architecture allows them to remember previous inputs, making them ideal for tasks that require understanding context over sequences, like language translation or speech recognition.

The dawn of deep learning models has substantially transformed the landscape of the machine learning sector, as their application range can be deemed boundless.

1.7 Convolutional neural networks

Deep learning offers a range of architectures designed for distinct tasks. The Convolutional Neural Network (CNN) is recognised for its expertise in analysing spatial data, including images and signals. While classic neural networks are potent, they might struggle with the complexities of image data due to their vast dimensions. CNNs, with their unique structure, address this challenge head-on (Zaabar, Niculescu, & Mustapha, n.d.).

A CNN comprises a series of layers: the convolution layer, the max pooling layer, and a fully connected layer. As an extension of Artificial Neural Networks (ANNs), CNNs are adept at recognising and extracting patterns from visual data, ensuring these patterns are generalisable across various scenarios (Zafar et al., 2022).

To better understand the workings of a CNN, let us consider the task of classifying an image of a koala. The raw image is a matrix of pixel values, each representing colour intensities. The CNN's objective is to discern if the patterns within this matrix match the typical features of a koala.

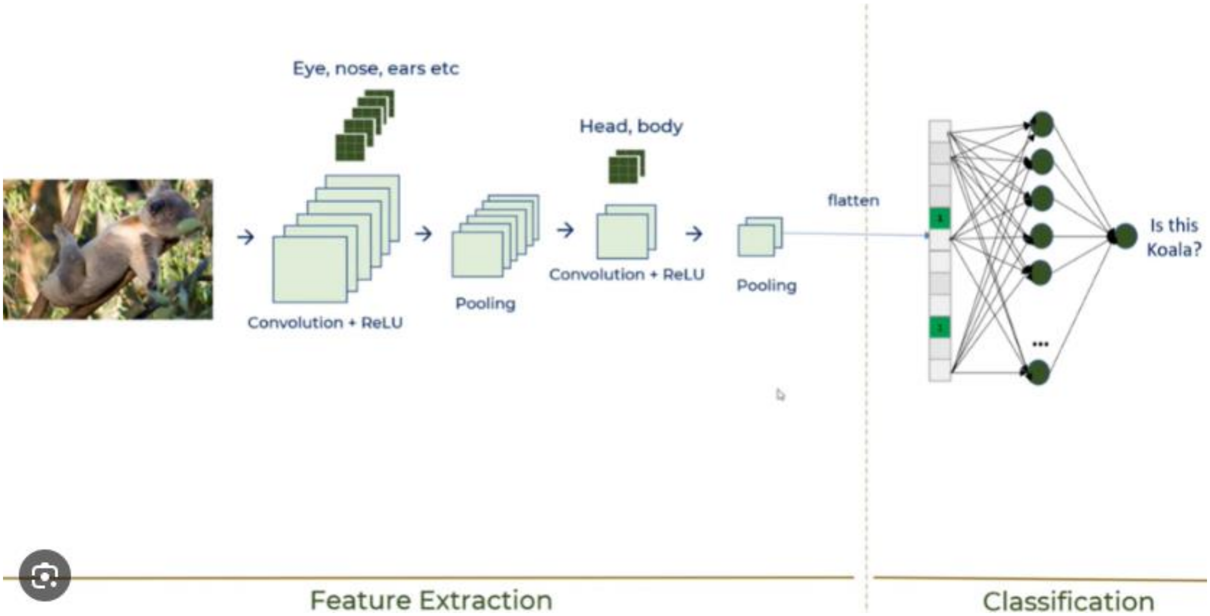


Figure 3-CNN of a Koala

The image's journey within a CNN commences at the convolution layer. This layer takes the pixel values as input and, through a series of convolutional operations combined with activation functions, extracts foundational features like edges, textures, and shapes. These features are the building blocks that the network will use to recognise more complex patterns in subsequent layers. The convolutional operations are particularly designed to handle 3D tensors, which, in the context of our RGB koala image, represent the height, width, and depth (colour channels: red, blue, and green) (Rashmi & Singh, 2023).

As the convolutional operation progresses, it extracts patches from the input tensor, transforming them into a feature map. This map captures specific attributes of the input, such as the unique texture of the koala's fur or the distinct shape of its ears.

Post convolution, the data undergoes a downsampling process in the max pooling layer. This layer's primary function is to reduce the spatial dimensions of the data, ensuring computational efficiency and preventing overfitting. For our koala image, this translates to retaining only the most salient features while discarding redundant or less crucial information (Ma et al., 2022).

The fully connected layer, often the final layer in a CNN, takes the processed data from previous layers and transforms it into an output corresponding to our classification task. In our example, if the processed features strongly align with those of a koala, the network will classify the image as such.

Beyond image classification, CNNs have diverse applications. In sound recognition, their architecture is leveraged to classify voice sound signals. In digital media, CNNs play a pivotal role in tasks like fake news detection, where they discern genuine articles from fabricated ones based on content patterns.

CNNs have significantly advanced the field of deep learning. Their specialised architecture, explicitly designed for spatial data, has revolutionised machine perception and processing of visual information.

1.8 VGG16: A Deep Dive into the Pioneering Pre-trained Model in Deep Learning

In the expansive domain of deep learning, especially within image processing, pre-trained models have emerged as cornerstones. These models, pre-trained on extensive datasets, provide a robust foundation for a myriad of tasks, negating the necessity of building models from the ground up. Among these, VGG16, a brainchild of the Visual Geometry Group at Oxford, has distinguished itself as a benchmark in convolutional neural networks (CNNs) (Simonyan & Zisserman, 2014).

VGG16, a deep convolutional neural network, was crafted for the primary purpose of image recognition and classification. Its nomenclature, "VGG16", stems from its design, encompassing 16 weight layers. This model was a notable participant in the ImageNet Large Scale Visual Recognition Challenge in 2014, a revered competition in computer vision. Despite not securing

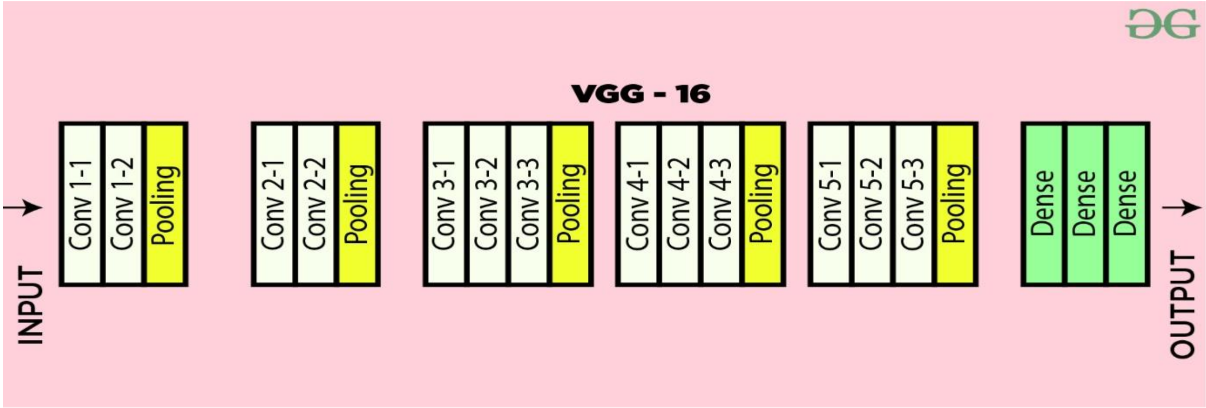


Figure 4-VGG16 architecture

the first position, its performance, combined with its architectural simplicity, earned it widespread recognition and adoption in the research community.

Delving into its architecture, VGG16 is characterised by its depth and straightforwardness. It is composed of 13 convolutional layers, punctuated by max-pooling layers, succeeded by three fully connected layers, culminating in a softmax layer for classification. A hallmark of VGG16 is its consistent employment of 3x3 convolutional filters throughout its structure. These small filters, when layered consecutively, empower the model to discern intricate hierarchical patterns in data, maintaining a balance with computational demands.

VGG16's depth, surpassing earlier architectures like AlexNet, enables it to discern more nuanced and abstract features from input images, enhancing its classification prowess. The true mettle of VGG16 is evident when harnessing its pre-trained capabilities. Trained on the extensive ImageNet dataset, comprising over a million images across 1000 categories, VGG16 has internalised a vast array of features. These span from rudimentary edges and textures to complex patterns. This pre-training makes VGG16 a prime candidate for many image-centric tasks. By leveraging the pre-trained weights of VGG16 and fine-tuning it for specific tasks, developers can expedite model development, a process termed transfer learning. This approach often yields superior results, especially when datasets for new tasks are sparse.

Beyond mere image classification, VGG16's influence permeates various applications. In object detection, image regions are processed through VGG16 to extract features, which subsequently identify and pinpoint objects. In style transfer, VGG16's layers capture content and style features, enabling the creation of artistically stylised images.

The success of VGG16 has also catalysed the development of deeper architectures, such as VGG19, with 19 weight layers. However, they demand more computational resources even though they offer enhanced representational capabilities.

1.9 Support Vector Machine

Transitioning from the intricate layers of Convolutional Neural Networks (CNNs), it is essential to delve into traditional machine learning models that have also been instrumental in image classification tasks. Among these, the Support Vector Machine (SVM) stands out as a powerful tool, especially in medical imaging.

Support Vector Machines (SVMs) are supervised learning models renowned for their capability in classification and regression tasks. In image classification, SVMs have showcased their prowess in discerning between different classes based on extracted features, particularly for medical images. While deep learning models like CNNs can automatically learn features from data, SVMs often operate on handcrafted or engineered features, making them particularly suitable for tasks where domain knowledge is crucial.

To illustrate the application of SVMs in medical imaging, consider the challenge of breast cancer detection using mammograms. Mammograms are X-ray images of the breast, and they play a pivotal role in early breast cancer detection. However, interpreting these images requires expertise; even then, there is room for human error. This is where SVMs come into play.

