

Detection of Cancer using Biopsy Images and Deep Neural Networks

Dwayne Francis D costa

Supervised by Prof Mahdi Maktab Dar Oghaz



A THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE MSc IN ARTIFICIAL INTELLIGENCE AND BIG DATA

SCHOOL OF COMPUTING AND INFORMATION SCIENCE
ANGLIA RUSKIN UNIVERSITY

05/05/2023

Declaration

I, **Dwayne Francis D costa**, declare that the work in this dissertation titled *“Detection of Cancer using Biopsy Images and Deep Neural Networks”* is carried out by me. This work has not been submitted to Anglia Ruskin University or any other educational institution for the award of a degree or educational qualification. I also declare that the information published in this dissertation has been obtained and presented in accordance with academic rules and ethical conduct. Any information obtained from other sources has been properly referenced.

Acknowledgement

This thesis work is dedicated to my mother, Dr. Isabel Dennis, who has been a constant source of support and encouragement during the challenges of graduate school and life. I am truly thankful for having you in my life.

Table of Contents

1	Introduction	1
1.1	Overview	1
1.2	Problem Background	2
1.3	Research Aim	3
1.4	Research Objectives	3
1.5	Contribution	4
2	Literature Review	5
3	Methodology	24
3.1	Overview	24
3.2	Research Framework	24
3.2.1	Data Acquisition	25
3.2.2	Data Processing	25
3.2.3	Model Training	26
3.2.4	Model Evaluation	27
3.2.5	Model Deployment	27
3.3	Dataset	27
3.4	Evaluation Metrics	27
4	Results and Discussions	29
4.1	Overview	29
4.2	Tabulation Summary	29
5	Conclusion	36
	Bibliography	39

List of Figures

3.1	Overall research framework	24
4.1	Graphical Representation of VGG16 performance	31
4.2	Graphical Representation of ResNet50 performance	32
4.3	Graphical Representation of InceptionV3 performance	33
4.4	Graphical Representation of EfficientNet performance	34
4.5	Tabulation of accuracies and losses across all models	35

Abstract

Cancer in layman terms is often defined as a condition that is brought about due to the chronic inflammation of tissues which are mostly concentrated in the regions surrounding the lymph nodes. If cancer is left to its own devices, it will fester overtime and spread to other parts of the human body. This is mainly brought on by the cancer hitching a ride on the infected individual's blood stream. This would in due course of time lead to metastasis if left unattended and would often result in damaged DNA. It has often been observed that cancer tends to get overlooked when patients visit a health centre for a regular health checkup due to the fact that the underlying symptoms of cancer often don't make themselves known. It has been roughly estimated by the W.H.O. that in the year 2020, cancer alone had amounted to about 10,000,000 deaths all across the world. Patients who have a high risk of getting cancer often reside in regions/countries such as North America, Australia, New Zealand, United Kingdom, France, Belgium, Norway and other Scandinavian countries. It has been observed that at times, even after a cancer survivor has received medication, the cancer tends to recur in that individual. It is during this time that many of the underlying factors in fact do tend to make themselves visible. Solely due to the fact that cancer cannot be identified, there exists a void when it comes to methodologies or mechanisms that are constructed for the detection of cancer. The methodology for creating a mechanism in order to detect cancer according to this research entails the acquiring of a good dataset. This is followed by splitting the training dataset into 80% training and 20% validation where the validation dataset is meant for hyperparameter tuning. It should be noted that the image classification models that are utilised in this research are pre-trained in order to reduce the amount of time and resources which would otherwise be wasted on training them from the ground up. These pre-trained models are then trained one after the other on the PCam dataset in order to obtain the highest validation accuracy. The model with the highest validation accuracy is then tested with the test dataset in order to generate a prediction and finally the model is deployed. During testing, it has been found that due to the existence of imbalances in classes, the accuracy that is obtained during the training of the model cannot be used and

therefore the f1 score is chosen as a better evaluation metric. It should be duly noted that this technology would result in people getting the right dosage of medicine at the right time. This is done in order to prevent recurrence and with the future adoption of this mechanism there could be a predicted gradual decline in cancer cases over a period of time.

Chapter 1

Introduction

1.1 Overview

The human body is a multi-cellular organism consisting of a finite number of cells which split to form new cells whenever there is either an injury to the tissue or it is routinely needed to replace old and dead cells. *What is Cancer?* (2019) tends to spread awareness about cancer in layman terms by terming it as an anomaly which occurs when certain mutations tend to intrude and try to disrupt the natural flow of the process of cell division. Due to the occurrence of the mutation, cells accumulate one over the other in a crazed manner. This rabid behaviour often results in the development of a mass of unwanted cells, most commonly known as a “tumour” or “carcinoma”. Once a “tumour” is born it is often classified as one of two types, on the basis of spreading. If the “tumour” increases in size and spreads onto different locations of the human body, then it goes by the terminology of being “malignant”, whereas if it can only increase in size but not spread, then it is termed as “benign”. These curable growths are further subdivided into a total of four types, namely “carcinoma”, “sarcoma”, “leukaemia” and “lymphoma” respectively. It’s a well known fact amongst the medical community, that tumours tend to utilise an organism’s own blood vessels to transfer itself to another part of the body during the multiple phases of its growth and this process goes by the terminology of “metastasis”.

During the process of “metastasis”, the growths travel around the human body all the way from the bones to the brain, but most of the time they tend to concentrate around areas which tend to have a high amount of “lymph nodes”. These forks often tend to lend a hand in the quarrel against a variety of inflammations. Cancer for most of the time is often difficult to diagnose as most patients don’t tend to have any of the predating traits. In order to alleviate this problem, a variety of tests have been designed, namely the “colonoscopy test”, “mammography” and the “Pap test”

which correlate to the cancer occurring in the bowel, breasts and cervix respectively. Finally, it should be noted that the only procedure that is used to ascertain or prove the diagnosis is known as “biopsy”. This procedure entails a minuscule extraction of a tissue from the affected region for a much deeper analysis.

1.2 Problem Background

Although metastatic cancer utilises the bloodstream of the human body to transport itself all over and simultaneously attack regions that possess the highest concentration of lymph nodes, over time a few methodologies have come about through which metastatic cancer can be diagnosed. *What is Metastasis?* (2022) tends to state that a minority of individuals tend to have the predating symptoms of metastasis when they go for a regular cancer check up. Another way to check for metastasis is to have a cancer specialist search around for any “recurrences” during the period of post-cancer treatment. If the recurring cancer has already undergone “metastasis”, then this stage of cancer is known in terminology as “metastatic recurrence” or rather “distant recurrence”. This period of malignancy has a variety of predating traits which can be identified, and based upon the observation, the malignancy can be further characterised into four separate groups. These involve locations such as bones, the human brain, liver, and finally the lungs. The predating traits for the occurrence of malignancy in bones usually involves large amounts of pain, muscular spasms, painful splintering of bones as well as issues involving the viscera respectively. Similarly, the predating traits which would aid in the identification of malignancies in the brain entail painful migraines, convulsions, disorientation, hallucination, morning sickness, muscle spasms as well as giddiness. Likewise, the predating traits that support the identification of a malignancy in the liver entails loss of cravings, tiredness, delirium, “jaundice”, inflammation of the feet, etc. In equivalence to the aforementioned categorisations, the predating traits that provide support in the location of malignancy in the lungs entail extreme cases of wheezing, coronary infarction, issues with inhaling air, etc.

It is good to know that the condition of “metastasis” can be cured by simply notifying the point of origin of the cancer which is also inclusive of the extent of its spread, the concerned individual’s current health status, as well as their preferred choice of cure. It is also good to note that obtaining a cure for “metastasis” and for growth/cancers are entirely two different things. Treatment for cancers normally includes the employment of “chemotherapy” which exposes the human body to small amounts of radiation on a regular basis. Another form of treatment apart from radiation-based chemotherapy would involve surgery and is usually carried

out on a trial and error basis. Although surgery and chemotherapy does provide temporary mental relief as it tends to ease the tumour, this treatment is not full-proof or permanent. There could be a “recurrence”, but with no complete cure as the treatment only slows down the growth of the tumours. The success of the treatments often depends on a list of varying factors, namely type and extent to which the malignance has reached, the number of layers that exist in the tumour, its growth rate, the concerned individual’s desire of cure and finally their present health conditions and survivability rate, etc. Although these factors facilitate only temporary healing, a series of medical trials are being carried out by researchers in order to provide a lasting cure for post-cancer survivors. The main aim of this research revolves around creating a more affordable method to detect and predict the existence of cancers in a blood smear sample that occurs in the dataset in the form of image scans.

1.3 Research Aim

The main aim of this research revolves around the creation of a more affordable method to detect and predict the existence of cancers in a blood smear sample which occurs in the dataset in the form of image scans. The goal of utilising machine learning to aid with the prediction of existence of cancer in the blood smear sample further aids the cancer specialist and helps in reducing time taken to generate overall results during a regular cancer checkup. It helps the individual to know whether they have cancer or not, much faster, thus generating a peace of mind. Not all models that are used to predict the existence of cancers in this research achieve a high accuracy, but a few exceptions do. Achieving a high accuracy through these exceptional image classification models will aid in obtaining higher results, thus bettering the chances of identification of cancer in blood smear samples.

1.4 Research Objectives

- To research and identify a variety of methods that can be used to realise the outcomes of the research.
- To research and obtain either data or a fully-fledged dataset which correlates with the aim.
- To design, develop and deploy a chosen set of DNN models to be trained on the dataset in order to achieve the prediction of the existence of cancer in various pathological image scans that occur in the dataset.

- Evaluating the performance of the chosen models by plotting the accuracy and loss of each of the models and pitting them against each other in order to obtain the best possible validation accuracy.
- To test the recently evaluated model with the highest validation accuracy on the test dataset in order to predict a probability that the centre of a 32x32 pixel region of the scan of the pathological smear contains cancer tissue.

