T.C.

ALTINBAS UNIVERSITY

Institute of Graduate Studies

Information Technologies

HYBRID DETECTION TECHNIQUES FOR SKIN CANCER IMAGES

Hasan Abed Hasan

Master of Science

Supervisor

Asst. Prof. Dr. Abdullahi Abdu IBRAHIM

Istanbul, 2021

HYBRID DETECTION TECHNIQUES FOR SKIN CANCER IMAGES

by

Hasan Abed Hasan

Information Technologies

Submitted to the Institute of Graduate Studies

in partial fulfillment of the requirements for the degree of

Master of Science

ALTINBAŞ UNIVERSITY

2021

The thesis titled “HYBRID DETECTION TECHNIQUES FOR SKIN CANCER IMAGES” prepared and presented by Hasan Abed Hasan was accepted as a Master of Science Thesis in Information Technologies.

Asst. Prof. Dr. Abdullahi Abdu IBRAHIM

Supervisor

Thesis Defense Jury Members:

Asst. Prof. Dr. Abdullahi Abdu IBRAHIM	School of Engineering and Natural Sciences, Altinbas University	_____
Asst. Prof. Dr. Doğu Çağdaş ATİLLA	School of Engineering and Natural Sciences, Altinbas University	_____
Assoc. Prof. Dr. A Mohammed ABUBAKAR	College of Business, Antalya Bilim University	_____

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Approval Date of Institute of Graduate Studies: ____/____/____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Hasan Abed Hasan

DEDICATION

I devote and pledge this research work to my supervisor who is salient for guiding me through whole research work as well as my family for always assisting me in my hard time.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor Asst. Prof. Dr. Abdullahi Abdu IBRAHIM for guiding and helping me along the way in writing this dissertation. Discussing my progress, problems, and ideas with my supervisor Asst. Prof. Dr. Abdullahi Abdu IBRAHIM a couple of times every week helped me tremendously in understanding the logic behind the research. It made me better realize the technical need for this research work.

ABSTRACT

HYBRID DETECTION TECHNIQUES FOR SKIN CANCER IMAGES

HASAN , Hasan Abed Hasan,

M.Sc., Information Technologies, Altınbaş University,

Supervisor: Asst. Prof. Dr.Abdullahi Abdu IBRAHIM

Date: March/2021

Pages: 55

According to W.H.O, skin cancer is one of the most common types of human malignancy in medical sector. The development of skin cancer detection system using CNN based on skin lesion diagnostic tasks is in hot pursuit, with special attention being given to the classification of benign and malignant. Given its high propensity to meta-size, going hand in hand with severe decreases in survival rates, and the high inter-patient variability in lesion appearance, as well as a strong requirement on the training of the physician/dermatologist properly diagnosing a melanoma can be considered a daunting task. The purpose of this research thesis is to use a deep learning for the detection and segmentation of skin diseases using with hybrid techniques in deep learning and imaging. The literature review of different papers was conducted with different data mining model architectures. Three custom models were created for the task, and deep learning techniques were used with different levels of fine tuning of hybrid deep learning models. Screening for skin cancer diseases to identify people at high risk has been under debate. In current discussion, the suggested screening procedure is bone density based including possible prescreening questionnaires. Development of new prediction models and automated diagnostics could contribute to general screening programs in the future. The custom deep learning model architectures were designed to represent different depths. The idea behind this is to analyze the effect of increasing

representational capacity to the results and visualizations for patients of skin cancer. Additionally, deep learning models were created to represent traditional data mining approach. Data mining has been used in some skin diseases, producing good results for example in segmentation of skin defected area structure and diagnosis of skin diseases. The study also attempts to find solutions to practical deep learning challenges such as low training speed and lack of transparency. There are two forms of skin tissue.

Keywords: Skin, cancer, deep learning, detection,CNN, Xception.

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

LIST OF ABBREVIATIONS

UVR : Ultraviolet Radiation

CNN : Conventional Neural Network

DNA : Deoxyribonucleic Acid

ROS : Reactive Oxygen Species

SVM : Support Vector Machine

GPU : Graphic Processing Unit

1. INTRODUCTION

1.1 INTRODUCTION

As depicted in Figure 1.1, the skin exists of three main layers. The bottom layer is the hypodermis, also referred to as subcutaneous tissue, consisting of loose connective tissue, primarily adipose tissue capable of storing fat. This layer mainly contains fat lobules, thus functioning as an energy reserve, insulation and protection layer, as well as nerves and larger blood and lymphatic vessels. On top of the hypodermis lies the dermis, a layer composed of collagen fibers and elastic fibers (i.e. connective tissue) and an extracellular matrix, all of which are produced by fibroblasts [1]. It contains blood and lymphatic vessels, sensory receptors, hair follicles and sweat glands. The upper skin layer is the epidermis, which can again be structurally subdivided in different layer. At the interface between dermis and epidermis lies the stratum basal, a columnar layer of basal models, primarily keratinocytes. The proliferative capacity of the skin has been observed to be restricted to the stratum basal, which is due to the presence of epidermal stem models [2]. Portraying both self-maintenance and self-renewal, they give upon (asymmetric) division rise to one daughter stem-cell and one transit-amplifying cell. The latter models have the capacity to undergo a limited number of cell divisions before producing fully differentiated keratinocytes, thus giving rise to a short-lived stem-cell lineage. The differentiated models move upwards into the supra-basal layers, which consists largely of squamous epithelial models, whereas the daughter stem cell remains in the stratum basal to maintain the stem cell pool [3]. Upon reaching the outermost skin layer, the keratinocytes have undergone a further maturation process and have lost their nucleus and cytoplasmic organelles. Stacked layers compose the outermost barrier of the epidermis, the stratum conium, which serves as a physical and chemical protection barrier and regulates skin hydration [4]. The non-viable coenocytes are continuously being shed through the process of desquamation and new keratinocytes enter from the underlying epidermal skin layers, thus preserving skin tissue homeostasis.

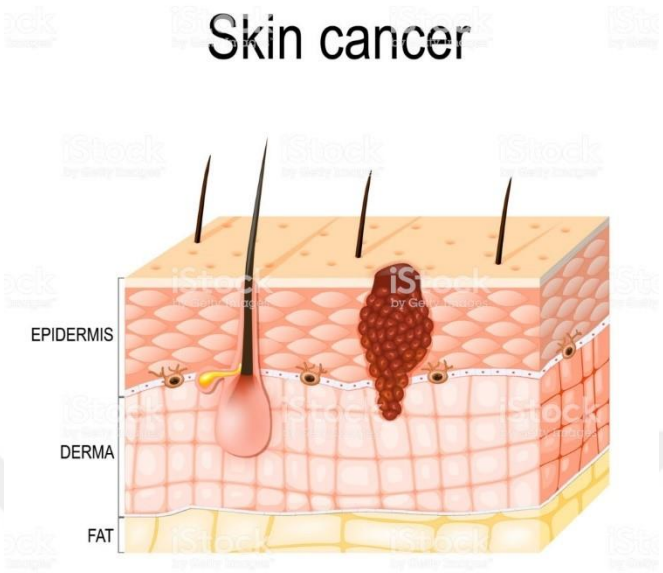

Figure 1.1: Structure of the skin [5].

The epidermis contains a number of other models, amongst others: Langerhans models which belong to the skin immune system, Merkel models which play a part in mechanoreceptor and are in contact with neurons and, of importance here, melanocytes [6]. Melanocytes are neural-crest derived models and can also be found in the hair, and iris, as well as the inner ear, nervous system and heart. Their main function is the production of melanin in dedicated organelles, known as melanosomes. Melanin is a natural pigment that comes in different forms: brown/black melanin, red/yellow pheomelanin and brown/black neuromelanin. The most prominent type in the skin, melanin, acts as a photo-absorber and has been reported to dissipate over 99.9% of absorbed visible and UV light through non-radioactive pathways [7]. The melanosomes are subsequently transported to the keratinocytes, where melanin accumulates in a strategically manner, providing a cap-like cover on the UV-exposed side of the models, Differences in skin pigmentation can be attributed to a difference in the amount of melanogenesis and the distribution, size and content of melanosomes, rather than to a difference in the number of melanocytes [8]. The melanogenesis in response to UV exposure is suggested to be under control of the keratinocytes themselves.

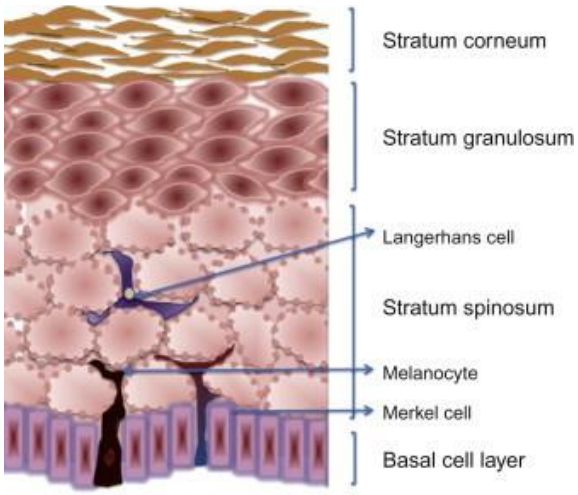

Figure 1.2: Structure of the epidermis [8].

1.2 MOTIVATION

The reason behind doing this thesis is classification of skin cancer based on the ISIC labeled skin lesion dataset which is a challenge to the human health. Skin cancer is the most common type of malignancy in the Caucasian population and incidence is on the rise. With only 1% of all skin cancers being comprised of melanomas, this constitutes the most dangerous form of skin cancer given its propensity to metastasize. Diagnostic processes based on visual assessment portray highly variable results and biopsies are required for confirmation. CAD-systems are therefore desired to allow early and accessible detection and enhance classification performance. This study focuses on the use of deep learning for classification of melanoma vs. non-melanoma skin lesions with a limited labeled dataset.

1.3 OVERVIEW

This work will focus on the use of deep learning CNN for skin cancer classification purposes. Rises in skin cancer, both benign and malignant, have been observed in the humans, with over 39 000 cases reported in 2019 as mentioned in [10]. Skin cancer incidence is furthermore expected to rise by another 5-9 % per year, depending on the type of skin cancer. Melanoma, representing only about 1 % of all skin cancer cases, is the most dangerous form given its propensity to metastasize as mentioned in [11]. Early detection and resection of melanoma, offers the. Patient excellent 5-year survival rates, but this drops severely as the melanoma progresses into the regional or distant

stage. In the current diagnostic practice for skin lesions, physicians rely on subjective systems by looking at specific characteristics of the lesions as mentioned in [12]. This renders the diagnosis highly dependent on the physician's training, previous experience and interpretation. Certainty is only obtained upon performing a biopsy, with many lesions thus being biopsied unnecessarily. In light of the need for early and accessible detection to improve survival chances as well as the high variability in the diagnostic process, instigating the performance of biopsies for finality, there is a clear window of opportunity for the use of CNN in detecting and classification based on ISIC skin lesion dataset.

1.4 SKIN CANCER RISK FACTORS AND PREVENTION

In the case of NMSC, there are strong personal and environmental risk factors. The personal risk factors include gender, genetic predisposition and age. The prolonged life expectancy encountered in current times may go hand in hand with an increased incidence of skin cancers as mentioned in [13]. Exposure to sunlight is the main environmental risk factor and should be the primary target of a prevention strategy. In the case of SCC, the effect of long-term, cumulative sunlight exposure will directly affect the risk. However, studies show a more intricate connection in the case of BCC's, where the risk seems to increase with excessive exposure during childhood, as well as indoor tanning. This thus offers an opportunity to put in place a prevention program focused on limiting direct sun exposure, supported by a strong educational component to inform the public. For MSC's, the relation with UVR exposure is less straightforward as mentioned in [14].

The second window of the prevention strategy is the establishment of an early detection program, which includes self-skin assessment and screening by the physicians [15]. In the case of melanomas less than 1 mm in depth, the 5-year survival rate is over 95% after local excision of the lesion as mentioned in [16]. As stated before, this drops tremendously as the melanoma transgresses in more advanced stages. This indicates the importance of early detection of melanomas. Furthermore, a correct classification of the type of skin tumor (benign or malignant and further subdivisions) is required for establishing the proper treatment plan, as different types require different handling as mentioned in [17].

1.5 SKIN CANCER PATHOLOGY

A cutaneous or skin lesion can be described as an abnormal change in skin tissue. It can be classified as benign, pre-malignant or malignant with respect to the nature of the lesion being non-cancerous, pre-cancerous or cancerous. In this respect, cutaneous or skin cancer refers to the proliferation of aberrant skin models as mentioned in [18]. A range of different factors is associated with skin cancer, amongst others particular environmental pollutants and arsenic in the drinking water. The most poignant cause however is exposure to ultraviolet radiation (UVR) (either from sun or artificial light).