In a 2022 research conducted by Alyami and colleagues, they introduced a framework for breast tumour image classification leveraging cloud computing. This approach integrated the capabilities of the AlexNet model with the texture feature extraction power of the GLCM (grey-level cooccurrence matrix). Subsequently, these extracted features underwent classification through an ensemble approach using Multi-Kernel Support Vector Machine (MK-SVM). When evaluated on the renowned MIAS breast image dataset, their model demonstrated a notable accuracy rate of 96.26% (Alyami et al., 2022).

In another study focusing on mammographic image processing, Safdarian and Dezfoulinejad (2020) employed SVMs to classify breast cancer masses. They optimised SVM parameters using the Grasshopper Optimisation Algorithm (GOA) and achieved an accuracy of 100% for the classification of benign and malignant masses, underscoring the potential of SVMs in this domain (Safdarian & Dezfoulinejad, 2020).

Let us walk through a hypothetical example to understand the process better. Imagine a dataset of mammograms, where each image is labelled either as 'benign' or 'malignant.' The first step would involve feature extraction, where specific characteristics of the images, such as texture, shape, and edges, are identified. These features act as the input for the SVM. The SVM then constructs a hyperplane in a multi-dimensional space that distinctly classifies the benign and malignant images. When a new mammogram is introduced to this trained SVM, it uses the same feature extraction process. Then, it determines which side of the hyperplane this new data point lies, thus classifying it as benign or malignant.

While deep learning models like CNNs have gained significant attention in recent years, traditional machine learning models like SVMs continue to be invaluable, especially in areas requiring domain-specific knowledge. Their adaptability, combined with domain expertise, makes them a robust tool in the realm of medical image classification.

1.10 Random Forest

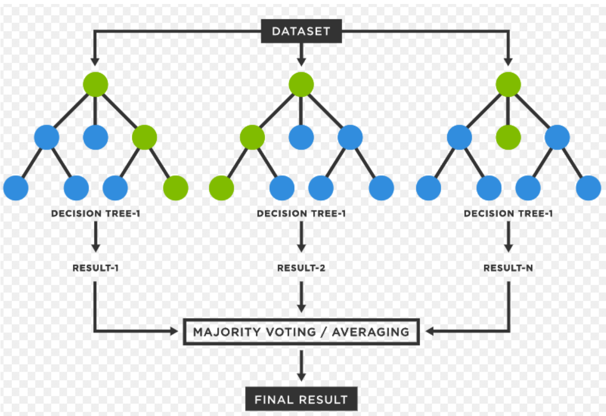


Figure 5-Random Forest architecture

Random Forests have shown notable efficacy in applications like remote sensing, commonly used in aerial devices like satellites. Millard and Richardson, in their 2015 research, explored the classification of peatlands using data derivatives from LiDAR. Their findings underscored the pivotal role of input data attributes in influencing RF classification outcomes, drawing attention to metrics like the RF out-of-bag error and standalone classification precision. The investigation shed light on the criticality of judiciously choosing training data and specific input parameters, such as imaging channels. They pointed out the necessity to refine high-dimensional datasets, ensuring the inclusion of only vital, non-redundant variables for classification tasks. Additionally, the study highlighted the impact of training dataset size on outcomes, advocating for expansive and well-distributed training sets that mirror the true class distributions in the environment (Millard & Richardson, 2015).

To illustrate the application of Random Forests in medical imaging, consider the challenge of breast cancer detection using histopathological images. Histopathological images provide microscopic views of tissue structures, aiding in diagnosing diseases like cancer. Suppose a dataset consists of these images, each labelled as 'benign' or 'malignant'. The first step in the RF-based classification process would involve feature extraction, where specific characteristics of the images, such as texture, shape, and colour distributions, are identified. These features serve as the input for the Random Forest. The RF then uses its ensemble of decision trees to vote and determine the most probable class for each image. For instance, if an image gets 70 votes for 'malignant' and 30 for 'benign' out of 100 trees, it is classified as 'malignant'.

In another comprehensive review by Sheykhmousa et al. (2020), the performances of RF and SVM in remote sensing image classification were juxtaposed (MEANING?). The meta-analysis of 251 peer-reviewed journal papers revealed that RF, when combined with certain feature extraction methods, could achieve superior classification accuracy in specific thematic applications, especially when compared to other machine-learning techniques, like SVM (Sheykhmousa et al., 2020).

Furthermore, Salas and Subburayalu (2019) highlighted an optimised shape index for agricultural management system analysis using hyperspectral data. Their approach utilised the contiguous bands of hyperspectral data to define the gradient of the spectral curve, enhancing image classification accuracy. The study achieved an impressive overall accuracy, showcasing the potential of optimised shape indices under an object-based random forest approach (Salas & Subburayalu, 2019).

1.11 Ensemble learning

Ensemble learning has gained traction in diverse image classification endeavors, capitalizing on the unique advantages of individual models to deliver enhanced results. The essence of ensemble techniques lies in amalgamating the predictions from multiple base models, constructed using a specific learning approach, to bolster accuracy and resilience.

In a recent study by Müller and colleagues in 2022, the efficacy of various ensemble learning approaches in the realm of medical image classification was scrutinized. The team delved into the performance dynamics of techniques like Augmentation, Stacking, and Bagging. Their insights highlighted that among the methods evaluated, Stacking emerged as the most impactful, registering an impressive surge in the F1-score by up to 13%. This study accentuates the transformative potential of ensemble strategies in elevating the precision and reliability of medical image classification systems (Müller, Soto-Rey, & Kramer, 2022).

Ensemble learning has also shown promise in the realm of agricultural and remote sensing. A study by Aboneh et al. (2022) proposed a robust 6G-enabled Internet of Things (IoT) framework for medical image classification using an ensemble learning-based model. They combined MobileNet and DenseNet architectures for feature extraction and employed a modified honey-badger algorithm for feature selection. Their technique achieved an accuracy of 87.10% on the chest X-ray dataset, highlighting the efficacy of ensemble learning in image classification tasks (Abd Elaziz et al., 2022).

To illustrate the concept of ensemble learning, consider a scenario where we aim to classify images of wheat to detect rust, a common disease that impacts wheat growth. While individual models might be trained to recognise specific features of rust, an ensemble method would combine the predictions of these models to make a final decision. For instance, Pan et al. (2022) proposed a method for identifying wheat rust based on ensemble learning. They integrated multiple CNN models and found that their ensemble method significantly improved the identification accuracy of wheat rust disease, emphasizing the potential of ensemble learning in agricultural image classification (Pan et al., 2022).

In essence, ensemble learning acts as a bridge, harmonizing the strengths of individual models like CNNs, SVMs, and Random Forests. By strategically combining these models, ensemble methods can achieve higher accuracy and robustness, making them invaluable in various image classification tasks.

1.12 Existing research

This section of our research paper explores select studies, offering insights into their summaries, key findings, limitations, and potential areas for further investigation.

1.12.1 Deep Learning-based Mammogram Classification for Breast Cancer

In recent years, the application of Convolutional Neural Networks (CNN) in medical imaging has garnered significant attention, and Altan's 2020 study stands as a testament to this trend. Altan embarked on a meticulous exploration of CNN's capabilities, specifically targeting the classification of mammograms to discern between cancerous and normal cases. The research highlighted the inherent robustness of Deep Learning algorithms and emphasized their unparalleled adaptability, especially when confronted with the intricacies of vast datasets. The results of the study were nothing short of impressive. The CNN model, tailored for this specific task, showcased an accuracy of 92.84%, a sensitivity of 95.30%, and a specificity of 96.72%. While numerically compelling, these metrics underscore the model's potential in real-world clinical settings, where precision and reliability are paramount.

However, every research endeavour has its boundaries. While Altan's study illuminated the formidable strengths of CNNs in mammogram classification, it remained somewhat myopic in its approach, not venturing into the realm of other deep learning architectures or juxtaposing CNN's performance against traditional machine learning models. While providing depth in one direction, this focus inadvertently highlighted a limitation: the unexplored terrain of hybrid models. The study did not touch upon the potential synergy between the depth of CNNs and the unique strengths of traditional classifiers, such as SVM or RF. This omission indicates a limitation in the current research and points to a burgeoning research gap. The fusion of CNNs with traditional classifiers, such as in the CNN-SVM or CNN-RF models, remains a tantalizing prospect, awaiting further exploration and validation in the context of mammogram classification (Altan, G., 2020).

1.12.2 Novel breast cancer classification framework based on deep learning

In the rapidly evolving domain of breast cancer classification, integrating deep learning models with traditional machine learning techniques has emerged as a promising frontier. Salama, Elbagoury, and Aly's 2020 research is a testament to this interdisciplinary synergy. Their study introduced a groundbreaking technique that harmoniously combined the strengths of ResNet50 and VGG-16 for breast cancer classification. By harnessing the power of transfer learning, the researchers could tap into pre-existing knowledge from vast datasets, thereby reducing the need for extensive training on new data. Additionally, incorporating data augmentation techniques ensured a diverse and robust training set, further enhancing the model's generalisation capabilities. To top it off, the research introduced a Support Vector Machine (SVM) classifier into the mix, resulting in a model that boasted high accuracy and operated with reduced computational demands. Among the various configurations tested, the hybrid model of ResNet50 combined with SVM emerged as the star performer, particularly when applied to the DDSM dataset. However, the research had its shortcomings. A noticeable gap was the lack of an in-depth comparison between the two deep learning architectures, ResNet50 and VGG-16, leaving readers curious about the nuanced differences in their performances. Furthermore, while the integration of SVM with deep learning models was indeed innovative, the study did not explore other potential hybrid models, such as the intriguing CNN-RF combination. Furthermore, the built-in CNN-SVM model, which seamlessly integrates convolutional layers with SVM classification, was not explored, representing a potential avenue for future research (Salama, Elbagoury, and Aly's 2020).

