Using Self-Supervised Learning Models to Predict Invasive Ductal Carcinoma from Histopathological Images

MSc Research Project
Data Analytics

Taher Abbas Poonawala
Student ID: X21118043

School of Computing
National College of Ireland

Supervisor: Professor Jorge Basilio

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Taher Abbas Poonawala
Student ID:	X21118043
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Professor Jorge Basilio
Submission Due Date:	15/12/2022
Project Title:	Using Self-Supervised Learning Models to Predict Invasive Ductal Carcinoma from Histopathological Images
Word Count:	7799
Page Count:	24

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Taher Abbas Poonawala
Date:	30th January 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Using Self-Supervised Learning Models to Predict Invasive Ductal Carcinoma from Histopathological Images

Taher Abbas Poonawala
X21118043

Abstract

The process of learning representations without annotated data is called self-supervised learning (SSL). SSL has found most of its applications and proof of concept in the field of natural images since its debut. These strategies have a lot of potential, especially in situations when data is scarce. While these algorithms would almost certainly benefit immensely from being researched and benchmarked on sparse medical datasets such as microscope images, they have gotten little attention so far. This research focuses on creating a framework for applying the SSL approach to low-data-regime scientific datasets. The dataset consisted of histopathological scans of invasive ductal carcinoma (IDC) with 96-pixel resolution images. Methods for adapting SSL protocols to operate with this data collection were thoroughly investigated. Employing the SimCLR framework, which uses a contrastive approach, to learn data representations with a focus on cropping algorithms. The fascinating structure of the dataset allows for significant changes. After training on an unlabeled dataset, the SimCLR achieved 84% accuracy as a prelude to logistic regression for image classification. The ResNet18 model, considered the baseline, could only properly predict 75% of the images. It outperformed a model trained only from supervision with 500 images per label by 8% using just a tenth of the labeled data.

1 Introduction

Most women over the age of 40 who are diagnosed with advanced breast cancer have infiltrating ductal carcinoma (IDC), (Samala et al.; 2019). The milk ducts serve as the initial staging area for lymphocytes before they spread to other areas of the breast and the body. Malignancies may be diagnosed using a variety of clinical modalities, including biopsies, MRI, US, and mammography. Tissue samples acquired from a patient and examined by a board-certified pathologist are the gold standard in diagnostics. The pathologist's ability to detect cancer and foresee its development is also taken for granted.

The effectiveness of Machine Learning(ML) relies on the models' capacity to acquire knowledge from data that has not been explicitly labeled. ML techniques were created with a specific emphasis on labeled data. ImageNet, cifar, MNSIT, and the fashion version of MNIST are just a few of the well-curated datasets that have been developed. Data labeling, however, is a time-consuming process. Identifying 14 million images in the ImageNet collection, required 22 human years of effort (Benbrahim and Behloul;

2021). Slowly but surely, difficulties began to emerge, such as "how to scale learning from unlabelled data?" While unlabeled data is readily accessible, it is impossible to identify everything in sight.

Given a dataset of 327,680 colored histopathological images of IDC, and using a supervised Convolutional Neural Network(CNN) model termed ResNet18 as a baseline, how effectively can a self-supervised deep learning SimCLR identify IDC using histopathology data in the field of medical imaging? How do hyper-parameters influence these representations?

The fundamental objective is to create a novel method for predicting the IDC by using a SSL model. When it comes to IDC forecasting, there has not been a lot of research on self-supervised systems. The IDC may be predicted quickly using this approach. It is not necessary to have specialized expertise to assess the data and draw conclusions. In this research, IDC is evaluated at a low cost and the appropriate moment, meaning that patients do not have to spend time and energy attempting to predict what will happen or postpone obtaining the therapy they need. The majority of the data set comes from multiple-angle micro-graphs of biopsy analyses. A potentially disastrous outcome might occur if pathologists examining the images fail to see the labels and thereby misinterpret the results. Input labels, a magnification setting, and user-guided cropping are required to train a model in the present study's CNN.

This study uses SimCLR, an SSL framework that leverages logistic regression as a binary classifier, and ResNet18, a baseline for comparison of SSL with supervised learning, for evaluating and verifying the efficacy of the contrastive learning technique. Methods that rely on SSL to maximize the learning, which enables quick tailoring for a specific classification issue. With SSL, a large dataset is usually not a problem. However, to train a model, it would be essential to manually identify each of those images using supervised learning. A manual process to identify the same amount of data would take a couple of months and would cost an arm and a leg. Therefore, using an SSL model with either no or limited labeled data can predict the output with ease. In a similar vein, just raw histopathological images of a patient are needed to predict IDC for that patient.

The roadmap for the paper is arranged as follows: Section 2 covers the state of the art in SSL, deep learning, and supervised ML techniques for IDC prediction using histopathology images. This section has been divided further into subheadings depending on the different ML techniques. Section 3 describes in detail the general methods used to acquire the findings. It has multiple subsections, one for each stage in the overall technique, that describe and justify the process used for that step. Section 4 discusses the architecture of the proposed SimCLR method that uses a contrastive approach for learning representations of the data and a supervised ML model. Section 5 will go into the specifics of putting the recommended solution into action. It focuses on the steps right from data loading to modeling. Section 6 refers to the methodology and outcomes of the testing will be discussed here. The components of an experiment vary from one to the next. Moreover, each experiment's overall findings are discussed at length in the paper's discussion section. Section 7 contains concluding thoughts and prospects for future research in the field are presented here.

2 Related Work

The potential to predict IDC using histopathological scans has been the primary focus of the research investigations and tests that have been conducted. SSL, ensemble approaches, deep learning, and transfer learning were the subjects of the research. The research was conducted with great caution and attention to detail, and a short summary of those findings is provided below.

2.1 Review of Work Based on Unsupervised Machine Learning Techniques

In order to accurately diagnose skin tumors high resolution images are required. When training the model with unlabeled data, the authors of the (Azizi et al.; 2021) study utilized a Multi-Instance Contrastive Learning (MICLe) that was self-supervised. In order to categorize the tissues, a variety of approaches are used, with semi-supervised algorithms serving as the major way. The model performs better when shown images at varying magnifications by 6.7% with top-1 accuracy and 1.1% in area under curve. The model's performance and accuracy were respectable when compared to those of conventional semi-supervised and supervised machine learning models. The author has laid up plans for the next studies that will use high-resolution images to make inferences about tumors at various magnifications.

The authors of the (Ghesu et al.; 2022) article have conducted research with the objective to train the model with over 100 million medical images, which has traditionally proven challenging. The author used a CNN that had been pre-trained in addition to an ML method. A small number of studies have investigated where SSL is evaluated on 10 million unlabeled images from various datasets including radiography, CT, MR imaging, and ultrasonography. Low-regime datasets nonetheless provide difficulties; without millions of data points, they are less dense than the other datasets and have unique difficulties. Typically, these datasets have been curated by professionals. The model has a sensitivity of 84.6 percent, which means that it can identify a significant portion of false positives. The pre-trained model has been less trained and on applying the multi-layer perceptron the scalability of the model has reduced.

A group of researchers have developed a model to determine if a person has the cancer-causing BRCA1 or BRCA2 gene mutations (Khaliliboroujeni et al.; 2022) centered on WSI of breast excision specimens and patients' BRCA gene status. Dimensionality reduction methods and combining good classification findings with tile-based classification may help provide reliable predictions. The magnification levels used for sorting range from 10x to 40x, with 50x being the most common. The authors could have employed an ML technique for selecting features.

2.2 Review of Work Based on Deep Learning Methods

The authors (Spanhol et al.; 2016) created a neural network that is divided into two phases. Because of the microscopic examination, learning may occur on several levels, from the level of breast cancer patches all the way down to the pixel level. The CNN network enables automatic resizing in terms of the dimensions of the input image, resulting in improved performance. This improves the model's retention of imperfect data. The model's findings were double-checked against the radiologist's. More than 100,000 images were used for training. This means that the condition might be diagnosed early on and treated successfully. Some restrictions exist on the idea, such as the authors not providing a mechanism for multiclass categorization and the lack of work at the cellular level.