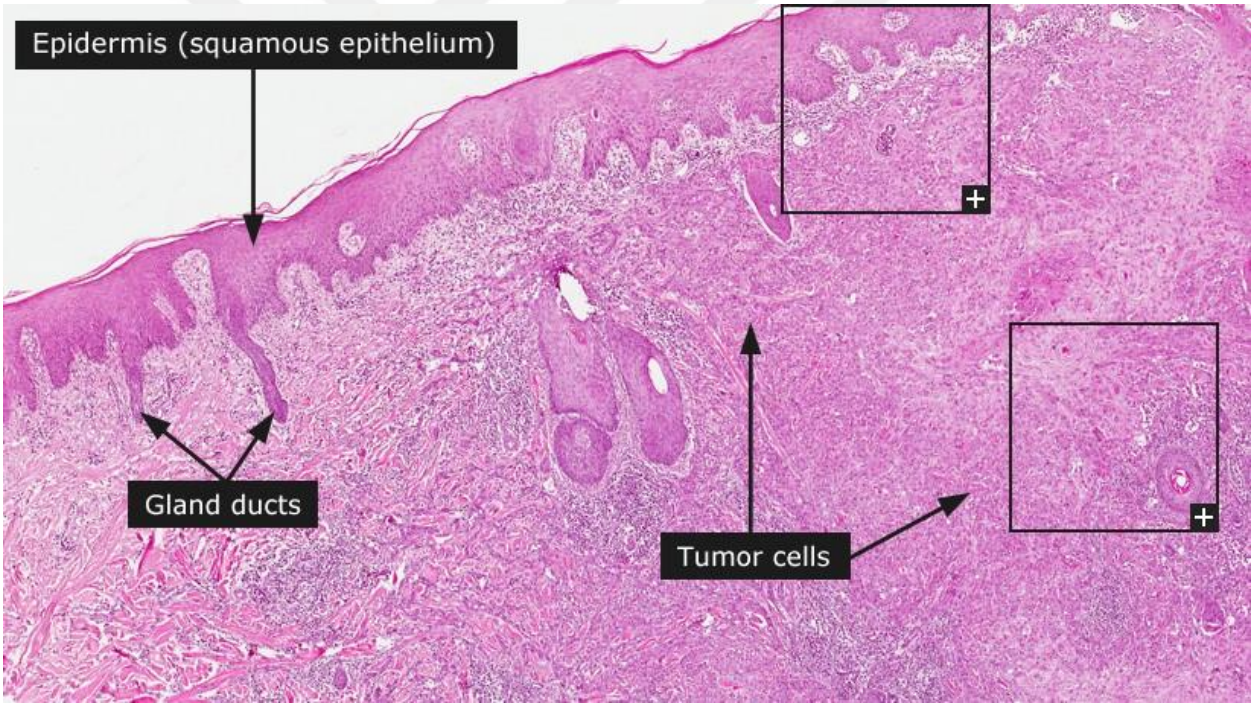

Figure 1.3: Skin cancer pathology with tumor models [19].

UVR is subdivided in three distinct categories according to the wavelengths it encompasses: (1) UVA with wavelengths of 315 - 400 nm, which makes up majority of UVR from sunlight as it is able to penetrate the ozone efficiently; (2) UVB in the 280 - 315 nm range is largely blocked by the ozone, only 1 - 10 % reaches the earth's surface (where increases in UVB exposure have been seen due to ozone depletion); and (3) the highest-energy UVC radiation (wavelengths of 100 - 280 nm) is completely filtered by the ozone layer and does not reach the earth's surface as mentioned

in [20]. Furthermore, UVA can penetrate the skin relatively deeply and enter the dermis. The higher-energy UVB radiation remains more restricted to the upper cell layers. as mentioned in [21]

Acute UVR exposure is associated with DNA damage, where direct effects are especially related with shorter wavelengths of UVB radiation. The photo-induced DNA lesions generally lead to covalent bonds between pyrimidine's, where the two primary types are: (1) pyrimidine (6-4) photoproducts and (2) pyrimidine dimers (CPD's) in a ratio of 1:4 (varying up to 1:10). Specific repair pathways of the affected models are put in place to remove the lesions [22]. If repair fails or the lesions are not repaired, they can result in sustained DNA mutations passed to daughter models upon cell division. If the damage is too severe or for models that have escaped repair, other mechanisms are called upon: (1) the cell can be arrested in the G1-phase of its cell-cycle, thus broadening the time window for DNA-repair before the cell enters the S-phase during which it replicates; (2) apoptotic pathways can be activated to induce cell death as mentioned in [23]. The key regulator in these processes is the protein, which is furthermore involved in a number of other protection mechanisms (e.g. inhibits angiogenesis, enhances correct DNA replication and repair, and maintains genomic stability). Acute effects also include sun burn erythema (inflammation due to photo-damage presenting as red skin), which is largely contributed to DNA damage from UVB exposure. Furthermore, the tanning process is initiated, which involves the construction of nuclear caps on the exposed cell-side to protect the underlying DNA, as mentioned before. Finally, UVR results in immuno-suppression of the immune system at work in the skin, which is composed of the Langerhans models, cytokines produced by the keratinocytes and lymph nodes [24].

Long-term effects of chronic or recurrent UVR have as clinical effect photo-aging of the skin. Whilst UVB radiation is highly efficient in causing direct DNA damage, UVA exposure causes the formation of free reactive oxygen species (ROS). ROS have been showing to interact with cell membranes, initiating a photo-aging pathway, and to indirectly damage the DNA, thus also being involved in skin cancer initiation as mentioned in [26]. The most prominent chronic UVR effect is associated with accumulation of DNA damage and sustained suppression. Respectively resulting in DNA mutations and enabling models to escape surveillance, these factors consequently form a highly favorable climate for carcinogenesis. This will be the primary subject of the following paragraphs, where the distinction is made between cancer affecting melanocytes or other skin cell types.

Different factors lie at the base of the carcinogenesis process. Mutations due to DNA damage generally result in an inactivation or under-activation of tumor-suppressant genes, whereas oncogenes are over activated. Furthermore, through new research the existence of cancer stem models has surfaced and indicated their impact on treatment strategies.

1.6 BIOLOGICAL VISUAL PERCEPTION

Biological visual perception was classically considered to occur through two separate retinocortical pathways. The existence of a ‘what’-pathway and a ‘where’-pathway was observed after the Second World War in veterans with cerebral missile wounds by researcher in [27]. Lesions in the parietal lobe impaired the maze-learning performance of the patients, whereas lesion in the temporal lobe resulted in a decreased closure ability, i.e. mentally reconstructing an image based on incomplete information. Further evidence for this model was acquired in which macaque monkeys were impaired with focal lesions in the cerebrum and investigated on an anatomical and electrophysiological level [28]. Their results indicated a distinction between two pathways originating from the primary visual cortex V1: a dorsal pathway projecting to the posterior parietal lobe and inferior ventral pathway to the temporal lobe, which are respectively involved in the spatial and perceptual functioning of biological vision.

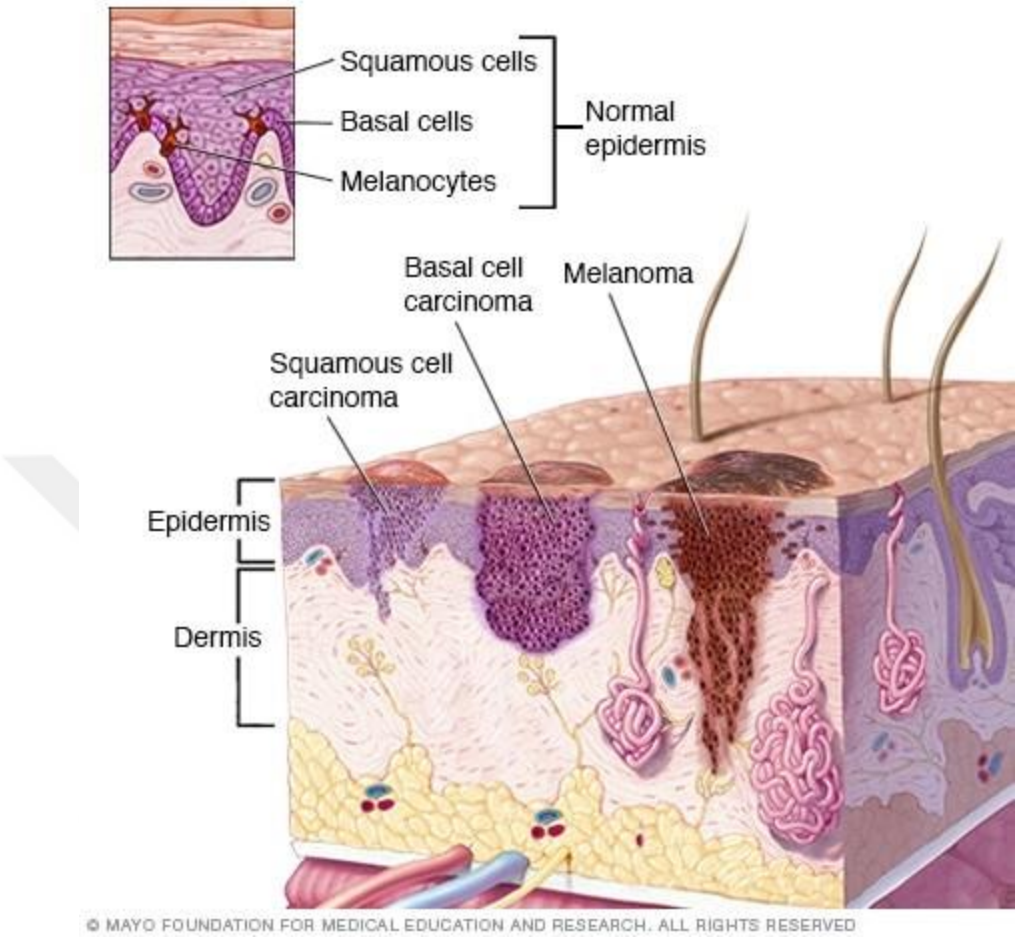

Figure 1.4: Simplified depiction of the visual streams of skin models in human skin [28].

As indicated in figure 1.4, both pathways receive their input from the photoreceptors at the retina and pass through many of the same cortical areas before parting ways spatially. They are however segregated on an anatomical level from as early as the ganglion models on the retina, which is reflected by the two pathways synoptically mapping at different layers in the same cortical areas they cross. At the lateral geniculate nucleus, the spatial system springs from larger neurons at the ventral side and is correspondingly known as the magnocellular-pathway, which processes luminescence-based spatial inputs. Contrarily, the parvocellular-pathway, which refers to the perceptual system stemming from smaller neurons at the dorsal side, conveys color-sensitive input without regard to spatial aspects.

Further distinctions can be made between the pathways with regard to their required outputs. The dorsal pathway projecting to the parietal lobe conveys the information required to execute an

action, i.e. a spatial description of the scene (position, orientation and structure), which is supported by the high temporal sensitivity of said pathway. The posterior parietal lobe receives input both sensitive to the spatial scene of objects and self-initiated movements, correlating the body's position to its surroundings and it further projects to the frontal lobe, where it thus plays a role in reaching movements of limbs and hand and finger grasping actions, as well as ocular control [29]. The ventral pathway on the other hand provides higher-resolution information more slowly to the temporal lobe with detailed perception as objective. The structures involved in the ventral stream play a role in visual perception, memory and learning.

1.7 CHALLENGES AND PROBLEMS

Through the availability of the ISIC Archive, CAD for skin lesion diagnosis has been a popular research area in recent years, as illustrated in the previous discussion. The diagnostic process for skin lesions generally includes an interview concerning the risk factors for different types, establishment of location, age, gender and arrangement of the lesions is taken into account and a visual screening is performed. With regard to melanoma's, the system, to detect early stage malignant melanoma's and is now used by physicians for a visual screening of pigmented skin lesions. The system relies on specific characteristics of melanomas, i.e. asymmetry, border, color and diameter. As rapidly evolving skin lesions may be indicative of melanomas, evolution was added as a criterion to this rule.

Even though this technique offers some semi-quantitative measures to assess skin lesion, physicians rely strongly on their own knowledge, experience and subjective perception for the classification of the lesions. Furthermore, the technique does not allow to assess depth of the skin lesion. Breslow depth or thickness, which can be defined as the depth of tumor invasion with respect to the granular layer of the epidermis, is however a feature with strong prognostic value and should thus be taken into account. Finally, deviant lesions that do not show the typical characteristics of melanomas will go unnoticed through this assessment. Other methods have since been proposed, which include the seven-point checklist.

1.8 AIM OF CONTRIBUTION

The research in this works aims to contribute to the following key aspects of the use of deep learning for skin lesion classification on ISIC dataset:

Evaluating the best-performing deep learning scheme for ISIC skin lesion dataset classification.

We ran ResNet50, VGG16, InceptionV3, VGG19, Xception, MobileNetV2, MobileNet architectures to find out the best model for that suits our type of dataset.