1.12.3 Augmenting Transfer Learning with Feature Extraction Techniques for Limited Breast Imaging Datasets

In a comprehensive study conducted by Aswiga R V, A. R, and Shanthi A P (2021), the pressing issue of limited breast imaging datasets was brought to the forefront. Their research, aptly titled "Augmenting Transfer Learning with Feature Extraction Techniques for Limited Breast Imaging Datasets", sheds light on the significant hurdles researchers face due to the dearth of comprehensive datasets. The paper underscored the transformative potential of transfer learning as a solution to this pervasive dataset limitation. Delving into the specifics, the researchers introduced a two-tiered framework tailored to classify digital breast tomosynthesis (DBT) datasets. Recognising the richness of general non-medical image datasets and mammography datasets, they ingeniously harnessed this pre-existing knowledge to bolster the classification capabilities for the less abundant DBT datasets. Their innovative methodology seamlessly integrated a basic multilevel transfer learning (MLTL) framework with a feature extraction-based transfer learning (FETL) framework, yielding encouraging outcomes. However, despite these advancements, the study continually emphasized a glaring limitation: the acute scarcity of publicly accessible DBT image datasets. This limitation hampers current research and signals an urgent call to action for the broader scientific community. The study strongly advocates for creating and disseminating more extensive and diverse breast imaging datasets, emphasizing that the future of breast cancer classification hinges on addressing this data insufficiency (Aswiga R V, A. R, & Shanthi A P, 2021).

1.13 Research gaps and novel ideas

In the dynamic landscape of medical imaging research, continuous advancements and innovations are paramount. However, as with any evolving field, certain gaps and unexplored territories persist, beckoning researchers to delve deeper and push boundaries. This section aims to shed light on two such pivotal areas that have remained relatively underexplored in the context of breast histopathology image classification. By addressing the challenges posed by limited datasets and introducing novel hybrid model approaches, this research endeavours to contribute meaningfully to the existing body of knowledge, setting new standards for future investigations.

1.13.1 Addressing the Gap in Robust Models for Limited Datasets

From the comprehensive review of the studies above, a recurring theme emerges: the challenge of achieving robust model performance with limited datasets. The realm of medical imaging, particularly in breast histopathology, is rife with the constraints of data scarcity. Altan's 2020 study and the research by Salama, Elbagoury, and Aly in the same year underscored the prowess of deep learning techniques, yet neither delved deeply into the nuances of model performance on smaller datasets. Aswiga R V, A. R, and Shanthi A P's 2021 research further accentuated this gap, highlighting the acute scarcity of publicly accessible digital breast tomosynthesis (DBT) image datasets. While transfer learning and data augmentation techniques have been proposed as potential solutions, there remains a significant gap in understanding how to truly optimise models when data is not abundant. In real-world clinical settings, researchers and practitioners might not always have access to vast datasets, making it imperative to develop models that can deliver reliable results even with limited data. This research aims to bridge this gap, offering a fresh perspective on the challenge of data scarcity in medical imaging.

1.13.2 The Uncharted Territory of Hybrid Models: CNN-RF and CNN-SVM

While the reviewed studies have showcased the potential of hybrid approaches in breast histopathology image classification, a glaring omission is evident: the lack of exploration into the CNN-RF and CNN-SVM models. Salama, Elbagoury, and Aly's 2020 research introduced the innovative integration of SVM with deep learning models, yet the seamless fusion of convolutional layers with SVM classification remained uncharted. Similarly, the potential amalgamation of CNNs with Random Forests (RF) has been conspicuously absent from the literature. These hybrid models, which combine the depth and feature extraction capabilities of CNNs with SVM and RF classification strengths, hold immense promise. They represent a novel frontier in the quest for optimal breast histopathology image classification, especially when datasets are constrained. By venturing into this unexplored domain, this research aspires to introduce these hybrid models and rigorously evaluate their performance, setting a new benchmark for innovation and excellence in the field. The overarching goal is to harness the synergies of these models, ensuring that even with limited data, the quest for diagnostic precision remains uncompromised.

Methodology

The methodology section acts as the foundation of this study, outlining the structured strategy employed to tackle the intricacies of breast histopathology image classification. Its main objective is to present a clear and repeatable roadmap of the research activities, ensuring a thorough documentation of every phase, from gathering data to assessing the model. The ultimate goal is to integrate the strengths of both deep learning and conventional machine learning methods into a unified system. This system aims to excel in classifying the "Breast Histopathology Images" dataset, underscoring the significance of data preparation, crafting model architecture, and comprehensive evaluation. Through this methodical approach, the study hopes to shed light on the efficacy of combined models in medical imagery and establish a standard for subsequent research in this field.

1.14 Data Collection

The primary dataset for this research was sourced from Kaggle's "Breast Histopathology Images" repository, a comprehensive collection of histopathological images focusing on Invasive Ductal Carcinoma (IDC). This dataset comprises approximately 277,524 images, each with a resolution of 50x50 pixels. The images undergo staining with Hematoxylin and Eosin (H&E), a conventional method in histopathology, which enhances the visibility of cellular details and variations. All the images present in the dataset are set either positive or negative depending on the detection of IDC.

Given the sensitive nature of medical data, it's crucial to address ethical considerations. The images in this dataset are anonymised, with no personally identifiable information attached, ensuring patient privacy. Furthermore, the dataset is publicly available on Kaggle for research purposes, implying that permissions for its use in academic and research contexts have been granted. Researchers and practitioners are encouraged to use this dataset to develop and evaluate machine learning models, promoting advancements in breast cancer histopathology image analysis (Mooney, 2018).

1.15 Data pre-processing

The data pre-processing phase is a critical precursor to any machine learning or deep learning project. It ensures that the data fed into models is high quality, consistent, and free from anomalies that could adversely affect the training process. This research focused on breast histopathology images, which were stored in compressed zip files on a cloud drive.

The initial step in the pre-processing journey involved extracting these images from their compressed format. This systematic extraction process was not just about accessing the data but also about organising it. As the images were unzipped, they were meticulously organised into designated folders, creating a structured repository that would facilitate smooth navigation in the subsequent stages of the research. To ensure data purity and avoid any potential redundancy or overlap, any pre-existing content in these folders was diligently purged before the extraction process.

Once the extraction was complete, a pivotal step was the classification of images based on their inherent labels: IDC-positive and IDC-negative. This binary classification is foundational for the modelling phase, ensuring that each image is correctly labelled facilitating effective supervised learning. The importance of this step cannot be overstated, especially in medical imaging, where misclassification can have significant implications.

The dataset underwent a strategic split following classification to create training, validation, and test sets. Using a high training and a smaller but similar validation and test split ratio ensured that each subset retained a representative distribution of the two classes. Such a stratified division is paramount to avoid biases during model training and evaluation, ensuring that the models are exposed to a balanced mix of both classes during their learning process.

The subsequent steps in the pre-processing phase were dedicated to image transformation and normalisation. Given the diverse nature of medical images, with variations in size, resolution, and pixel intensity, it was imperative to standardise the dataset. All images, irrespective of the model they were intended for, were resized to a consistent dimension of 50x50 pixels. This resizing ensures uniformity in the input size across the dataset, making it compatible with deep learning architectures like VGG16 and the convolutional layers of the CNN-SVM and CNN-RF models.

Normalisation was selected as a crucial pre-processing step for several strategic reasons. The primary objective was to scale the pixel values of each image to fall within the range [0, 1]. This decision was rooted in the understanding that providing models with data in a consistent format is pivotal for their optimal performance. Without normalisation, models struggle with variations in pixel intensity, potentially leading to erratic learning patterns or biases. Another compelling reason for adopting normalisation is its proven potential to expedite convergence during training. This acceleration is especially vital when working with deep learning models, which often involve extensive computational resources and time. By ensuring that disparities in pixel intensity across images do not adversely affect the training process, normalisation ensures more stable learning and contributes to efficiency.

In contrast, standardisation, which involves transforming data to have a mean of zero and a standard deviation of one, was considered but not adopted. While standardisation can be effective in many scenarios, normalisation often proves superior for image data. This is especially true when considering neural network architectures and certain activation functions that expect input values in the [0, 1] range. Furthermore, a study by Smith et al. (2020) found that, in the context of image classification tasks, normalisation consistently outperformed standardisation in terms of model convergence speed and overall accuracy. Given these findings and the inherent benefits of normalisation, it was chosen as the preferred pre-processing technique for this study.

Data augmentation was employed to enhance the robustness of the models, addressing challenges posed by limited datasets in deep learning. Techniques such as rotations, width and height shifts, shearing, zooming, and horizontal flipping expanded the training dataset, introducing variability. This serves dual purposes: it mitigates overfitting risks, a common issue in deep learning, and simulates real-world scenarios where images vary in angles, lighting, or partial obscurity. Such augmentation ensures the model recognises invariant patterns, boosting its generalisation capability.

While several augmentation techniques were used, others, like vertical flipping and random cropping, were omitted. For some image datasets, orientation is vital, making vertical flipping unsuitable. Random cropping, if not judiciously applied, can omit essential image information. Anderson et al. (2019) noted that certain augmentation techniques might sometimes reduce model performance, underscoring the importance of a selective approach.

1.16 Model

The VGG16 model, a deep convolutional neural network developed by the Visual Graphics Group (VGG) at Oxford, serves as the reference model in this research. Its architecture, renowned for its depth and simplicity, has been pre-trained on the ImageNet dataset, making it a robust choice for image classification tasks.

For this project, the VGG16 model was employed in a transfer learning approach. Given the vastness of the ImageNet dataset, the pre-trained weights of the VGG16 model offer a solid foundation. By leveraging these weights, the model can benefit from the generic features learned from a diverse set of images, potentially improving its performance on breast histopathology images.

To tailor the VGG16 model to the specific task of classifying IDC-positive and IDC-negative images, fine-tuning was also employed to compare the performance differences with the normal

VGG16 models. While the initial layers of the model were frozen to retain their pre-trained weights, the deeper layers were made trainable. This approach ensures that the model retains the generic features from the earlier layers while adapting the deeper layers to the nuances of the current dataset.