(Samala et al.; 2019) did more research on breast cancer subgroups. When working with SE-ResNet, when squeeze-and-excitation is used in conjunction with residual models, only the bare minimum of hyperparameter adjustment is needed as a learning rate. Cell overlap and unstable color images generated by different methods are two examples of current areas of use. On the other hand, the author has hit a pinpoint target. The author could have focused on batch size, since it could influence the model performance.

Research by (Roth et al.; 2016) used CNN to provide forecasts regarding breast cancer metastasis to the lymph nodes. Before using the images to train the model, they were transformed to greyscale. A possible downside is that it may be difficult to detect whether or not lymph nodes have metastasized after clinical surgery.

Data storage constraints make it difficult to train a CNN model on full-size whole-slide images. High-throughput Adaptive Sampling of Histopathological Images provides a vital paradigm for overcoming huge volume dataset(HASHI), (Cruz-Roa et al.; 2018). The algorithm predicts that during the next 24 hours, almost 6 million images will be shot. The biggest problem is that it can not predict the tumor at stage zero, the earliest potential stage.

The authors (Siegel et al.; 2017) have constructed a neural network that consists of two steps. Learning from breast cancer patch level to pixel level under microscopic observation. Auto adjustment in terms of image dimension is achieved through the ResNet network to get high performance. This helps the model adhere to noisy input. The output of the model was verified and tested against the radiologist's output. More than 100,000 images were used to train. Hence, the identification of breast cancer is possible at an early stage. There are a few drawbacks of the model, which are that the authors have not laid down how multiclass classification can be achieved, and very little work has been performed at the microscopic level.

The prediction of high-resolution images of cells is accomplished using CNN by the author (Diao et al.; 2021). Feature extraction is carried out using the random forest approach. After each iteration, the model was supplied with labels, and a 95% confidence interval was obtained. To increase the model's performance, the author used the Benjamin-Hochberg approach to merge the P-value from the current step with the previously constructed decision tree.

Predictions of cancer outcomes are made using patient-reported clinical data, genomic data, and histological images. The outputs of a CNN using a Sigmoid optimizer are fed into a random forest classifier, as described by the research group referenced in (Arya and Saha; 2021). Researchers may employ a number of different pieces of information gleaned from histopathological scans to enhance the precision of forecasts and the chances of survival. Although the authors do employ CNN, it is possible that a recurrent neural combined with a CNN or feed-forward approach might be more effective for tumor prediction.

2.3 Review of Work Based on Supervised Machine Learning

The authors of (Islam et al.; 2017) employed ML techniques such as K-nearest neighbors(KNN), and a support vector machine (SVM). It was at the tissue level that the tumor was primarily classified. The authors state that a large number of tests were run, and the findings of several different models were taken into account when determining the level of accuracy achieved. When comparing SVM with random forest for data classification, logistic regression and decision trees could have been compared as well.

The authors (Singh et al.; 2021) used supervised ML classifiers like random forest and XGBoost on IDC data. To smooth out the information flow and get rid of outliers, the standard deviation was lowered during preprocessing. Classifiers were trained on a dataset consisting of 2,178 occurrences and 12 predictors. The fact that the model was applied without first identifying the most important traits is one criticism.

A framework driven by XGBoost and random forest has been developed by the authors (Thongkam et al.; 2008) to identify breast cancer survivability. The model performance is measured in terms of the Area Under the Curve (AUC) in order to facilitate the process of disease progression forecasting. In order to offer the most accurate approximation that is practically possible, the framework makes use of the Efron approximation for the learning functions and gradient equations. According to understanding from the authors' work, there does not seem to be any clear indication of when the rate of loss will reach a constant.

AdaBoost ultrasound images were utilized to differentiate between cancerous and benign breast cancers by the team of researchers in (Yang et al.; 2022). The authors used bi-clustering mining to determine the image's recurrent theme. Research revealed the PC-DFS problem, which was fixed by using the FSDND technique for gauging distance. The study is not perfect; for example, many cases of cancer are misdiagnosed, and mistakes are still made when it comes to detecting very tough tumors.

To conclude the literature review, previous studies have shown that the choice of which attributes to include in a model may significantly affect the model's final results. To boost the model's accuracy and precision, it may be necessary to experiment with the model with different image augmentations techniques. The study's novel component is its classification of IDC based on histopathological images of patients using 327,680 color images taken from a range of angles. In order to combine a wide range of augmentation activities, a projection head, a large batch size, and various training strategies, it is proposed to build a contrastive learning SimCLR model based on SSL and baseline fully supervised learning such as ResNet18 to compare the results.

3 Methodology

To carry out a machine learning activity that involves inspecting a dataset and extracting relevant patterns from it, a certain architecture must be employed. Sequential iterative techniques are shown in figure 1. The first stage involves an unlabeled dataset being loaded into a pre-processing and transformation pipeline. The model must first be trained and tuned before moving on to the evaluation and its monitoring.

3.1 Dataset Description

To better learn prominent features, the contrastive technique takes use of differences in visual characteristics across image patches When examining images at a high resolution, it might be challenging to obtain visually distinguishable patches for digital histopathology. Selecting image patches from the same WSI or sampling images from the same dataset will result in a less diverse dataset than selecting images with different resolutions, tissue

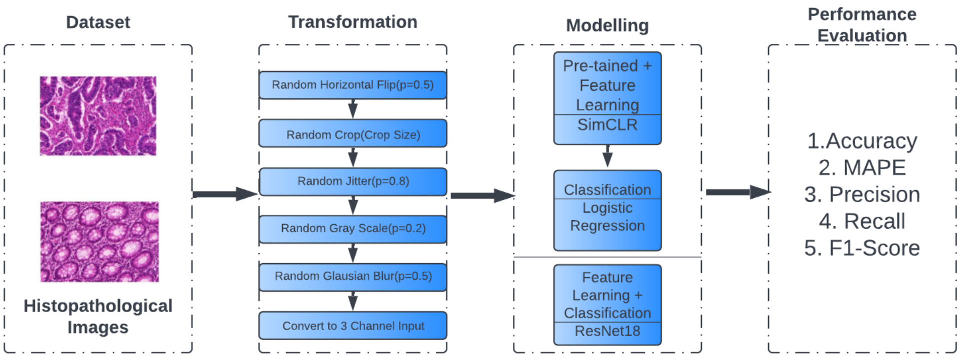

Figure 1: Methodology

types, and varied staining. Figure 2 depicts the dataset, which contains 327,680 breast histopathological colored images(96 x 96px) of patients, all of which are obtained from an open-source repository[1] during the data collecting phase.

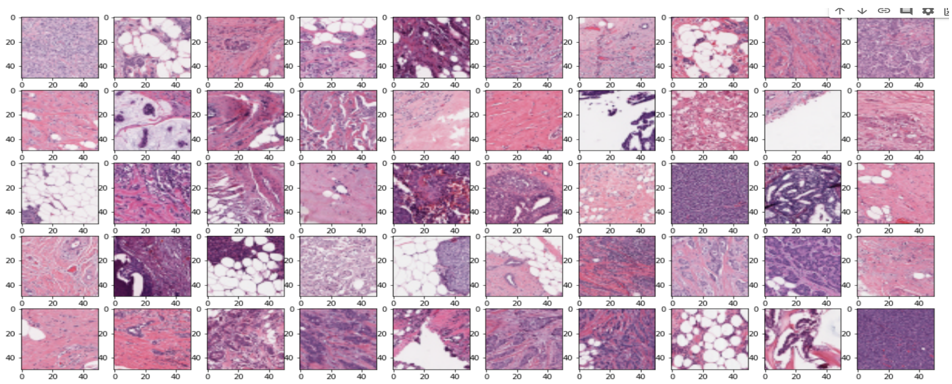

Figure 2: Dataset Overview

3.2 Data Pre-processing and Transformation

For a batch size of $2N$ in each pretraining step, there is corresponding two images generated that are one positive image and negative image $2(N\text{-}1)$, which corresponds to the transformation step. As studied in the research given by (He et al.; 2020), the research will use the negative instance of the images from the batch instead of generating negative samples additionally, as this will not only cut the model cost but also optimize the $4N^2$ feature vectors in each batch during the pretraining step. The main parts of the SimCLR framework are, in brief,

- The formation of a pretext task relies heavily on a composition of numerous augmentation activities;
- Using a projection head, or non-linear head $g(.)$, refer figure 3 for an example of how much of an improvement in representational quality it makes;
- Important for representations are high batch sizes and extended training steps;

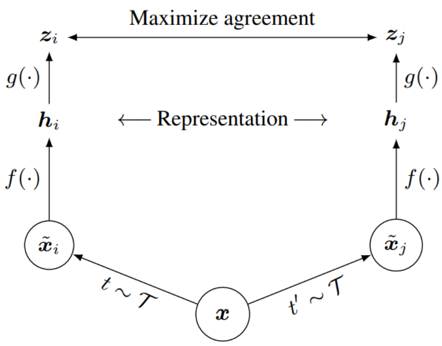

Figure 3: Self Supervised Learning: Highlevel Architecture

In the natural image dataset, SimCLR uses random crop followed by resizing as a critical transformation in the augmentation pipeline. however, in the case of histopathological images resizing should not be considered as it will degrade the micro-structure in the image, as per the analysis provided by (Chen, Kornblith, Swersky, Norouzi and Hinton; 2020). The standard augmentation pipeline used for histopathological images is as shown in figure 4.