The development of skin cancer detection system using seven hybrid CNN based models on skin lesion diagnostic tasks is in hot pursuit, with special attention being given to the classification of benign and malignant.

Exploring different strategies to deal with class imbalance and early over-fitting on the ISIC dataset.

Proposing a seven hybrid techniques with convolutional neural network to exploit the use of labeled data to get better results as no body tried this approach before in this domain.

Increasing the accuracy of classification and comparing it with the previously done work in this research industry using CNN.

2. LITERATURE REVIEW

Deep learning for skin lesion classification has long been a 'hot' topic with many attempts at improving the diagnostic performances of such systems [30]. Past approaches have been incorporating two main steps: (1) hand-designing a feature extractor after image processing and (2) training a classification algorithm on these features. The bottleneck in these approaches is clearly the strength of the hand-crafted feature extractor, which tend to imitate the diagnostic tools physicians use to visually assess lesions. However, the high variability in diagnostic performances of physicians portrays that they might not be able to establish the most characteristic features of skin lesions. Therefore, recent trends in research have portrayed a shift towards deep learning in this domain, with the aim of letting the algorithm itself now decide on features, possibly discovering ones that humans have not been able to notice.

2.1 MACHINE LEARNING FOR SKIN LESION CLASSIFICATION

Summarizing the previous discussion, it is possible to state that early detection of melanomas increases survival rate, visual diagnosis is complicated due to inter-patient variability in lesion appearances and subject to the physician's prior knowledge/experience and final diagnosis requires skin biopsies as mentioned in [31]. As dermatoscopy provides high-resolution, magnified images of skin lesions, computer-aided diagnosis has surfaced as a tool to support skin lesion diagnosis.

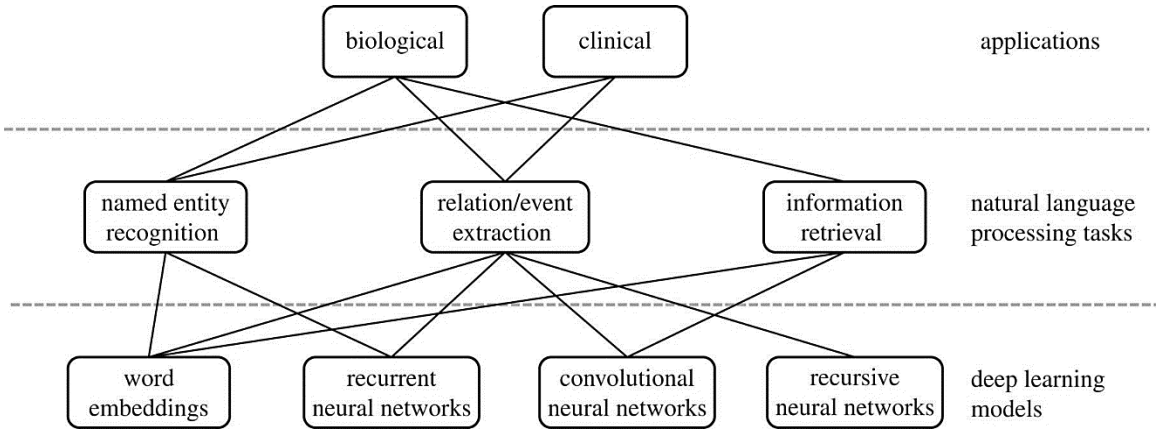

Figure 2.1: The biological and Clinical approaches for skin cancer detection exploited in machine learning and deep learning [31].

2.2 CLASSICAL MACHINE LEARNING APPROACHES

Researchers in [32] represented a thoughtful procedure regarding to skin cancer detection. They proposed a method to detect melanoma by using image processing. The images they used as input of their system were of skin lesions which was suspected to had skin cancer. At the first step, they pre-processed their images and enriched the image quality. For image segmentation they used edge detection and automatic thesholding technique as mentioned in [33]. After that, they extracted the geometrical features of the images and performed lesion area analysis. The reason they used geometrical features was as it was the most dominant features of the lesions. They used the extracted features to classify skin lesions whether it had cancer or not by comparing the features extracted with the predefined thresholds. They created skin lesion mask and used it as input images to get the segmented images. The segmented images of skin lesions could detect melanoma quite precisely.

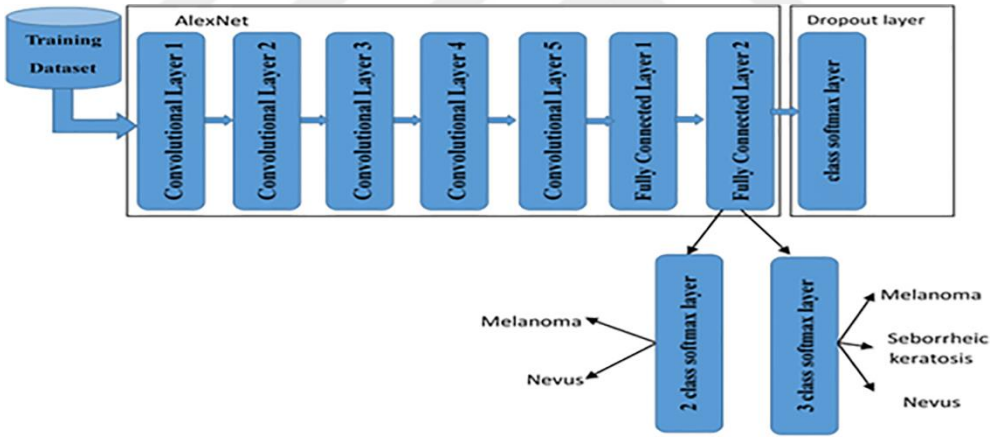

Figure 2.2: Previously held literature on lesion classification using transfer learning and Alexnet [33].

Researchers in [34] proposed a methodology of a fully automatic system to determine lesions using dermatological infection recognition, an approach by machine in opposed to detect the lesions visually by medical expert. His model was designed in three phases. The first phase was to compromising of data collection and data augmentation. As the image pre-processing technique they performed shade removal, glare removal and hair removal and in that way they were able to identify features like shape, color, size and texture as mentioned in [35]. The second phase was about to design their architecture and it consisted of feature extraction and segmentation where

segmentation was done by three techniques which are Otsu segmentation technique, modified Otsu segmentation technique and Water shed segmentation technique as mentioned in [36]. The last and most important phase was to predict the lesions by designing and training the architectures. They used Back Propagation Algorithm, SVM and CNN to train their dataset.

2.3 RELATED WORK

Researchers in [37] proposed an approach of deep learning system using a server with graphic processing unit (GPU), which was proposed for recognition of melanoma lesions. In their proposed system, they used input images which have illumination and noise effects. They used their system to reduce those occurrences. In that way, they increased the quality of the images and subsequently used those images in a pre-trained convolutional neural network architecture. Their proposed method performed better in finding accuracy than existing latest pre-trained CNN architectures as mentioned in [38]. They built an automatic diagnosis system using deep learning methods to detect melanoma in skin. Since their dataset had low number of images, they performed data augmentation such as resizing, cropping and rotating the images. The output result of their proposed system increased exceptionally by it.

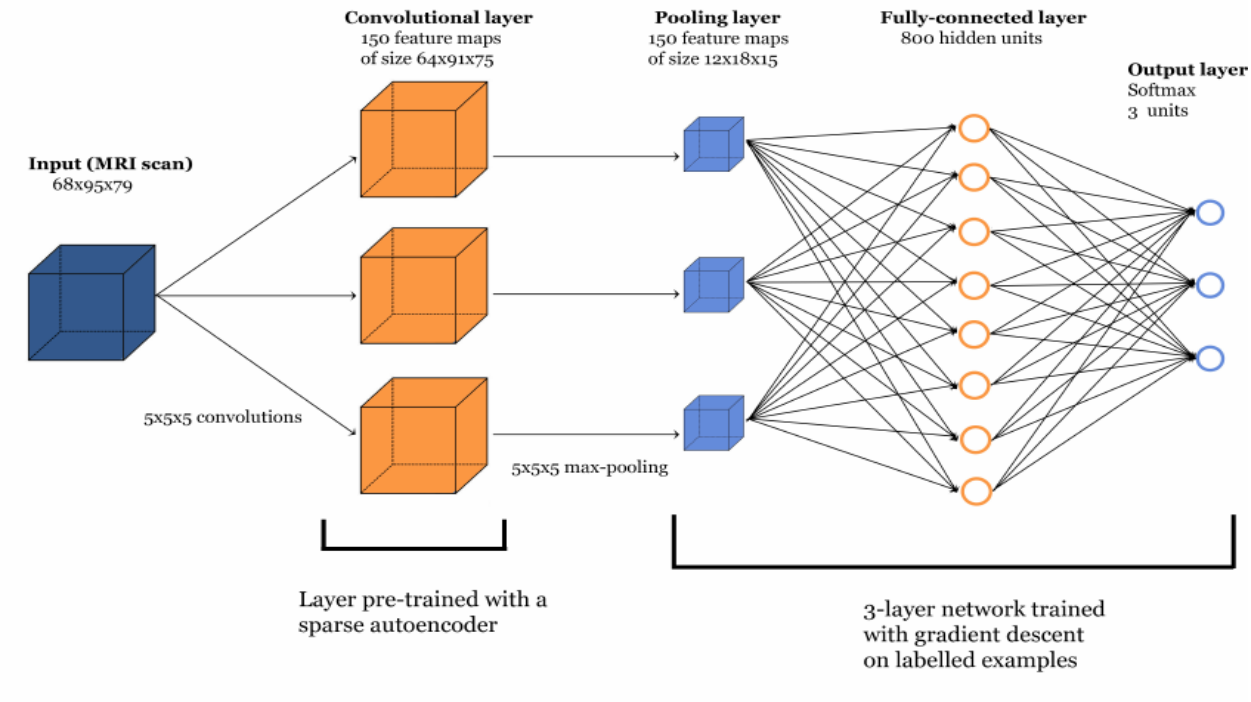

Figure 2.3: A robust pattern recognition using the convolutional neural network [38].

Researchers in [39] proposed a method for precise extraction of lesion region which is based on deep learning methodologies. To lessen noisy artifacts, the input image, after pre-processing, was applied to CNN. CNN produced a segmentation mask which detected the area of the lesion. Using some post processing operations, the quality image of the mask is being further improved. Their input images were produced by customary cameras; henceforth, those were pre-processed in order to manage noisy artifacts. They used a filter to decrease the noise of the images as mentioned in [40]. They created two patches, one is local patch and another is global patch. The local patch is illustrated as a window around each pixel. It shows the local texture around the middle pixel as mentioned in [41]. Whereas, the global patch is used to show the global structure of the area. After that, both the local and global textures are sent into the proposed CNN architecture as the training input. The experimental outcomes showed that their suggested method can outpace other architectures to detect lesions.

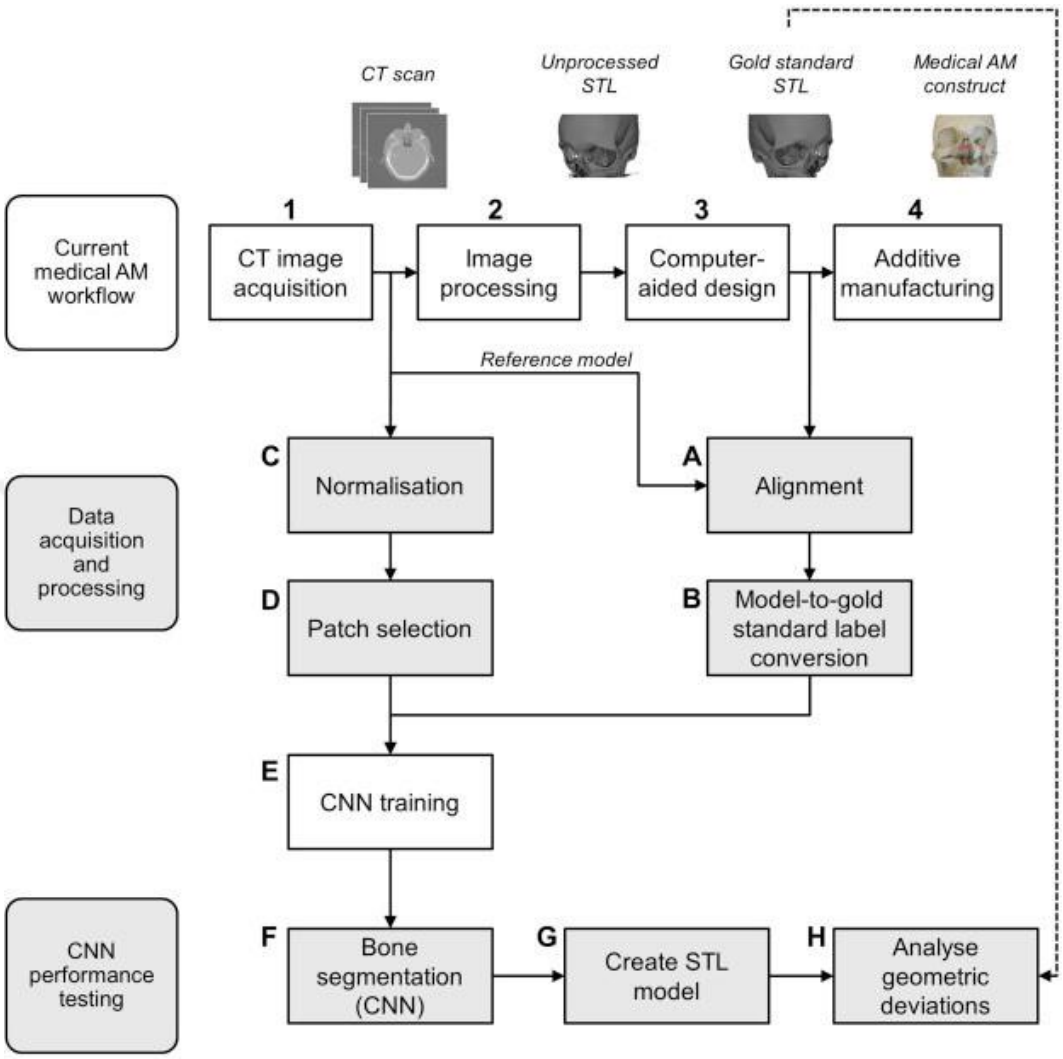

Figure 2.4: The disturbed skin intracellular signaling process for detection of cancerous sample with CNN [41]