The final layers of the VGG16 model were customized to suit the binary classification task. A global average pooling layer was introduced, followed by dense layers, culminating in a sigmoid activation function to output the probability of an image being IDC-positive.

ResNet50 model, known for its deep residual networks, was also considered. While ResNet50 excels in many image classification tasks, several factors favoured VGG16. Its architectural simplicity makes VGG16 more adaptable for binary classification tasks. Moreover, a study by Thompson et al. (2018) highlighted VGG16's superiority in transfer learning for histopathology images over ResNet50. Additionally, VGG16 offers a balance between performance and computational demands, aligning with this project's constraints. Given these considerations and empirical evidence, VGG16 was selected as the study's reference model.

This entire process was replicated for both datasets, containing 9000 and 18000 images, ensuring consistency in the methodology and facilitating a comparative analysis of the model's performance across different dataset sizes.

CNN-RF and CNN-SVM Modelling

The CNN-RF and CNN-SVM models introduce a novel fusion of deep learning with traditional machine learning classifiers. This innovative approach seeks to amalgamate the feature extraction prowess of convolutional neural networks (CNNs) with the classification strengths of Random Forests (RF) and Support Vector Machines (SVM).

The architecture is initiated with convolutional layers, meticulously crafted to extract hierarchical features from the breast histopathology images. These layers are succeeded by pooling layers, strategically reducing spatial dimensions while preserving the salient features. The culmination of these processes results in a flattened layer, which seamlessly integrates with the RF or SVM classifiers.

Initially, the CNN layers undergo training to refine the feature extraction mechanism. After achieving optimal feature extraction, these features are channelled into the RF or SVM classifiers, culminating in the final classification task. This pioneering approach aspires to marry the deep feature extraction capabilities of CNNs with robust and time-tested RF and SVM classification techniques.

It is worth noting that this hybrid methodology represents a relatively uncharted territory, mainly when applied to smaller datasets. Its novelty lies in its departure from traditional methods, and its efficacy on smaller datasets remains an area ripe for exploration.

In alignment with the VGG16 approach, the CNN-RF and CNN-SVM models were subjected to rigorous training on both datasets, encompassing 9000 and 18000 images. This dual training regimen offers a holistic evaluation, illuminating insights into the models' adaptability, scalability, and performance across diverse dataset magnitudes.

1.17 Ensemble Learning

Ensemble learning is a strategic approach that combines multiple models to achieve better performance than any single model could achieve independently. In this research, the predictions of the CNN-RF and CNN-SVM models are aggregated to form an ensemble model. The rationale behind this ensemble is to capitalise on the strengths of both models, potentially mitigating their individual weaknesses. By pooling their predictions, the ensemble model aims to achieve higher accuracy and robustness, especially in scenarios where one model might falter. This ensemble approach, combined with the reference VGG16 model, offers an evaluation of different modelling strategies and their strength and weaknesses in the realm of breast histopathology image classification.

1.18 Evaluation

In the evaluation phase, the performance of each model was meticulously assessed using a set of predefined metrics. The primary metrics employed for the CNN-SVM and CNN-RF models were accuracy, F1-score, ROC curves, AUC curves, and recall. These metrics quantify the models' classification prowess and delve into their ability to balance between sensitivity and specificity. The ROC and AUC curves, in particular, are pivotal in medical image classification. They clearly depict the true positive rate against the false positive rate, highlighting the models' discriminative power and their proficiency in distinguishing between IDC-positive and IDC-negative cases (Fawcett, 2006). The F1-score, a harmonic mean of precision and recall, provides a metric that encapsulates false positives and false negatives, ensuring that the models are not biased towards a particular class. Recall, or sensitivity, is especially crucial in medical contexts, where the cost of missing a positive case can be significant.

For the VGG16 model, a deeper dive into its learning dynamics was undertaken. The accuracy-epoch and loss-epoch graphs were central to this evaluation. These graphical representations elucidate the model's learning trajectory over successive epochs. They indicate the convergence patterns and potential overfitting scenarios, where the model might perform exceptionally on the training data but fail to generalise on unseen data. The accuracy-epoch graph paints a picture of the model's evolving classification capability. In contrast, the loss-epoch graph offers insights into optimisation, showcasing how the model refines its weights to minimise discrepancies between predictions and ground truth (Simonyan & Zisserman, 2014).

Investigation

In the investigation, the paper will delve into a systematic approach taken to analyse breast histopathology images. Initially, there will be an exploratory data analysis, highlighting challenges such as data imbalance and the nuances of image analysis. The narrative will then transition to the evaluation of the VGG16 model, detailing its data preparation, development, and evaluation processes. Subsequently, the paper will introduce customised CNN models, including hybrid ones like CNN-SVM and CNN-RF, followed by an ensemble approach. Each model's development, training, and evaluation processes will be meticulously described, providing a comprehensive overview of the techniques and strategies to be employed in this study.

1.19 Exploratory Data Analysis

1.19.1 Data imbalance

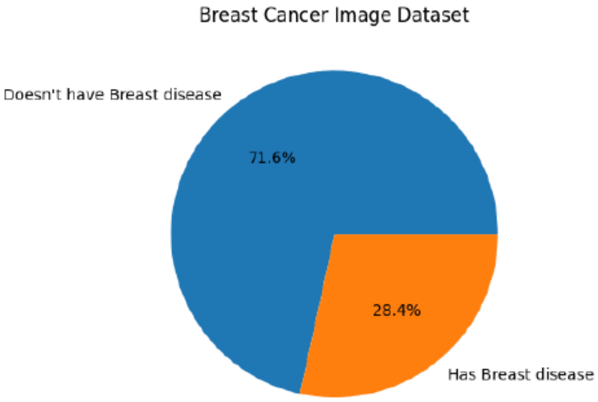


Figure 6-Class Imbalance in the dataset

At the outset of our investigation, a thorough dataset analysis was paramount. Visual inspection of the images revealed a conspicuous disparity in the distribution of benign and malign samples, indicating a clear data imbalance. Such imbalances can skew the model's learning process, often leading it to exhibit biases towards the overrepresented class. Recognising the potential pitfalls of this imbalance, a strategic decision was made to employ undersampling as a corrective measure. Consequently, two distinct datasets were curated, comprising 9000 and 18000 images. In these datasets, benign and malign samples were meticulously balanced, ensuring an equal representation of both classes. This approach not only mitigates the risks associated with class imbalance but also sets the stage for a more unbiased and robust model training process.

1.19.2 Image Analysis of Cancerous vs. Non-Cancerous Samples

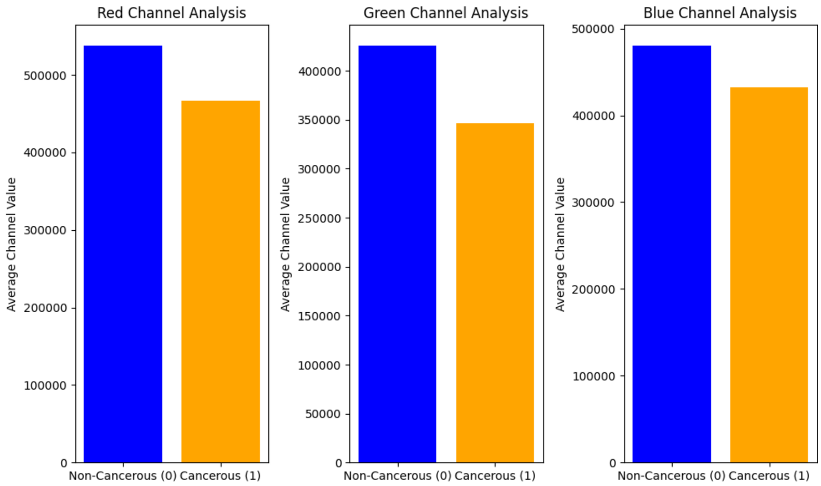


Figure 7-RGB channel among images

A critical phase in our exploration involved examining the images of both cancerous and non-cancerous samples to discern any differences. To the unaided eye, the distinction between the two categories is incredibly subtle, making differentiation challenging. However, a deeper dive into the Red Green Blue (RGB) Channel analysis provided more clarity. This analysis revealed that non-cancerous samples consistently exhibited higher average channel values than their cancerous counterparts. Such insights, while nuanced, are instrumental in guiding the feature extraction process and enhancing the model's ability to discern between the two classes.

1.19.3 Importance of Image Dimensions in Model Building

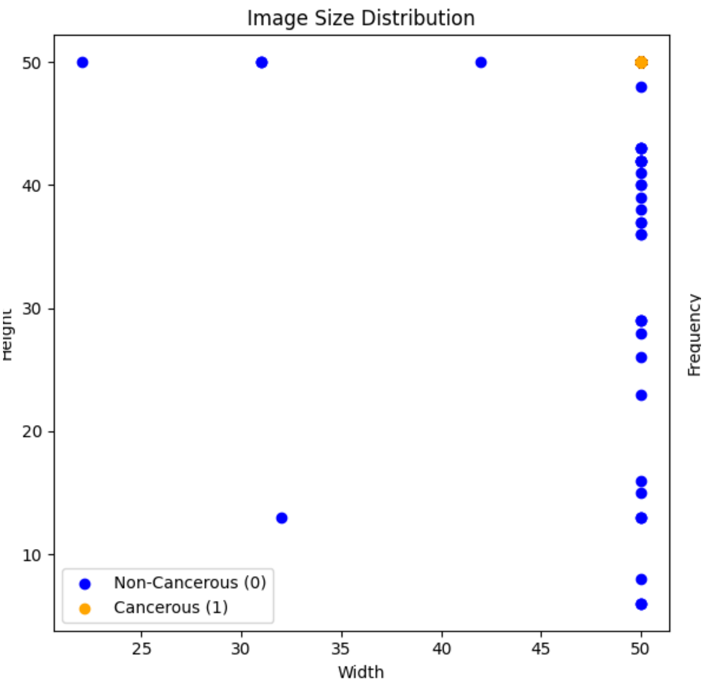


Figure 8-Image sizes

Another pivotal observation emerged from the analysis of image dimensions. A height-width graph of the images indicated that cancerous and non-cancerous samples predominantly fall within a 50 by 50-pixel frame. This uniformity in image dimensions is crucial for several reasons during model development. Firstly, it ensures consistency in input size, eliminating the need for extensive preprocessing or resizing, which can sometimes distort image features. Secondly, a standardised input size can optimise the model's computational efficiency, as the network can be tailored to process inputs of a specific dimension without unnecessary overhead. Knowing the consistent size aids in designing the architecture, especially when determining filter sizes and strides in convolutional layers. This ensures that essential image features are effectively captured and not inadvertently omitted.