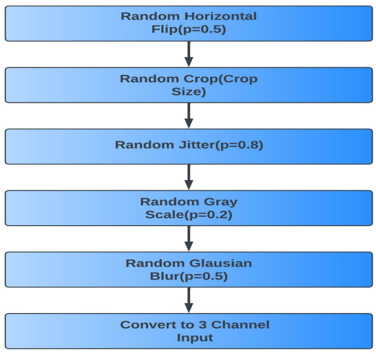

Figure 4: Image Augmentation Pipeline

3.3 Feature Extraction

Images include a color histogram that goes from 0 to 256, but for usage with data mining techniques, this must be transformed into a range between 0 and 1 so that different classification methods may be used. Also, the images are converted to grayscale and padded. Both overfitting and underfitting of the model, which may occur with large datasets are addressed by preprocessing the data to make sure the final dataset is balanced. When compared to the surrounding healthy tissue, malignant growths seem darker and denser. Is this something that often happens in ductal tissues or is it more prevalent in cancer cells? Though mostly in good condition, there are a few strikingly violet areas. It would be fascinating to hear the pathologist's opinion on what abilities are most important. Milk ducts might be exposed via the tissue holes. Tumors that are more likely to be malignant are often greater in size and more numerous than those that are not. The model is said to be able to identify even minute differences that are indicative of the

[1]https://zenodo.org/record/1494286#.Y35QvHbP07E

images' different states, despite its inability to discern the two kinds apart based on look alone. There is a considerable standard variation when comparing the total number of image patches created by each individual. Typically, patients have a broad range in the average number of image patches. It's only reasonable to question whether the resolution of tissue cells seen in images from various patients, refer figure 5, varies at all. More than 80% of the patient's patches have IDC in previous circumstances. Either all of the tissue is infected with cancer or the breast slice used to analyze IDC cancer was too small to detect significant disease in those areas.

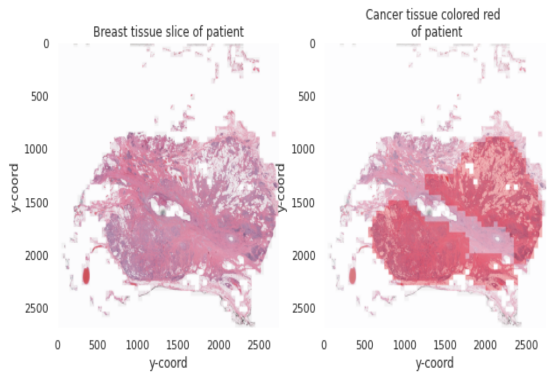

Figure 5: Cancer Patches

Pathologists rely on medical images for illness detection and diagnosis, however, these images need to be enhanced for the benefit of the pathologists. It is important for images to be of high quality before being processed using an automated deep-learning model. The research provided by (Zhuang and Guan; 2017) mentioned a technique called Histogram Equalization basically improves the brightness and contrast of an image without altering the actual data within the image, this technique is variance and mean of the original image. For the purpose of improving contrast and brightness, histopathology image data is processed using the Histogram Equalization technique in this study. Because of the tiny nature of the analysis, histopathological images are stained and might be difficult to interpret. When an image is normalized, the contrast is enhanced. Therefore, it improves the precision and efficiency of unsupervised learning.

3.4 Modelling

SSL also known as unsupervised learning, refers to a circumstance in which a learning algorithm is given input data without the labels required for conventional supervised training. Yet, there is a plethora of information that can be extracted from this data, such as the degree to which the images differ from one another. How do different types of patterns define various images? Can the images be grouped together? It's important to keep going in the same direction. A benefit of SSL is that it is often straightforward to amass a large dataset.

As a first step in solving image identification issues, a model is trained using an unlabeled histopathology image collection. It is standard practice to only employ a small fraction of available images throughout each training iteration. Two versions of each image result from using the data transformation pipeline. A 1D feature vector is extracted from the images using a CNN-based ResNet and then fed into a single-layer multi-layer perceptron (MLP). In the training step, the model tries to pull the similar feature vector of the same image while it repels the vectors belonging to different images. . This framework is useful for supervised learning because it compels the model to focus on visual properties, such as images, that are unaffected by the augmented data. Images are classified as IDC using a logistic regression model built on top of the SimCLR model, and the results are compared to those of the fully supervised ResNet18 model. Also, to avoid overfitting the classification model and baseline model after 2 and 10 epochs respectively the model is evaluated.

3.5 Evaluations Criteria

In machine learning, evaluation metrics are used to measure how well a model performs on inferred data. There are several ways to measure a model's efficacy, but one common method is to compare its test set performance versus the training set. The effectiveness of the model is as follows:

3.5.1 Loss Function

Loss functions are used to train the weights of the network by comparing the model's prediction to the label of the input data (also known as ground truth). While Mean Squared Error (MSE) is utilized for regression, Cross Entropy is often employed for multi-class classification. This is vital knowledge since MSE is employed as a loss function in durability prediction evaluations. $\mathcal{L}_2$ loss is another name for MSE. When dealing with a two-class situation, the cross entropy is defined by, $\mathcal{L} = -(y \log(p) + (1 - y) \log(p))$. Also, the loss is computed for each class label per image in a multi-class classification issue, and then added together. In recognition of:

$$\mathcal{L} = -\sum_{m}^{c} y_{i,c} \log(p_i, c)$$

Here, y is a binary variable (0 or 1), M is the total number of categories for observation i and that has a probability prediction of p_i that it belongs to class c if and only if the class label c is the correct classification.

The formula of the regression MSE is:

$$\mathcal{L} = \frac{1}{n} \sum_{n=i}^{n} (y_i - \hat{y}_i)^2$$

where $\hat{y}_i$ is label predicted for the corresponding i^{th} sample, y_i is ground truth for i^{th} sample, and n is number of samples.

In order to compare and contrast various techniques and hyperparameters in the classification job, the Accuracy metric is used.

$$Accuracy\,(y_i, \hat{y}_i) = \frac{1}{n} \sum_{n=1}^{n} [[y_i = \hat{y}_i]]$$

where, y_i and $\hat{y}_i$ are the labels that really happened and the ones that were anticipated, whereas [[]] represents the lverson bracket. 1 is returned if the expression within the bracket evaluates to true; 0 otherwise.

For comparison of regression task, Mean Absolute Percentage Error (MAPE) is used as a metric and defined as below;

$$MAPE = \frac{1}{n}\sum_{n=i}^{n} |\frac{y_i - \hat{y}_i}{y_i}|$$

where $\hat{y}_i$ and y_i are predicted value and ground truth respectively and n is number of samples, .

The following instances have been identified as IDC-positive predicted cases:

$$recall = \frac{TruePositives}{TruePositives + FalsePositives}$$

Acquiring a positive IDC forecast with precision is the quality referred to as precision.

$$Precision = \frac{TruePositives}{TruePositives + FalseNegatives}$$

The f_1 score is determined by calculating the harmonic mean of the recall and precision scores:

$$f_1 = \frac{2}{\frac{1}{recall} + \frac{1}{precision}}$$

4 Design Specification

In this section, the framework of the system and the major features of the investigation are shown. As seen from the figure 6 that the model is broken down into two distinct phases: the first deals with feature extraction and SimCLR training, while the second focuses on classification. The unlabeled histology dataset is loaded into the model at the commencement of the process. After the data loader step is completed the transformation pipeline is triggered, relevant features from the dataset are extracted in preparation for training the model using SimCLR, and the checkpoint is stored. Using the saved checkpoint the logistic regression model is used to perform the classification task. ResNet18 is used as a baseline model to evaluate the result of SSL.