1.5 Contribution

This research just like any other cancer research *Early cancer diagnosis saves lives, cuts treatment costs* (2017) tends to improve the awareness and perception surrounding not only cancer, but also its many survivors. The research mentioned in this project is intended to be more of an early warning mechanism to people who have many of the predating symptoms of cancer and intend to go for a checkup. It works by incorporating machine learning and deep neural networks into the detection and prediction of cancer tissue from the blood smears that are obtained from various health centres and are merged together to form a dataset for training the models. This way it can inform the patients who have cancer in advance so that they can avail medical treatment without allowing for the further inflammation of the tumours.

Chapter 2

Literature Review

DNNs are indeed difficult to train mostly due to a few underlying problems, the primary one being issues concerning gradients. The usage and deployment of a few activation functions sometimes tend to lead to the disappearance of the gradients altogether. The research presented in the paper(He et al. 2016) confirms the notion that deep neural networks are in fact very difficult to train. Apart from this, it proposes an alternate method which attempts to facilitate a much more simpler architecture which is utilised to reduce the overall burden on the process of training the neural networks. It has also made sure to convert all the neural layers into remaining functions with the sole reason that in doing so one would be able to easily optimise said networks and attain the maximum possible accuracy. One of the 3 major datasets utilised in this research paper is the ImageNet dataset where remaining or left over networks are appraised up to a magnitude of at least 159 layers, but an ensemble of these leftover layers tends to achieve about a miniscule 3.57% error rate on this particular dataset. Apart from the ImageNet dataset, many others such as the CIFAR-10 have also been utilised but up to an increased level of magnitude. The main component of the research in this paper tends to revolve around image identification wherein the magnitude is of great importance. The final dataset that tends to flaunt its importance is the COCO visual identification dataset. Over the recent years, it has been proven time and again that deep neural networks are an integral part of machine learning and often help in advancements of certain fields such as visual categorizations. This is because it has become a known fact that these deep neural networks or DNNs tend to merge lower, mid and higher tier-based facets along with many classifiers in an end-to-end structure. Because of this, a confusion regarding whether learning neural networks are comparable to the addition of one-or-more layers one above the other does crop up. The major issue in dealing with this problem is definitely the case of the “exploding gradient”, which tends to reduce forking from its initiation. But overtime, this situation has been abetted

with the help of induction based regularisation. This would in turn aid countless layers in forking all for SGD in conjunction with back-propagation. Sometimes when neural networks begin forking, a new problem known as “degradation” often comes to light. This mainly occurs when the magnitude of the neural network keeps increasing thereby resulting in extreme levels of plunging of the accuracy. This would lead on to indicate that the process of optimisation isn’t favourable to all models. The research presented in this paper does offer some respite for the problem of “degradation” after all. The implication is directed to the fact that all the layers are in fact allowed to be fit to a mapping of leftovers. If the issue is moderate no longer, then it would be best to move the leftover to 0 instead of fitting it alongside a stack of non-horizontal layers.

The authors of this research paper (Radford et al. 2021) have proposed an effortless method to increase the efficiency in learning of the “SOTA” method which is used to learn image representation on an enormous dataset that is obtained legally through the internet. This particular method is studied after appropriately measuring it against the data obtained from a lump sum of datasets that mainly reference computer vision, wholly inclusive of a variety of tasks such as movement recognition, geo-constraintment, as well as a much granular form of object categorisation. The newly proposed model tends to pass on insignificance to a plethora of tasks and intends to be competitive without the underlying need to be trained. Such a well qualified method came into the works because it was often known that an existential access to average number of pre-trained models that exist within the clusters of web-scaling tend to outright surpass that of an elevated quality of NLP datasets. That being said, this doesn’t seem to be true in the case of computer vision as pre-trained models seem to be still following the age-old practice of being on crowd-labelled datasets which mainly include those such as the ImageNet dataset. After a thorough investigation it has been proven that there could exist a possibility to shift challenging assignments from a “pre-trained web-scaling” method from NLP onto another. By using this technique, one can obtain zero-shot which is a method of learning in the field of computer vision which can be in turn used to transfer the proposed methodology into a variety of already pre-existing datasets. At a nominal position, the performance of the models has been elevated although it might have a chance to be elevated in the near future.

The research that is presented in this paper (Gu et al. 2022) proposes a novel method for the generation of labelled occurrences which is done with the help of a paradigm that is used for pseudo labelling which is a method used for labelling a data model in order to predict unlabeled data. This newly proposed structure indicates that freshly distributed noise is somewhat relatively closer to human labels,

as in comparison to the process of random flipping which is solely dependent upon conditions in classes. Labelled noise is used as a controlled experiment in order to study the negative impact of labelled noise which is problematic in itself. Numerous pre-existing algorithms are additionally benchmarked to be utilised for learning along with the previously mentioned noise labels. Their behaviour can be further compared with the behaviour of learning algorithms on “synthetic datasets” as well as those datasets which have noise injected into them at any particular instance. The data that is generated using the newly proposed structure is then further utilised in order to develop a novel technique known as “Label Quality Model or LQM”, which functions by holding the features of the interpreter which aids in prediction and correction of numerous noisy labels. The main aim of “LQM” is to increase the model’s performance by being utilised to correct the various labels one level at a time preceding the application of already existing algorithms. The process of training machine learning models for a variety of applications are often noisy, are data hogs, are time consuming and often lead to an overall reduction in performance when it comes to training human generated models with labelled data. This has often led to these datasets being nicknamed as labelled noise, which by definition would indicate that they are error-prone and would most likely be impossible to be controlled, hence it would be difficult to study. As a result, most of the recent research is conducted via simulations. The research proposed by the authors tends to advocate for a new technique which isn’t random, can be executed promptly as well as possesses the ability to use all available data to generate clean data which is devoid of all labelled noise. This would enable individuals to obtain a much smaller dataset with cleaner data as compared to its original. Finally, it must be noted that along with the merits there also exist a few demerits. Most notable ones include the one being the existence of a variety of labelled noise which would in due course need their own unique kind of simulation, continuous maintenance of datasets which mostly include labelled noise would require an extremely fine tuned architecture as well as there being an understanding that possession of both noise and clean labels in the dataset is an underlying requirement that is necessary for LQM, which in reality isn’t the case for most of the applications.

The novel methodology that is proposed in this research paper(Huang et al. 2017) is intended to affirm the nature that CNNs tend to be more precise when it comes to training models than normal neural architectures. This happens due to the presence of miniscule links in between the many tiny layers which tend to appear to be much closer to the input and output layer. This notion is further proved right by the proposal of “DenseNet” which tries to link the input and output layers of the neural networks via a “feed forward” design. In order to calculate

the number of connections in DenseNet, it should be duly noted that all the facts of prior layers are intended to be utilised as specifics for the forthcoming ones. “DenseNet” has many pros over the traditional neural network architectures such as increasing the probability of reusing a variety of facets, reducing the total number of necessary pre-conditions, further reinforcement of the methodology of propagation of features, etc. After constant experimentation, it has been proved that the DenseNet architecture doesn’t eat up much resources during the training of models but enables some to obtain a peak performance. This is further tested by using pre-existing benchmarks which most commonly include the CIFAR-10, CIFAR-100, SVHN and ImageNet. One thing to be kept in mind is that unlike the ResNet architecture which does share some features, it isn’t entirely identical. The main difference being that “DenseNet” doesn’t merge all the facets via aggregation prior to sending them to the next layer. While performing the research mentioned in this paper, the researchers resorted to hyperparameter tuning in order to achieve a consistent hike in accuracy. It has further been predicted that in order for DenseNets to achieve a greater accuracy, there must be a more extensive hypertuning. Apart from all this, one should also understand that “DenseNets” have the ability to automatically fuse alongwith the specifications of mappings, inclusive depths, etc. This directly permits the reusability within the neural networks, thus generating highly precise models.

The researchers in this paper (Mu et al. 2022) have proposed a hypothesis on the implementation of an automated form of supervised learning that can be utilised to aid in learning through visual representation. This is because, up until now it has been proved that automated supervised learning tends to have an edge over normal supervised learning especially in the department of advancements. The model pre-training phase involves the official merger of both SLIP and CLIP. Post this process, an assessment is conducted which includes an examination of the quality of portrayal as well as the overall performance. This is then tallied alongside the CLIP and automated supervised learning under a parameter triad; i) zero-shot transfer, ii) linear classification, iii) end-to-end tuning. SLIP has been gauged by running it on the ImageNet dataset as well as a plethora of others. After continuous training using the pre-existing criterion, it has been noted that SLIP witnessed a hike in accuracy. Following this, there was a stringent validation process where a number of experiments were conducted pertaining to sizes of the model, different plans for model training as well as prior training of datasets. As a consequence of the aforementioned parameters, SLIP tends to have a duality in accuracy, one of zero-shot and the other of linear classification. After vigorously conducting a variety of experiments on ImageNet and other datasets, it is known that self-supervised

learning tends to be a rather give and take.

The researchers in this paper (Veeling et al. 2018) have proposed a new prototype that can be fully utilised for the purpose of decomposition of digital pathology, which is overtly based upon histopathological pictures, which are largely uniform in nature after the processes of gyration and backscattering have been carried out on them. The research carried out by the authors of this paper tends to put forth an evaluation of images of a dataset that consists of the progression of lymph nodes. They also intend to prove the amount of improvement in performance of the model in tumour detection. The final result is a derivative of the original dataset whose main job is to initiate a thorough differentiation between a variety of automated learning models which occurs in conjunction with predefined standards. This new derived dataset will further end up being a cornerstone for new benchmarks in the field of research with respect to machine learning. Over the recent decades, digital pathology has made many strides not only in the fields of development, but also in new advancements of hardware designed for the improvement of microscopic imaging. Another big achievement would be the complete digitisation of all earlier physical glass slides into images or WSIs. The completion of this procedure has in turn facilitated the automation of diagnostics with the help of algorithms which are specifically designed for visual analysis. A perfect answer to this would be the implementation of CNNs. This methodology has been proven effective in excelling a lot more than regular pathologists. This methodology basically includes run-of-the-mill CNN models which are a bit modified as it increases its effectiveness, this happens largely because of even distributions of all the conditions equally throughout all the layers. This is known as “translational equivariance” and it successfully takes advantage of rotational conformity which is one of the main traits in naturally occurring images. As a summarisation, the researcher-proposed methodology out-classes its predecessors which are none other than conventional CNN models. As a result of continuous experimentation, the final result is more or less a derivative of the original dataset which thereby permits extreme forms of evaluation. It further proves the effectiveness of “rotational equivariance” as it has drastically raised the overall dependability of the prototype in the field of medical visual analysis.