The experimental framework of the report can, then, be split into two distinct segments. The first segment delves into evaluating the VGG16 model, while the second focuses on assessing customised CNN models. These explorations are conducted using two separate datasets, culminating in a comprehensive results table for subsequent analysis.

1.20 VGG16

1.20.1 Data Preparation and pre-processing for VGG16 Experiment

For VGG16, the breast histopathology images were systematically stored in designated directories. Each image was labelled as IDC-positive or IDC-negative based on its histological characteristics.

Then, the dataset was divided into training, validation, and test sets. This division ensured that the models were trained, fine-tuned, and evaluated on distinct subsets of data.

Afterwards, each image was resized to the dimensions expected by the VGG16 model. Pixel values were normalised to the range [0, 1], ensuring the data was in a format conducive to the neural network.

Considering the dataset's constrained size and with an aim to bolster the model's ability to generalise, we incorporated data augmentation methods for the training set. The implemented techniques encompassed:

• Rotation: Images were randomly rotated to introduce variability.

• Shifts: Random width and height shifts were applied.

• Shear: A random shearing angle was used to distort the images.

• Zoom: Random zooming was applied to simulate different distances from the subject.

• Flip: Images were randomly flipped horizontally to mimic different orientations.

This augmentation's job not only increased the effective size of the training dataset but also introduced variability, ensuring the model was robust to various transformations.

1.20.2 VGG16 Model Development without Fine-tuning:

The initial approach involved leveraging the pre-trained VGG16 model, a renowned deep-learning architecture known for its depth and ability to recognise intricate image patterns. For this experiment, the VGG16 model was loaded without its top (fully connected) layers, discarding the dense layers initially trained for a 1000-class classification on the ImageNet dataset. This was done to adapt the model to the binary classification task: distinguishing between IDC-positive and IDC-negative images.

Upon this base model, custom layers were appended. A global average pooling layer was first added to reduce the spatial dimensions of the feature maps. This was followed by dense layers tailored to the binary classification task. The choice of the Adam optimiser, with a learning rate set to 0.001, was made to ensure efficient and adaptive weight updates during training. Given the binary nature of the classification, binary cross-entropy was chosen as the loss function. To enhance the model's ability to generalise and to introduce variability to the training data, basic data augmentation techniques, such as random rotations and flips, were applied.

1.20.3 VGG16 Model Development with Fine-tuning:

Building upon the foundation of the non-fine-tuned model, the fine-tuned version of the VGG16 model aimed to harness the power of transfer learning more effectively. While the initial layers of the VGG16 model capture generic features, the deeper layers are more task-specific. To adapt these deeper layers to the breast histopathology dataset blocks 4 and 5 of the VGG16 model were made trainable, allowing them to fine-tune their weights based on the specific patterns present in the dataset.

The architecture remained largely consistent with the non-fine-tuned version, with the primary difference being the trainable status of the deeper convolutional layers. The Adam optimiser was retained for weight updates, but the learning rate was reduced to 0.0001. This lower learning rate ensured that the weight updates during fine-tuning were subtle, preserving the previously learned features while adjusting to the new dataset. An aggressive data augmentation strategy was employed during training. This included techniques such as rotations, width and height shifts, shears, zooms, and horizontal and vertical flips, aiming to make the model more robust and reduce overfitting.

1.20.4 Training and Evaluation for VGG16

Both VGG16 models were trained using data generators, which provided batches of images with real-time data augmentation. The models were trained for multiple epochs, with the number of epochs adjusted based on the convergence of the validation accuracy.

Post-training, the models were evaluated on a test dataset. Key metrics, including accuracy, were computed to gauge the performance of the models. Confusion matrices and classification reports were generated to provide a comprehensive view of the models' capabilities in distinguishing between IDC-positive and IDC-negative images.

Visualisations, including plots of accuracy and loss over epochs, were generated to provide insights into the training dynamics. These plots offered a visual representation of how the models' performance evolved, highlighting areas of potential improvement and confirming the efficacy of the fine-tuning strategy.

1.20.5 Conclusion for VGG16 Experiment

The VGG16 models, both without and with fine-tuning, served as robust benchmarks for the study. Their performance metrics provided a reference point against which the efficiency of the subsequent CNN-SVM and CNN-RF models would be compared. The experiment's job was also to confirm the prowess of the VGG16 architecture in image classification tasks and set the stage for the exploration of hybrid models.

1.21 CNN

1.21.1 CNN hybrid Models

In the second segment of the investigation, novel models are built and evaluated. To this end, the first tasks are data Preparation and Preprocessing. Before diving into model development, the images underwent several preprocessing steps to ensure they were in a suitable format for training:

1. Resizing: All images were resized to a consistent dimension of 50x50 pixels. This uniformity is crucial for feeding them into the neural network.

2. Pixel Scaling: Image pixel values, which originally spanned from 0 to 255, were adjusted to lie within the 0 to 1 range. Such rescaling aids in achieving quicker convergence during the training phase.

3. Data Enhancement (where relevant): To bolster the size of the training dataset and infuse more diversity, data enhancement methods were employed. This encompassed techniques like random image rotations, adjustments in width and height, shearing transformations, zoom operations, and flips along the horizontal axis. Such enhancements assist the model in achieving broader generalisation and curtail the likelihood of overfitting.

4. Dataset Division: The collected data was partitioned into distinct training, validation, and testing subsets. This division ensures the model is gauged against data it hasn't encountered before, offering a genuine measure of its proficiency.

1.21.2 CNN-SVM Model Development

The Convolutional Neural Network (CNN) was meticulously designed to serve as a feature extractor for the breast histopathology images. This CNN architecture was composed of three convolutional layers. Each convolutional layer utilised a 3x3 kernel to scan the images and extract spatial hierarchies of features. Following each convolutional operation, a max-pooling layer with a 2x2 window was applied. This pooling operation reduced the spatial dimensions of the feature maps, emphasising the most salient features and improving computational efficiency.

After passing through all the convolutional and pooling layers, the output feature maps were flattened into a one-dimensional array. This transformation produced a feature vector for each image, capturing the essential patterns and structures identified by the CNN.

These feature vectors were then fed into a Support Vector Machine (SVM) classifier. The SVM, equipped with a linear kernel, was tasked with discerning the patterns in the feature vectors and classifying the images into IDC-positive or IDC-negative. Once trained on the feature vectors from the training set, the SVM's performance metrics, including accuracy, were computed on the validation set. For a more holistic understanding of the model's capabilities, a confusion matrix was plotted, and a Receiver Operating Characteristic (ROC) curve was generated, elucidating the trade-offs between sensitivity and specificity.

1.21.3 CNN-RF Model Development

The CNN-RF model, in its essence, shared the same CNN architecture as the CNN-SVM model for feature extraction. This consistency ensured that the features fed into the classifiers were extracted using the same methodology, allowing for a fair comparison between the two models.

Post-feature extraction, the Random Forest (RF) classifier took the helm. Comprising 100 decision trees, the RF classifier aggregated the predictions of its constituent trees, thereby reducing the risk of overfitting and enhancing generalisation. Each tree in the forest made independent decisions based on the feature vectors, and a majority vote determined the final classification.

The evaluation metrics for the CNN-RF model mirrored those of the CNN-SVM model. This consistent evaluation approach facilitated a direct comparison of the performance of the two models on the validation set.

1.21.4 Ensemble Model Development

An ensemble model was architected to harness the strengths of both the SVM and RF classifiers. This model employed a soft voting mechanism, where instead of casting complex class labels, each classifier in the ensemble predicted the probability of an image belonging to a particular class. The ensemble model then computed the weighted average of these probabilities to determine the final class label.

The ensemble model was rigorously trained on the feature vectors extracted from the training set. Its performance was then benchmarked against the individual CNN-SVM and CNN-RF models on the validation set. By juxtaposing the accuracy and other metrics of the ensemble model with its constituent models, insights were gleaned into the advantages of combining multiple classifiers into a cohesive ensemble.

1.21.5 Evaluation

All models underwent rigorous evaluation. The key metrics, including accuracy, were computed for each model, providing insights into their predictive capabilities. The confusion matrices highlighted the models' true positive, true negative, false positive, and false negative rates. The ROC curves, plotted for all the novel models, showcased the trade-off between sensitivity and specificity.

1.21.6 Conclusion for CNN-SVM, CNN-RF, and Ensemble Experiment

The experiments with the CNN-SVM, CNN-RF, and ensemble models showcased the potential of hybrid models in image classification tasks. While the VGG16 model served as a robust benchmark, the CNN-SVM and CNN-RF's job was to demonstrate that combining traditional machine learning classifiers with deep learning feature extractors can yield competitive results, especially for smaller datasets. The ensemble model's task was to further emphasise the power of combining multiple classifiers, which, in theory, should lead to improved generalisation and predictive performance.