4.1 Models

4.1.1 Pre-trained: SimCLR

When classifying images, a pre-trained model is fed into a logistic regression model, and the features are obtained from the SimCLR model. In contrast, the results of the self-supervised model, a fully supervised model ResNet18 are also built. The contrastive learning strategy proposed in (Chen, Kornblith, Norouzi and Hinton; 2020) is shown in figure 7, and it centers on optimizing the degree to which two representations of the same image agree with one another. The objective is to discover a way to compare and contrast the normalized feature representations of two augmentations of the same image (i) an encoder or neural network $f_\theta(.)$, (ii) an auxiliary projection layer $p_{\hat{\theta}}(.)$ with parameters , (iii)for an input image i the augmented images are z_i and z_j, and (iv)with parameters θ a probabilistic augmentation function $f_{avg}(.)$: $z_i = p_{\hat{\theta}}(f_\theta(f_{\text{avg}}(i)))$. Concurrently, a

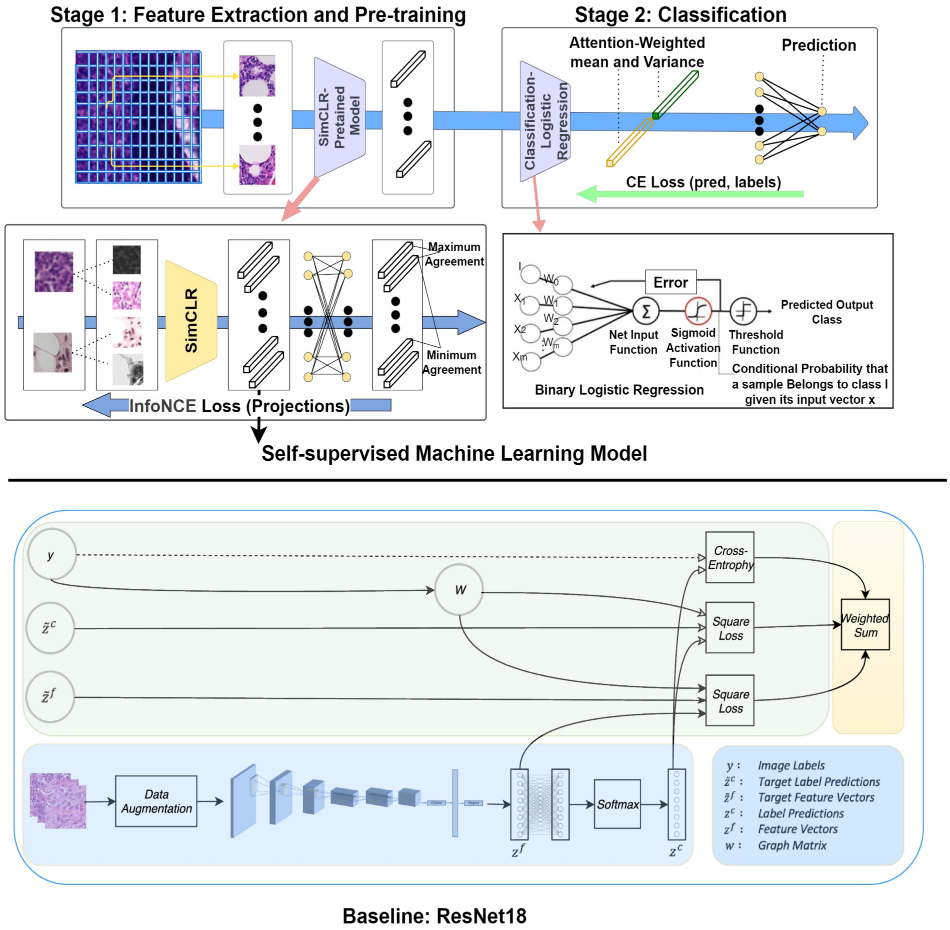

Figure 6: System Design

contrastive loss function called infoNCE loss is constructed that makes other images in a batch distinct from the image i as prescribed by (Oord et al.; 2018).

$$\mathcal{L}_{i,j} = -\log \frac{\exp\left(\text{similarity}\left(z_i, z_j\right)/\tau\right)}{\sum_{k=1}^{2N} 1_{[k \neq i]} \exp\left(\text{similarity}\left(z_i, z_k\right)/\tau\right)}$$

where the similarity function is a distance metric between two l_2 representations, 1 is an indicator function that outputs 0 when $k = i$ else 1, and τ is a temperature parameter that assists in the weighting of dissimilar situations to complete hard negative mining and for this study, it is set to 0.07. In a single hidden layer of multi-level perceptron(MLP) refer figure 3 the lower embedding space is fed pre-activation layer

z_i depicts the pre-activation layer output of the function $f_\theta(.)$, and the MLP output as the result of the function $p_{\hat{\theta}}(.)$ in figure 7. The pre-activation layer's output is projected into a lower embedding space using an auxiliary projection layer, which is a single hidden layer Multi-Layer Perceptron (MLP) refer figure 3. It was discovered that simply comparing the outputs of the pre-activation layer wasn't the most efficient method for learning representations. Instead, it was found that comparing z_i and z_j was more beneficial. Using the cosine similarity measure for the experiments, which can be expressed as similarity$(x,y) = x^k y / \|x\| \|y\|$. The authors of the study (He et al.; 2015) experimented and found that using infoNCE was more effective in teaching better representations than using other loss functions such as margin or logistic losses. NT-XENT was proven to be effective for the natural dataset as per the analysis

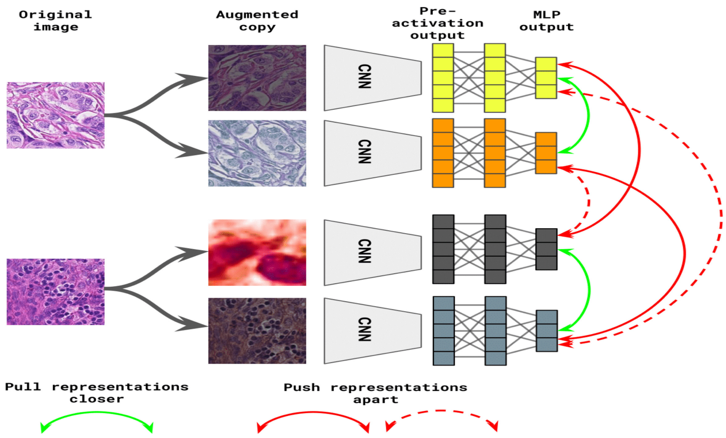

Figure 7: Research Methodology

4.1.2 Classification: Logistic Regression

Logistic Regression (LR) is a well-liked option for binary classifiers. It is a very effective learning model that can estimate the probability that a given piece of input data belongs to class X or Y. LR employs a linear function to weighted-blend input data to predict an output. The sigmoid function is then utilized to transform the LR's results to a value between 0 and 1. The training of an LR binary classifier consists of improving the weights of the LR function applied to the input data by undertaking a maximum-likelihood estimate that minimizes the error in probabilities predicted by the LR model (Hirra et al.; 2021). It is expected that the optimum weights will get the network's prediction very close to being exactly 0 or 1. If a classification job returns a yes or no, and a threshold value is set to 0.5. It means the prediction is for class Y if the probability is less than 0.5 and class X if it's higher than 0.5.