The researchers in this paper (Dubois et al. 2021) have proposed the need for the usage of augmentations of data in order to attain a high performance on all the predictions that are carried out by a particular group of changes. The researchers also believe that if they stick to the methodology that they have proposed, they can build upon it and lay down new objectives for training neural networks based on unsupervised learning. If the newly proposed objectives are used correctly, it would greatly increase the rate of savings wherever performance of models is concerned.

This paper gives the methodology of data augmentation a lot of spotlight due to the fact that around the world tons of data is getting trafficked and collected, and the only time it sees the light of day is when a model is being trained on a dataset within which it resides. Although this is technically the usual case, software solutions such as ZIP tend to maintain a loyal perceptuality rather than be interested in specific particulars. It has become a quite common phenomenon in the current years that there has been a gradual decrease in the availability of data for models to be trained on. The use cases tend to vary, mostly from climate prediction to autonomous driving, etc. All genuine “lossy” condensers are really not up to the mark because their main purpose of existence is reconstruction of data for human consumption although the majority of it isn’t needed as basic image classification needs only about 1 kb of storage. The main goal of the researchers in this paper is to consistently maintain a very high performance on the execution of prediction tasks.

The research presented by the authors in this paper(Cohen & Welling 2016) intended to propose a contemporary methodology of “Group Equivariant Convolutional Networks” which tends to exploit uniformity which in turn leads to the reduction of complexity of an appraised dataset. G-convolutions, which is a new layer that is often utilised by the proposed G-CNNs, tends to enjoy a few perks in comparison to traditional convolutional layers which mainly revolves around an elevated degree of weight sharing. These new convolutions are extremely easy to utilise as well as be able to be executed with zero mathematical costs and are often generated through various translations, rotations, etc. It tends to achieve an above average performance when it’s tested on MNIST and CIFAR-10 datasets respectively which are used to train machine learning models as well computer vision algorithms. Over due course of time, Deep CNNs have been proven to be tremendously sturdy models mainly consisting of snapshots, films as well as sound files. Despite there being a lack of design for a neural network, there exists a plethora of proof which affirms the fact that the sharing of weights as well as depth is of utmost importance when it comes to increasing performance. “Translation Symmetry” which is a term coined for the process of transferring an object from one place to another without any rotation of said object. This exists in a variety of tasks, further compliments the sharing of weights and thus makes them very successful. In contrast with a normal layer, a convolutional one tends to utilise identical weights to perform an analysis of images within the dataset. Alongside this, it also tends to preserve the age-old ability of training with a variety of changes. The concept of “translational equivariant” can be loosely defined as the process of dislocating a photograph and inserting it past a number of layer can be further equated to the process of insertion of the native photograph past a number of quantifiable identical sheets which is then

followed by a sudden change in the finalised “feature map”. In layman terms, this would imply that the structure is protected by each of the layers, thus giving it the allowance to be completely exploited. In conclusion, it would be apt to mention that this research paper proposes the introduction of the concept of G-CNNs which is none other than a hypothesis of CNNs which overtime would gradually raise the amount of conditions. Through the exploitation of the said structures, the G-CNNs tend to achieve the most favourable results whilst being trained on a pivoted MNIST dataset in conjunction with the CIFAR-10. After arduous amounts of experimentation, it has been duly proved that “G-convolutions” can be utilised as a replacement for dimensional complexities in the generation of new designs for networks, which in turns brings about a massive boost in performance.

The researchers in this paper(Graham et al. 2020) have proposed a new iteration of the CNN model known as DSF-CNN whose main purpose is to use G-CNNs along with a variety of continuously changing sieves which are basically none other than complexities that are stretched out in a straight line, most of which include manoeuvrable sieves which in turn help in the reduction of conditions that are put on training a model. The researchers have not only proven the merits of hard encoding of continuously changing frameworks into contemporary ones, but also depict a much more extensive differentiation for continuously changing but uniform CNN models which are utilised in histological clinical visual analysis. Over time, the research that has been carried out and penned down in this paper has proved that the new proposed model exhibits a much greater improvement in its performance levels especially since it has fewer conditions for training the model when it is run on three different datasets concerning histopathological visual analysis which are segmentations of breast tumours, classification of colon glands and the classification of multiple-tissue nuclei. It also should be noted that researchers in this paper decided to go forward with the propositions of DSF-CNN mainly due to the fact that contemporary CNNs tend to be “data hungry” and pick up facets during every positioning. Over the years, due to rigorous advancements in the field of clinical glass staining, there have been numerous strides in the applications of DNNs to achieve the result after completing a visual analysis of the glass stain, the most common one used being CNN. A sub-facet of CNN known as “translational equivariance” is one of the main properties of these images that helps for the easy deployment of CNN as it often tends to get rid of the ability of learning at the various points of a given object. The reason for the employment is additionally due to the fact that it can act as a counter to “translational equivariance” by not needing any learning capabilities to study any data to perform analysis from different sieve positions mostly due to the application of “CPath”. In retrospect, the research in this paper talks about

employment of CNN as it doesn't need to learn data from the dataset to perform analysis owing to the fact that "translational equivariance" is present on the image samples and is additionally a sub-facet of CNN itself. The researchers further prove that they are able to achieve the best performance levels across multiple datasets with a miniscule number of conditions. For future improvements, it also has been proposed that multiple uniformities will be augmented within the G-CNNs as it will lead on to raising the consistency of performance and give more impetus to clinical research-based statistical pathology.

The researchers in this paper(Lafarge et al. 2021) have proposed a novel structure of SE(2) grade to be utilised in CNNs which would in turn lead to the development of relocation as well as uniform revolution through the introduction of SE(2) which is none other than a "Euclidean motion group" that make sure that the results which are generated are constant beneath a very carefully picked collection of revolutions. It should be made clear that applications of clinical research have a most prized choice when it comes to analysis of clinical visuals using machine learning and is none other than the feature of constant revolutions. Traditional methodologies involving the concept of constant revolutions have the habit of continuously relying on enhancement of data, but this might not always be a good thing. This is because there exist no assurances that there will be any improvements in the final result once the revolutions have come to pass. On the other hand, traditional CNNs would instead require a lot of time and effort in order to reach their optimum levels of performance. The research that is presented in this paper tends to lay an emphasis on critical visual analysis of histopathological-based datasets not being caught by any of the several already existing machine learning models. The novel methodology that is proposed in this research paper tends to evaluate the final data that is obtained after continuous testing on three variants of histopathological datasets which include the discovery of mitosis, nuclei and tumour respectively. After the process of evaluation has been completed, a proper analysis of the tasks is generated along with visualisations which depict peak performance levels achieved across all three datasets. As mentioned before, constant revolutions is one of the most sought after details within the confines of machine learning. There is a possibility of utilising it to analyse a variety of histopathologic clinical visuals in order to further select and segment various modular structures, structures that occur within a particular range can be allowed to possess three-dimensional layouts, and the tissues can be removed surgically. If all of this is taken into consideration, then the final result changes with respect to the inputs which in return tends to generate a lot of ambiguity. It should be kept in mind that the final resulting output should always remain consistent while undergoing the process of rotation. A new structure known as the "trainable

convolutional network” kind of permits traditional CNNs to study the various facets in any form. One of the many ways to acquire a constant revolution is by the process of continuously improving the dataset through the creation of randomised samples. Although one of the many methodologies by which an individual can create constant revolution is through the creation and addition of data to the dataset, this process does not come with any warranties that would enable the traditional CNNs to start becoming revolutionary-constant. Additionally, with these types of approaches it has become a much common operation of aggregating all the predictions of a model that has already been trained in a given set at a particular time. This kind of approach has its own merits and demerits, being a peak in performance of the model that is used and the cost of resources that have been consumed during the processes respectively. The researchers in this paper have proposed the utilisation of “euclidean motion group” as a replacement to traditionally occurring CNNs. This novel framework tends to reminisce the fact that observations that have been studied are always uniform with the structures for the input when it comes to revolutions. Edging within a straight line on the kernels of the “euclidean motion group” tends to enable above and beyond performance of hard encoding at points that lie perpendicular to the straight line. In retrospect, the methodology of SE(2) which has been proposed by the researchers in this paper is done on the basis of it being the best option when it comes to the analysis of histopathological visuals. This novel architecture introduces the ability for recently trained models to become consistent with respect to the constant revolutions which seem to be part and parcel of translational structures for histopathological medical visuals. The researchers in this paper have tried to depict that G-CNNs which are empowered by the SE(2) architecture tend to yield a much better performance than traditional CNN models at the expense of resources consumed. The researchers have also maintained that any futuristic applications of this novel methodology would revolve around a much more deeper analysis of the connection between frameworks that are related to the tuning of hyperparameters and its influence on the performance on newly trained models.