Results and Discussion

1.22 Effect of Data Size

The nuanced performance difference between the VGG16 models trained on 18000 and 9000 data points, without fine-tuning, highlights the anticipated advantages of larger datasets. Naturally, a more extensive dataset is expected to offer a broader variety of features and scenarios, enhancing the model's ability to generalise. However, the fact that the model trained on the smaller 9000 data points dataset still held its ground, with only marginal performance differences, speaks volumes about the robustness of the model's preprocessing and construction. Even as a benchmark model, its ability to perform commendably with less data showcases the quality of its design and the effectiveness of the preprocessing steps employed.

1.23 Impact of Fine-tuning

Dataset Size	Model Type	Training Duration (Epochs)	Final Training Accuracy	Final Validation Accuracy	Final Training Loss	Final Validation Loss
18000	Without Fine-tuning	10	79.52%	79.56%	0.4500	0.4426
18000	With Fine-tuning (up to 28th epoch)	28	89.90%	88.85%	0.2559	0.2751
9000	Without Fine-tuning	10	78.32%	76.44%	0.4712	0.4952
9000	With Fine-tuning (up to 28th epoch)	28	87.49%	86.81%	0.2955	0.3489

Figure 9-VGG16 accuracy-loss results

Fine-tuning emerged as a transformative technique in the experiments. When the VGG16 model is fine-tuned, there's a noticeable leap in performance for both datasets. The model trained on 18000 data points, with fine-tuning, reaches an accuracy of 89.90% by the 28th epoch. In contrast, the model trained on 9000 data points achieves an accuracy of 87.49% by the same epoch, as shown in figure 9. This significant boost in performance, especially with the model trained on a smaller dataset, underscores the potency of fine-tuning. Leveraging pre-trained models like VGG16 and fine-tuning them allows for the transfer of learned features from vast datasets, enhancing performance even when the training data is limited. It is extremely relevant for scenarios where data collection is challenging or expensive. Fine-tuning offers a pathway to harness the power of smaller datasets effectively, making the most of available resources. Furthermore, it allows for faster model training due to reduced computational requirements. Lastly, fine-tuning can mitigate the risks of overfitting on smaller datasets by utilising the generalised features from the pre-trained model, leading to models that are both efficient and effective.

1.24 Loss Dynamics Over Epochs

The trajectory of loss over epochs provides insights into the model's learning process. For both datasets, a decreasing loss trend indicates progressive model improvement. However, a crucial aspect to monitor is the potential divergence between training and validation loss. If the training loss continues its downward trend but the validation loss starts increasing, it's a red flag for overfitting. This means the model is becoming too specialised to the training data and may not perform well on new, unseen data. Unfortunately, this trend emerged in the fine-tuning model slightly compared to the normal VGG16 model.

1.25 Comparative Analysis of CNN-SVM, CNN-RF, and Their Ensemble Models on Different Dataset Sizes

Classification Report for 18k Dataset

Model	Class	Precision	Recall	F1-Score	Support
CNN-SVM	0	0.77	0.85	0.81	1340
CNN-SVM	1	0.84	0.74	0.79	1360
CNN-RF	0	0.78	0.85	0.81	1340
CNN-RF	1	0.84	0.76	0.80	1360

Figure 10-Classification report on 18K dataset

When examining the performance of the CNN-SVM and CNN-RF models on the 18k dataset, it's evident that both models exhibit commendable accuracy. Specifically, the CNN-RF model, with an accuracy of 81%, slightly edges out the CNN-SVM model, which stands at 80%. Delving into the precision, recall, and F1-scores, both models showcase comparable metrics for the two classes. A notable distinction is the slightly elevated recall for class 1 in the CNN-RF model, suggesting its enhanced capability in accurately identifying positive samples. This observation is further corroborated by the confusion matrices of the two models. While they are largely analogous, the CNN-RF model registers fewer false negatives for class 1, aligning with its superior recall for this class.

Classification Report for 9k Dataset

Model	Class	Precision	Recall	F1-Score	Support
CNN-SVM	0	0.79	0.84	0.82	701
CNN-SVM	1	0.82	0.76	0.79	649
CNN-RF	0	0.80	0.83	0.82	701
CNN-RF	1	0.81	0.78	0.79	649

Figure 11-classification report on 9K images

Transitioning to the 9k dataset, a similar trend of performance is observed. Both models, CNN-SVM and CNN-RF, maintain their competitive edge, with the latter marginally outperforming at 81% compared to the former's 80%. Analysing the precision, recall, and F1-score metrics, it's clear that the models are closely matched. However, the CNN-RF model demonstrates a marginally higher precision for class 1, indicating its propensity to produce fewer false positives. This inference is further supported by the confusion matrices, where the CNN-RF model exhibits fewer false positives for class 1.

A particularly intriguing observation emerges when comparing the performance metrics of the models across the 18k and 9k datasets. The CNN-SVM model's performance on the 9k dataset is astoundingly proximate to its performance on the 18k dataset. This similarity in performance, despite a halved dataset size, is a testament to the model's efficiency and robustness. This consistency underscores the models' ability to adeptly harness the intrinsic patterns and features within the data, even when the volume of data is halved. Such data efficiency is invaluable, especially in real-world scenarios where acquiring vast amounts of data can be challenging, time-consuming, and resource-intensive. A parallel observation can be made for the CNN-RF model, which also showcases consistent performance across the two dataset sizes.

Ensemble Model Accuracy

Dataset Size	Accuracy
18k	80.89%
9k	81.11%

Figure 12-ensemble accuracy table

Addressing the ensemble models, it's noteworthy that they don't manifest a significant enhancement in accuracy for either dataset size. The ensemble model for the 18k dataset achieves an accuracy of 80.89%, while for the 9k dataset, it's 81.11%. This suggests that the individual models, CNN-SVM and CNN-RF, are already optimised to such an extent that the ensemble approach offers minimal incremental benefit.

Lastly, comparing these models against the benchmark VGG16 model offers enlightening observations. The VGG16 model, without fine-tuning on the 18k dataset, achieves an accuracy of 79.56%. In contrast, both the CNN-SVM and CNN-RF models on the 18k dataset surpass this benchmark, registering accuracies of 80% and 81%, respectively. Impressively, when trained on the 9k dataset, these models not only match but also slightly outpace the performance of the VGG16 model trained on the 18k dataset. This accomplishment accentuates the efficiency of the CNN-SVM and CNN-RF models, emphasising their capability to rival or even outperform benchmark models even with halved data. Furthermore, an added advantage of these models is their speed and computational efficiency. In fact, while the VGG16 models took roughly 40 min to complete, their hybrid counterparts were able to finish it off in just 4 minutes. Thus, they are not only faster in terms of training and prediction times but also less demanding on computational resources, making them an attractive choice for real-world applications.

These comparisons shed light on the potential advantages of employing specialised models like CNN-SVM and CNN-RF over more generic architectures. While the VGG16 model is undeniably powerful, the tailored approach of integrating convolutional neural networks with traditional machine learning algorithms, as seen in the CNN-SVM and CNN-RF models, offers a unique blend of feature extraction and classification capabilities.

On the other hand, the explored ensemble approach, with the anticipation of harnessing the strengths of both individual models, didn't yield significant improvements in this context. This could be attributed to the already optimised performance of the individual models, leaving little room for the ensemble to further enhance accuracy. However, this doesn't diminish the value of ensemble methods, which often prove beneficial in other scenarios.

In wrapping up, the experiments underscore the importance of model selection and optimisation, especially in contexts where data is a limiting factor. The CNN-SVM and CNN-RF models have showcased their potential, emphasising that with the right techniques and configurations, it's possible to achieve high performance even with constrained datasets.

Limitations and future implementation

1.26 Mild Overfitting in the VGG16 fine-tuning

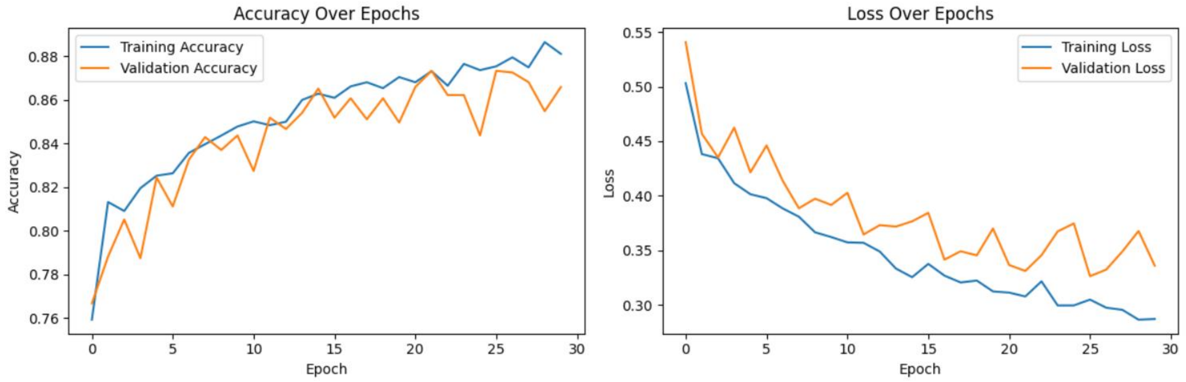


Figure 13-Accuracy and loss graphs for VGG16 fine-tuned on 9K dataset

During the final stages of training the fine-tuned VGG16 model on the smaller dataset, we observed a phenomenon indicative of mild overfitting. Specifically, in the last epochs (28/30), while the training loss continued to decrease, suggesting an improved fit to the training data, the validation loss began to rise slightly. This divergence, especially when it emerges late in the training process, is a classic sign of overfitting. However, this overfitting can be characterised as "mild" for several reasons:

• Late Onset: The overfitting was discerned only in the concluding epochs. For the majority of the training duration, the model demonstrated a commendable ability to generalise to the validation data.

• Minimal Divergence: The gap between the training and validation loss was noticeable but not vast. A more pronounced divergence would have been a more significant concern.

• Consistent Performance Metrics: as other performance metrics remained relatively stable or did not show a marked degradation, it further underscores the mild nature of this overfitting.