Researchers in [42] proposed an innovative automated methodology to detect and recognize skin lesions using CNN. They performed three steps. The first step was to enhance the contrast of the images using first local laplacian filtering with HSV color conversion. The second step was to extract lesion boundary using XOR operation of color CNN methodology. The final step was to extract the features by applying transfer learning by using InceptionV3 in order to feature fusion using hamming distance technique as mentioned in [43]. They introduced a feature selection technique by controlling entropy which was presented for the collection of the most discriminant types. Their proposed system was verified by different datasets. They achieved excellent accuracy using their proposed system.

2.4 TYPES OF SKIN CANCER

2.4.1 Non-melanoma skin cancer

Non-melanoma skin cancer (NMSC) is the most frequently diagnosed type of skin cancer in the Caucasian population and consists primarily of keratinocyte tumors, subdivided in basal cell carcinoma's (BCC) and squamous cell carcinoma's (SCC) as mentioned in [45]. To give an indication of their prevalence, we report the statistics documented by the American Cancer Society in the US: an estimated of 5.4 million BCC's and SCC's were diagnosed in 2016 [46], spread out over 3.3 million Americans. BCC is the most common type and is responsible for up to 80% of all NMSC cases. They occur mostly on sun-exposed parts of the body, such as the neck and the head, but are however slow-growing and unlikely to spread to other parts of the body. SCC represents 19% of skin cancer cases and it also affects sun-exposed areas such as hands, ears, arms and legs, as well as damaged skin (e.g. by chemicals or radiation or scar tissue). Even though it is more likely to spread to other areas, with 2-5 % of SCC's portraying this behavior, it still remains rather unlikely. Rarer types of NMSC, together contributing only to 1% of all NMSC cases, include Merkel cell carcinoma, cutaneous lymphoma and Kaposi sarcoma [47].

Given that NMSC's can predominantly be found in sun-exposed areas, the role of UV radiation is especially pronounced in these cases [44]. The accumulation of DNA-damage generally results in mutations, both in tumor suppressant genes (which tend to be inhibited) and oncogenes (which are subsequently overexpressed). In over 50 % of all human skin cancers (and more broadly, in over 50 % of all cancers), mutations can be found in TP53 gene, the tumor suppressant gene responsible for producing, which is as discussed before a key player in maintaining the cell in a healthy state under external stress [43, 47]. More specifically, about 50 % of all BCC and up to 90 % of all SCC have a TP53 mutation. Other genes are implicated in the NMSC carcinogenesis, including: CDKN2A locus (encoding for two different promoters that arrest the cell cycle) and RAS oncogenes [45].

2.4.2 Melanoma skin cancer

Even though invasive melanomas only make up 1% of all diagnosed skin cancers in the US, they are responsible for over 75% of all skin-cancer induced deaths. According to the American Cancer

Society, melanoma of the skin is the 5th cancer with highest incidence in the US, following breast cancer, lung and bronchus cancer, prostate cancer and colorectal cancer [48]. An estimated number of 87,100 of new cases were observed in 2017, as well as 9730 melanoma-induced deaths. The trends over the years in the US show an increase in incidence of melanomas, whereas the mortality rate remained quasi-stable. The increase in incidence can be related to changes in human behavior (e.g. sun tanning), as well as the aging population. The number might however be biased by an increase in screening methods as well.

The documented incidence in Turkey reported the diagnosis of 2635 new cases of malignant melanoma in 2016 [44]. At that time, it was the 4th most frequent cancer in females and 7th most frequent cancer in males. Projections about cancer incidence for the period 2014-2024 in Turkey point out a similar increase in malignant melanoma incidence. An increase of about 49% in the yearly incidence is expected, which in absolute numbers entails an increase of 2925 to 4356 new cases each year [49].

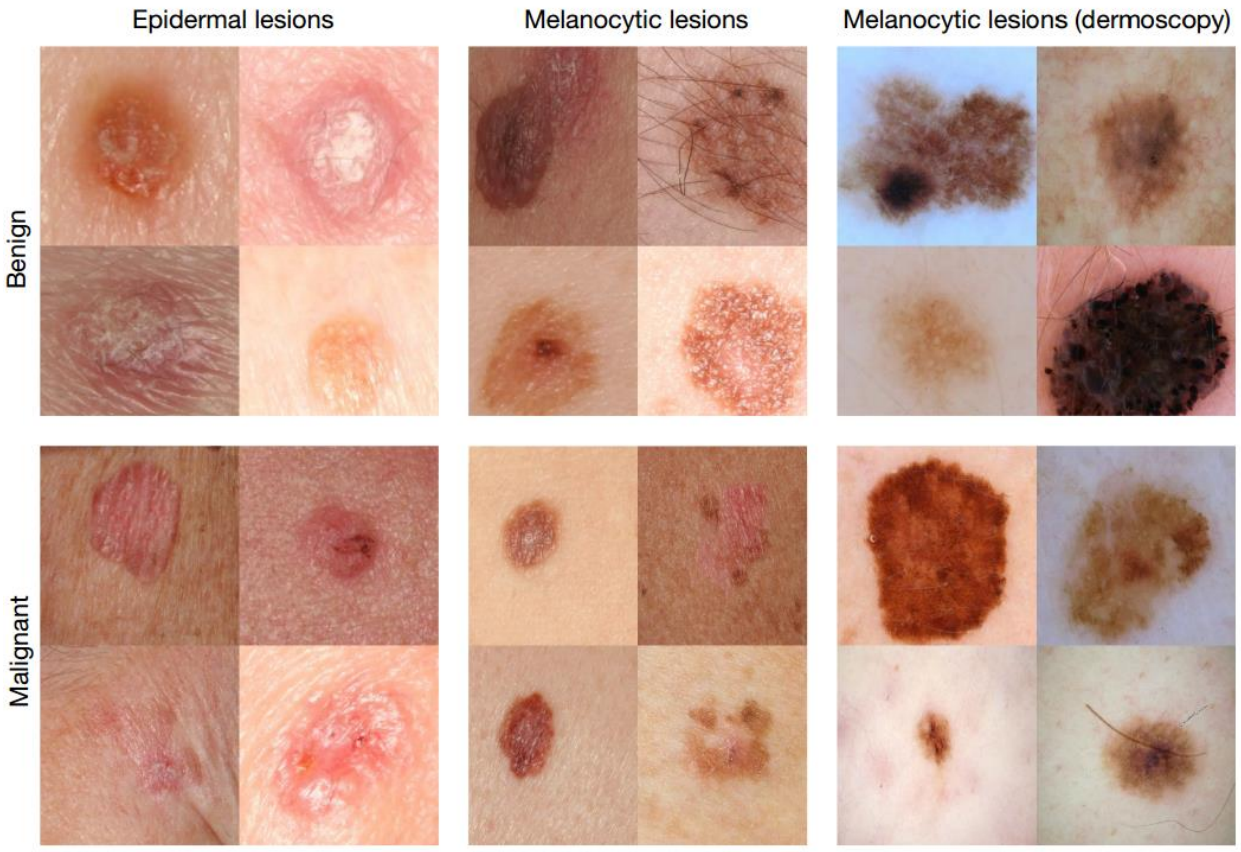

Figure 2.5: Types of skin cancer including benign and malignant [45].

Two major genetic processes underlie the propensity of aberrant melanocytes to proliferate and metastasize. Firstly, gain-of-function mutations in genes responsible for handling melanocyte proliferation and survival are over activated. Identified oncogenes include BRAF (mutated in 50-70% of melanomas and) and NRAS (mutated in 15-30% of melanomas) which stimulate a sustained ERK activity. ERK is a cytosolic protein kinase involved in the pathway and its hyperactivity results in augmented melanocyte proliferation and survival [35]. Furthermore, mutated BRAF genes facilitate tumor maintenance and neo-angiogenesis. New studies show potential hyper activation of the pathway, involved in cell survival, proliferation, growth and motility, in many melanomas. Furthermore, melanocyte functioning seems to be critically interwoven with the concentrations of MITF (microphthalmia-associated transcription factor), where only at intermediate concentrations proliferation occurs, and it can collaborate with the mutated BRAF to stimulate proliferation [39]. Secondly, genetic alterations in tumor suppressor genes can allow melanocytes to evade entering a state of senescence (i.e. complete stop of cell proliferation), which normally occurs upon cell exposure to potentially oncogenic stress. Identified gene loci include the gene, responsible for producing (amongst others) the tumor suppressor protein, and mutated functioning, both causing an inactivation of these tumor suppressant pathways.

Through a disturbed intracellular signaling process, melanocytes escape the strict control of the keratinocytes and consequently start to proliferate and spread. The different phases of this process as portrayed, shows an initial proliferation confined to the epidermis, dermis or both, resulting in a typically benign nevus or dark spot. The radial growth phase (RGP) is generally considered as the primary malignant phase and results in some local micro-invasions of the dermis. Upon reaching the more dangerous vertical growth phase (VGP), clusters of proliferated models can be found in the dermis, which can easily metastasize to another region. Their propensity to metastasize is exactly what makes these melanomas so dangerous. The 5-year survival rate for melanomas in the localized state is 99%, this drops to 63% as the melanoma enters the regional stage and further to 18% upon reaching the distal stage [50].

3. METHODOLOGY

It is notable that the skin cancer has successively used deeper networks but advanced deep learning techniques will provide great efficiency due to its architectural flexibility. Having more data models possess greater representational power making it possible to approximate more complex functions. Also, fast development in the field continuously introduces new techniques and architectures expanding the toolbox for design. However, science has always embraced simplicity, and it is reasonable to argue that the network should be kept as simple as possible. A study has been using the CNN technique even questioned the common practice of using pooling and sophisticated activation functions in data mining. Instead, they suggest using just recurrent layers with properly selected stride and filter size parameters

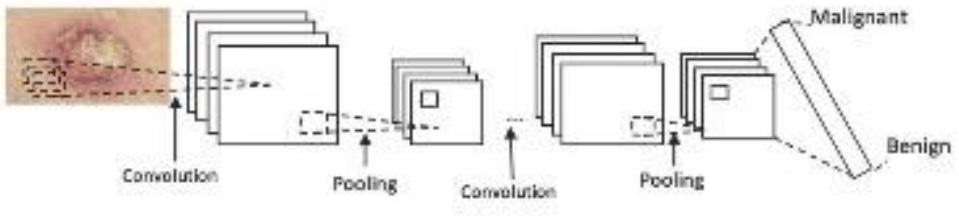

Figure 3.1: Proposed system depicted by flowchart using CNN.

3.1 SKIN LESION DATASET

The International Skin Imaging Collaboration (ISIC) has been created in an international attempt to construct a large, publicly available database for skin lesions images, both clinical and dermatoscopic. The images in the dataset are all taken with the same scanner device, so they are relatively standard. There is also a calibration process involved when replacing devices, so in theory, the images should be comparable to ones taken with other devices. Still, there is some variance in the images regarding size and positioning. Some images have blurriness, e.g. due to high amount of fat in the patient's body. Also, the patient must hold very still while the image is taken or movement artefacts may occur. The width of an image is standard 300 pixels, but the height depends on the length of the scanning process, varying from 134 to 186 pixels in the dataset. Furthermore, the collaboration focuses on creating standards for current dermatoscopic/clinical imaging practices, as they tend to vary in type of technology, technique and terminology while taking into account privacy and possibility for sharing the data between different platforms. A

yearly challenge has been set up since 2017 which has, based on a selected training, validation and test set from the ISIC Data Archive, a three-fold goal: (1) lesion segmentation, (2) dermatoscopic feature classification task and (3) disease classification.

