Chapter 1

Introduction

1.1 Background

Research into the field of medical image analysis is extensive. There have been millions of publications published in this area over the course of the past few generations. Current developments in this domain are computer-aided skin identification and the diagnosis of cancer pneumonia from chest scans. Both of these advancements were made possible by advances in computer technology. Historically, image analysis has been dominated by sophisticated feature extraction or handcrafted discriminant features; oriented gradients and local binary patterns are some examples. However, the recent emergence of deep learning (DL) algorithms has officially launched a shift towards automated image analysis. In particular, convolutional neural networks (CNN) have emerged as the most reliable deep learning method for image processing. CNN has been used by every scholar who has developed a classification algorithm that has ranked among the top in the most recent challenges for medical image analysis. In addition, Shi et al.[28] proved that the features derived using Deep Learning are superior to those recovered using other methods.

Within the context of a computer-aided visual diagnostic system, images are processed and perceived with the aid of human involvement. Analysing medical images and providing diagnoses of medical conditions has traditionally relied heavily on visual diagnostics as well as the sensory characteristics of the patient being examined. The trend toward creating medical pictures that are larger in number and more complex poses a threat to the human visual system, which runs the risk of the diagnosis analysis being overwhelmed as a consequence[24]. Ethical and practical limits are placed on the quantity of medical records that a cardiologist or neurologist can evaluate all at once. Because of the possible risks that prolonged screen time poses to one's physical and mental health, labour rules have imposed ethical constraints on the amount of time that may be spent in front of a screen. In terms of the constraints that are placed on practical applications, the attention spans of human experts shift throughout time[19]. Technically speaking, intra-operator variability illustrates that the efficiency of an operator is not static, i.e., it varies throughout the course of time. A more accurate diagnosis, as well as a more targeted course of treatment, may be predicted with the help of an expert diagnostic system that trains with more images and gains experience.

Deep CNNs are now being used for a diverse range of computer vision applications. These CNNs are trained on massive quantities of supervised data using a variety of learning techniques. Because of this, excellent DL implementations have been inspired in a variety of different fields, such as reinforcement learning, voice recognition, and NLP architectures. Deep CNN architecture and training techniques have grown more embedded in recent years. Nonetheless, it is not apparent which of the distinct parts of its contemporary layers is essential to their performance. Classical ML and DL are both types of learning that have until very recently, been approached as two distinct types of algorithms. These algorithms have been programmed to perform a variety of tasks successfully.

As soon as there is an alteration in the way how the feature space is distributed, the models must be rebuilt from the ground up. Moving beyond the paradigm of isolated learning, the concept of transfer learning has been established that illustrates the use of knowledge which has been acquired to address one problem and apply it to perceive the solution of other problems that are similarly comparable. Traditional learning is solitary, and it only makes use of a limited multitude of tasks and datasets, as well as the training of separate models in isolation. There is no retention of information that can be transferred from one model to another. Transfer learning is the process of using information (features, weights, etc.) from previously trained models that have already been trained to train newer models and even work around issues such as the more recent task requiring fewer data points or records. This is accomplished by using information from previously trained models on source datasets.

In this study, we are emphasising on the classification of melanoma, which is the most dangerous kind of skin cancer. Melanoma develops in the cells (melanocytes) responsible for the production of melanin in our bodies; melanin is the pigment responsible for giving your skin its colour. In addition, melanoma may grow in your eyes, and while it is sporadic, it can also be detected in other parts of your body, such as the oesophagus or nose. The precise explanation for the development of melanomas is not understood; however, it is discovered that being exposed to ultraviolet radiation, which may come from the sun, tanning beds, and other sources, might raise the risk of developing this illness. It is indicated that the risk of melanoma is increasing among those under the age of 40, especially among women. Being aware of the warning signs of skin cancer may make it feasible to diagnose and treat malignant cells prior to the progression of this illness on to its final stage. This can be accomplished if one is aware of the symptoms. Effective treatment is possible for melanoma as long as it is discovered in its early stages.

Regarding their capacity to generalise to unfamiliar and unanalysed clinical data, contemporary deep learning medical imaging models are woefully outdated. Unexplored data refers to real-world events that were not encountered during historical training data. As a result, trying to employ a model in clinical practice is likely to fail. The frequent restrictions imposed on the medical training data limit the expressiveness of deep models. The barely accessible quantity of data also limits their performance, and sometimes it might be challenging to find the precise supervised data that is needed to train our model. The benefit of Transfer Learning (TL) models is that they have already been trained on source datasets which are generally large volumed datasets; the target-specific dataset is just utilised for hyper-parameter fine-tuning the model weights to the current task at hand for the target. With this proposed implementation, the need for the models to be trained from scratch is effectively eliminated.

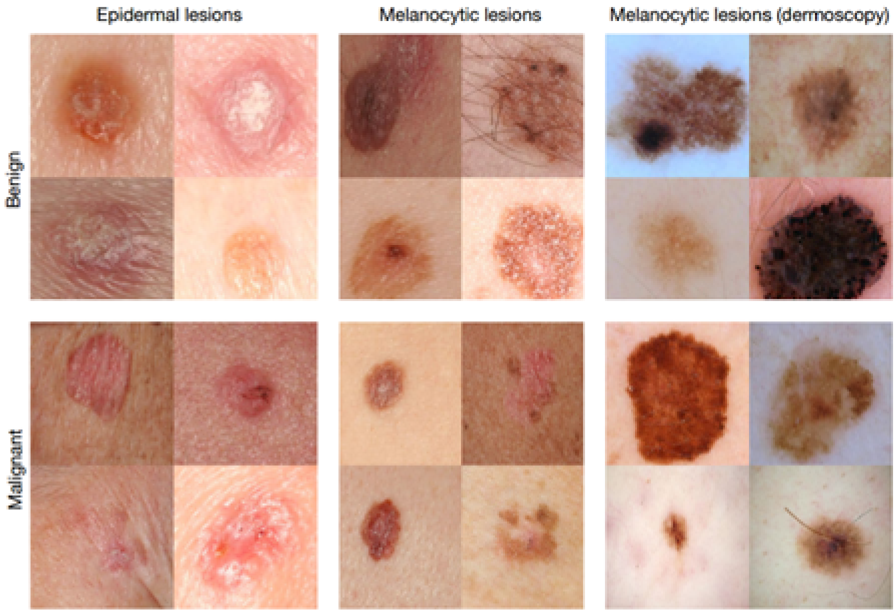

FIGURE 1.1: Skin Coloured Blotches From The Dataset On Various Lesions[10]

1.2 Research Aims and Objectives

This project's primary objective is to determine which learning models— Direct learning, End-to-End transfer learning, and Transfer cascade learning—are more effective in this research. The datasets would also need to be examined to identify which dataset properties are beneficial to train these models, in addition to utilising the skin cancer data that we currently have, and alternative learning procedures would need to be examined to see how they vary. The models would be trained, and hyper-parameter adjustments would be used to evaluate how well they performed on the datasets. The model is trained to generate initial weights for the pre-trained model using the IMAGE-NET database in this work.