4.2 Baseline Model: ResNet18

(He et al.; 2015), created a series of models in 2016 that included the deep residual network (or ResNet for short). For the most part, training a deep neural network is a time-consuming process that can only go so many layers deep; this architecture was created to circumvent these challenges. The baseline model is based on the ResNet18 conventionally-supervised learning model. Training data undergoes image augmentation, and resnet incorporates more factors than the data points itself to guarantee a fair comparison. In addition, the ResNet model is trained entirely from scratch, without the use of any pre-existing models. As demonstrated by (He et al.; 2015), models with several hidden layers may be trained via residual learning. The following describes what a ResNet residual block is, y=$F(x, W + x)$. For a given layer representation (x, y), a residual map function F may be calculated. The residual block on ResNet may be finished if the input and output data dimensions are the same. ResNet-18 networks also include multiple levels inside each ResNet block, with the first two layers using the convolution of size 77 and max-pooling of size 33 with a stride of 227, much like GoogleNet, (Aljuaid et al.; 2022). This investigation is focused on a single model that downscales images to a 96x96 grid and is built on ResNet18 architecture. ResNet's weights are initially set using ADAM and the default momentum values.

5 Implementation

The below section consists of the details of the implementation:

5.1 Computational Details

The thesis implementation code may be accessed at the following address: `https://github.com/taherpoonawala/self-supervised-learning`. PyTorch[2], Torchvision [3] is used during the construction of the model, to plot the graphs matplotlib library is used, while TensorBoard is the tool that is utilized for logging purposes. Each and every experiment is carried out using a collection of machines known as the Elastic Compute Cloud(EC2) Resource, which is offered by Amazon Web Service(AWS). The overview of the AWS-EC2 configuration is 48 vCPUs and 4 x NVIDIA V100-SXM2 GPUs of 16 GB memory each per compute node.

5.2 Contrastive Learning Data Augmentation

A data loading strategy is developed, wherein random augmentations are introduced to each sample in the dataset, to assist efficient training. The quickest and easiest method to produce this effect is the augmentation pipeline which generates two distinct views of the same image. However, for the most effective instruction, just two are often required. The next stage is to decide which changes will be implemented. Refer to figure 8 for an illustration of the many types of augmentation that may be given to an image. The human brain can distinguish or differentiate between the foreground and background of an image on the fly. It is also possible to identify a given object in any orientation. Similarly, Augmentation is critical in making an SSL model robust in generalizing properties of input images. The authors (Chen, Kornblith, Norouzi and Hinton; 2020) demonstrated the use of augmentation using the dog image refer figure 8 and its benefits in assisting the system in learning more about the image. When the image is cropped and resized, only the tail is visible, necessitating a flip to comprehend the entire image. If this image is fed into the system, the background color has more influence than the object in question, so color distortion is required. Adding noise and blur to make the model more resistant to such images. Additionally, rotating the image at a random angle improves model learning.

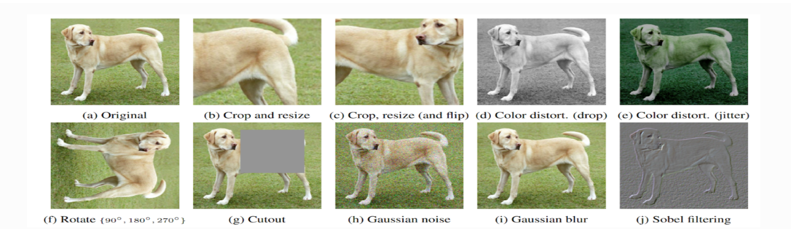

Figure 8: Augmentation Pipeline Output: figure by (Chen, Kornblith, Norouzi and Hinton; 2020)

While they may all be beneficial, the ability to resize and crop images, as well as alter their colors, stands out as especially powerful tools. However, as per (Chen, Korn-

[3]`https://aws.amazon.com/pytorch/`
[3]`https://pytorch.org/vision/stable/index.html`

blith, Norouzi and Hinton; 2020), they are only effective when used together. One can distinguish between

1. an image A that has been randomly cropped and resized to display a nearby area of (b) an image B that has also been randomly cropped and scaled,
2. two images C and D that have been arbitrarily cropped and enlarged to display surrounding areas of the same image, refer figure 9

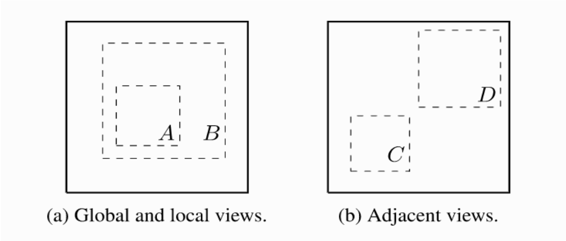

Figure 9: Cropping and Resizing: figure by (Chen, Kornblith, Norouzi and Hinton; 2020)

In case 1, the model just has to learn how to make crops to make them identical. In case 2, on the other hand, the model needs to learn how to understand the image which has been cropped out is the most difficult task. On the other hand, the model could make advantage of a defect without causing any color distortion. This imperfection is the fact that diverse image cropping tends to seem quite similar in terms of color space. It is not easy to discern that the two patches are from the same image just by looking at how close the colors of the fur and the background are to one another. The backdrop has a greenish tone, refer to figure 8. In such a case, the model may wind up paying too much attention to the color histograms of the images while paying too little attention to other parameters that may be used in a broader sense. Any change in the cut out of any image is isolated with the transformation the model will not be able to learn the representation. As a result, the model can only match two patches since it must first acquire generalizable representations before it can do so. This is produced by a procedure that mixes random cropping with color distortions.

The images used in this research were randomly rotated 90 degrees, cropped, their colors distorted, converted to grayscale, and blurred using a gaussian filter. The research used a dataset with substantial fluctuations in brightness and showed that using this design led to better performance, shorter training time, and higher stability.

5.3 Importing the Dataset

According to the criteria described in the section 3, the dataset consists of both carcinogenic and non-carcinogenic examples of image data types. The collection has 500 annotated images for each class and the images themselves are of a higher resolution (in terms of pixels). Additionally, it includes an augmented collection of tagged IDC histopathology images that are visually comparable to the training images. Thus, the dataset is ideal for illustrating the merits of SSL. Note that additional space is needed to store this dataset since it is larger and has a somewhat higher resolution (7.7 GB). This study generates two data loaders using the aforementioned contrastive transformations for use in the first discussion of SSL and SimCLR; one will be used to train the model on unlabeled data using contrastive learning, and the other will be used as a validation set for that model. After importing the dataset, the outcomes of data augmentation, such as arbitrary cropping, grayscaling, gaussian blur, and color distortion is demonstrated in the snapshot, as shown in figure 10. As a result, the model continues to struggle with the task of matching two independently enhanced parts of the same image.

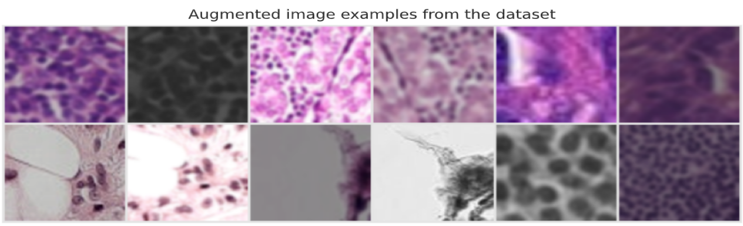

Figure 10: Augmented Cancer Patches

5.4 SimCLR Implementation

Now that the data-loading pipeline is in place, the next step is to build SimCLR. For each input image x, the data loader generates two output images, x_i and x_j, as shown in figure 3, which are changed differently with each iteration. The objective is to maximize the similarity between these two images and the rest of the batch, which has been represented by a one-dimensional feature vector. The encoder system is made up of a base encoder network, which is denoted by *f(.)*, and a projection head, which is denoted by *g(.)*. The output of ResNet-18, a well-known neural network design, and is used as the basis for the subsequent tests, and its equation $f(\bar{x}_\mathrm{i}) = h_\mathrm{i}$ will serve as the guiding principle. In order to determine the degree of similarity between two vectors by use of the contrastive loss, it projects the representation h into space by making use of the projection head *g(.)*. The article (Chen, Kornblith, Norouzi and Hinton; 2020), is consistent with the study's definition of SimCLR as a ReLU activation with two-layer MLP in the hidden layer. After the contrastive learning training is complete, the projection head *g(.)* may be taken off and the system can be used as a pre-trained feature extractor *f(.)*. It has been proven that representations generated by the projection head *g(.)* are inferior to those generated by the base network during training. Due to the invariance of the learned representations to factors such as hue, which may prove crucial in later challenges. Thus, *g(.)* is not required at any point in the learning process. Cosines may have a similarity between +1 and -1, with +1 being the most common. Due to this limitation, it is expected that z_i and z_j should be different as the CosSim will lead to convergence between(0,-1).