3.2 CONVOLUTIONAL NEURAL NETWORK

In this architecture, the CNN are responsible for feature extraction, where the features to which they will respond are determined through a learning process of the variable input connections. Simple features, such as specific orientations of lines and edges, will be extracted by lower-level CNN models, while higher-level CNN models extract more global features (e.g. parts of learned patterns). The CNN model receive their inputs through fixed connections from CNN models in the previous layers and are responsible for a robust pattern recognition (i.e. decreasing sensitivity to a deformation or location shift of a pattern). Each CNN model receives inputs from a number of CNN models (’the input window’) in one cell plane, i.e. from CNN models that extract the same feature, but at different spatial locations. The models will fire, if it receives an input from at least one of the models. Consequently, the feature will be detected even under a shift in the location of the input features, rendering the system less sensitive to the exact feature locations. The behavior of the models can however also be interpreted from an alternative point of view. The input windows for the different models strongly overlap thus models can be considered as performing a spatial blur on the excitatory signals they receive from the models. This spatial pooling is obtained by averaging these signals from the input window, which are models that perceive the same feature at slightly different locations. Furthermore, the excitatory cell input window is often framed by a small inhibitory region. This enables the models to distinguish a continuous line from the end-point of the line and allows the network to still recognize features after blurring. A consequence of the fixed input windows of the models is a reduction in size of the CNN-planes, thus allowing the network to compress data as it progresses to higher stages.

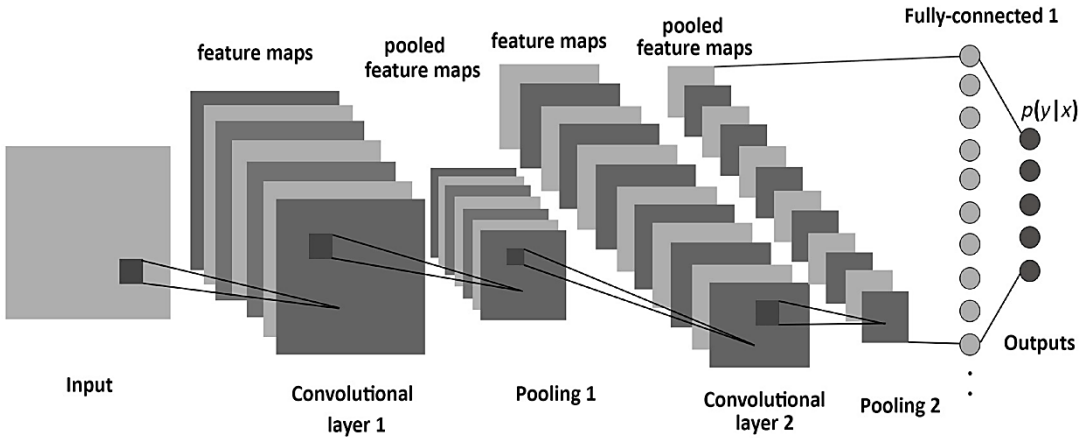

Figure 3.2: Illustration of a deep convolutional neural network.

CNN stands for convolutional neural networks originated from this work whose network has a similar architecture cognition. It translates the behavior of the models in mathematical formulations, respectively as convolutions and subsampling. The network primarily distinguishes itself however from the cognition in its training process. Whereas the neo-cognition relies on a layer-wise, unsupervised training process of the cell layers and supervised training of the output layer, CNN is trained to find a global minimum over all the parameters through back-propagation. The key notions that mimic biology and render CNN's so apt for image classification.

3.3 POWERFUL CNN'S FOR SKIN DISEASE CLASSIFICATION

3.3.1 Key concepts of CNN's

Convolutional layer as images contain highly correlated 2D information (generally in three RGB-channels), the idea of local connections has been exploited by creating units in a convolutional layer which receive weighted input from local patches in the previous layer, referred to as their receptive fields (RF's). By relying on local RF's, this introduces sparse connections, rather than the whole-scale image statistics passed on to units and causing high-dimensional problems in fully-connected feed-forward networks. The weight vector of a unit is also known as filter bank and through training the weights will be updated to filter for specific features, giving the units the behavior of filter detectors. The weighted inputs and additive bias are then passed through the

activation function of said unit, where ReLU's have portrayed much faster learning in deep networks compared to other activation functions.

Furthermore, images tend to portray similar features at different spatial locations. Accordingly, units with receptive fields at different locations will have the same weight vectors to extract similar features, which is known as parameter sharing and further reduces the dimensionality of the problem.

On a structural level, the outputs of units with the same weight vector are organized in feature maps. Multiple feature maps (with different weight vectors) in each convolutional layer enable scanning for different features in each local image patch. The optimal number of feature maps needed to capture the image dynamics tends to increase with the complexity of the input images.

3.3.2 Pooling layer

The feature maps in the convolutional layer serve as input for the next stage, being a pooling layer. Given that the relative location of features with respect to each other contains significantly more information than their precise pixel-location, further stressed by the notion that small shifts in the precise feature locations are to be expected in different images, the idea of local pooling of features to reduce spatial resolution is implemented at this stage. This results in a network invariant to small shifts in the images and the stages of convolutional/pooling layers equip the network to handle variable size inputs.

Pooling units will scan the feature maps with (non-)overlapping p1 x p2 receptive fields and output a single value. Typical choices for pooling units are subsampling, where the output of a unit in the jth layer is the average of units in the receptive field, supplemented with additive bias b and passed through an activation function and max pooling.

3.3.3 Deep hierarchies

Finally, the use of many layers represents the different stages in the simplified model of the ventral stream. Features at higher-level stages results from combinations of lower-level features, thus enabling deep structures to capture the compositional hierarchy of natural images.

3.4 HYBRID MODEL FOR SKIN IMAGES

An implementation is based on hybrid models of CNN from various approaches. The visualization method proved to be helpful in interpreting the predictions, although it produced different results than expected. The expectation was that the CNN model would learn to identify small signs of skin disease risk. Instead, with custom CNN models they learned wider patterns that possibly predict resistance to fractures. This can of course be seen as the other side of the same thing. The experiments show that deep learning techniques can be used with imaging techniques to produce predictive ability different from the standard clinical methods currently in use. However, due to limited amount of data and wide range of architectural possibilities, their true potential in this context remains to be unfolded. It is also likely that the data analysis alone cannot provide the best fracture risk prediction, regardless of the model. It could contribute to more comprehensive prediction method that utilizes a wider set of input data to cover different aspects of the skin disease risk. Using hybrid deep learning techniques produced the best skin disease prediction results in the experiments. The results support the advertised benefits of hybrid deep learning, such as faster training speed and suitability for larger datasets.

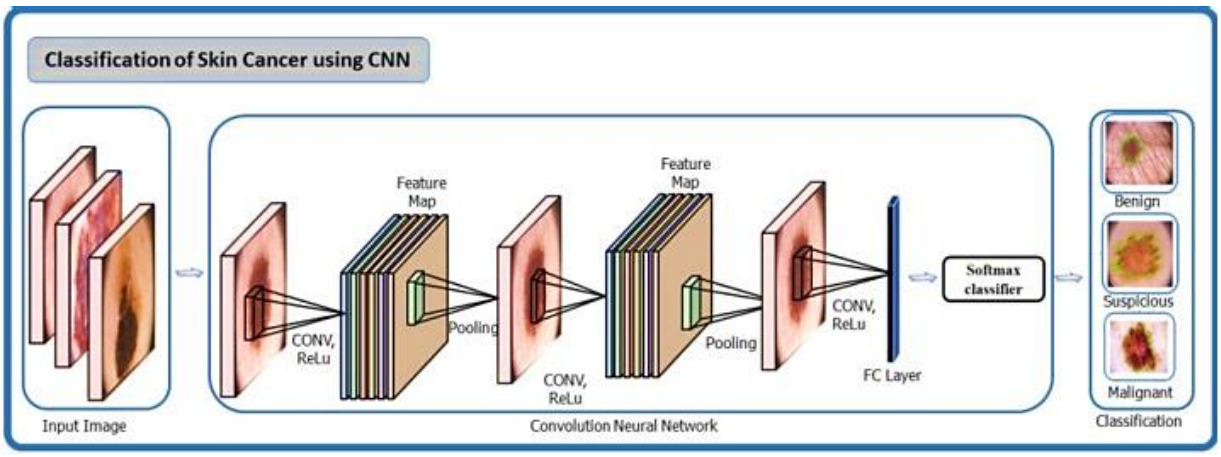

Figure 3.3: The hybrid CNN model for skin disease detection.

Given the current hype of deep learning, optimizations of CNN architectures and designs is in hot pursuit. This section will therefore be devoted to some of the powerful CNN models that have been developed in the past few years and will be used throughout this work. Specific, revolutionary design-elements that have assisted these improvements are marked throughout this discourse.

3.4.1 Cross Validation

The cross validation is used to validate the data in artificial intelligence based systems, interpretable machine may even teach humans about how to make better decisions. To address the transparency problem, visualization techniques can be used to identify the image regions most important for the model's prediction. It is also possible to highlight the shapes and textures the network sees by visualizing pixel-wise impact on prediction score. Another visualization technique is called guided backpropagation. It produces an illustration of the detected features that contributed to the skin class prediction. This is done by taking account only the neurons that had a positive effect on the output, while others are set to zero. Combining different visualization techniques can produce effective visual explanations for the CNN model's predictions. To gain better coverage of the problem space, the number of training samples can be artificially expanded. This method called cross validation means modifying samples with random transformations such as rotation and flipping, varying properties like brightness and contrast or by adding random noise.

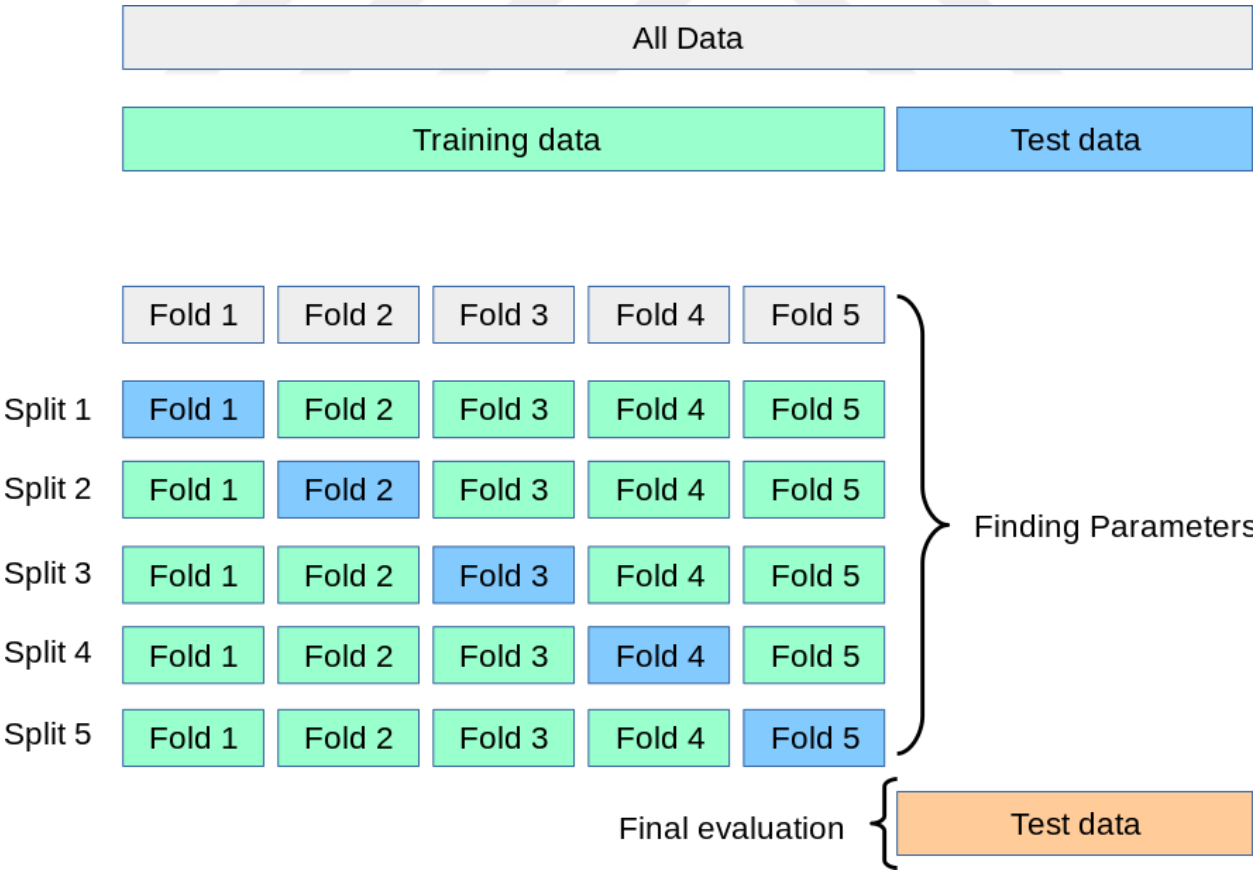

Figure 3.4: An example of dividing dataset for cross validation.