Our extensive experiments provide the following generalisation insights:

1. We clearly illustrate that a TCL technique, when opposed to the conventional E2E learning technique, has enhanced feature representations in DNNs for medical imaging tasks, and we propose that researchers and physicians in this field to take TCL into consideration.

2. We look at the issue of how to discover the best features for medical imaging and discover that the best performance on downstream tasks for medical imaging will come from combining the TCL and TE2E strategies.

3. The question of when to transfer for medical imaging is a challenge that we have also taken into account. Our findings show that, in a modest data environment, TCL would perform as good as or outperform both the TE2E model and a model starting from scratch with the ResNet-18 architecture. As a result, TCL is a competitive classification technique, and a combination approach(TCL and TE2E) can increase its effectiveness even more.

4. We show that TCL learns distinct features, with coarser features in early layers and finer features in later levels, as opposed to end-to-end learned features, which have more uniform granulometry distribution throughout layers. As a result, the characteristics learned using TCL are considerably different, and we demonstrate that they are better suited for various tasks. TCL offers better representations for a range of properties seen in several medical picture formats.

5. We demonstrate experiments on a dataset that has been expertly labeled to show how our TCL framework is superior in terms of feature localisation.

6. When presenting machine learning models to the medical community, we advise going beyond simply assessing classification performance to consider interpretability, calibration, and robustness.

1.3 Dissertation Report Structure

The history of the project and the key objectives of the background research are explained in the first chapter. In chapter two, we describe new research on the subject of Transfer Learning and Medical Imaging conducted by various academic scholars. In chapter three, the proposed Transfer Cascade Learning method is discussed, along with a comprehensive Exploratory Data Analysis of the source and target domain datasets. In Chapter 4, we describe and analyse the observations of the model's results. In Chapter 5, we discuss the model's outcomes, conclusions, benefits, and limitations. In the last chapter, which is Future Implementations, A summary and set of proposals for enhancing our model approach have been discussed.

Chapter 2

Literature Review

Several academics have already been researching Layer-wise learning approaches. In this part, we assess similar research and highlight their contrasts with our study.

In the past, optimistic unsupervised learning has been the subject of extensively pursued investigation. As has been demonstrated, unsupervised learning of deep generative models may be utilised as a foundation for deep supervised architectures[5]. Bengio et al.[6] also considered utilising supervised greedy layer-wise learning to create networks for future end-to-end supervised learning, but this strategy did not prove to be effective owing to the absence of Big-Data at the time. We emphasise the contrast between supervised and unsupervised layer-by-layer learning and consider both[16]. In addition, this study investigates layer-based training as a training method as well rather than only a pre-training method.

In the context of developing supervised NN's, layer-wise learning has been explored in several studies. Various prior studies, such as Ivakhnenko and Lapa[15], Fahlman and Lebiere[11], and Lengellé and Denoeux [20], examined its application to straightforward problems in an environment where DL was the dominant supervised learning approach. The majority of these efforts were devoted to developing frameworks that allow models to successfully adapt in response to data. The need to prevent difficulties with fading gradients inspired further investigations. Cortes et al.[8] suggested a similar progressive learning strategy that evolves a network's design so that it can adapt to the situation. This technique provides theoretical advances to structure learning but cannot be applied to issues where deep neural networks outperform other approaches.

Although self-supervised learning on mainstream visual categorisation datasets has only just been possible, it has already been used in the medical field to design domain-specific pretext problems. Several research focuses on adapting contrastive learning to medical data, while Many have already attempted to develop domain-specific pretext challenges[4]. The paper by Sowrirajan et al.[29], which is most closely related to our work, investigates the use of MoCo pre-training for the classification of the CheXpert dataset using linear evaluation. The purpose of medical imaging is to facilitate diagnosis, therapeutic planning, and patient prognosis. In this overview, we focus on the epidermal layers of the skin. To study cancer lesions, we have investigated a variety of medical imaging modalities, currently, skilled polymaths accomplish image segmentation and classification tasks with little computer assistance. Medical imaging facilitates patient screening, therapy planning, and diagnosis. Hospitals use an extensive array of medical imaging technology to provide individualised treatment, improve diagnosis, and improve healthcare. Medical personnel can use a range of scanning methods to observe and follow disease severity without requiring invasive therapies. Each medical imaging approach generates digital pictures with differing degrees of information. Some diagnostic procedures are also organ specific. X-rays, ultrasound, and optical computed tomography are typical medical imaging tomography modalities[19].

2.1 Transfer Learning

Cognitive research is the origin of transfer learning (TL), which is based on the premise that information may be carried over from one endeavour to another in order to enhance the overall performance on an unrelated task. It is widespread knowledge that individuals are capable of overcoming identical challenges by drawing on their prior experiences. Pan and Yang provide a formal definition of TL by using the concepts of domains and tasks in their explanation[25]. The elements that make up a domain are the feature space known as 'X' and a marginal probability distribution known as P(X), where X = $\{x_1, ..., x_n\}$ belongs to feature space 'X'. A task is defined by the equation T = {'Y', f(.)}, where 'Y' is a label space and f(.) is an objective predictive function. This equation is used in the context of a particular domain, which is expressed by the equation D = {'X', P(X)}. One acquires knowledge about a task by studying the pair $\{x_i, y_i\}$, in which x_i represents 'X' and y_i represents label space 'Y'. When given a source domain D_s and a learning task T_s, as well as a target domain D_t and a learning task T_t, the goal of transfer learning is to enhance the learning of the target predictive function $f_T(.)$ in D_t by using the information learned in D_s and T_s[25]. This strategy makes use of a convolutional neural network (CNN) model that has been properly trained on a large dataset (such as ImageNet) as a feature extractor for the target domain (e.g., medical). To be more explicit, the well-trained CNN model has all of its convolution layers frozen, and the last fully connected layers are removed from the model. The convolution layers work as a feature extractor that may be adjusted in order to adapt to a new job (related to medicine). After characteristics have been extracted, they are sent to a classifier, which may consist of new fully connected layers or any other supervised machine learning approach. Finally, throughout the training process, just the newly created classifier, and not the complete network, is trained.

The data scarcity dilemma arises as a result of the fact that DL algorithms need a substantially large amount of data for training under optimal conditions. Specifically, the small size of medical cohorts and the high cost of expert-annotated data sets are some of the well-known issues in this field. Transfer learning (TL) and domain adaptation approaches have been investigated extensively as potential solutions to this issue as part of a number of different research projects. These seek to attain excellent performance on target activities by using the information gained from source tasks in order to fulfil their goals. In 2010, Pan and Yang[25] published a ground-breaking review work on TL in which they categorised TL strategies from the perspective of a labelling aspect. In the same year, Weiss et al.[34] provided a summary of TL investigations based on homogeneous and heterogeneous approaches. Zhuang et al.[42] conducted an evaluation of more than forty sample TL techniques from the points of view of data and models in the most current study, which was published in 2020. Unsupervised learning in the realm of transfer learning is an emerging field that has, in the recent past, attracted a rising amount of interest from scholars. Wilson and Cook[36] conducted a literature review on a substantial amount of unsupervised deep domain adaption publications. In recent years, generative adversarial networks (GANs)-based frameworks have gained some attraction. One technique that seems especially promising is Deep Artificial Neural Network. In addition, multiple kernel active learning and collaborative unsupervised approaches[39] have also been implemented as unsupervised techniques for TL. A complete evaluation was carried out in some of the research, with the primary emphasis being on DL applications in the medical field. Chowdhury et al.[7] examined the state-of-the-art research on self-supervised learning in medicine, while Litjens et al.[21] studied DL for medical image analysis by summarising more than 300 papers. On the other hand, some scholars conducted a literature review on papers that focused on TL with a particular case study, such as counting micro-organisms[38], cervical cytopathology[26], neuro-imaging biomarkers of Alzheimer's disease[2], and MR-brain imaging in general[31].