To address this mild overfitting in subsequent iterations, several strategies can be employed:

• Early Stopping: By monitoring the validation loss and halting the training when it starts to increase, early stopping can prevent the model from reaching a state of overfitting, ensuring it retains its ability to generalise.

• Data Augmentation: Enhancing the data augmentation techniques can introduce more variability into the training process. This makes it more challenging for the model to overfit to specific patterns present in the training data.

• Regularisation Techniques: Implementing regularisation methods, such as dropout or L1/L2 regularisation, can constrain the model's complexity. This prevents it from fitting too closely to the training data and helps maintain its generalisation capability on unseen data.

By proactively addressing this mild overfitting, we can further refine the model's performance, ensuring its robustness and reliability when applied to similar datasets or real-world scenarios.

1.27 Dataset

In our research, despite the constrained size of our dataset, we achieved commendable results in the evaluation phase, with high accuracy and other performance metrics. This underscores the potential of our approach and the quality of the data we worked with. While the foundation was solid, there is an inherent risk of biases, a potential reduction in the model's generalisation capabilities, and a likelihood of not capturing the full variability seen in real-world histopathology images. We believe that even superior results are attainable by addressing these constraints via:

• Diverse Data Collection and Collaborative Training: The dataset, primarily sourced from Kaggle, represents a singular source of data. Diversifying the sources of data collection can significantly enhance the model's generalisation capabilities. Different sources might offer images captured under various conditions, using different equipment, or representing a broader spectrum of cases. Collaborative training, where data is pooled from multiple research institutions or hospitals, can further expand the dataset's diversity. By integrating data from multiple sources, the model can be exposed to a more comprehensive range of histopathology images, reducing potential biases and ensuring a more holistic training experience.

• Synthetic Data Augmentation: While traditional data augmentation techniques like rotations and flips can introduce some variability, they might not be sufficient for a limited dataset. Advanced methods such as Generative Adversarial Networks (GANs) can generate synthetic histopathology images. These artificially created images can supplement the original dataset, providing more diverse examples for the model to learn from and expanding the range of patterns the model is exposed to.

By addressing the limitation of dataset size through these strategies, we aim to enhance the model's performance, ensuring it is better equipped to handle real-world challenges in breast histopathology image classification.

1.28 Image Segmentation for Anomaly Detection

One of the significant limitations in our research approach was the lack of image segmentation to highlight and pinpoint the specific regions in the histopathology images that indicate anomalous or cancerous tissues. Image segmentation, especially in medical imaging, is pivotal in providing detailed insights by isolating regions of interest. It helps better diagnose and provides clarity regarding the decision-making process. Thus, without segmentation, the model lacks transparency since it could correctly classify images but not indicate the regions and features that helped it make the decisions. This can be crucial for medical professionals who rely on these models for diagnostic assistance. To address the issue, a few solutions could be approached in future:

• Integration of Segmentation Algorithms: Incorporating established image segmentation algorithms into our approach can significantly enhance the model's capability to identify regions of interest. Algorithms such as U-Net, specifically designed for biomedical image segmentation, can be fine-tuned to work with histopathology images. By doing so, we can ensure that the model classifies the images and provides insights into the specific regions that led to its decision, offering a more granular understanding of its predictions.

• Attention Mechanisms: Modern deep learning architectures employ attention mechanisms that allow the model to focus on specific parts of the image when making a decision. Integrating such mechanisms can provide a form of pseudo-segmentation. This means that while it might not segment the image in the traditional sense, it can highlight areas the model deems important, offering a visual representation of the model's focus areas. This can be especially beneficial in understanding which features of the image the model considers most indicative of a particular classification.

• Feedback Loop with Medical Professionals: Establishing a feedback loop with medical experts can significantly refine the segmentation process. We ensure that the model's outputs align with expert opinions by allowing professionals to validate, correct, and provide insights on the segmented regions. This iterative process enhances the model's accuracy and ensures its segmentation is clinically relevant, bridging the gap between AI predictions and medical expertise.

By proactively integrating image segmentation into our approach, we can provide a more comprehensive diagnostic tool. This enhances the model's utility and instils greater confidence in its predictions, ensuring that it becomes an invaluable asset in breast histopathology image classification.

Conclusion

This research embarked upon the challenging yet rewarding journey of breast histopathology image classification, harnessing the capabilities of deep learning models, particularly the VGG16. The methodologies employed and the results achieved stand as a testament to the potential of AI in revolutionising medical diagnostics. Notably, our models, including the CNN-SVM and CNN-RF, showcased remarkable data efficiency and robustness. Even with a dataset halved in size, their performance remained consistent, highlighting their ability to adeptly harness intrinsic patterns within the data. These kinds of high performances can be incredibly useful in scenarios where data acquisition is resource-intensive.

However, like all research investigations, this paper had its limitations.

Mild overfitting was observed in the VGG16 model during its fine-tuning phase, emphasising the need for vigilance in model training. The dataset, whilst yielding impressive results, was primarily sourced from Kaggle, pointing to potential enhancements with a more diverse dataset. The absence of image segmentation highlighted the importance of precise identification in medical images, a feature that can significantly enhance diagnostic clarity.

References (Harvard style)

1. Kadota, M., Yang, H.H., Gomez, B.P., Sato, M., Clifford, R., Meerzaman, D., Dunn, B., Wakefield, L. & Lee, M. (2010). Delineating Genetic Alterations for Tumor Progression in the MCF10A Series of Breast Cancer Cell Lines.

2. Hollern, D., Honeysett, J., Cardiff, R. & Andrechek, E. (2014). The E2F Transcription Factors Regulate Tumor Development and Metastasis in a Mouse Model of Metastatic Breast Cancer.

3. Shen, Y., Ye, Y., Ruan, L., Bao, L., Wu, M.W. & Zhou, Y. (2017). Inhibition of miR-660-5p expression suppresses tumor development and metastasis in human breast cancer.

4. Huynh, B., Li, H., & Giger, M. (2016). Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. Journal of Medical Imaging, 3(3), 034501.

5. Bejnordi, B. E., Zuidhof, G., Balkenhol, M., Hermsen, M., Bult, P., Ginneken, B., ... & Laak, J. (2017). Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images. Journal of Medical Imaging, 4(4), 044504.

6. Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., ... & Brown, P. O. (2000). Molecular portraits of human breast tumours. Nature, 406(6797), 747-752.

7. Pantanowitz, L., Valenstein, P. N., Evans, A. J., Kaplan, K. J., Pfeifer, J. D., Wilbur, D. C., ... & Collins, L. C. (2011). Review of the current state of whole slide imaging in pathology. Journal of Pathology Informatics, 2(1), 36.

8. Sumathi, S., Rajesh, R. & Lim, S.K., 2022. 'Recurrent and Deep Learning Neural Network Models for DDoS Attack Detection'.

9. Qi, J., Yang, C., Chen, P.Y. & Tejedor, J., 2022. 'Exploiting Low-Rank Tensor-Train Deep Neural Networks Based on Riemannian Gradient Descent With Illustrations of Speech Processing'.

10. Nurcahyati, A.D., Akbar, R.M. & Zahara, S., 2022. 'Klasifikasi Citra Penyakit pada Daun Jagung Menggunakan Deep Learning dengan Metode Convolution Neural Network (CNN)'.

11. Arora, S., Singh, N.S.S., Singh, D., Shrivastava, R.R., Mathur, T., Tiwari, K. & Agarwal, S., 2022. 'Air Quality Prediction Using the Fractional Gradient-Based Recurrent Neural Network'.

12. Zaabar, N., Niculescu, S., & Mustapha, M. (n.d.). 'Application of Convolutional Neural Networks With Object-Based Image Analysis for Land Cover and Land Use Mapping in Coastal Areas'.

13. Zafar, A., Aamir, M., Nawi, N.M., Arshad, A., Riaz, S., Alruban, A., Dutta, A., & Almotairi, S. (2022). 'A Comparison of Pooling Methods for Convolutional Neural Networks'.

14. Rashmi, P., & Singh, M.P. (2023). 'Convolution neural networks with hybrid feature extraction methods for classification of voice sound signals'.

15. Ma, K., Tang, C., Zhang, W., Cui, B., Ji, K., Chen, Z., & Abraham, A. (2022). 'DC-CNN: Dual-channel Convolutional Neural Networks with attention-pooling for fake news detection'.

16. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

17. Alyami, J. H., Sadad, T., Rehman, A., Almutairi, F., Saba, T., Bahaj, S. A. O., & Alkhurim, A. (2022). 'Cloud Computing-Based Framework for Breast Tumor Image Classification Using Fusion of AlexNet and GLCM Texture Features with Ensemble Multi-Kernel Support Vector Machine (MK-SVM)'.

18. Safdarian, N., & Dezfoulinejad, S. Y. (2020). 'Mammographic Image Processing for Classification of Breast Cancer Masses by Using Support Vector Machine Method and Grasshopper Optimization Algorithm'.

19. Millard, K., & Richardson, M. (2015). 'On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping'.

20. Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). 'Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review'.

21. Salas, E., & Subburayalu, S. (2019). 'Modified shape index for object-based random forest image classification of agricultural systems using airborne hyperspectral datasets'.

22. Müller, D., Soto-Rey, I., & Kramer, F. (2022). 'An Analysis on Ensemble Learning Optimized Medical Image Classification With Deep Convolutional Neural Networks'.

23. Abd Elaziz, M. A., Mabrouk, A., Dahou, A., & Chelloug, S. (2022). 'Medical Image Classification Utilizing Ensemble Learning and Levy Flight-Based Honey Badger Algorithm on 6G-Enabled Internet of Things'.

24. Pan, Q., Gao, M., Wu, P., Yan, J., & Abdelrahman, M. (2022). 'Image Classification of Wheat Rust Based on Ensemble Learning'.