3.4.2 Xception Model

The name Xception stands for Extreme version of inception with depth-wise separable convolution. Depth-wise convolution means the channel-wise nxn spatial convolution. If a layer has five channels, then there will be 5 nxn spatial convolution. Xception is considered better than Inception v3 on various datasets. There are two types of depth-wise separable convolution, one is original depth-wise separable convolution and modified depth-wise separable convolution. Original depth-wise separable convolution is the depth-wise convolution followed by a pointwise convolution. Pointwise convolution is the 1x1 convolution to change the dimension. Depth-wise separable convolution does not perform convolution across all its channels like conventional approach. For this reason, this is a convenient approach since the runtime gets optimized and the model shows better accuracy also the model gets lighter. This modification is inspired by 1x1 convolution done by Inception v3 before performing nxn spatial convolution. Besides, to show the mean accuracy prediction when multiple objects appears in a single image, Xception is the most fruitful approach. Figure 3.5 shows the original depth-wise separable convolution in Xception.

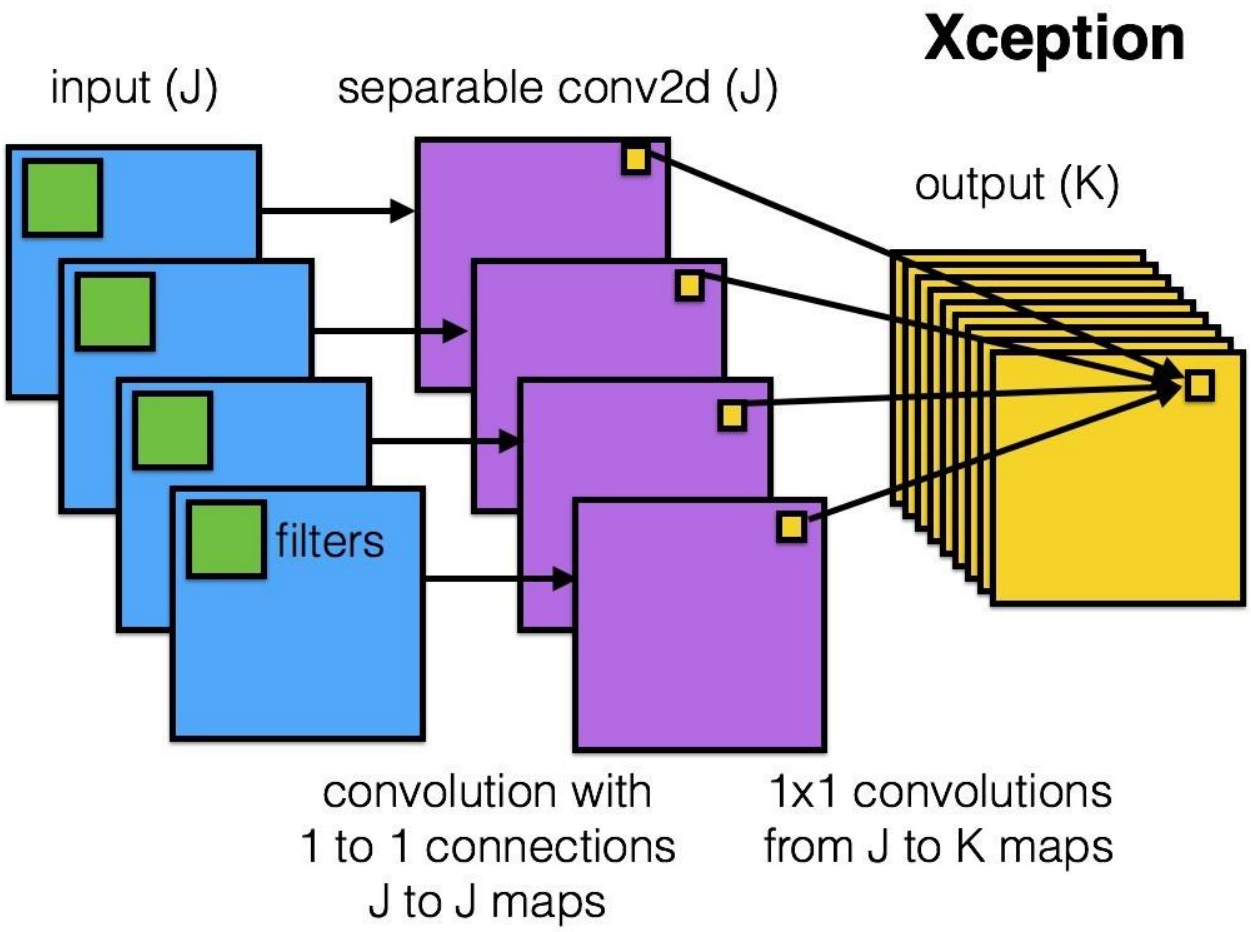

Figure 3.5: The modified depth-wise separable convolution in Xception.

The final classification layer was replaced by the skin cancer classification task (disease classes, generated from a disease partitioning algorithm) and the whole CNN was fine-tuned. For the three-class classification task between benign, malignant and non-neoplastic lesions (inferred from the broader disease partitions), a nine-fold cross-validation scheme portrayed an overall accuracy of 85.3 ± 0.9%, which is the averaged accuracy over the individual classes. Furthermore, they reported an AUC-score of 85% for melanomas.

3.4.3 Training and Validation

As deep learning techniques take numerical data as input, image analysis might seem like a simple case of feeding the network with raw image data. After all, images are nothing but grids of numbers representing pixel intensities. Actually, this approach can work well for simple recognition tasks

with small, normalized images. The problem is poor scalability. When the input size and number of layers' increase, the amount of required computation explodes.

Table 3.1: The number of epochs per training instances.

Number of Epochs	Loss	Accuracy	Validation Loss	Validation Accuracy
5	0.3301	0.8492	0.6911	0.5360
10	0.2418	0.9047	0.6942	0.5360
15	0.1891	0.9350	0.7985	0.5360
20	0.1702	0.9426	0.9698	0.5379
25	0.1384	0.9606	0.4709	0.7803
30	0.1246	0.9625	0.3569	0.8220
35	0.1020	0.9768	0.3543	0.8295
40	0.0809	0.9829	0.3598	0.8371
45	0.0747	0.9853	0.3662	0.8295
50	0.0678	0.9867	0.3709	0.8314

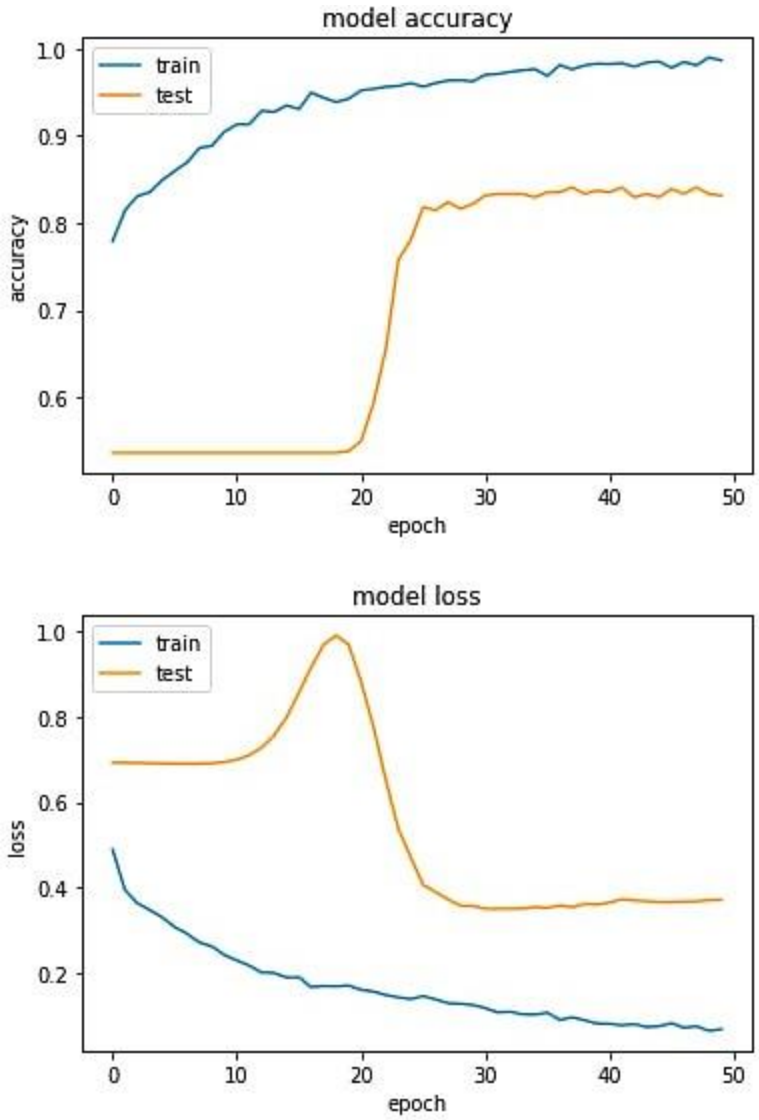

Figure 3.6: Xception Model Accuracy (accuracy vs epochs) Model Loss (loss vs epochs)

3.5 IMAGE PRE-PROCESSING IN PYTHON

Possible sources of variation include the skin diseases, disease status, topography, skin diseases properties, skin conditions, detection methods and other classification influences. Due to such local variations, the task of accurately differentiating between several field skin diseases can become ambiguous. This makes it difficult to create robust image processing algorithms that recognize all kinds of field instances without overfitting to the specific scene. In addition to the variation within a single scene, satellite images can show even stronger intra-class variations. This applies to images from different skin regions, from different dates or growing seasons, or under different

atmospheric and illumination conditions. Also, in view of increasing public availability of high quality skin image datasets, the use of additional and more heterogeneous ground truth and skin image data (e.g. from multiple study areas in parallel) could improve the general model accuracy, robustness and transferability. Additional data augmentation (e.g. random scaling, distortions, color offset or jitter, introduction of image noise) might also help mitigate overfitting and increase the robustness to different detection settings as well as image lightning conditions.

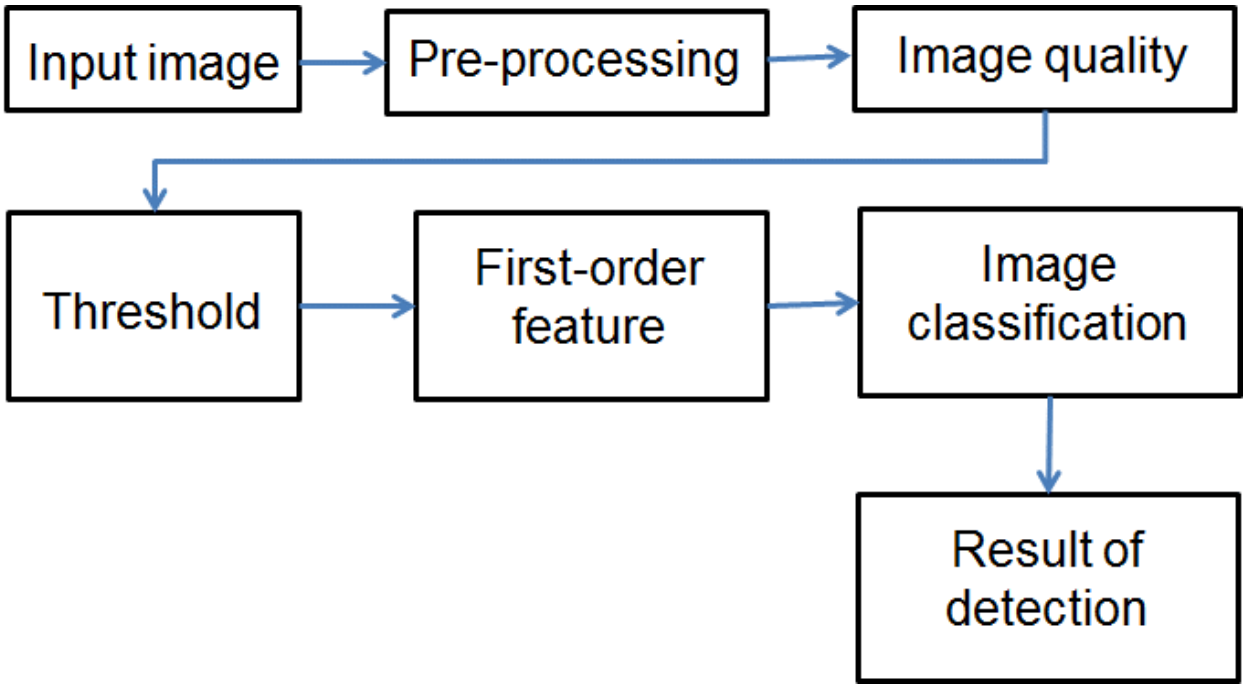

Figure 3.7: The flowchart approach for skin cancer detection.