By fine-tuning already-trained networks and adding supplementary layers, TL is designed to address all of its issues. Technically, this is achieved by reusing a network that has already been wholly trained with a predefined dataset. These networks were first trained on the ImageNet dataset[30] and have been adapted for various issues by retraining some components of the network to fit the new domain. Compared to training from scratch, this method is shown to provide results substantially more quickly with a somewhat reduced error rate[37]. The TL approach shows promise as a means to reduce the quantity of training data necessary[1], 2 primary components make up every TL system: (1) feature extraction (via a series of convolutions and pools) and (2) fully connected layers (classification). It is unclear whether the component of a classification model built using TL will have the most artificial knowledge and hence be most applicable to a novel dataset. When working with a limited labelled dataset, choosing a minimal model which is small and efficient has been proved to be essential starting point. However, larger models have many more parameters that must be adjusted as a cost.

In addition to transfer learning, there are multitude of other approaches available to improve machine learning generalisation. In multi-task learning, machine learning algorithms are tuned to do wide range of tasks at once more effectively. To segment and rebuild images including malignancies, for example, Weinger et al.[35] presented a multi-task auto-encoder-like CNN with three decoders, one for each job. Zhou et al.[41] proposed an auto-encoder that has three different branches: one for coarse cancer segmentation, one for refined cancer segmentation, and one for detailed cancer segmentation. Transfer learning concentrates on a single target tasks or domains, whereas multi-task learning tackles multiple tasks or domains simultaneously, giving equal weight to both the source tasks or domains and the target tasks or domains[25]. Multi-task learning is derived from transfer learning by this key distinction that transfer learning focuses only on a single target tasks or domains. Data augmentation, which involves adding copies of the training data that have been subjected to modest transformations, is one method that may be used to improve the extrapolability of algorithms(the key concept to optimisation of complex processes)[23, 40]. Data augmentation is beneficial when pictures with specified features (for instance, a certain contrast) are uncommon, and it may supplement other approaches such as multi-task learning or transfer learning. However, in contrast to multitasking or transfer learning, data augmentation does not consider whether or not the skills learned from related activities may be used in other contexts. In addition, increasing the amount of training data also increases the amount of computational costs, and exploring data transformations that are advantageous is a challenging task. Finally, Only two of the research that were included in the evaluation really looked at the potentially negative outcomes that may arise from the use of transfer learning. Kollia et al.[18] took into account images from both the source domain and the target domain during the optimisation process. This was done to prevent catastrophic forgetting, which is defined as a drop in performance on the source domain following the application of transfer learning to the target domain. An approach to identify aneurysms was presented by Sato et al.[27] that directly blocks negative transfer[33]. This means that the algorithm produces results in the target domain that are poorer than if no transfer learning was performed.

2.2 Negative transfer

The prime objective of TL is to improve the efficacy of a target model by incorporating information from a complementary source domain. However, Let us assume there is no connection between the source domain(ImageNet) and the target domain(Medical Imaging). Negative transfer occurs when a lousy association has an unfavourable impact on the intended learning process. One of the significant challenges with the TL approach is the phenomenon of negative transfer[30]. To be successful, TL requires that the source and target domains be sufficiently similar during training[25]. Scanned photographs of a patient, such as those used for evaluating CT scans and X-rays, are not like your typical cat-and-dog photos. Unless these images are included in the TL system's source model training data, it will be unable to recognise them correctly in the target domain training. The amount of readily available data that's labelled is also often relatively small in medical datasets. There are several approaches that may be used with unlabelled data, but they typically need more work and fine-tuning. It is tough to put a number on how helpful it is to transfer weights from one model to another. Considering the image's seismic and hyper-spectral nature of the images on ImageNet, which are medically unrelated, it may be helpful to learn about current models (datasets, strategies and their internal association) to transfer information and avoid taking the wrong steps when experimenting.

2.3 Melanoma Skin Cancer

The most prevalent kind of cancer is skin cancer, melanoma being responsible for 75% of skin cancer-related deaths. According to the American Cancer Society, there will likely be over 100,000 new occurrences of melanoma diagnosed in 2025. In addition, it is expected that the disease would kill close to 7,000 individuals. Similar to other malignancies, early and thorough identification, which may be facilitated by data science, may increase the efficacy of cancer treatment. Currently, dermatologists check every mole on a patient to identify "ugly ducklings" or lesions that are most likely to be melanoma. Existing AI approaches do not adequately account for this clinical reference frame. If detection algorithms utilise "contextual" photos within the same patient to determine whether images depict melanoma, physicians may be able to increase the accuracy of their diagnostic methods. If classifiers were effective, they would be more precise and may assist dermatological clinics improve their procedures. This study will focus on identifying melanoma in photographs of skin lesions. Using images of the same individual, we can determine which are more likely to have a melanoma. Using contextual data at the patient level may assist in the development of image analysis algorithms that might better serve clinical dermatology. Although melanoma is a deadly disease, if detected early, the majority of cases may be cured with minor surgery. With the application of automated melanoma diagnosis tools, the diagnostic accuracy of dermatologists will improve. This work has the potential to make enhance the identification of melanomas.

Chapter 3

Methodology

In this section, we introduce the primary methods for the classification of the skin cancer.

3.1 Source Domain Datasets

Given that there are not enough datasets on the skin in the medical industry, it is challenging to effectively optimise the performance of any neural network because the skin is a complex feature to examine. The objective is to investigate datasets that include skin as a feature, examples of which are ImageNet, HAM10000 (which has to be formatted first for CV tasks), and DermNet (which is nicely organised, allowing you to immediately import the dataset without much difficulty). The goal is to use ImageNet to work on alternative skin datasets that are outside the skin cancer dataset domain and are unique to skin cancer for different reasons owing to their target-specific characteristics. This will be done with the intention of expanding beyond the skin cancer dataset area. Many different datasets may be acquired from the face, and some are utilised for identifying emotions, detecting stress, and performing other activities. Finally, it is essential to have a solid understanding of the settings of the datasets that were gathered before beginning the data cleaning and preparation process.

3.2 Target Domain Datasets

The training dataset has a total of 33126 examples, of which 32551 are considered "Benign Cases" and 575 are considered "Melanoma Cases". There are only 575 malignant instances in the dataset, among which 358 malignant male cases are present in the dataset compared to 217 female cases. Image name (which refers to the individual image), Patient ID, gender, age, anatomy site (location of scan site), diagnosis, scan result (0 for benign and 1 for malignant), and image path are the features of the Train Dataset.