The concepts of G-CNN have been introduced in order to improve upon traditional CNNs by providing the former versions with new frameworks. It has been proven that G-CNNs tend to utilise a new concept known as group theory in order to achieve the necessary upliftment of the mapping of a variety of features to peak level representations. Keeping this aside, one thing to be noted is that the real and quantifiable applications of the G-CNNs tend to be extremely concentrated only in small groups. The researchers in this paper(Bekkers 2019) intend to put an end to the demerits and put forth a novel methodology that is utilised for the creation

of new G-CNN-based “lie groups”. Through particular novel methodology, the research tends to make readers understand that lie groups are utilised to increase kernel methods in CNNs on a tiny format of B-Spline. The further effects of the newly proposed methodology and the standardisation that it received after it was trained on a couple of standardisation-based datasets such as one involving revelation of cancers revolving around “histopathology” within which revolutional-based equivariance does play a main part. This also is inclusive of knowing the location of an individual’s facial characteristics around different buildings in the city. This should be duly noted that frameworks that had been constructed on the basis of G-CNNs tend to best their two-dimensional equivalents. Over the years, it has officially been accepted that G-CNNs have made multiple improvements over the years. This research paper further talks about the proposal of a methodology sans restrictions during construction and deployment of “lie groups”. This new proposed methodology which is none other than another iteration of G-CNN depends solely upon the denotation of what “B-Spline” really is as it mostly involves augmentation as well as the evaluation of a complex set of kernels. This would result in the creation of a totally pliable structure of “B-Spline” as it aids with the construction of G-CNNs involving certain features that couldn’t be obtained by the normal “fourier series”. Two datasets namely the PCam and CelebA datasets are utilised in the detection of cancer and facial recognition near buildings in the city respectively. The only similarity that exists between these two is that the G-CNN-based models are outperforming their two-dimensional equivalents as stated before. In retrospect, the research that is penned down in this paper does aid in the construction of novel iterations of the erstwhile CNN model such as G-CNN while utilising “B-Splines” as the foundation stone. “B-Spline” is actually a truly magnificent way of portraying utterly complex kernel methods that can’t ever be achieved by the traditional fourier functions. These features mostly include ones such as current location, dark complexities as well as those kernels that are disfigured or deformed. Subsequent research on this novel methodology would usually entail the thorough searching of new ways by which novel applications can tend to benefit through uniform conditions that are observed across all the applications.

The researchers in this paper(Romero et al. 2020), have proposed a novel methodology known as “Attentive Group Equivariant Convolutions” which is basically none other than a sub-concept of G-CNNs, wherein the application of “attention” is only initiated during the course of process of contortion. This is done in order to focus on underlining the significance of uniformity groupings and reduce the significance of those that do not conform to this way. The researchers have made it known that previous studies that are carried out on image scrutiny can best be recounted as ex-

traordinary instances of the proposed methodology and be described by verification that the novel methodology continuously one-ups regular G-CNNs when it comes to standardisation of visuals in datasets. Apart from this, the researchers enable an easier understanding of a variety of concepts through the process of image conception of a plethora of mappings. Another fact to be understood is that in spite of G-CNNs being able to entirely understand tough visualisations which are designed upon unique structures, it should be noted that these neural networks in fact lack the basic ability to form bonds between one another. The recent advancements of a variety of machine learning algorithms including CNN have been a result of continuous experimentation with a variety of perceptions. Although this is true for all machine learning algorithms, CNNs are known to have the highest rate of performance. This unusually high rate of performance is CNNs's most redeeming factor as this machine learning algorithm possesses a plethora of downsides, most of which are associated with standardisation and it does need a vast amount of improvement in the fields of hiking the efficiencies of its training models. The concept of "equivariance" is considered the weakest link which can be utterly exploited since the model utilises this concept in order to draw inspiration from humans. This notion of "equivariance" is strongly intertwined with uniformity. Although changes tend to target objects in different locales, dimensions, perspectives as well as surroundings, as long as it doesn't overtly change the basis of the object, these changes should be treated as a singular concept. Over the years countless procedures have been identified and accepted in order to protect uniformities. A summarisation of this research paper would entail the introduction of G-CNNs which were used to call attention to uniform structures. After continuous experimentation, it has been duly noted that G-CNNs continuously and steadily beat traditional CNNs while providing an equivariant attention mapping system on the side which enables the model to depict all the data as visualisations which it has learned over the course of time.

Over the recent decades medical visual classification has made great strides especially with many advancements in the field of deep learning. The most desired methodologies are the ones that have capabilities for the handling of noise labels which may otherwise lead to overall reduction in classification model accuracy. As an answer to this need, the researchers of this paper(Liu et al. 2021) have proposed a novel methodology which involves a newly designed structure for visualisation of medical visual data which is christened as "co-correcting". The newly proposed concept of co-correcting has a singular main purpose which is none other than gradually raising the accuracy of classification models. Other than hiking accuracies, it can sometimes also include the finding of a variety of highly accurate noise labels, generation of estimates for labels as well as the correction of the labelling

for any sort of curriculum. This newly proposed methodology has been run on a couple of datasets namely the MNIST image classification dataset as well as the ISIC-Archive Dataset. Continuous experimentation involving at least six methods are carried out on this dataset and the final result reveals the best accuracy that a model could achieve along with an extremely high rate of generalisation. Clinical visual classification is one of the hottest trends in modern day medical fields as it involves the assimilation of the field of machine learning into medicine. A variety of machine learning methodologies mostly associated with deep learning have been employed in order to achieve unprecedented boosts in performance mostly due to the growing presence of a large number of datasets which are built for medical image classification. The large presence of medical image classification-based datasets does correlate to the employment of DNNs for the training of models on them. But since these datasets are extremely large, the collection of clinical visuals does raise a few red flags especially when it comes to the spheres of having to explain a variety of arduous tasks as well as the individual's privacy. There do exist a couple of ways by which one may acquire a lot of data for medical classification. Most notable would include rummaging through already existing medical reports which generally share a lot of correlation among each other as well as procuring pre-existing noise labels and finding new ways to boost their performances. The second method of collecting data would be the most suitable as datasets that are obtained through rummaging often tend to display the developments of high levels of noise labels and are therefore unsuitable to be utilised for classification. The research presented in this paper does bring to light the fact that high noise labels do corrupt the labels at the initial level which in due course of time does have a huge impact upon its accuracy. An efficient solution to this might be to train the model on a dataset which consists of a very selective set of data. The proposed novel methodology of co-correction tends to train a couple of neural networks at the same time. The process is rather simple as it includes tiny pieces of data in each dataset, and understands the amount of loss that can be generated in each iteration which sometimes leads to the fitting of labels with high noise. But as ill luck would have it, many times the samples often get overlooked mostly pertaining due to their accuracy and performance. LNL is another methodology that is used to interpret clinical visual taxonomy. There are stark contrasts that exist between clinical visual taxonomy and a much more generalised visual taxonomy. These mostly revolve around a few pointers such as clinical datasets being tiny, problems arising out of misclassification of certain visuals and the non-availability of methodologies that have been created with clinical applications in mind. In retrospect, it should be duly noted that the research that is penned down in this paper discusses proposing a novel methodology

of co-correction or co-teaching wherein a couple of models would be trained simultaneously on a dataset. This methodology when trained on the MNIST dataset which is a rather popular one when it comes to image classifications, does deliver a high end performance and accuracy.

“Tensor Networks” is the new focal point for the research in this paper (Selvan & Dam 2020), as it has been instrumental in the analysis of quantum-based systems. Apart from accepting that tensor-based neural networks work as mules in order to realise the goals of physicists around the world, the researchers have stressed on the need for familiarisation with MPS which finds its uses in clinical visual classification. The researchers intend to concentrate much of their focus upon the generation of customization of two-dimensional visuals which tend to utilise traditional classification methodologies. Utilising propositions for unsegmented LoTeNet models, the researchers do intend to depict that these “tensor networks” are more than capable of reaching peak performances which in itself can be likened to much more modern DNNs. The final phase would involve the evaluation of the resulting model on a couple of free and open-source datasets which are primarily designed for clinical visual categorization in conjunction with displaying the levels of rendition that is carried out. It has been made sure to make this phase very cost effective as a lot of trouble is taken to see that this process utilises very low amounts of computational resources. In recent decades, the concept of SVMs has been thoroughly popularised in machine learning since the onset of new concepts known as kernel methods which always intended to better the identification of guidelines. The newly encouraged concept of tensor neural networks can be an effective methodology in doing away with peak spaces of proportions. These tensor networks are none other than mathematical expressions of a variety of quantum-based states entirely dependent on their internal structure. Over the recent years, there has been a consistent rise in the usage of tensor neural networks in the field of machine learning mainly due to the fact that these methodologies tend to reconvert two-dimensional images into one-dimensional images before the process of hard encoding gets initiated. There does exist some amount of association when it comes to minuscule visuals in commonly occurring visual categorisation-based datasets such as MNIST, etc., which can be later utilised through the implementations of tensor neural networks. The researchers in this paper intend to implement methodologies of tensor neural networks when it comes to the categorisation of clinical visuals. Hence, they had decided to propose LoTeNet which can be used to perform categorisation in a straight line of peak performance-based proportions as well as tending to further optimise these proportions by back propagating to the tensor neural networks. In this paper, the final resulting models are trained on the “PCam dataset” which totally consists of histopathology-based

visuals and the “LIDC-IDRI dataset” which consists of CT scan visuals. The final model tends to liken itself to a deep learning model since there exist a very small amount of conditions and ends up using a very minute amount of computational resources in comparison to its traditional counterpart. In retrospect, the LoTeNet model deploys ranking system-based order wherein it tends to retain its former and core framework. The research also does shed light upon the various capabilities of the model, especially the one which pertains to performing categorisation on the two aforementioned datasets, produce a comparable or much greater levels of performance when it comes to deep learning models as well as the utilisation of minuscule amounts of computational resources.

The researchers in this paper(Kaczmarzyk et al. 2022) have proposed the implementation of ChampKit which is a novel methodology that can be utilised in widespread gathering of clinical images for classification purposes and the most important thing is that it can work for a variety of cancers. It also tends to record the new performance levels which are obtained after models have been improved and retrained. It is also important to note that “ChampKit” method’s datasets and its own source code is available to all, free of charge. The researchers in this paper decided to go forward with the introduction of this methodology as the already existing methods rendered by “histopathology” often does not focus upon any changes made by the generalisations of the tasks, but sticks solely to the age old criteria of utilising machine learning models for the optimisation of singular methods. “Histopathology” has rightfully earned the moniker of being considered as a “gold standard for cancer diagnosis” over the years. Histopathology has time and again been utilised in conjunction with other “computer vision” methodologies and as a result has risen to prominence with the new induction of “deep learning” concepts. The novel concept of deep learning has in turn paved the way for better understandings and the ability to prophesied the presence of tumours, detect whether these tumours have infiltrated lymphocytes as well as the detection of “microsatellite instability”. The ImageNet dataset which consists of large amounts of medical images used in clinical visual classification is the sole dataset that is utilised to train newly developed models. The newly proposed solution of ChampKit consists of three main phases, selection of the models and pre-training, the training of models on multiple scenarios and lastly the evaluation of obtained data respectively. The initial phase basically involves the selection of any model such as ResNet in conjunction with a pre-trained model such as ImageNet. This then followed by training the models chosen on a variety of tasks mostly involving the classification of tumours, detecting whether the said tumours have invaded the lymphocytes and the task of identification of microsatellite instability in the location of cancers in different parts of the body.