Often cited to demonstrate the power of CNN's for image recognition tasks is the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Consisting of a publicly available dataset, labeled through crowd-funding approaches, spanning 1000 different object classes, it yearly sets up a competition and is often used as bench mark for new CNN architecture performances for object recognition. In the proceeding discussion, the ILSVRC will serve as a guideline to discuss the most poignant results obtained with deep CNN's with indication of some remarkable new elements. To give an indication of the performance of these models, the top-5 error rate on the ImageNet dataset of the aforementioned nets is presented in Table 3.2. It has to be noted that the submitted nets, training and testing schemes offer a myriad of other hyper-parameters, which were tuned to achieve these results (e.g. a 7-network ensemble of VGG-architectures). From these results, it can

be observed that ResNet, as well as other more recent nets (e.g. InceptionV3, InceptionResNetV2), succeed in surpassing the documented human performance (5.1 %) on this dataset.

Table 3.2: Number of layers in each pre-trained model with respective error rate.

CNN (Pre-trained)	No. of layers	Top-5 error rate
AlexNet [27]	7	16.4 %
VGG (16-19)	16 - 19	7.3 %
GoogleLeNet [29]	19	6.67 %
ResNet50	152	3.57 %

3.6 COMPILING HYBRID MODELS

The research focuses on compiling the delineate details and classifies skin image while distinguishing between skin diseases instances of the same class. Instance segmentation lies at the intersection of skin disease detection (predicting the bounding box and class of a variable number of skin images, but not segmenting them) and semantic segmentation (labeling each image pixel with a semantic category label, but not distinguishing between skin instances of the same category). With semantic segmentation it would be possible to find single instances of skin disease fields if they were completely surrounded by areas of different image classes. However, skin disease field are often directly adjacent to each other and have touching boundaries. In this case, skin disease segmentation would yield the disease detection. Skin instance segmentation is able to distinguish between these connected or overlapping instances of the same class. Deep learning has not reached mainstream adoption in the remote sensing community yet.

These datasets usually skin from manual tracing of field boundaries. However, manual tracing of skin image is heavily dependent on the used imagery and supplementary data. It also is a highly subjective task, inevitably leading to inaccuracies and ambiguities depending on the priorities of the operator. The ground truth dataset can be complemented by existing property parcel information that could be indistinguishable based on the imagery alone. Furthermore, a single field skin disease can potentially show several subfields due to unreported land use changes within the growing season. Given these constraints, even a very powerful automated image processing

algorithm can only become as good as the ground truth data but will never be in perfect alignment with reality. Although the model can generalize within certain limits (local generalization), its transferability is mainly limited by the variability of the training data.

3.6.1 Motivation for Hybrid CNN Models

However, the ISIC 2017 dataset poses a unique problem with millions of annotated images in which each class is sufficiently rep resented, thus rendering it highly dissimilar from the often scarce and unbalanced real-life datasets. Furthermore, the training of the aforementioned deep CNNs generally required multiple GPU's and was a time-consuming process (up to 3 weeks). Thus, it is clear that training the same, high-performance CNN architectures on real-life data is in most cases unfeasible. The CNNs trained to recognize the ImageNet dataset have nevertheless a strong capability to extract very distinctive features from natural images, rendering them useful for datasets of other images. This is supported by the idea that the lower levels of these architectures extract general features from the images and become increasingly more task-specific deeper in the network. From this the idea of transfer learning sprung forth, which aims to kick start the process of the random weight initialization by using the pre-trained weights of deep CNN's to facilitate the training process.

3.6.2 Key Notions Concerning Hybrid CNN Models

Transfer learning entails training a source network on a source dataset for a source task, transferring the learned knowledge (e.g. features or weights) to a new network for a different learning task and/or on a different target domain.

Transfer learning is used with the objective of increasing performance on the target task by exploiting the trained knowledge of the source task. It can either serve as a 'Kick-starter' for the process improving initial performance, increase final performance or speed up the training of the target task. Three main types of transfer learning can be distinguished with respect to similarities in the source/target domains/tasks: (1) inductive transfer learning with the goal of inducing a generalizable, predictive model for the target task from a different source task; (2) transfer learning where a model should be derived in the target domain for the target task lacking labeled data, based

on the source task containing a substantial amount of labeled data; and (3) unsupervised transfer learning which targets unsupervised learning in the target domain based on a different source task.

For purposes of image classification, the approach of inductive transfer learning has gained momentum in the past years. Given the large training time and requirements on labeled data, architectures are pre-trained on the ImageNet (source domain) for 1000-class classification purposes (source task) and knowledge is transferred to a labeled dataset of choice (target domain) for e.g. classification as target task. Thus, it primarily exploits the learned feature-representations in the source domain and aims to transfer these to the target domain with the aim of improving classification performance whilst making use of deep, data-greedy CNN architectures.

The transferability of different levels of features was touched upon in this work with results confirming that the deepest layers are most task-specific and better results can be obtained by using features from levels upwards in the model. This was confirmed by the work, in which furthermore was suggested that pooling of features as well as combinations of features may improve performance. We added to the evidence by documenting higher performances for CNN's initiate with pre-trained weights and fine-tuned for the task at hand, rather than through random weight initialization and training from scratch. Furthermore, they investigated the transferability of specific feature layers and reported co-adaptation of intermediate feature layers, where splitting between frozen and trainable layers at this level might result in a drop in performance.

4. RESULTS

Guided by the research questions and founded on the methods as described in Methodology Chapter, the obtained results will be presented throughout this chapter. Focus was placed throughout the previous chapter on the three implemented frameworks in relation to the choice of pre-training of CNN architectures. To facilitate readability, we will adopt in the respective sections the same structure and answer the research questions in the corresponding sections. Skin cancer detection using the deep convolutional neural network on ISIC dataset with data augmentation in the data space, which was employed to double the dataset with randomly augmented images leads to a drop in performance. As can be observed, skin cancer detection using the convolutional neural network on ISIC-Archive dataset with data augmentation in the data space, which was employed to double the dataset with randomly augmented images leads to a drop in performance. Additional data augmentation (e.g. random scaling, distortions, color offset or jitter, introduction of image noise) might also help mitigate overfitting and increase the robustness to different geographic settings as well as skin lightning conditions. Although the model can generalize within certain limits (local generalization), its transferability is mainly limited by the variability of the training data. Furthermore, higher spatial resolution would certainly improve the exact delineation of the field boundaries, especially for small objects. At the time of writing, deep learning presents the highest resolution of freely available satellite imagery. This makes it difficult to create robust algorithms that recognize all kinds of field instances without overfitting to the specific scene. In addition to the variation within a single scene, skin images can show even stronger intra-class variations. This applies to skin images from different skin regions, from different dates or growing seasons, or under different skin diseases and illumination conditions. This research chose to circumvent this skin cancer detection problem by redefining a top model with the desired number of trainable layers based on the CNN architecture and transferring the pre-trained weights and biases to this architecture. Omitting the step of skin data augmentation, features are extracted from desired, frozen depth of the CNN model which runs in inference mode and the top model can be trained on these features.

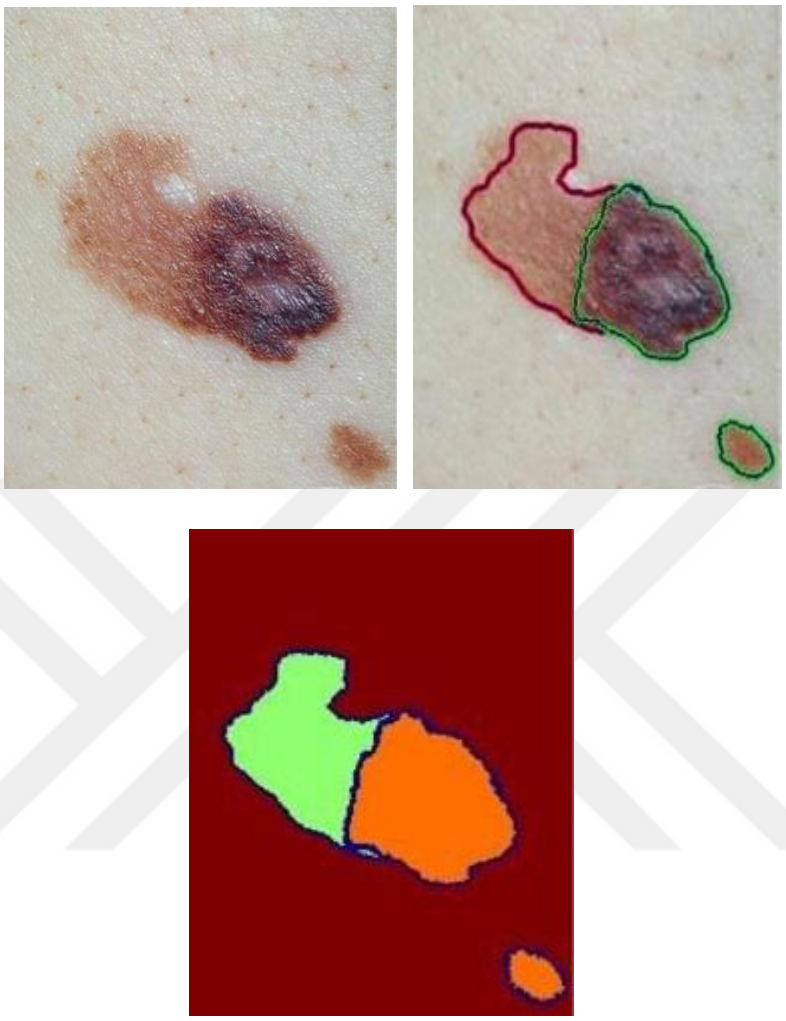

Figure 4.1: The input image, detection and classification using the Xception network.

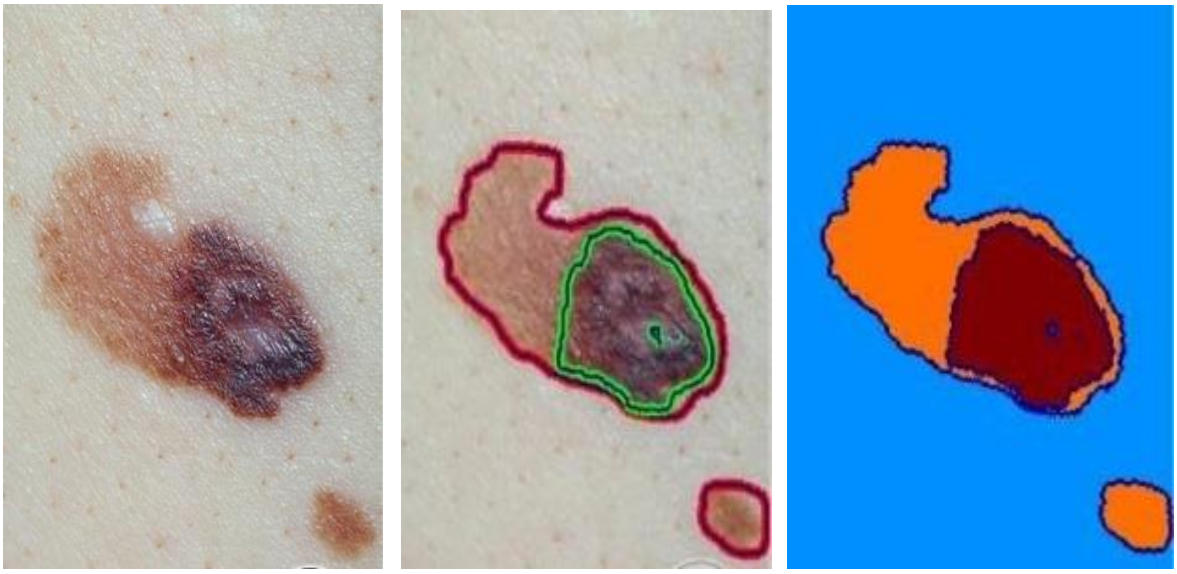

Figure 4.2: The input image, detection and classification using the Xception network.

True Positive and False Negative, or alternatively the False Positive as their harmonic mean, can be further optimized by choosing appropriate decision thresholds. By default, the CNN-classifier will assign a sample to a class upon 85 % probability. However, according to the ROC-curve, a new threshold can be selected based on the calculated True Positive and False Negative at different thresholds. It is important to note that this thresholding might offer opportunities in a medical framework to select those criteria that have the highest diagnostic value and tune the classifier as such. As an illustration, thresholding is performed to optimize the sum of False Positive and True Negative.