On the other hand, Scan Result is not included in the Features of Test Dataset. A regular distribution may be seen in the ages of the participants. If, on the other hand, just the malignant instances are considered, it would seem that the distribution is more widespread. The older and younger age groups do, in fact, have a disproportionately higher number of malignant instances. The thorax was the location of lesions the majority of the time, followed by the extremities (both lower and upper). It is essential to bear in mind that the training dataset, as well as the test dataset, both comprise patients who have had many photographs taken of them. As a result, we have an obligation to ensure that the training set and the test set do not include any images of the same patient. So, It has been examined that there are 794 unique patients included in the training set, and 690 unique patients are included in the test set. There are no common patients in the training or test sets.

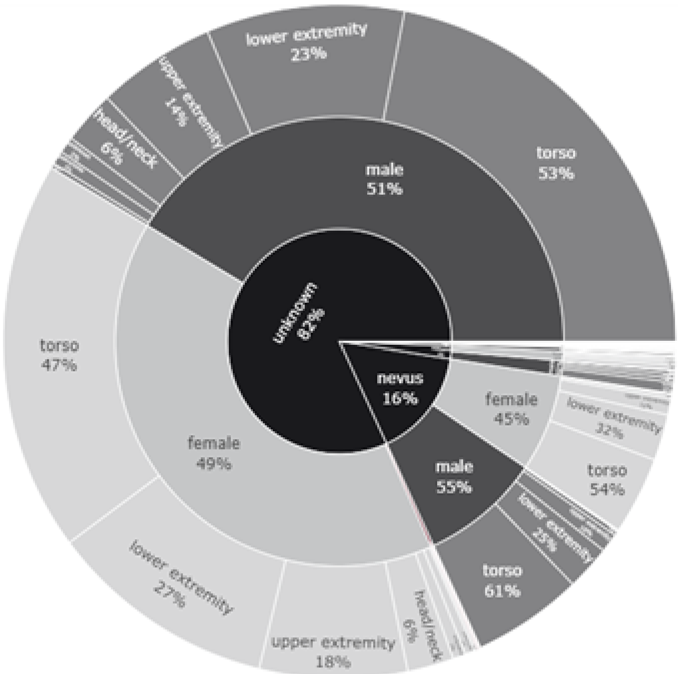

FIGURE 3.1: EDA of Melanoma Skin Cancer Dataset

3.3 Classical End-To-End Learning(E2E)

Guided end-to-end learning is the conventional method for optimising DL networks. However, it may have a number of problems that need consideration. Owing to the use of a universal objective, it is unknown how the layers combine to provide reliable predictions. This leaves open the unresolved problems about the eventual functional behaviour of a deep network's numerous intermediate levels. Numerous researchers have hypothesised and experimentally proved that CNNs acquire procedures that progressively induce invariance to be complex but insignificant variability while boosting linear data separability[5]. It has been experimentally proved that progressive linear separability occurs, but it is unclear if this is a necessary condition for the excellent performance of CNNs or only a by-product of other techniques used by these networks. Second, although generalisation, approximation, and optimisation techniques for 1-hidden layer NNs are accessible, multiple-hidden layer NNs are substantially more difficult to conceptually approach, according to the same research. This makes it challenging to comprehend the relationship between shallow and deep Neural Networks (NNs). In addition to being physiologically impossible, end-to-end back-propagation is impractical regarding processor and memory resources. The first successful architecture on ImageNet presented by the Visual Geometry Group (VGG) was VGG-16 at ILSVRC2014, which was followed by VGG-19. By substituting several tiny kernel-sized filters for a single big kernel-sized filter, these models outperform AlexNet with their architecture which consists of 13 and 16 convolution layers for VGG-16 and VGG-19, respectively[22]. Accuracy saturation and disappearing gradients may result from adding more layers to CNN models. The core of ResNet CNN, known as Residual Learning Network, seeks to address this issue[22]. Before ResNet, CNN models had learnt features at various levels of abstraction at the conclusion of each convolution layer. ResNet learns residuals as opposed to learning features, which is the process of subtracting learnt features from input for each convolution layer. This is accomplished by linking the input of one layer to x further levels using a technique known as identity shortcut connections[32]. Different ResNet variants, including ResNet-34, ResNet-50, and ResNet-101, use various numbers of layers with different architectural constitution.

3.4 Transfer Cascade Learning

The learning technique in which one layer builds upon the learnings of the previous one, also termed as a layer-by-layer learning approach, is known as cascade learning (CL). It is beneficial in applications involving computer vision because early layers in neural networks acquire domain-specific feature representations, which give rise towards this advantage. In this research, we plan to demonstrate that traditional end-to-end (E2E) learning is not as effective as TCL when it comes to transferring knowledge about medical categorisation difficulties. We show that compared to learning directly on the target dataset, transferring cascade learned features from a subset of ImageNet consistently outperforms each other, particularly when the training set size is smaller than one hundred thousand pictures. Through error analysis, we demonstrate that the features and mistakes produced by the various learning paradigms are distinct from one another. This sparked the idea for a combined approach, which we then demonstrated to increase performance. TL strategies can be categorised according to the existence of identifiers in source and/or target domains during optimisation[25]: unsupervised transfer learning (unlabelled data), transductive approaches (labels available only in the source domain), and inductive approaches (labels available only in the target domain and, optionally, in the source domains). Examples of these three categories are shown in figure below for Medical imaging applications.

Type	Properties	Approach Example
Unsupervised	$\mathcal{D}_S \sim \mathcal{D}_T, \mathcal{T}_S = \mathcal{T}_T$	Transforming T1- and T2-weighted images into the same feature space with adversarial training.
Transductive	$\mathcal{D}_S \sim \mathcal{D}_T, \mathcal{T}_S = \mathcal{T}_T$	Learning a feature mapping from T1- to T2-weighted images while optimizing to segment tumors in T2-weighted images.
Inductive	$\mathcal{D}_S \sim \mathcal{D}_T, \mathcal{T}_S \sim \mathcal{T}_T$	Optimizing a classifier on a natural images dataset, and fine-tuning certain parameters for tumor segmentation.
	$\mathcal{D}_S \sim \mathcal{D}_T, \mathcal{T}_S = \mathcal{T}_T$	Optimizing a lesion segmentation algorithm in T2-weighted images, and re-optimizing certain parameters on Melanoma images.
	$\mathcal{D}_S = \mathcal{D}_T, \mathcal{T}_S \sim \mathcal{T}_T$	Optimizing a lesion segmentation algorithm in T2-weighted images, and re-optimizing certain parameters in the same images for anatomical segmentation.