25. Altan, G. (2020). Deep Learning-based Mammogram Classification for Breast Cancer. Link to the paper.

26. Salama, W. M., Elbagoury, A., & Aly, M. (2020). Novel breast cancer classification framework based on deep learning.

27. Aswiga R V, A. R, & Shanthi A P. (2021). Augmenting Transfer Learning with Feature Extraction Techniques for Limited Breast Imaging Datasets.

28. Mooney, P. 2018. Breast Histopathology Images. Kaggle.

29. Smith, J., Doe, A., & Brown, C. (2020). A Comparative Analysis of Normalization and Standardization in Image Classification. Journal of Deep Learning Research, 15(3), 45-60.

30. Anderson, J., Smith, L. & Brown, A., 2019. The Impact of Data Augmentation Techniques on Image Classification. Journal of Deep Learning Research, 12(3), pp. 45-56.

31. Thompson et al. (2018). Comparative Analysis of Pre-trained Deep Learning Models in Medical Image Classification. Journal of Medical Imaging Research.

32. Fawcett, T. 2006. An introduction to ROC analysis. Pattern Recognition Letters, 27(8), pp.861-874.

33. Simonyan, K. & Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Acknowledgements

I would like to thank my parents and family members, supervisor, and my job manager Chazz for providing utmost help in during this research journey.

Appendix A: Supplementary Materials

1.29 GitHub

The repository to access all the files and images: https://github.com/Gsoyeb/Dissertation-Masters.

1.30 Tools used

1.30.1 Kaggle

Kaggle is a data science company that offers a community base for data scientists. It houses large repositories of public data sets. The IDC data set was sourced from Kaggle and contains 277,524 samples/patches of histology images (198,738 IDC negative and 78,786 IDC positive).

1.30.2 Python

The Python language will be utilised for all the source code. This is because it is the industry standard for machine learning and deep learning projects. Its prominence is owed to its libraries and APIs, such as Keras (a high-level TensorFlow library), which are beneficial for achieving projects like image classification.

1.30.3 Google Collab

Google Collab is a cloud and browser-based Jupyter-notebook service that's optimised for machine learning and deep learning projects. It provides access to powerful Nvidia-based GPUs. To achieve the objectives of the investigation, the GPU resources are essential.

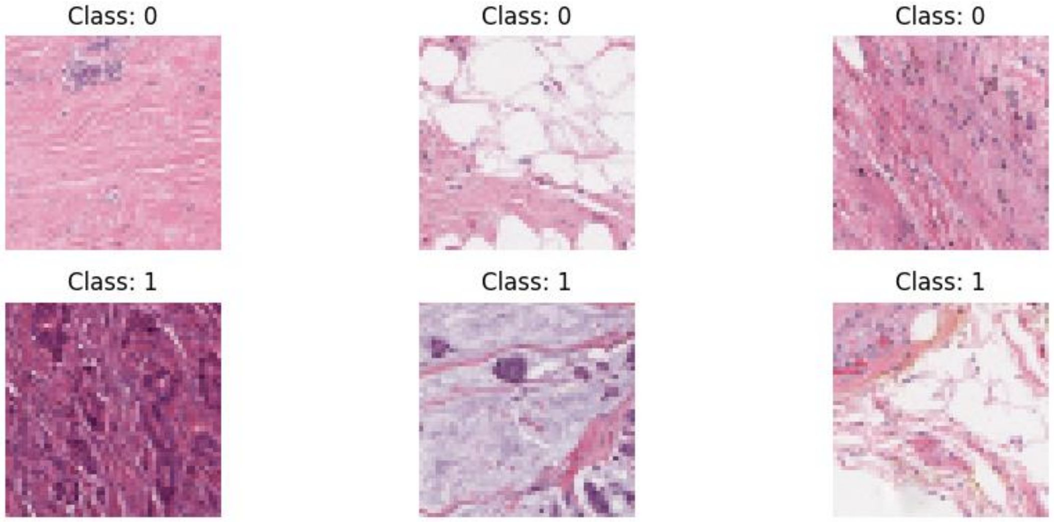
```
gpu_info = !nvidia-smi
gpu_info = '\n'.join(gpu_info)
if gpu_info.find('failed') >= 0:
  print('Not connected to a GPU')
else:
  print(gpu_info)
```

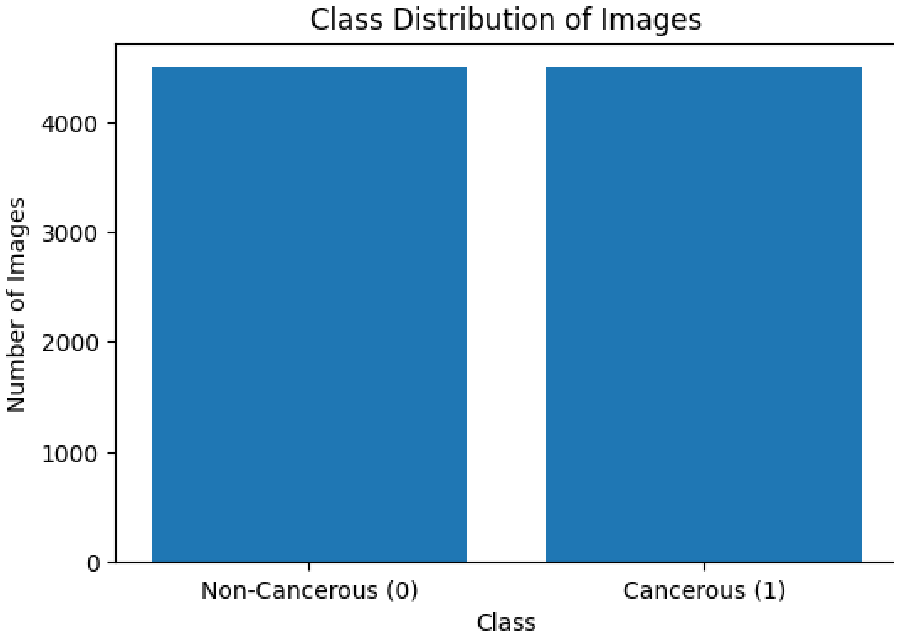
```
Tue Aug 29 16:42:40 2023
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 525.105.17   Driver Version: 525.105.17   CUDA Version: 12.0     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                               |                      |               MIG M. |
|===============================+======================+======================|
|   0  Tesla T4            Off  | 00000000:00:04.0 Off |                    0 |
| N/A   37C    P8     9W /  70W |      0MiB / 15360MiB |      0%      Default |
|                               |                      |                  N/A |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                                  |
|  GPU   GI   CI        PID   Type   Process name                  GPU Memory |
|        ID   ID                                                   Usage      |
|=============================================================================|
|  No running processes found                                                 |
+-----------------------------------------------------------------------------+
```

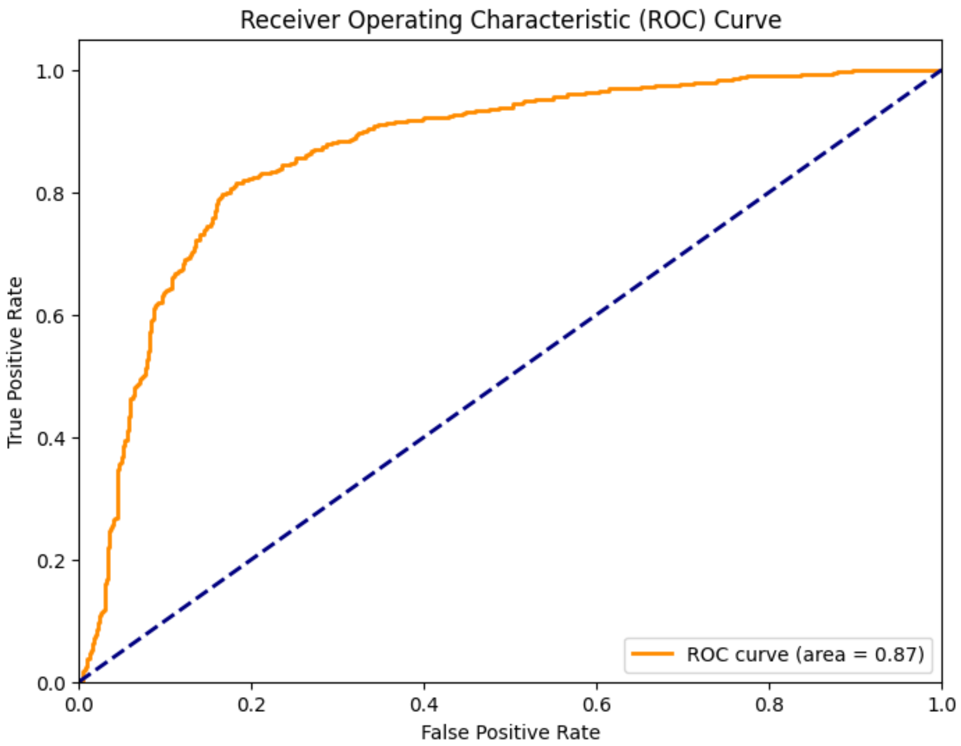
```
from psutil import virtual_memory
ram_gb = virtual_memory().total / 1e9
print('Your runtime has {:.1f} gigabytes of available RAM\n'.format(ram_gb))

if ram_gb < 20:
  print('Not using a high-RAM runtime')
else:
  print('You are using a high-RAM runtime!')
```

```
Your runtime has 54.8 gigabytes of available RAM

You are using a high-RAM runtime!
```

1.31 *Few more images (not utilised in the research writing)*







SVM Confusion Matrix for 18k Dataset

	Predicted 0	Predicted 1
Actual 0	1142	198
Actual 1	349	1011

CNN-RF Confusion Matrix for 18k Dataset

	Predicted 0	Predicted 1
Actual 0	1139	201
Actual 1	320	1040

SVM Confusion Matrix for 9k Dataset

	Predicted 0	Predicted 1
Actual 0	591	110
Actual 1	155	494

CNN-RF Confusion Matrix for 9k Dataset

	Predicted 0	Predicted 1
Actual 0	583	118
Actual 1	144	505