5.4.1 Training SimCLR

After integrating SimCLR and configuring the data-loading pipeline, the model must be trained. Training will continue with the same set of training functions as before. This measure is recommended for picking the best model since it is generally less noisy than the top-1 metric. To have the model learn from the representations more precisely *batch_ size* should be chosen ideally. When the batch size is increased, each image may be compared against more negative samples, resulting in flatter loss curves. Nonetheless, 256 batch sizes were found to be enough for this application.

5.5 Logistic Regression

The performance of the models tends to improve with larger batch sizes in contrastive learning. Each image may be compared to more negative samples when the batch size is larger, which leads to flatter loss curves. However, it was discovered that a batch size of 256, with a learning rate of 5e-4, the epoch was set to 500 since the loss value showed a consistent value. The h should be represented in such a way that it must explain characteristics for the job if the model is to perform effectively, regardless of whether the training strategy modifies the underlying network *f(.)* or not. Also, there will be only very few factors in the training process that overfitting should not be a problem. That being the case, it stands to reason that the model will function well with little input. In the first step, using a Logistic Regression, which makes advantage of the fact that assuming the image feature vectors already exist. In the event of data scarcity, at the time of training, the task of encoding images should be performed in order to undertake data augmentations. The method used here, however, is much more efficient and easy to teach. Not only did data upgrades not help much in this restricted setting, but they really hurt things. The trials will use even smaller datasets than the training dataset, which only has 500 labeled images per class. In this research, we examine the effects of different training dataset sizes on a Logistic Regression model (10, 20, 50, 100, 200, and 500 samples per class). Some insight into the transferability of contrastive learning representations to this particular classification issue in image recognition is provided by this finding.

5.6 Baseline: ResNet18

To compare the SSL with the ResNet18 model, a similar augmentation pipeline should be developed and applied on the dataset. The findings will show how using contrastive learning on unlabeled data might be more beneficial than using just supervised training. Since the touchvision library provides the ResNet architecture, putting the model into action is simple. ResNet overfits easily to the number of parameters allocated in the model due to the huge metadata being generated(Aljuaid et al.; 2022). Use data augmentations similar to what was done in the contrastive learning technique in the previous section, so that it can be competently contrasted with contrastive learning models. The color distribution of an image is an established factor in classification, therefore conventional color distortions are avoided. Accordingly, it seems that augmentation does not show any improvement in the model performance. The reason for this difference is that in contrastive learning, it was sufficient to establish whether two patches were from the same image, but in classification, the model had to recognize the whole object. As a consequence, subsequent improvements sometimes prove less effective than they would have been in a contrastive learning setting. Training ResNet is quite similar to training a standard Logistic Regression model. You may recall that initially ResNet was allowed to check for overfitting just once every two iterations through validation.

6 Evaluation

The following experiments were conducted, and the outcomes were analyzed using the assessment metrics stated in subsection 3.5. This section contains a comprehensive analysis of all experiment outcomes.

6.1 Experiment 1

The experiment aims to model a fully supervised ResNet18 model and a self-supervised SimCLR model to predict the IDC on histopathological images without any image augmentation.

6.1.1 Model

This experiment compares the results of the SimCLR model and contrastive learning without any augmentation to the baseline supervised ResNet18 model, offering insight into the relevance of augmentation in the context of image processing. In this experiment, the SimCLR model is given an unaltered dataset to analyze, and a linear logistic regression is used to classify the images, with the SimCLR model acting as the baseline against which the experiment's outcomes may be assessed.

6.1.2 Result

The SimCLR model performed at a MAPE of 0.57% and accuracy of 71.80% without any additional features. The Mean absolute percentage error (MAPE) for the baseline

Table 1: SimCLR and ResNet18 Model Output

Model Name	Class	Precision	Recall	*f1*_score
SimCLR	0	0.713	0.7308	0.721
	1	0.723	0.706	0.715
ResNet18	0	0.655	0.895	0.757
	1	0.834	0.529	0.648

model is 1.50%, which corresponds to an accuracy of 75.34%. While ResNet18 is 5% more accurate than SSL, the confusion matrix paints a different image of the model's efficacy. Recall, Precision, and *f1*_score statistics are also shown in table 1. Figure 11 shows a loss curve, also both graphs show a dip in the accuracy and a rise in loss which denotes that the model is unable to grasp the unaugmented dataset. While the most reliable projections for each coupling style are shown in figure 12.

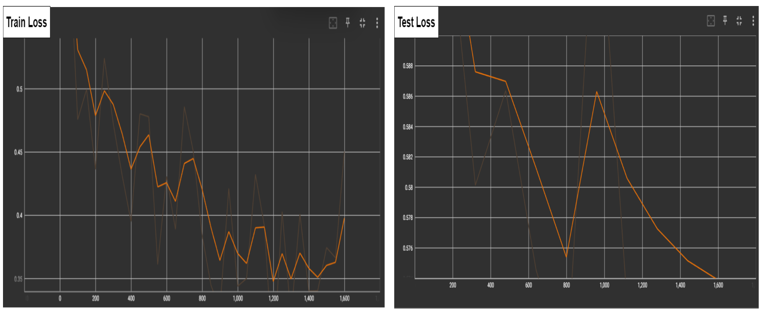

Figure 11: Loss Comparison: SimCLR

Figure 13 contrasts expected and achieved values, the accuracy tends to be exponential as the model tends to extract much information from the raw data. While confusion

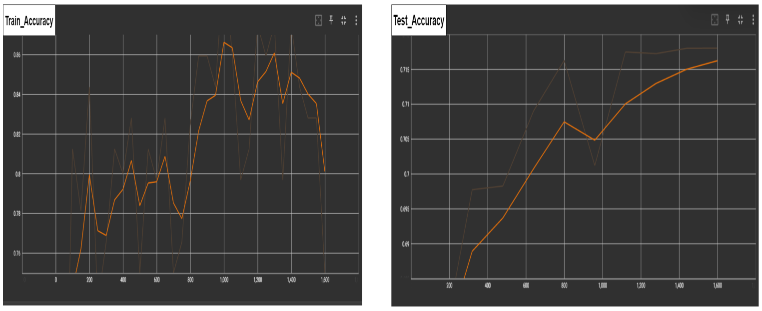

Figure 12: Accuracy Comparison: SimCLR

matrix comparison findings are shown in figure 14. Since the precision is high for the SimCLR model as compared with ResNet18, it denotes a low false positive rate, also similar analytics follow for recall value.

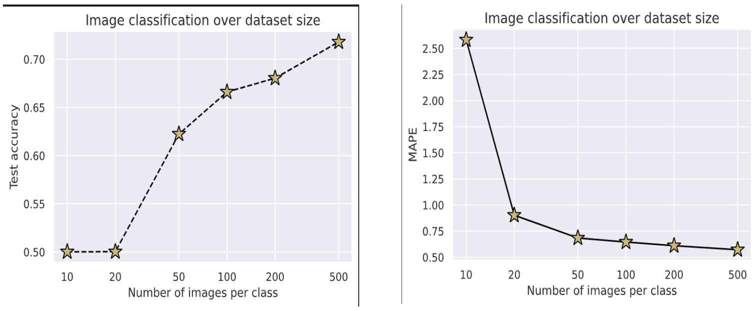

Figure 13: Output of Self-Supervised Model

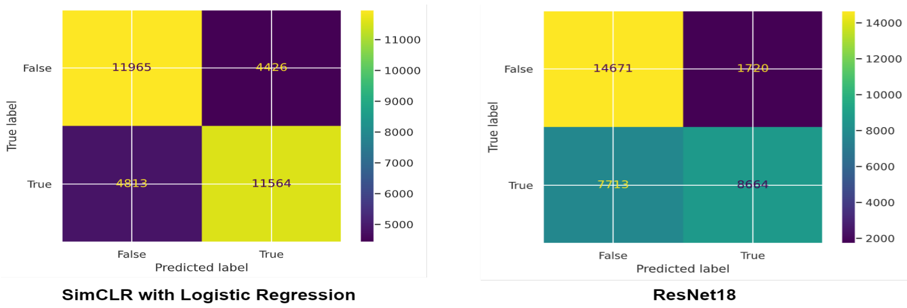

Figure 14: Confusion Matrix of Self-Supervised and ResNet18 Model

6.2 Experiment 2

The experiment's goal is to train a fully trained ResNet18 model and a self-supervised SimCLR model to predict the IDC on histopathology images without using mix augmentations such as random horizontal flip, random resizing crop, random color jitter, random grayscale, and gaussian blur.