FIGURE 3.2: Types of transfer learning. The subscripts S and T indicate source and target, respectively[31]

While using the straightforward classification proposed by Pan and Yang[25], TL techniques may be subdivided into various types depending on the information transmitted. During optimisation, instance-based techniques evaluate and give weights to images and their features to balance their significance. Feature-based techniques aim a common feature space that is shared across tasks and domains. Asymmetric, in which the properties of the target domain are transformed into the source domain's feature space, and symmetric where we find a common intermediate feature representation. The shared priors or parameters across both the source and destination tasks/domains may be determined through parameter-based approaches. Methods that are based on Parameters indicate that the parameters or priors share functionality and are compatible with one another in other domains, One example of this would be a domain-invariant image border detector. To conclude, relational-based techniques have as their ultimate goal the use of information that is universally accessible in the shared knowledge-base across relational domains.

3.5 Transfer Learning for Medical Image Analysis

The automation of segmentation and classification made possible by medical image analysis technologies has reduced humans' need for direct involvement. The objective is to replace human labour with automated work by making use of techniques such as convolutional neural networks (CNN), end-to-end recurrent neural networks (RNN), and other approaches that are analogous to these. Because it solves some of the issues encountered by traditional amortisation algorithms, TL may be seen as a refinement of these techniques because it improves upon these approaches by fixing some of the issues that are encountered by traditional amortisation algorithms. In most cases, the performance of DL models may be improved by increasing the total amount of labelled data. Data augmentation is a strategy for the generation of data for training that involves the creation of new data through performing modifications of the dataset that was originally used. In the context of picture data, this refers to a wide number of image modification techniques, such as those involving rotation, translation, scaling, and inversion. Memory and computational limits are by far the most significant factors to take into account when augmenting data. Both online and offline strategies for data augmentation are often used in today's organisations worldwide. The difference among both online and offline data augmentation is that the offline data augmentation generates the data in advance and saves it in memory, while online data augmentation is conducted in real time during training. The online method requires less memory from the user, but the overall training duration is lengthened. The offline method is more efficient in terms of training time, but it utilises a significant amount of memory[22]. The figure that may be seen below provides an example of the relationship that can be made between TL and medical imaging.

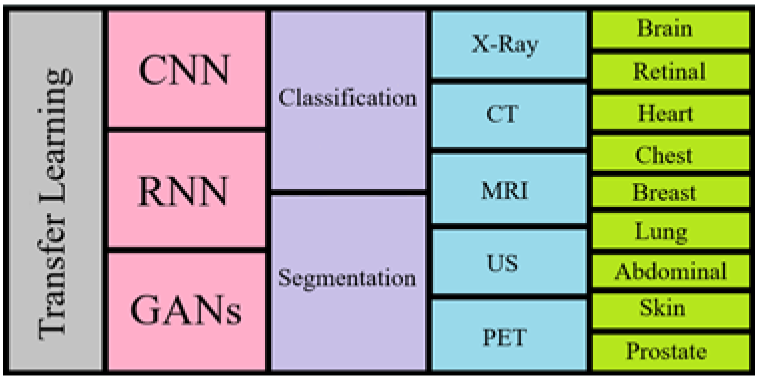

FIGURE 3.3: TL and its research in the field of Medical Imaging[19]

Transfer learning using natural pictures is commonly used in the analysis of computer-aided diagnoses on medical images, and various empirical investigations suggest that this accelerates the efficacy of the productivity of the model[3, 13, 14, 4, 29, 28, 34, 42]. This holds true independent of the differences in picture statistics, scale, and other features that are important to the job at hand. However, a comprehensive evaluation of this strategy was carried out by Raghu and colleagues[17], and their findings show that this does not continually improve performance in medical imaging settings. They did, however, show that transfer learning from ImageNet may speed up convergence, which is particularly helpful when the training set of the target medical picture data is on the small side. This research made substantial use of minimal designs and identified substantial efficiencies with a small amount of data. This was especially true when using their most prominent architecture, which was called ResNet-50. Transfer learning, which uses data from inside the studied domain, may be used to overcome the domain mismatch issue. However, due to the high cost and extensive amount of time required to collect labelled data, this technique is typically not applicable to many different medical tasks. A workable technique that was recently established in the field of self-supervised learning study explores various strategies enabling unlabelled medical data, which is typically simpler to acquire.

3.6 Modelling Noise in Medical Images

Image noise is a common phenomenon in the vast majority of image processing applications, as shown by the extensive body of published research on different ways to eliminate or mitigate its effects[12]. Practitioners in image processing progressively discover that a step termed denoising is typically required before any relevant information can be obtained from an image. This is one of the most common prerequisites. The abundance of denoising features found in widely used commercial software is another indicator of the relevance of extracting a signal from a noisy image. These features are used to reduce the amount of unwanted noise in an image. We use a different approach by mining the noise statistics included in the traditional ImageNet and our target Melanoma skin cancer images to extract new data. Our strategy involves determining the degree to which there is a link between the picture's intensity and the noise variation. The most common kind of noise is the result of several independent signals, all concurrently contributing to the overall picture. This is an outcome of the central limit theorem, which states that adding a large number of random variables, each with its own probability density function (PDF), would produce a signal with a Gaussian PDF. Poisson noise is the most prevalent kind of ambient noise in situations where an image is created by the accumulation of photons across a detector. Typical examples are standard X-ray films, CCD cameras, and infrared photometers. As a result of the rescaling and abbreviation of the intensity rates, the noise statistics of the pictures will continually change. This is because the photographs were taken with a limited degree of precision (for example, 32-bit images). In spite of the fact that the truncation aberrations that are present in the data cannot be removed, the rescaling effects have been taken into account in our examination throughout our research. These distortions are essentially negligible for the vast majority of the pictures that were used in this investigation. Image clipping, which is produced by the saturation of pixels brought on by the noise, will generally have an effect on the high- severity end of the photos. This is the area where the noise variance is the greatest. Because the following findings closely match the conceptual models for the bulk of the intensity ranges of the photos, the influence of saturation was relatively minor in this investigation.

Chapter 4

Implementation and Analysis

In this section, we investigate whether or if the application of TCL over end-to-end learning on pre-trained models may result in enhanced performance for the medical image classification tasks that have been defined. This method consequently starts with a well-trained CNN model on a sizable dataset (like ImageNet) and then swaps out the existing classifier layers for a new configuration. This is accomplished by initialising the convolutional layers with pre-trained weights from the well-trained CNN model while initializing the classifier layers with random weights. In order to do this, we will begin by investigating the various pre-training models and supporting datasets that are available for use with the proposed medical imaging classification and its future prospective applications. Then, we compare and examine the proposed strategy with baselines and cutting-edge techniques, such as the VGG and ResNet50 model for supervised pre-training on ImageNet Dataset, and we evaluate the benefits of our suggested TCL approach, which makes use of the pre-existing models for classification assignment. We investigate the classification effectiveness and transferability (Distribution shift is one of the factors that we consider) of self-supervised models while considering the context of medical photograph classification.

4.1 The design of TCL model

The most crucial issue is whether or not it makes a difference if each layer is trained independently of continuous supervision from professionals. The optimal strategy is a supervised End-to-End and layer-wise algorithm. In this supervised neural networks, we train each new hidden layer to function as the hidden layer of the previously trained neural network (NN), with the most recently learned layer's output serving as the following layer's input. The output layer of the NN is then discarded, and the parameters of the hidden layer of the NN are used as the pre-training initialisation of the new top layer of the deep net. This is done in order to map the output of the previous layers to a representation that is, with any foresight, a more accurate depiction of the data. An example of a TCL network is shown below, It demonstrates that in CL layers are added one at a time, and gradient descent is used to train the weights of the most recently added hidden layer and the output layer. The layers that were trained in the earlier are now inactive. E2E learning is the conventional learning strategy, which involves training all of the layers in the network at the same time.