The final phase does involve the evaluation of the recently obtained data. The researchers have introduced ChampKit with the best of intentions as they feel that it can amplify the speed of intense research when it comes to the domains of “deep learning” and “histopathology” in pursuits of achieving new heights in the field of medicine and clinical research.

The research that is presented in this paper(Worrall & Welling 2019) proposes a novel generalisation of CNNs which is known as DSS, otherwise known as “Deep Scale Spaces”. The DSS methodology can be utilised for the exploitation of uniform layouts of normal visual taxonomic procedures. In layman terms, this would imply that the template of visuals is often inalienable from the perspectives which it is viewed from. The research that is penned down in this paper proposes to construct an augmentation to a traditional CNN which is normally known as “cross-correlation”. This procedure can be utilised on modern day DNN architecture as an autorun program. The researchers also tend to show the behaviour of their models after training them on the “Patch Camelyon” and “CityScapes” datasets respectively. This is done in order to make known the variety of uses as well as the process of carrying out experiments on them. In an ideal environment, the models that run on a variety of datasets should have an in-built understanding of proportions. The field of computer vision that falls under the umbrella of artificial intelligence does possess a few methodologies which help it evade the concept of proportions. These are known as aggregation of proportions, selection of proportions, and the expansion of proportions respectively. The new methodology of DSS has been implemented only because the earlier mentioned methodologies all lack the capability of fusing data of different proportions together. “Translational Symmetry” is utilised for classifying structures, detecting them as well as labelling them. This is done in order to reduce the number of conditions as well as keeping the uniformity of the layers below intact. Another function of it would be reduction of generalisation as well as building the various convolutions. In retrospect, the researchers in this paper have proposed a novel methodology called DSS which is utilised to generalise the traditional CNN networks. The new concept of DSS is widely based upon a “cross-correlation”.

In this paper(Pagliardini et al. 2022), the researchers have put forth the concept of D-BAT algorithm which is a much needed change as it has the potential capability to realise a consensus amongst models which are being trained using the training data while simultaneously being in opposition to the OOD. This is done mainly due to the fact that “gradient-based learning algorithms” tend to possess a miniscule amount of bias which hampers choices when it comes to picking of predictors. To add insult to injury, it can also affect the portability of models undergoing training by intending to favour simplistic data along with the attachment of a tiny remnant

of data which entirely consists of perceptive facets. The newly proposed concept of D-BAT has multiple new facets, most notable include reduction in procedures for learning, further inflating unpredictability as well as the ability to identify OOD. The miniscule bias that is generated by the gradient descent algorithm is known as “Simplicity Bias” which is the main reason behind a largely superior conceptualisation of resources. On the one hand, it is often known that DNNs don’t usually overfit, it has been proved through rigorous research that the bias is known to cause overfitting. The bias can be rather severe by forcing predictors to be simplistic in nature thereby drawing data only from the training dataset and excluding the testing dataset. The introduction of the D-BAT algorithm gives the chance to predictors to be more diverse over the choice of data as it mainly aids in the portability of ensembles. Future improvements in this particular field could entail the further development of sample choice from OOD and the construction of a much tougher hypothesis. This newly proposed algorithm also can find many applications which do not feature in this research paper, such as inconsistency detection or a general test for detection of the presence of any prejudices in any of the trained models.

The authors of this research paper(Wenger et al. 2020) intend to propose an entirely new procedure whose main job is the modification of the level of trustability of a commonly occurring classification method. This is done with the objective of mainly reducing the overall errors in classification of data. It is often seen as a counter to already existing methodologies as it tends to deploy a new form of tuning which is designed to cater to the needs of a classification that consists of mixed classes. It is to be duly noted that this new procedure is not only limited to neural networks, but also applicable in enabling the increase in trustability rates of a number of locally available classifiers. During the process of conducting a variety of experiments, it is often depicted that it has a consistently high performance rating across a plethora of classifiers and varying yardsticks, although this would most likely be in the case of neural networks. Machine learning has made a multitude of advancements in various fields such as computer vision, utilising NLP to recognise an individual’s face or their voice. Apart from this, there have been leaps in the designing of new architectures which are totally complicated and are used to train on exponentially huge datasets. These classification methodologies also find their application in self-driving cars where it is utilised to increase the safety of the rides. Credible unpredictability is one of the many ways which can aid in the surge of said classifier’s prediction capabilities. “Gaussian Processes” often abbreviated as GP are novel classification methodologies which tend to have the accuracy as well as computation power gradually regress although they tend to outmatch others when it comes to generating excellent estimations of unpredictability, which is way better

than the ones that are generated by CNNs as they often fall victim to widespread miscalibrations. The process of calibration tends to reflect the approximate validity of classifiers which is directly affiliated with the ultimate forecast being accurate. To summarise, the researchers have recommended a novel methodology for calibration of classifiers which is based upon a changeable deduction. These novel methodologies for calibration are often executed in applications involving fields such as computer vision and all the tests are conducted on the availability of pre-set conditions or standards. The new model also shows an immense rise in performance ratings when it is trained on a variety of datasets.

The researchers have proposed a novel architecture variant of CNN in this paper (Marcos et al. 2017). Every complex segregation is iteratively applied at a plethora of variations and in return it yields parameters for indicators which are identifiable with the extent as well as the inclination of the apex score at a given point. The researchers in this paper have put forward a proposal to develop a highly distinct and modified CNN fixer which is dependent on the above theory in order to receive DNN-based frameworks. The CNN-based framework that is proposed by the researchers in this paper goes by the name of “RotEqNet” which is scrutinised thoroughly by continuously testing this CNN model on a variety of datasets and registering the various responses generated by it in a variety of scenarios. These scenarios would include ones such as the matching of patches, clinical visual partitioning, the categorisation of visuals as well as the assessment of the positions. The researchers also managed to depict that across all use cases, the RotEqNet proposes an overly dense model when a variety of conditions come into play. Additionally, it does generate results in league with the models to a much higher extent. The main reason that researchers used to propose this methodology is that sometimes during the execution of certain computer vision scenarios, the overall dimensions of the model gradually reduces owing to the fact that due to harsh encoding, complexities tend to be reduced instead of being treated as variants. One of the many consequences of researching complexities is a steep fall in the total number of conditions. The process of sharing weights over a large surface area is known as “weight-tying” which tends to decrease the overall number of conditions, thus making them proportionate to the total number of pixels that exist in a visual. This later does encode translational uniformity within the model that is used for testing. The newly proposed CNN-based methodology of RotEqNet does depict similar meritorious features especially when it comes to dealing with revolutions. This implies that through the process of encoding, there comes about a decrease in the total number of conditions that can be utilised. This tends to go hand-in-hand with the provision of a rather consistent or improved accuracy across various fields. In retrospect, it should be

noted that the researchers have proposed a novel CNN-based methodology known as RotEqNet which involves the process of encoding certain predefined behaviours into CNN concerning revolutions. RotEqNet tends to achieve this by the applications of consecutive sieves one above the other, withdrawing a variety of mapped traits and finally initiating the process of encoding. A plethora of experiments carried out on categorisation and separation tend to depict whether this methodology is best suited for this type of issue.

The research presented in this paper (Selvan et al. 2020) pertains to an improvement on an already existing form of tensor neural networks which is popularly abbreviated as MPS. This particular tensor neural network is designed to function on 1D visual images, but it can also be utilised on 2D and 3D visuals. The pre-training phase of MPS would often involve the overly squeezing of a lot of data into one particular image. The new upscaled proposed version of MPS is known as “LoTeNet” which doesn’t forcefully consume a lot of computing capital, in order to achieve the best performance which would then put it in the same league as its counterparts. It has come to light that over the years SVMs as well as other methodologies that have something to do with kernels have induced a new era of acumen which is totally based upon supervised machine learning. It is quite a known fact that SVMs are highly volatile when it comes to certain kernel-based conditions. It is also one of the main reasons as to why their inception has been greatly reduced in cases which involve enormous datasets which contain extremely complex visuals. This paper mentions a lot about “Tensor networks” which are basically a domain of “wave functions” which are utilised in the research of “quantum systems”. In the current decade, it has been noticed that problems involving image classification which always employ supervised machine learning have a great desire to deploy “tensor networks”. The constant existence of a small number of conditions, the chosen methodologies have a particular rule that is enforced which is that it has to only deal with visuals of either tiny or single dimensional clarity. There have been numerous attempts to increase said visual clarity dimensions as the association of crispness of the pixel greatly diminishes. This would only work well if the images were tiny but would yield completely opposite results if the images were large. Another issue would be the continuous decrease of the image clarity which would in turn have a great impact on the accepted dimensions of the image. Finally summarising the research that has been penned down in this paper, it is noteworthy to recap that a tensor-based model that has been proposed to implement a taxonomy of two-dimensional and three-dimensional clinical visuals which are certainly of a very high grade. A much more advanced version of MPS known as LoteNet is utilised in a very iterative manner which tends to preserve the universally accepted

structure of the clinical visual. After conducting various experiments on a total of 3 datasets, it has come to light that the aforementioned models after training would return aggressive levels of rendition. The amount of GPU required for this task is only a tiny morsel of the amount that is used up by CNN-based models.



Chapter 3

Methodology

3.1 Overview

As mentioned before, the main aim of this research revolves around creating a more affordable method to detect and predict the existence of cancers in a blood smear sample that occurs in the dataset in the form of image scans. The goal of utilising machine learning to aid with the prediction of existence of cancer in the blood smear sample further aids the cancer specialist and helps in reducing time taken to generate overall results during a regular cancer check up.

3.2 Research Framework

The following machine learning pipeline is a diagrammatic representation of the step-by-step procedure of the method used to predict the existence of cancer tissue in the histopathological screening of a blood smear.