6.2.1 Model

This experiment sheds light on the necessity of augmentation in the context of image processing by comparing the outcomes of the SimCLR model and contrastive learning with mixed augmentation to the baseline supervised ResNet18 model. Linear logistic regression is utilized to categorize the images in this experiment, with the SimCLR model serving as a reference point against which the results are evaluated.

6.2.2 Result

The SimCLR model performed at a MAPE of 0.36% and accuracy of 84.97% without any additional features. The Mean absolute percentage error (MAPE) for the baseline model is 1.50%, which corresponds to an accuracy of 75.34%. Recall, Precision, and *f1* _score statistics are also shown in table 2. Figure 15 shows a loss curve, the model shows a trend of less loss value and high accuracy, which denotes that augmentation has helped the model to learn from the unlabeled dataset. While the most reliable projections for each coupling style are shown in figure 16.

Table 2: SimCLR and ResNet18 Model Output

Model Name	Class	Precision	Recall	*f1* _score
SimCLR	0	0.819	0.898	0.857
	1	0.887	0.802	0.842
ResNet18	0	0.680	0.956	0.795
	1	0.926	0.551	0.691

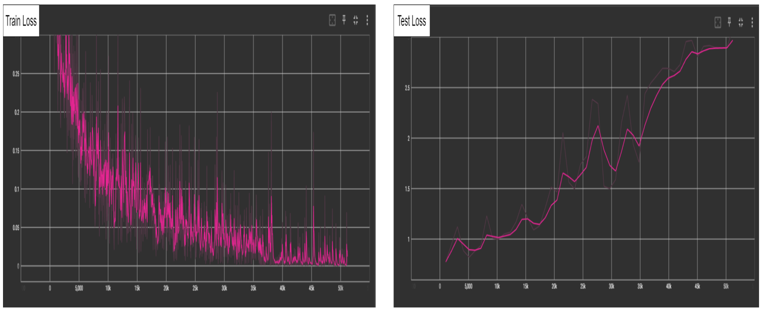

Figure 15: Loss Comparison: SimCLR

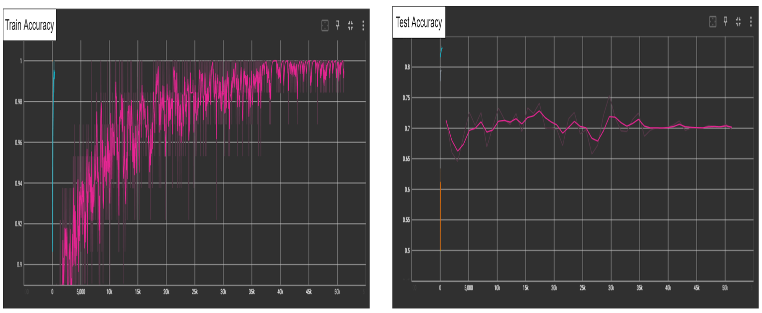

Figure 16: Accuracy Comparison: SimCLR

Figure 17 contrasts expected and achieved values. Confusion matrix comparison findings are shown in figure 18. SimCLR and logistic regression showed an improvement of 9.97% performing augmentation over baseline model ResNet18, the confusion matrix displays a different image of the model's efficacy. It's clear that SimCLR has a low false

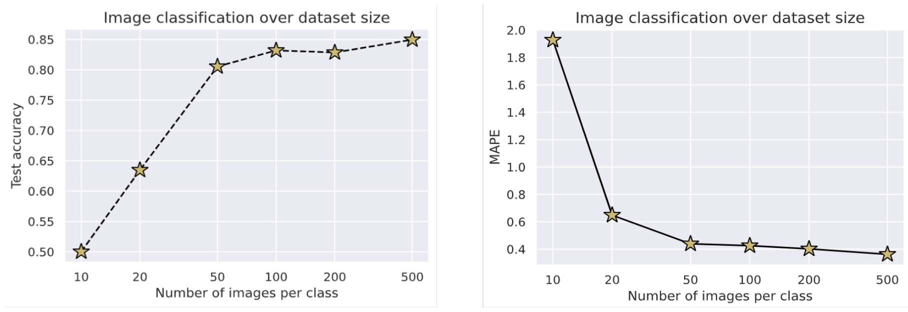

Figure 17: Output of Self-Supervised Model

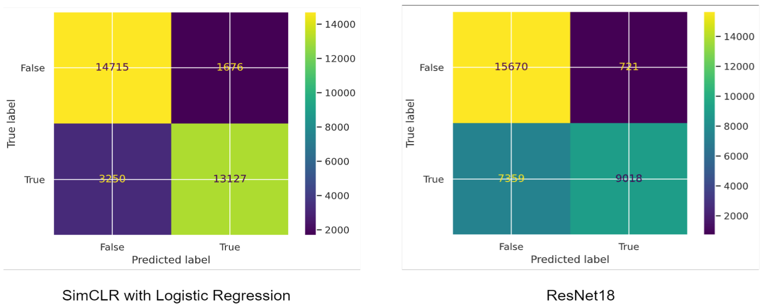

Figure 18: Confusion Matrix of Self-Supervised and ResNet18 Model

positive rate since its precision is so much higher than ResNet18's, and the same logic applies to its recall.

6.3 Discussion

In section 6.1, it was tested whether randomized augmentation techniques are preferable to predetermined ones. The accuracy of the learned representation was evaluated using a linear test of the classifier's ability to sort physically-present sample classes. Random augmentation strategy improved accuracy by 13% compared to using a fixed augmentation strategy, as this was implemented by (Spanhol et al.; 2016). This is because local information may be lost at the fixed borders while processing patches in fixed augmentation, but the random crop in random augmentation encompasses the whole image resolution. Next, a baseline was established, and methods for randomly cropping the image were developed. Better model performance by 15% was achieved by focusing on the importance of understanding how uncertainty impacts the quality of the mentioned representation. Higher the cropping level, the worse the quality of the image. A useful learning signal is achieved by optimizing for consensus amongst viewpoints that do not overlap.

The enhancement in feature representations obtained by various cropping strategies is compared. In addition to the upstream process of linear evaluation for physical sample-wise classification, the downstream task of label count per class is performed. When compared to its companion category, this one clearly shows a 15% improvement in representational quality as compared to the accuracy achieved by (Khaliliboroujeni et al.; 2022). Even though multi-instance generally outperforms, a random cropping strategy is ideal in the absence of label data. With enough time and data, multiple projector heads may be used to train the model with various cropping configurations and then evaluate its performance.

In section 6.2, studied how different pipeline enhancement techniques affected results. For this, it was adjusted to the range of possible augmentations from 0% to 100%, with a 20% spread. Surprisingly, it was found that the cropping and flipping augmentation instance technique produced the best representations as compared to when augmentations were disabled, but the mix augmentation strategy required augmentation for optimal performance. Formerly, it was thought that the mix augmentation method and the standard SimCLR technique were more crucial to multi-instances than the augmentation function itself. For this reason, it seems sensible to reevaluate the role of augmentation, even if doing so thoroughly was out of the question for this thesis. To learn more about the impact of hyper-parameters, researchers looked into determining the optimal projector and crop size for pretraining in SimCLR. Important as it may be to learn a similar feature representation for a smaller crop as the default size of (96,96). The ideal representation for the projector's output size may be learned with a vector size of (1,256), however, this is only slightly better than the other options. Therefore, it is unable to ascertain whether or not the output of a projector is significant. It has been seen that after adding the augmentation to the images there is an improvement of 15% accuracy from the previous experiment and also there's an improvement of 5% accuracy compared to the output of (Arya and Saha; 2021). Hence the accuracy of the SimCLR is 84.97% and the baseline model that is ResNet18 is 75.34%. When compared to the work of (Roth et al.; 2016), these metrics show a 3% drop in accuracy, but a significant reduction in false positive rates.