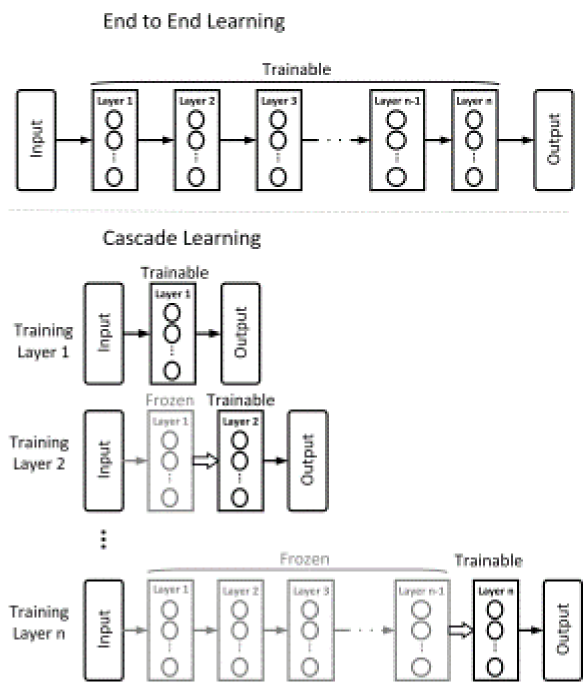

FIGURE 4.1: Schematic illustration of cascade learning (CL) and end-to-end (E2E) learning[9].

4.2 TCL model Implementation

We would want to eliminate the necessity of determining the right amount of training iterations for each layer while using the layer-wise training strategy. The notion that each layer is pre-trained to model its input, disregarding the impact of higher layers, would be helpful if we could train the model by actively adding the layers one at a time rather than having to train all layers concurrently from scratch. Experiments have shown that the functionality of this version is at least comparable to that of the traditional approach. The benefit is that we must now comply to a unified stopping standard (for the whole network).Although hyper-parameter optimisation requires less time, the complexity of the computation is significantly higher. This is due to the fact that we perform more computations on the top layers at the beginning of the process. Before the lower layers converge on a suitable representation, these calculations may be wasted.This version could be more desired for training on extensive data sets, even if one occasionally iterates back to the learning process. Unsupervised pre-training helps contribute to the layer-wise training strategy for effective classification by minimising the complexity of deep networks' challenging optimisation problems. During the unsupervised pre-training phase, the weights of all layers are initialised more precisely.

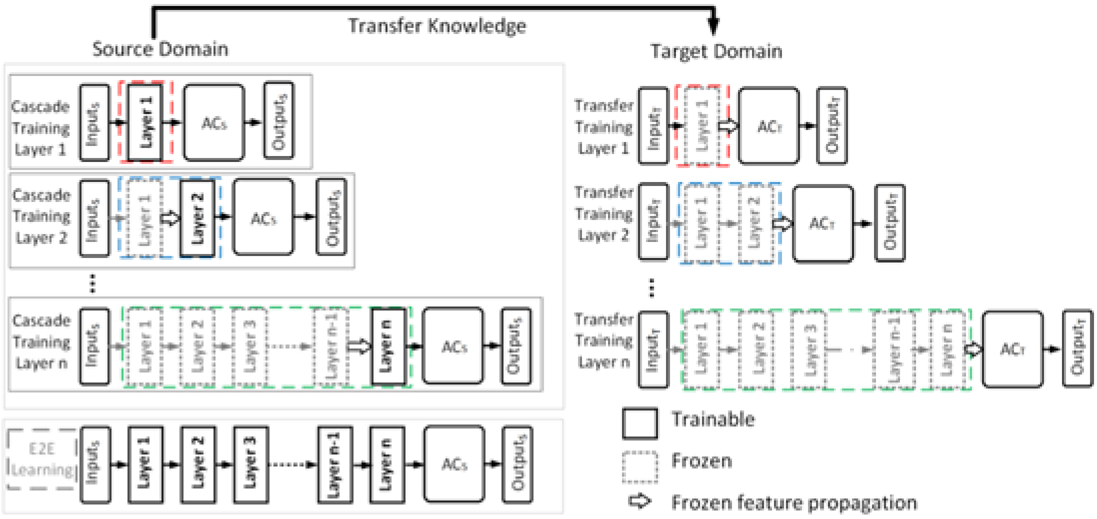

FIGURE 4.2: TL incorporated into Cascade Learning setting[32].

4.3 Performance Evaluation of TCL

In order to pave the way for our investigation, we chose several neural network architectures and assessed how certain they performed transfer learning from ImageNet while (1) training from an unbalanced target dataset and (2) training on a balanced target dataset. We assess ResNet50 and VGG models to compare our model performance, which have both been widely used in medical transfer learning applications, using the standard ImageNet architectures. We trained using the family of significantly state-of-the-art convolutional neural networks, which have previously achieved phenomenal performance on medical tasks, and our distinctive high-performing TCL designs have proven valuable for medical image analysis. We are using F1 Scores as a metric for comparison of different model architectures in this study. When the True Positives and True Negatives are more relevant, the accuracy metric is employed, but the F1-score metric is used when the False Negatives and False Positives are more essential. F1-score is a superior metric in situations when there is an imbalance between classes in the given dataset, such as the one discussed in this study. The majority of real-world classification problems include an uneven distribution of classes; hence, the F1-score is a more appropriate statistic to use when evaluating our model.

Chapter 5

Results and Discussion

5.1 Results

According to what can be observed through table 1, the model does not become much preferable beyond layer 3 and 4. In addition, we are able to determine that the model has not yet become very knowledgeable from the data used for training it as this is clear from looking at the loss percentage that the model achieves, on comparing results on both the training and the validation data sets. At the 89th epoch, the model had the most efficient loss production. The trends followed by the model suggest that our model is not very generalisable since some of the validation curves exhibit a large amount of variance. It is observed from the accuracy plot that the model is incapable of acquiring new information beyond epoch 96. According to the data depicted in the table, layers 2, 4, and 5 of the CL model have the best accuracy scores out of the whole model. On the other hand, the accuracy of the pre-trained model is the second lowest of all models, coming in at 68%, just behind layer 7 of the cascade learning model, which has the lowest accuracy of all models, 65%. However, since the dataset we are using is very unbalanced, the F1 score is more significant than it was previously thought to be, as stated earlier in the methodology; layers 2 and 3 have the best performance in this regard. Based on the fact that layer 8 has a low F1 score, we can draw the conclusion that the model is more effective at identifying cancerous patterns in its earliest stages, as opposed to adding multiple layers arbitrarily. However, the low F1 score may also be the result of the model's inability to learn sufficient patterns from the dataset. This has been reflected on by taking into consideration the fact that the accuracy for layer 8 is also low and that there is not a significant gap between the F1 score and accuracy in this case. Layer 3, on the other hand, achieves the most outstanding results in terms of the F1 score, and because it is 10% higher than its accuracy, it could mean that our performance in the minority malignant class is significantly better than our performance in the majority benign class. This is supported by the nearly identical accuracy and precision scores as well as the significantly higher recall score compared to the majority