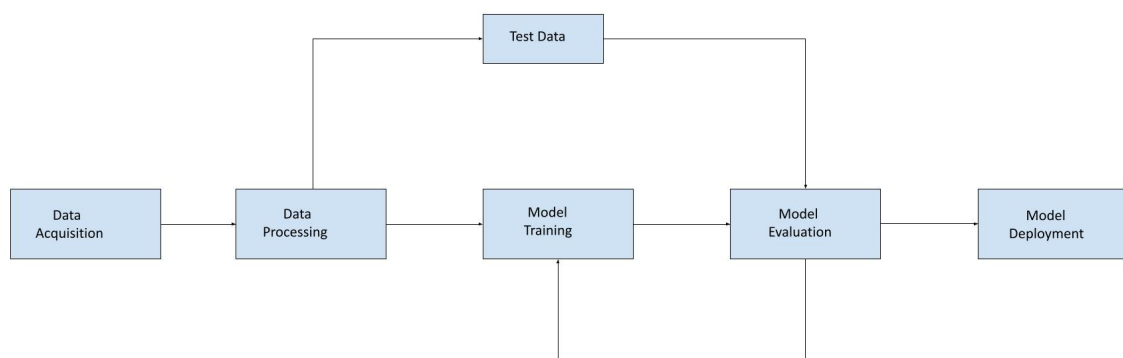


Figure 3.1: Overall research framework

3.2.1 Data Acquisition

The initial phase of the machine learning pipeline is about acquiring a dataset that is not only relevant to the research, but also consists of a large amount of data which by far is the best solution for overfitting as the desired validation accuracy that is generated during the training of the model would be totally unachievable otherwise. The phase of acquisition of data doesn't only involve itself with obtaining data, but also a variety of other steps. These include importing the necessary libraries that were needed to initiate the machine learning pipeline. Apart from this, it should be duly noted that already existing libraries should have their versions upgraded in order to ensure a seamless and uninterrupted execution. This particular stage is later followed by loading the "train_labels.csv" dataset into the dataframe of "train_data", while "train_data.head()" is a method which is often used in the python programming language in order to return the initial rows and columns along with all the data that they possess from the dataset.

3.2.2 Data Processing

The secondary phase in the machine learning pipeline mostly involves itself with the processing of the data which was acquired in the primary phase. This phase normally consists of steps such as the addition of affirmative values, the erasure of incorrect data, docketing of data, etc. But with respect to this research, it mostly revolves around the partitioning of the training dataset "train_labels.csv" into two independent pieces of training and validation, each having a size of 80% and 20% respectively. In this project, the training dataset is split into training and validation instead of the usual splitting of training and testing, due to the fact that training set is utilised for the generation of predictive models while the validation set is utilised to check out the performance levels of these predictive models. The validation dataset is used to initiate the tuning of hyperparameters. It also helps in the reduction of overfitting as it can be easily put to a stop the moment it is detected using the procedure of early-stopping. Since the main aim of this research is to predict the existence of cancer tissue within the scan of a blood smear, image classification models are considered to be the best choice in order to achieve this feat. In order to further improve the quality of training data, the concept of augmentation of images is utilised. This implies that this procedure aids in the development of novel scans on the basis of already existing ones which in turn make the job of the machine learning engineer much easier as it automates the processing of data by the inclusion of haphazard flips. After this process has been completed, it has been made sure that the augmentation of the image data hasn't ended up in the creation of weird

amalgamations of previously existing image scan data.

3.2.3 Model Training

This process of training a variety of chosen models on the PCam dataset forms the third phase in the machine learning pipeline. The chosen image classification models are usually pre-trained in order to save time as well as computer resources, which otherwise would lead to wasteful consumption if the model is built ground up. The list of pre-trained image classification models includes VGG16, ResNet50, Inception_V3 and EfficientNet. These four pre-trained models are then trained on the PCam dataset in order to generate validation accuracies. VGG16 is one of the premier and most sought after pre-trained visual categorizational model which consist of many layers and has many iterations as well such as VGG19, where 19 refers to the number of layers. The only downside to the implementation of VGG16 is that it takes way too long to train enormous datasets. The Inception_V3 model is a third generation iteration of the original Inception image classification model. It is the second most sought after model just behind VGG16. Moreover it consists of about 42 layers while VGG16 contains only 16. This means that it has an error rate of 4.2%. The ResNet is an image classification model that came into being in the year 2015. The main purpose of the ResNet model architecture is to have the accuracy obtained during training to keep decreasing in correlation with how deep the model goes. Unlike VGG16, ResNet isn't into stacking layers as it would rather fit a residual map instead. Like the previous image classification models, ResNet too has many iterations with the latest one being ResNet152 and produces a 5% error rate. EfficientNet which is an image classification model that comes straight in at number 4 is an offspring of Google and it utilises a novel scaling methodology that goes by the name of Compound Scaling. Although this novel scaling technique can be implemented by a basic CNN model, the image classification model of EfficientNet has its own progenitor model known as EfficientNetB0. The version of the EfficientNet model that is used in this research is its seventh iteration, The model with the highest validation accuracy is then declared most suitable to be used for testing. Another piece of useful information is how much ever one might train their models, they can never achieve a 100% validation accuracy. After each model is trained, the weights of the model are saved using ".h5py" which is none other than a package interface to HDF5 (a binary format) which allows compression of large amounts of data as a single file. Each model after being trained in the model training phase of the machine learning pipeline plots its own performance through an accuracy and loss graph. At the end of the training phase, all the graphs are tallied side-by-side, and the one with the highest validation accuracy is considered

to be the most suitable to be utilised in model testing/evaluation.

3.2.4 Model Evaluation

The model evaluation phase involves running the model with the highest validation accuracy that is obtained from the previous phase, against a testing dataset in order to check for efficiency and accuracy. The process of prediction of the existence of cancer tissue in the blood smear takes place in the phase of evaluation of models. Following the penultimate phase of model evaluation comes the ultimate phase of model deployment.

3.2.5 Model Deployment

The phase of model deployment is the last phase in the machine learning pipeline which only gets executed after a certain level of performance has been achieved by the model during its phase of evaluation.

3.3 Dataset

The models in this research are trained on a miniature version of the original “PCam dataset” (Cukierski 2018) which itself is a derivative of the “Camelyon 16” dataset. This was done for the sole reason that the original “PCam (PatchCamelyon) dataset” is way too large as it consists of multiple copies of the same medical visual scans. The PCam dataset (Veeling et al. 2018) is a dataset that consists of clinical visual images of “metastatic cancer” derived from much larger scans. It contains approximately about 327,680 coloured visuals of the size 96x96 pixels which are extracted from the histopathological screening of a variety of areas with high concentrations of lymph nodes.

3.4 Evaluation Metrics

The main evaluation metrics for image classification that is utilised in this research is accuracy and f1-score, more commonly known as “the harmonic mean”. The main goal is to obtain all the validation accuracies of the models that are involved and tally them side-by-side in order to find the best suited model that could be tested in order to achieve the necessary prediction. From amongst these two evaluation matrices, the f1 score does have an edge over accuracy as in the case of disruptive classes the value of the accuracy is often consigned to the dustbin. Lastly, accuracies don’t seem

to hold to well when it comes unequal bifurcation of data into three independent datasets.



Chapter 4

Results and Discussions

4.1 Overview

Given below are the graphical representations of the total number of accuracies and losses of all the pre-trained image classification models that are implemented in this research. These graphs depict accuracies and losses against a total number of eight epochs after being trained for ten epochs with one hundred steps each.

The VGG16 is an image classification model which after being trained on the dataset has achieved a validation accuracy of 86.04% and acquired a validation loss of 34.25%. This is depicted in Figure 4.1.

The ResNet50 is an image classification model which after being trained on the dataset has achieved a validation accuracy of 66.85% and acquired a validation loss of 69.16%. This is depicted in Figure 4.2.

The InceptionV3 is an image classification model which after being trained on the dataset has achieved a validation accuracy of 81.43% and acquired a validation loss of 41.58%. This is depicted in Figure 4.3.

The EfficientNet is an image classification model which after being trained on the dataset has achieved an accuracy of 81.34% and acquired a validation loss of 40.87%. This is depicted in Figure 4.4.

4.2 Tabulation Summary

The accuracies and losses have been neatly compiled into a tabular column Figure 4.5. It has been duly noted that the image classification model of VGG16 tends to achieve the highest validation accuracy of 86.04% while being trained on the PCam dataset and therefore gets chosen to undergo evaluation. There often exists a confusion between validation accuracy and testing accuracy. The easiest way to understand this is that the former concerns the validation set while the latter

concerns the test set. The existence of the validation set does aid in overcoming overfitting by utilising a procedure known as early stopping. Validation loss on the other hand is the paramount loss that is generated during training on a validation set. During the application of cross-validation, the training dataset is split between training and validation instead of the usual training and testing, which ultimately ends up in the formation of three independent datasets. This new validation dataset has only one function and that is the tuning of hyperparameters whereas the testing dataset is utilised for the calculation and determination of the final evaluation metrics. When it comes to the topic of evaluation metrics the accuracy doesn't hold a candle to the f1 score. When it comes to application of f1 score during the phase of model evaluation it is deemed to be quite apt as accuracies may often tend to be misleading due to sudden disparities when it comes to classes. If a disparity in classes does occur then the resulting accuracy can be deemed as useless. Although the accuracy may tend to have a lot of proficiencies going for it as it can be easily understood by any lay man without needing any explanations, but when it comes to topics relating to unequal segregation of data into three different datasets, its deficiencies begin to emerge.