After carefully evaluating the developed methodologies and assessing the impact of various augmentations and hyper-parameters, it is crucial to determine whether the learned representations aided in predicting the regression value of weariness lifespan in IDC's histopathology dataset. This statement's veracity is measured by the regression value prediction error. Because the MAPE error is still more than the constant projection, it seems reasonable to infer that models cannot estimate the weariness value. It may not be feasible, given the current state of representation and pretrained models, to compare how well it performs on the physical sample categorization, or how well it performs on the lifetime prediction. The two main causes are: Similar to the multi-instance classification problem (a) the physical sample classification problem and (b) a classification scheme based on the labels is then constructed. Therefore, this assessment lacks the objectivity necessary for serious consideration. The accuracy of the microstructure is crucial to the prediction model's weariness. Because the fatigue loads imposed on a material also depend on its mechanical properties at the microscopic level.

7 Conclusion and Future Work

This thesis modified the SimCLR method to enable SSL in order to improve feature representation in IDC histopathology images. Specifically, augmentation should be used as it has proven advantageous. The findings demonstrate superior performance on downstream classification tasks and a significant increase in accuracy compared to the gold standard technique. Using just one-tenth of the labeled data, the model outperformed the supervised model with 500 images per label by 8%. The study not only assessed the uncertainty over unlabeled datasets but also analyzed the augmentation process and hyper-parameters. Along with the described hyperparameters, model size seems to have a role in contrastive learning. It is possible that bigger models, given a substantial quantity of unlabeled data, may be able to provide outcomes that are very close to their supervised baselines when the augmentations pipeline is disabled. The learned representations benefit from the stochastic component of the augmentation process, although the augmentation effect is small when using the multi-instance approach, calling for more investigation. Finally, the study checked whether representations gained by different augmentation techniques may be used to forecast model weariness. Compared to the standard supervised method, model weariness has a much lower error rate, but it's still rather sizable.

The future work of this research work can be to assess the effect of a more robust model, such as ResNet34 or ResNet50. More parameters may be required for the downstream model's regression result to have less room for error. The single projector head atop the encoder used in the current research can be replaced by two alternatives: using multiple projectors to give the positive and training the encoder jointly by averaging the loss from multiple projector outputs. It is expected that with continuous development, overall performance would increase. Simply said, the method described in this dissertation offers a significant development in the area of unsupervised learning for medical datasets.

Acknowledgement

I'd want to express my gratitude to Professor Jorge Basilio for all the time and effort he put into clearly outlining the scope and approach of this research. I'd like to show my appreciation to the National College of Ireland for this opportunity.

References

Aljuaid, H., Alturki, N., Alsubaie, N., Cavallaro, L. and Liotta, A. (2022). Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning, *Computer Methods and Programs in Biomedicine* **223**: 106951.
URL: *https://www.sciencedirect.com/science/article/pii/S0169260722003339*

Arya, N. and Saha, S. (2021). Multi-modal advanced deep learning architectures for breast cancer survival prediction, *Knowledge-Based Systems* **221**: 106965.

Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., Loh, A., Karthikesalingam, A., Kornblith, S., Chen, T., Natarajan, V. and Norouzi, M. (2021). Big self-supervised models advance medical image classification, *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3458–3468.

Benbrahim, H. and Behloul, A. (2021). Fine-tuned xception for image classification on tiny imagenet, *2021 International Conference on Artificial Intelligence for Cyber Security Systems and Privacy (AI-CSP)*, pp. 1–4.

Chen, T., Kornblith, S., Norouzi, M. and Hinton, G. (2020). A simple framework for contrastive learning of visual representations.

Chen, T., Kornblith, S., Swersky, K., Norouzi, M. and Hinton, G. (2020). Big self-supervised models are strong semi-supervised learners.
URL: *https://arxiv.org/abs/2006.10029*

Cruz-Roa, A., Gilmore, H., Basavanhally, A., Feldman, M., Ganesan, S., Shih, N., Tomaszewski, J., Madabhushi, A. and González, F. (2018). High-throughput adaptive sampling for whole-slide histopathology image analysis (hashi) via convolutional neural networks: Application to invasive breast cancer detection, *PLOS ONE* **13**(5): 1–23.

Diao, J. A., Wang, J. K., Chui, W. F., Mountain, V., Gullapally, S. C., Srinivasan, R., Mitchell, R. N., Glass, B., Hoffman, S., Rao, S. K., Maheshwari, C., Lahiri, A., Prakash, A., McLoughlin, R., Kerner, J. K., Resnick, M. B., Montalto, M. C., Khosla, A., Wapinski, I. N., Beck, A. H., Elliott, H. L. and Taylor-Weiner, A. (2021). Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes, *Nature Communications* **12**(1): 1613.

Ghesu, F. C., Georgescu, B., Mansoor, A., Yoo, Y., Neumann, D., Patel, P., Vishwanath, R. S., Balter, J. M., Cao, Y., Grbic, S. and Comaniciu, D. (2022). Self-supervised learning from 100 million medical images.

He, K., Fan, H., Wu, Y., Xie, S. and Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning, *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9726–9735.

He, K., Zhang, X., Ren, S. and Sun, J. (2015). Deep residual learning for image recognition.

Hirra, I., Ahmad, M., Hussain, A., Ashraf, M. U., Saeed, I. A., Qadri, S. F., Alghamdi, A. M. and Alfakeeh, A. S. (2021). Breast cancer classification from histopathological images using patch-based deep learning modeling, *IEEE Access* **9**: 24273–24287.

Islam, M. M., Iqbal, H., Haque, M. R. and Hasan, M. K. (2017). Prediction of breast cancer using support vector machine and k-nearest neighbors, *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, pp. 226–229.

Khaliliboroujeni, S., He, X., Jia, W. and Amirgholipour, S. (2022). End-to-end metastasis detection of breast cancer from histopathology whole slide images, *Computerized Medical Imaging and Graphics* **102**: 102136.
URL: *https://www.sciencedirect.com/science/article/pii/S0895611122001069*

Oord, A. v. d., Li, Y. and Vinyals, O. (2018). Representation learning with contrastive predictive coding.
URL: *https://arxiv.org/abs/1807.03748*

Roth, H. R., Lu, L., Liu, J., Yao, J., Seff, A., Cherry, K., Kim, L. and Summers, R. M. (2016). Improving computer-aided detection using convolutional neural networks and random view aggregation, *IEEE Transactions on Medical Imaging* **35**(5): 1170–1181.

Samala, R. K., Chan, H.-P., Hadjiiski, L., Helvie, M. A., Richter, C. D. and Cha, K. H. (2019). Breast cancer diagnosis in digital breast tomosynthesis: Effects of training sample size on multi-stage transfer learning using deep neural nets, *IEEE Transactions on Medical Imaging* **38**(3): 686–696.

Siegel, R. L., Miller, K. D., Fedewa, S. A., Ahnen, D. J., Meester, R. G. S., Barzi, A. and Jemal, A. (2017). Colorectal cancer statistics, 2017, *CA: A Cancer Journal for Clinicians* **67**(3): 177–193.

Singh, S., Prasad, J., Prasad, S. and Naik, S. (2021). Breast cancer prediction using supervised machine learning techniques, *2021 IEEE 18th India Council International Conference (INDICON)*, pp. 1–5.

Spanhol, F. A., Oliveira, L. S., Petitjean, C. and Heutte, L. (2016). Breast cancer histopathological image classification using convolutional neural networks, *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 2560–2567.

Thongkam, J., Xu, G. and Zhang, Y. (2008). Adaboost algorithm with random forests for predicting breast cancer survivability, *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 3062–3069.

Yang, J., Ju, J., Guo, L., Ji, B., Shi, S., Yang, Z., Gao, S., Yuan, X., Tian, G., Liang, Y. and Yuan, P. (2022). Prediction of her2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning, *Computational and Structural Biotechnology Journal* **20**: 333–342.

Zhuang, L. and Guan, Y. (2017). Image enhancement via subimage histogram equalization based on mean and variance, *Computational Intelligence and Neuroscience* **2017**: 1–12.