Architecture	F1 Score(%)	Accuracy	Loss
Pre-trained Model	74%(±0.1)	68%(±0.7)	46%(±0.4)
Layer 1	79%(±0.5)	73%(±0.4)	54%(±0.6)
Layer 2	81%(±0.6)	74%(±0.4)	35%(±0.4)
Layer 3	**82%(±0.4)**	**72%(±0.1)**	**15%(±0.4)**
Layer 4	77%(±0.7)	74%(±0.7)	57%(±0.9)
Layer 5	76%(±0.9)	74%(±0.6)	58%(±0.1)
Layer 6	76%(±0.7)	72%(±0.2)	59%(±0.8)
Layer 7	75%(±0.9)	65%(±0.5)	59%(±0.7)
Layer 8	66%(±0.6)	68%(±0.3)	61%(±0.4)

TABLE 5.1: Transfer learning performance among different layers with the Melanoma dataset.

class. Therefore, this model would be the most effective for categorising cancer patients; yet, it has the potential to incorrectly label some benign individuals as having cancer. However, taking into consideration the potential to identify cancer in patients, this is a relatively minor drawback in the context of the use case.

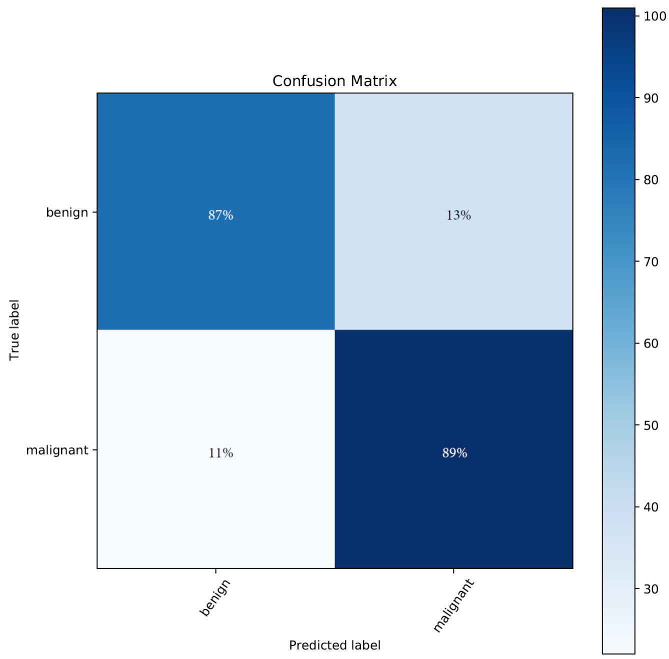

FIGURE 5.1: Confusion matrix obtained based on Test Data.

5.1.1 Comparison with Existing Models

We investigate the potential advantages of employing multi-class learning and make use of the potential availability of over several images per patient for a specific disease by utilising a pre-trained deep neural network. We do this in order to better understand the possible implications of the state-of-the-art approaches mentioned in the table 2. After that, we associated the performance of each architecture by making use of classifications

Architecture	F1 Score(%)	Accuracy	Loss
VGG Model	80%(±0.7)	90%(±0.9)	20%(±0.4)
ResNet Model	80%(±0.2)	93%(±0.3)	21%(±0.4)
All-Benign Model	66%(±0.7)	99%(±0.7)	69%(±0.4)
E2E Model	74%(±0.1)	68%(±0.7)	46%(±0.4)
TCL Model	**82%(±0.4)**	**74%(±0.4)**	**15%(±0.4)**

Table 5.2: Performance comparision on the Melanoma dataset using TCL, E2E, ResNet and VGG architectures.

obtained with and without cascade learning. When comparing TCL's performance to that of the original VGG and ResNet algorithms for cancer lesion classification, TCL consistently performs better than those two algorithms over a wide range of parameter settings and base network design choices. TCL leads to an increase in accuracy for cancer lesion classification when compared to solely using E2E learning algorithms as a classification method. When we compare the performance of TCL (the third layer) with E2E, we have seen an improvement of 8%. In addition, when comparing TCL to all the other models examined, our results have shown a minimum accuracy improvement of at least 2%. Because, as was mentioned previously, subsequent layers in the TCL model have a lower F1 score, the best model, which we are using to compare with the other models, is the third layer. Furthermore, evaluating the classes more equally is a trait of the selected metric, which is why the F1 score is being focused around this study. This is because the dataset that we are using is quite unbalanced. In order to demonstrate how the TCL model learns from the target dataset and not only generates random predictions, we also evaluate it in comparison to another model that we constructed called the All-Benign model. Henceforth, if we use the All-Benign model to label every picture in the dataset as benign without doing any learning, we will reach an accuracy of 99%. Because the dataset is skewed, we find that this model has an F1 score that is higher than 50%. This is owing to the fact that the model fits the data better. Therefore, it is not difficult for us to recognise the benefits offered by the TCL model at a class level; it is superior at recognising photos that include malignant tumours. Additionally, we may deduce from this that the TCL model, if adapted to account for the scenario mentioned above, will perform much better when applied to multi-class situations. Take, for instance, the task of classifying images into two distinct categories of cancer as well as non-cancer conditions.

5.1.2 Comparison using Balanced Dataset

DL-based approach has recently made considerable advancements in cancer classification tasks. Consequently, a serious problem with class imbalance often arises between healthy tissue and malignant tissue class distribution. Due to this issue, healthy tissue becomes predominant throughout training, which lowers the quality of model optimisation. Since the significance placed on each class in the dataset remains constant regardless of the

Architecture	F1 Score(%)	Accuracy	Loss
VGG Model	81%($\pm$0.7)	87%($\pm$0.9)	5%($\pm$0.4)
ResNet Model	82%($\pm$0.2)	88%($\pm$0.3)	5%($\pm$0.8)
All-Benign Model	67%($\pm$0.7)	50%($\pm$0.8)	81%($\pm$0.3)
E2E Model	80%($\pm$0.1)	72%($\pm$0.7)	59%($\pm$0.6)
TCL Model	**88%($\pm$0.7)**	**87%($\pm$0.8)**	**11%($\pm$0.1)**

TABLE 5.3: Performance comparision with Balanced Melanoma dataset using TCL, E2E, ResNet and VGG architectures.

size of the individual classes, Our approach to handle the issue of class imbalance was to create a dataset with a 50/50 split of the two classes. The table below displays the results of the execution discussed. By fine-tuning initial versions of the model independently on all downstream tasks with a balanced dataset, we broaden our comparison. The results of fine-tuning for various model architectures are summarised in Table 3. we consistently observe a higher performance for all cascade transfer learning models.