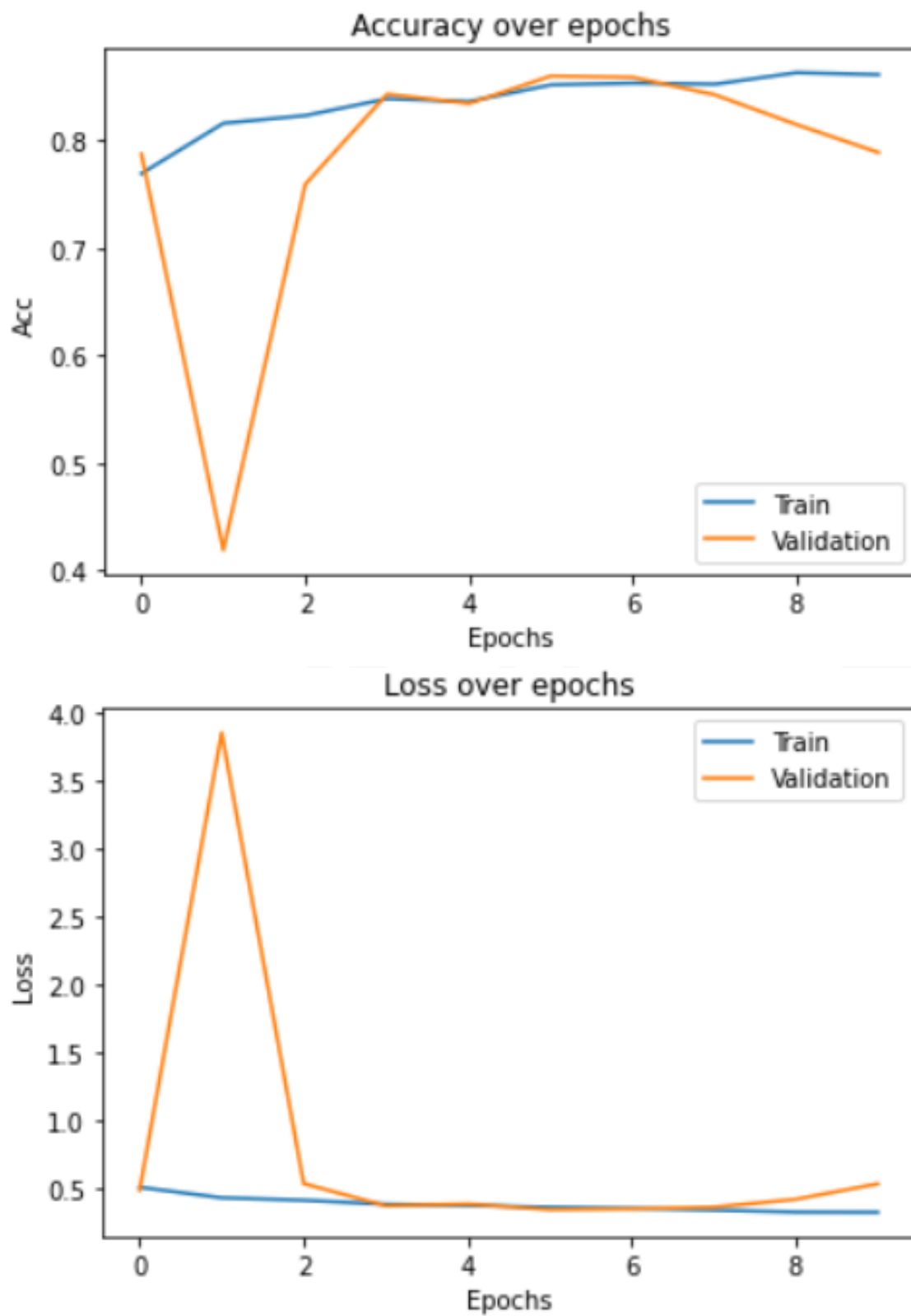


Figure 4.1: Graphical Representation of VGG16 performance

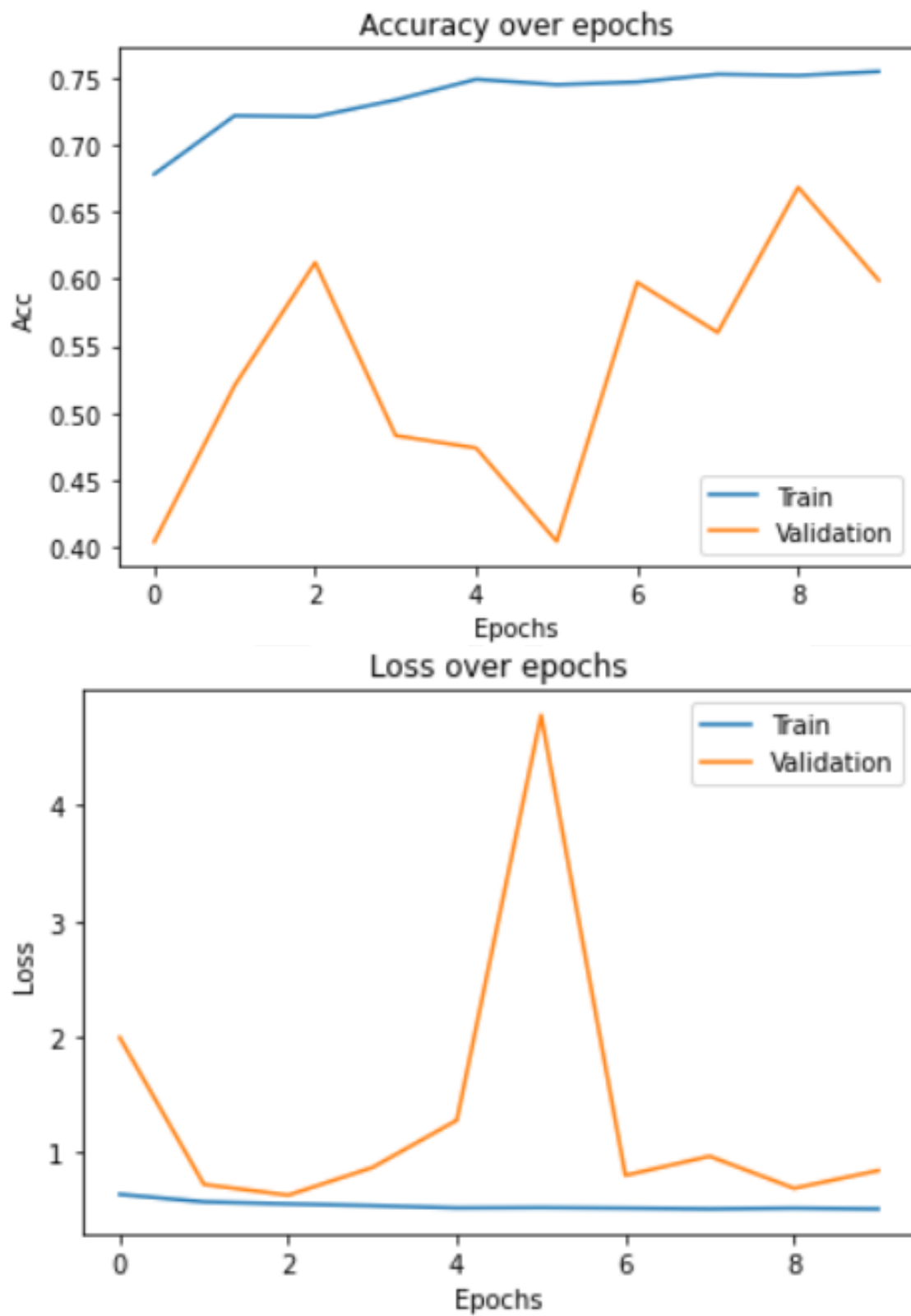


Figure 4.2: Graphical Representation of ResNet50 performance

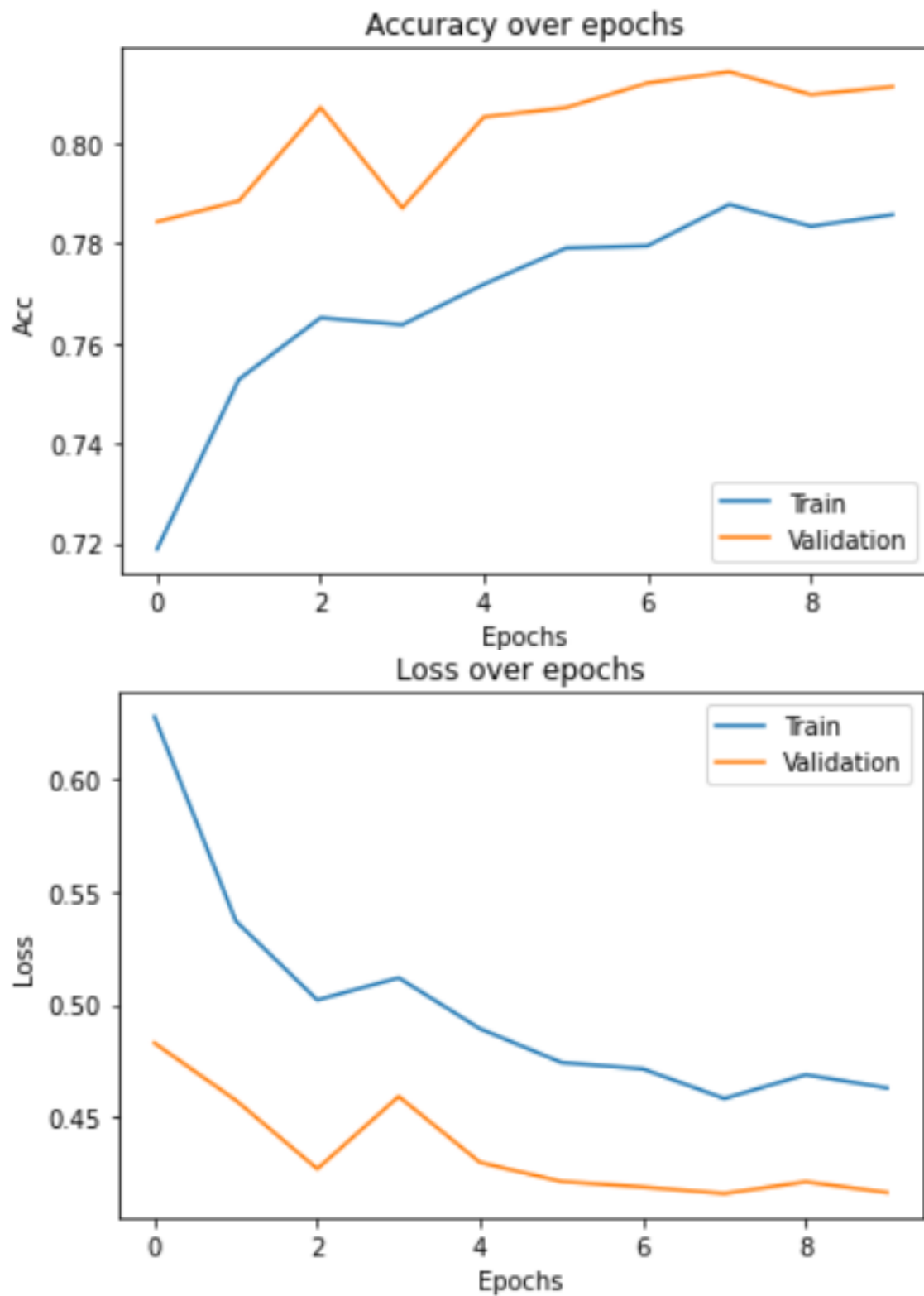


Figure 4.3: Graphical Representation of InceptionV3 performance

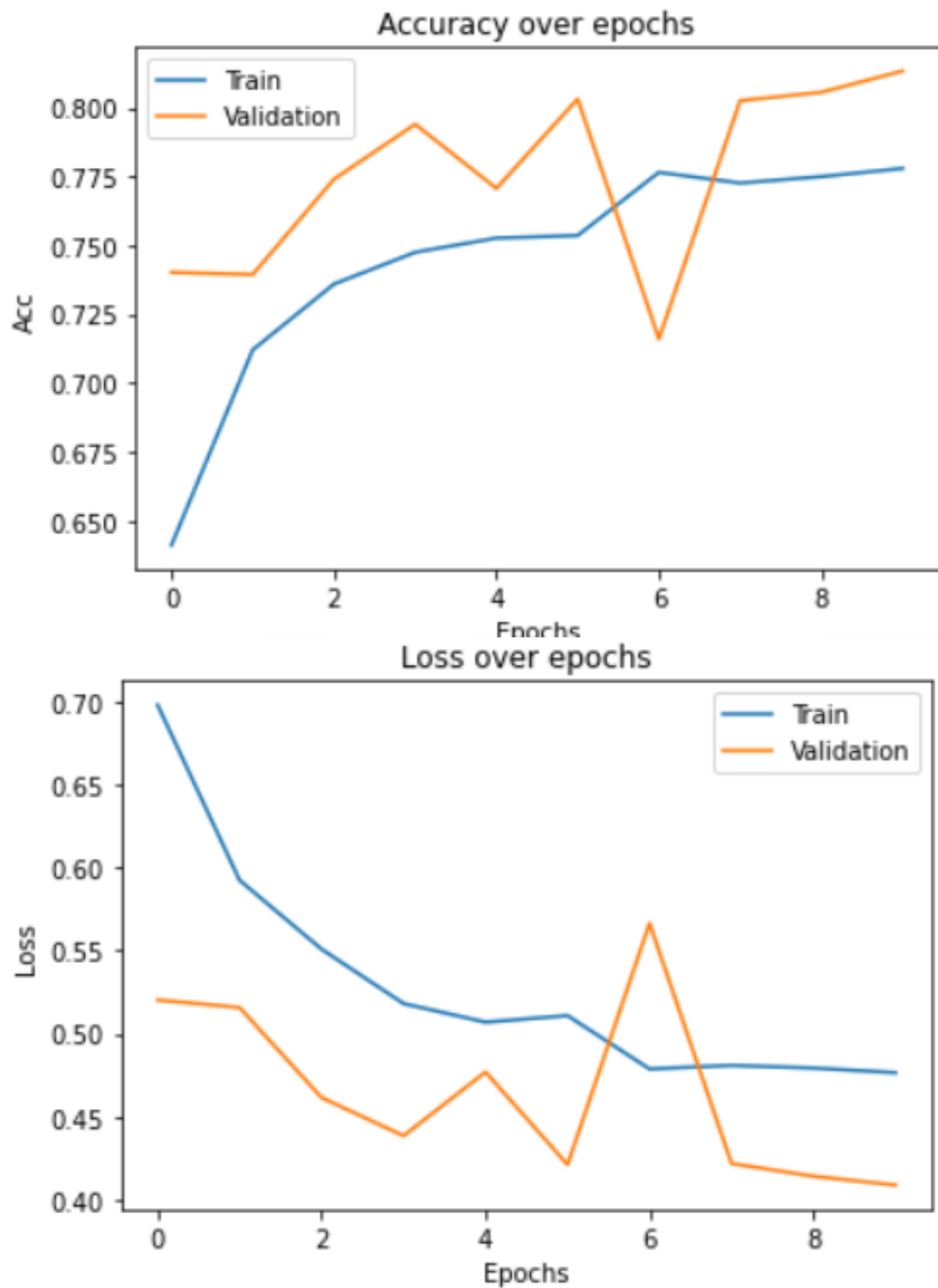


Figure 4.4: Graphical Representation of EfficientNet performance

	VGG16	ResNet50	Inception_V3	<u>EfficientNet</u>
Accuracy	85.22%	75.20%	78.78%	77.80%
Validation Accuracy	86.04%	66.85%	81.43%	81.34%
Loss	36.05%	51.96%	45.81%	47.65%
Validation Loss	34.25%	69.16%	41.58%	40.87%

Figure 4.5: Tabulation of accuracies and losses across all models

Chapter 5

Conclusion

The research in this project intends to fill a gap in the area of existence of mechanisms or methodologies for the detection of cancer. This research has one main primary goal which is to aid cancer specialists to provide a quick diagnosis to expecting patients by incorporating machine learning and deep neural networks into the process in order to speed things up. This mechanism detects the existence of cancer tissue in a sample of blood that exists in the dataset in the form of medical image scans. After going through the works of other authors a few concepts like co-correction did stand out. Unlike the image classification models that are trained on the dataset which are considered to be the best in their league, co-correction is on another level as it can single handedly amplify the accuracies of the trained models.

Bibliography

- Bekkers, E. J. (2019), ‘B-spline cnns on lie groups’, *arXiv preprint arXiv:1909.12057* .
- Cohen, T. & Welling, M. (2016), Group equivariant convolutional networks, in ‘International conference on machine learning’, PMLR, pp. 2990–2999.
- Cukierski, W. (2018), ‘Histopathologic cancer detection’.
URL: <https://kaggle.com/competitions/histopathologic-cancer-detection>
- Dubois, Y., Bloem-Reddy, B., Ullrich, K. & Maddison, C. J. (2021), ‘Lossy compression for lossless prediction’, *Advances in Neural Information Processing Systems* **34**, 14014–14028.
- Early cancer diagnosis saves lives, cuts treatment costs* (2017).
URL: <https://rb.gy/ui9vo>
- Graham, S., Epstein, D. & Rajpoot, N. (2020), ‘Dense steerable filter cnns for exploiting rotational symmetry in histology images’, *IEEE Transactions on Medical Imaging* **39**(12), 4124–4136.
- Gu, K., Masotto, X., Bachani, V., Lakshminarayanan, B., Nikodem, J. & Yin, D. (2022), ‘An instance-dependent simulation framework for learning with label noise’, *Machine Learning* pp. 1–26.
- He, K., Zhang, X., Ren, S. & Sun, J. (2016), Deep residual learning for image recognition, in ‘Proceedings of the IEEE conference on computer vision and pattern recognition’, pp. 770–778.
- Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. (2017), Densely connected convolutional networks, in ‘Proceedings of the IEEE conference on computer vision and pattern recognition’, pp. 4700–4708.
- Kaczmarzyk, J. R., Kurc, T. M., Abousamra, S., Gupta, R., Saltz, J. H. & Koo, P. K. (2022), ‘Evaluating histopathology transfer learning with champkit’, *arXiv preprint arXiv:2206.06862* .

- Lafarge, M. W., Bekkers, E. J., Pluim, J. P., Duits, R. & Veta, M. (2021), ‘Roto-translation equivariant convolutional networks: Application to histopathology image analysis’, *Medical Image Analysis* **68**, 101849.
- Liu, J., Li, R. & Sun, C. (2021), ‘Co-correcting: noise-tolerant medical image classification via mutual label correction’, *IEEE Transactions on Medical Imaging* **40**(12), 3580–3592.
- Marcos, D., Volpi, M., Komodakis, N. & Tuia, D. (2017), Rotation equivariant vector field networks, in ‘Proceedings of the IEEE International Conference on Computer Vision’, pp. 5048–5057.
- Mu, N., Kirillov, A., Wagner, D. & Xie, S. (2022), Slip: Self-supervision meets language-image pre-training, in ‘Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXVI’, Springer, pp. 529–544.
- Pagliardini, M., Jaggi, M., Fleuret, F. & Karimireddy, S. P. (2022), ‘Agree to disagree: Diversity through disagreement for better transferability’, *arXiv preprint arXiv:2202.04414*.
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J. et al. (2021), Learning transferable visual models from natural language supervision, in ‘International conference on machine learning’, PMLR, pp. 8748–8763.
- Romero, D., Bekkers, E., Tomczak, J. & Hoogendoorn, M. (2020), Attentive group equivariant convolutional networks, in ‘International Conference on Machine Learning’, PMLR, pp. 8188–8199.
- Selvan, R. & Dam, E. B. (2020), Tensor networks for medical image classification, in ‘Medical Imaging with Deep Learning’, PMLR, pp. 721–732.
- Selvan, R., Ørting, S. & Dam, E. B. (2020), ‘Locally orderless tensor networks for classifying two-and three-dimensional medical images’, *arXiv preprint arXiv:2009.12280*.
- Veeling, B. S., Linmans, J., Winkens, J., Cohen, T. & Welling, M. (2018), Rotation equivariant cnns for digital pathology, in ‘Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11’, Springer, pp. 210–218.

Wenger, J., Kjellström, H. & Triebel, R. (2020), Non-parametric calibration for classification, in ‘International Conference on Artificial Intelligence and Statistics’, PMLR, pp. 178–190.

What is Cancer? (2019).

URL: <https://rb.gy/6e5ak>

What is Metastasis? (2022).

URL: <https://rb.gy/d2yea>

Worrall, D. & Welling, M. (2019), ‘Deep scale-spaces: Equivariance over scale’, *Advances in Neural Information Processing Systems* **32**.