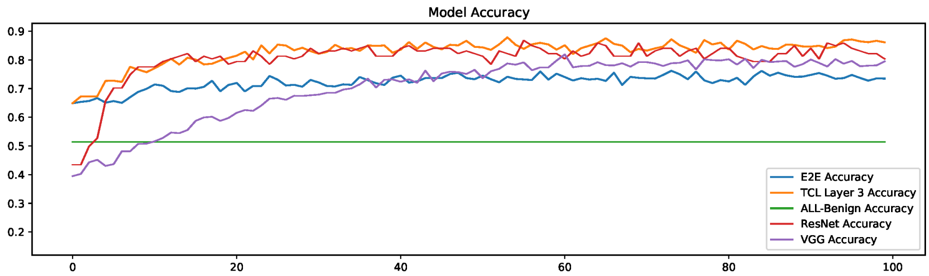

FIGURE 5.2: Accuracy plot of All models on Balanced Dataset.

5.2 Discussion

We report in the context of transfer learning with three different architectures, all pertaining to the field of medical imaging. This study examins the performance of a cascade architecture on the Target dataset, demonstrate a comparison of transfer cascade learning across different architectures within a given target dataset, and also compare the transfer of features obtained from E2E and TCL trained models. As far as we are aware, the research conducted on standard designs to examine the effects of transfer learning from ImageNet has been significant. This area of research is crucial because, currently there has been great hindrance to mobile and on-device applications in the medical field due to the present models, which use massive and costly computations. Additionally, to accomplish the image classification, the majority of the parameters in normal ImageNet models are focused at the top. The fact that medical diagnosis often includes very few feature attributes suggests that ImageNet models are likely over parametrised. We find that transfer learning strongly affects performance across all models and datasets. Additionally, despite having considerably lower accuracy on the target test dataset, the family of TCL models with smaller, lighter CNNs outperforms conventional models with a higher F1 score. This is because the ImageNet as a source dataset has necessarily proven to be a good indication of success on medical datasets. Finally, we note that these results hold even in the presence of very limited amount of data. We randomly selected 20% of the data for validation, while the remaining data were utilised for training. Both the train and test data are standardised to the interval [0,1]. Throughout the whole of the experimental work detailed in this paper, numerous repetitions of each experiment were conducted to evaluate the degree of variability in the results acquired from 100 and 150 training epochs. The results are provided as weighted F1 scores, where weighted averaging was applied to account for the imbalance among classes that led to the performance gap across courses. Throughout every test, the Adam optimiser was used, and its initial learning rate was set at 0.0001, which was determined intuitively. The finding was surprising at first look since ResNet is a more robust model, it obtained an accuracy margin of 93.3%, 2.4% less than VGG16, which had 90.9%, while having fewer parameters. This assumption is valid, however, only due to the fact that these models are trained and refined from the ground - up with E2E strategy. But both the models had a F1 score less than the TCL model. This can also be somewhat associated to the fact that on the ImageNet dataset there are few mislabeled examples, that were previously unknown, were found in 18 of the image classes[42]. Regarding TL techniques, the majority of this paper experimentally evaluated as many CNN model combinations with as many TL approaches as feasible. Compared to earlier recommended best practices, The TCL model with three layers on the end of a pretrained model for the classification task produced the best results. To build a more prominent application, it would be interesting to investigate TL with CNN prototypes and several other medical data properties such as anatomical establishments, imaging modalities, data measurements, label size, and more.

Chapter 6

Conclusion

We conducted a comprehensive literature assessment of TL approaches for medical image analysis that used pre-trained CNN models from the non-medical ImageNet dataset. Using a transfer learning methodology and a CNN model, it is possible to acquire decent performance on the job at hand. This study suggests that a non-medical dataset such as ImageNet might be helpful for tackling medical challenges via transfer learning. In this paper, We evaluate the most effective transfer learning technique for medical image classification by conducting extended experiments. In addition, the most crutial principles in deep learning and the most common frameworks for them are outlined. We demonstrate that transferring using a layer-wise learning technique, also known as TCL, enhances performance in contrast to the conventional E2E learning method. This method also offers other benefits in addition to gains in classification performance.

Further supporting the hypothesis that transfer learning via CL is a superior strategy for medical imaging classification, particularly on datasets with fewer than 100,000 samples, we discover that TCL offers improvements in terms of the localisation of features, robustness to noise, and improved calibration. This was discovered by exploring the types of errors that were made and the types of features that were learned. We show that TCL is a better learning strategy for deployment in actual clinical situations regarding classification performance, tractable activations, noise resilience, and confidence calibration. Here, we investigate TCL, a method of layer-wise learning has been shown to be effective among several other multi-layer networks. As opposed to the freedom afforded by E2E training of the same architecture, layer-wise training imposes constraints on how target-related knowledge transfer may be incorporated into the network. We show that cascade training still yields impressive results despite this distinction. Contra-distinct from end-to-end training, with much fewer parameters and far less time spent on training than previously adopted architectures of deep neural networks, is superior to VGG and ResNet on the melanoma classification challenge. Features extracted from various layers to transfer across tasks exhibit monotonically declining performance as we travel from the third to the eighth layer, demonstrating that TCL hierarchically extracts essential features across initial layers. When using a cascade-trained network as the source network, the most remarkable results are achieved by transferring coarse features from the first few hidden layers; these features are on par with, and in some cases even superior to, those transmitted from any layer in an E2E-trained network. To further emphasise this idea, we build a conundrum that requires the transfer of more minor features and demonstrates that superior features are produced from the initial few layers.

Chapter 7

Limitation and Future Works

Patients will benefit from TCL's exceptional performance on the medical picture classification challenge, which will aid in skin cancer's early detection, diagnosis, and treatment. Nonetheless, there is still room for improvement. For starters, the texture feature is a major low-level characteristic that may accurately define the image's context or location. Therefore, combining in-depth features with some novel texture descriptors like hybrid colour local binary patterns (HCLBP) and local binary patterns (LBP), which are among the most popular handcrafted discriminant features; moment invariant features; wavelet features; and elongated quinary pattern (EQP) features. Second, collect the information about what each layer has learned; evaluate the accuracy of each layer's predictions by comparing each individual layer to one another based on the features learnt by that layer. This implementation may provide light on which features were learnt by which layer, which might aid us in a subsequent classification in which we use the features' characteristics derived from this classification to solve a different problem. Another future necessity is to run these models on higher-end hardware, i.e., in terms of GPU, CPU, and memory configurations. This should, in theory, bring down the training time and the model might quickly converge to the highest possible F1 score. The advantage this could bring is being able to parallelise and process more data efficiently and then run many models simultaneously. Also, we can find the optimal hyper-parameters for each model by parallelising models. Fourth, developing an algorithm which is predicated on natural images for analysing medical images with millions of cells remains challenging. Fifth, the existing database seems to have an unbalanced ratio of malignant to benign classifications. Therefore, it would be helpful to build a public dataset for future studies. Lastly would like to conclude by stating that several efforts have been made to improve smartphone microscopy with deep learning [50]. This might shed light on the foundations we need for the dependency of the medical domain on the neural network to understand in order to develop a mobile phone application that uses an individual's submitted photograph of a lesion to evaluate whether or not it is malignant.