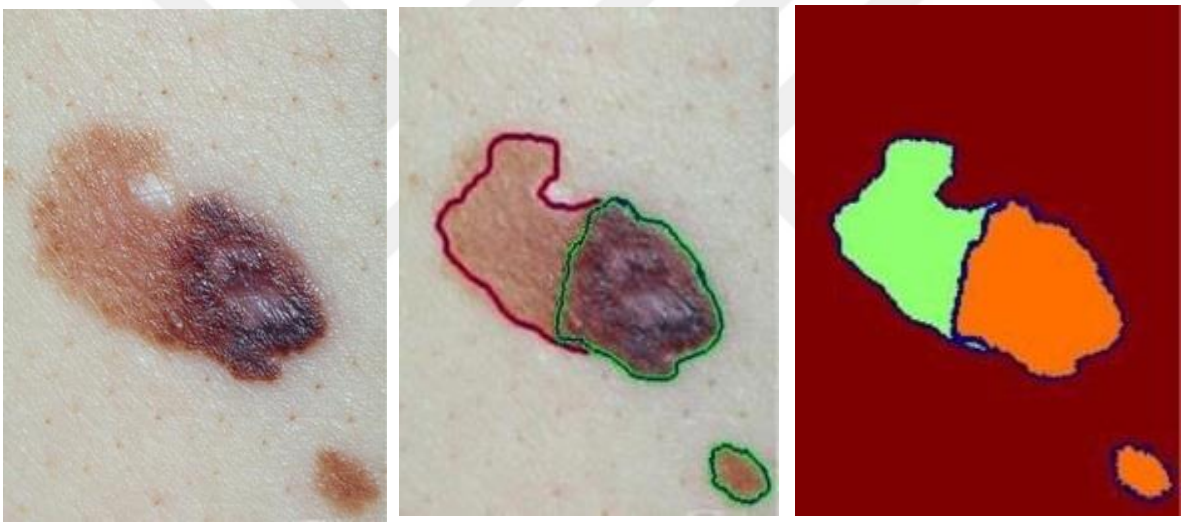

Figure 4.3: The input image, detection and classification using the Xception network.

We chose to circumvent this skin cancer detection problem by redefining a top model with the desired number of trainable layers based on the CNN architecture and transferring the pre-trained weights and biases to this architecture. Omitting the step of data augmentation, features are extracted from desired, frozen depth of the CNN model which runs in inference mode and the top model can be trained on these features.

5. DISCUSSION

The development of skin cancer detection system using CNN based on skin lesion diagnostic tasks is in hot pursuit, with special attention being given to the diagnosis of melanomas. Given its high propensity to metastasize, going hand in hand with severe decreases in survival rates, and the high inter-patient variability in lesion appearance, as well as a strong requirement on the training of the physician/dermatologist properly diagnosing a melanoma can be considered a daunting task. Furthermore, there is no non linearity between the blocks. Modified conv blocks are implemented throughout the model like Inception module.

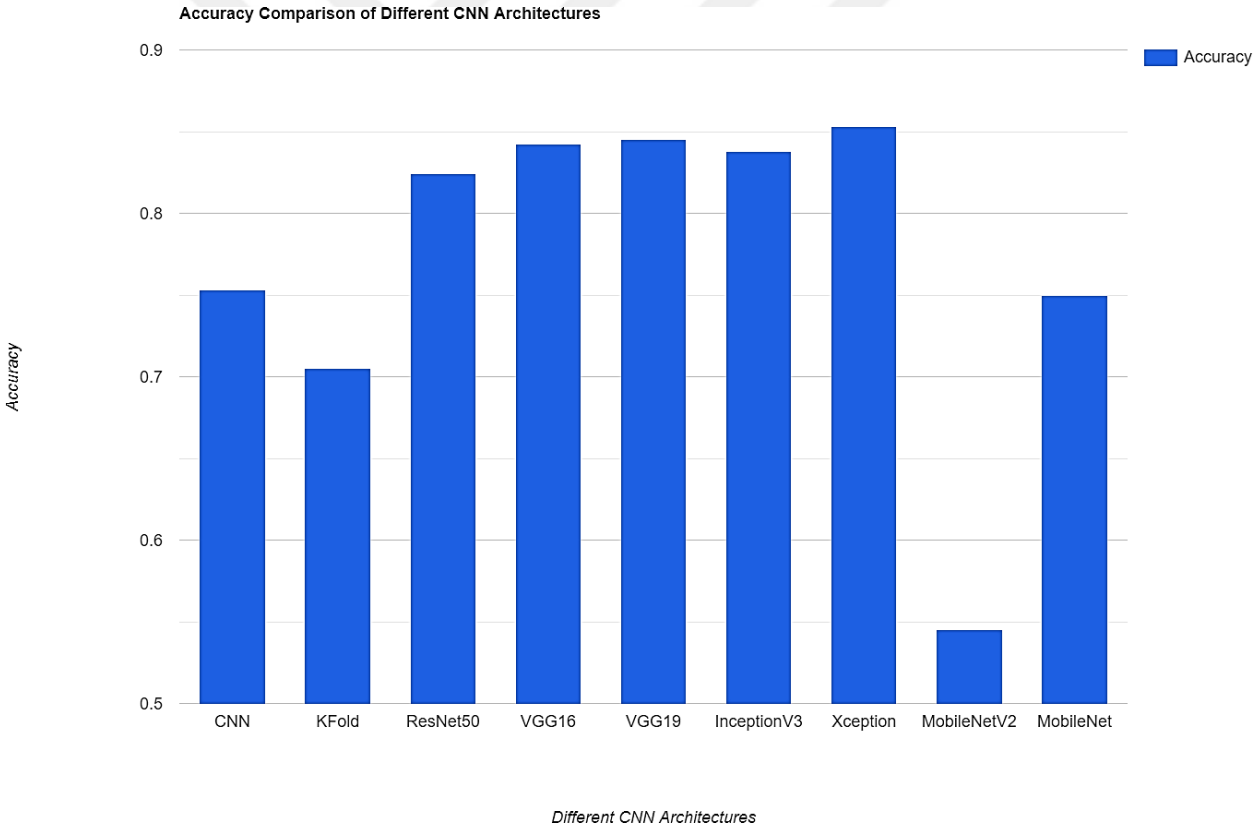

Figure 5.1: The variance accuracy score of all the architectures.

As we have performed 7 CNN architectures along with our model and k-fold on the dataset, a huge number of variance in prediction is witnessed on those architectures. Out of those nine architectures, the worst accuracy we got was from mobilenetv2. It was only 54.54%. Also the highest accuracy we got was from Xception which was 85.30%.

The reasons behind this variance of accuracy are discussed below: Firstly, MobilenetV2 which is a lightweight model was faster in training than most other architectures since it has bottleneck residual block that takes a lower dimensional vector as input and passes the same for the output. In this way MobilenetV2 reduces computational hazard and has an improve runtime. But unfortunately it ended up resulting in an unexpected accuracy of 54.54%. The accuracy during the epochs was higher than other CNN architectures and it reached 80% in the last epoch. Since MobilenetM2 does not have any activation function in its output block, this could be a reason behind this accuracy drop.

In k-fold we have got a validation accuracy of 70.53% with 3 folds. Since the model was a simple one with sequential 4 layers in it only, it could not extract enough features to learn about the dataset. The highest accuracy we have got was in fold 2 where the accuracy was 80.43%.

MobileNet was the fastest of all the predefined architectures that we have used in our dataset. This architecture was trained in 1/3rd time of other architectures like ResNet50 or Xception. But the accuracy was only 75% which is lower than the simplest model we have built at the beginning. MobileNet does not have any pooling layer from the beginning through the very end. It has a global average pooling before the output layer. The input shapes were reduced with stride convolution. Depth-wise separable convolutional block was the reason behind the improved run-time.

However, we have built a very simple sequential model with only 4 layers where there are two conv2d block with max-pooling and dropout and an output layer followed by a dense layer. Consequently, this model was the fastest in the training. The model was 12 times faster than VGG models. But we have got a marginal accuracy of 75.30%. Since the model was not deep enough, it could not extract enough features to distinguish between the melanoma of skin with a better accuracy.

ResNet50 is a deep network with 50 layers. This network builds residual block be- tween the layers to solve the gradient vanishing problem and also for better learning in the training phase. This was a popular approach before the idea of depth-wise separable convolution. This network uses max-pooling throughout the network with just one global average pooling in the output block. This model gave out 82.42% of accuracy in our dataset. InceptionV3 was the fastest model out of all the deep models we have used with our dataset. It was almost as fast as MobileNet and

MobileNetV2 model. Also the accuracy was more than ResNet50. Since we know Inception is a multilevel feature extractor model which means 1x1, 3x3 and 5x5 conv blocks are computed within the same module. For this reason, this architecture was pretty deep with 48 layers and also computation time is faster. Inception uses max-pooling in those conv blocks and concatenates the output blocks with average pooling which means the best features are first selected from each block and then those features were generalized. This is why InceptionV3 gives out an accuracy of 83.78% with a comparatively lower run time.

VGG16 and VGG19 both are sequential architectures. Furthermore, they are quite identical in their structure. The only difference is that VGG19 has three more convolution blocks. Although these architectures are not deep enough, they can predict the output more accurately than many deeper models since data is learned and pass through the model in a sequence extracting the core features by using max-pooling. The accuracy of these two architectures is also on a margin of 30%. VGG16 gave out the accuracy of 84.24% and VGG19 was 84.54%. The only drawback VGG architectures have is that they are pretty much slower in terms of training and evaluating. On our dataset, these two architectures have the 2nd highest runtime.

Xception is one of the latest additions in CNN architectures. This architecture delivered the best accuracy out of all the architectures we have used on our dataset. There are certain reasons behind this. This architecture was built with the concept of previously defined architectures for example ResNet and Inception to be specific. This model uses a modified depth-wise convolutional block which starts with a point-wise convolution.

Table 5.1: Comparative Analysis of Accuracy and Loss on Different CNN Architectures.

Model Name	Accuracy	Loss
CNN	0.7530	0.2470
KFold	0.7053	0.2947
ResNet50	0.8242	0.1758
VGG16	0.8424	0.1576
VGG19	0.8454	0.1546
InceptionV3	0.8378	0.1622
Xception	0.8530	0.1470
MobileNetV2	0.5454	0.4546
MobileNet	0.7500	0.2500

A methodology consisted of benign/malignant classification with CNN's on features extracted from state-of-the-art models (Xception). Best performance was achieved with deep skin cancer based-features, followed by FC1-features from Xception. Combining features of different layers was investigated for the Xception, but did not enhance classification performance. Furthermore, combining features from different networks through concatenation was investigated and an increase in performance was observed when combining Xception features, possibly indicating that both nets extract slightly complementary features.

After an inspection of fine-tuning inside skin, our focus shifted to the fine-tuning of multiple blocks to investigate the depth of fine-tuning and its possible enhancement of performance. We fine-tuned up to 4 blocks in the Xception model and the most relevant experimental results are shown in Table 5.2 showing the Averaged Accuracy-scores for all three datasets. An increase in performance can be observed when incorporating the first block in fine-tuning process.

Firstly, the random weight initialization of the new 3-neuron softmax layer requires 'pre-tuning', i.e. separate training of the said layer with the entire pre-trained base model in frozen state in CNN. This can of course be attributed to the gradients over this last layer being too high initially, thus being highly disruptive to full training of the model. Secondly, even after fine-tuning of different

layers, the obtained AUC-scores were substantially lower than what was achieved with CNN. Consequently, and in light of the promising results of the Xception model, the latter was chosen as further framework to investigate the depth of fine-tuning research questions.

Table 5.2: Different layers of CNN with Averaged Accuracy-scores for skin images.

Layers	Recall	F1-score	AUC	Precision
Input-layer	0.816	0.815	0.822	0.813
Weighted	0.816	0.823	0.824	0.815
FC-layer	0.799	0.79	0.79	0.76
Pooling	0.799	0.79	0.79	0.76
Augment	0.806	0.817	0.817	0.812

For the CNN-classifier on FC-layer features from using Xception pre-trained model, as visualized, both kernel-type and class imbalance handling were investigated through a 10-fold cross-validation scheme. The obtained means and skin cancer detection of five different performance measures, the overall accuracy and specifically for the positive class the F1-score, precision, recall and the AUC-score, are reported in Tables 5.2. The accuracy is mentioned for the sake of completeness, but comparing the accuracy and F1-scores of higher-scoring models clearly portrays that a higher accuracy does not entail a higher recall/precision, with the latter being desirable as it specifies how the models cope with class imbalance.

6. CONCLUSION

6.1 CONCLUSION

This master's thesis focuses on detection and classification of skin diseases using deep learning. The deep learning challenge originates from the accuracy of the ground truth data sets that are used as validation and/or training data for automated image processing algorithms. These datasets usually stem from manual tracing of field boundaries. However, manual tracing of skin image is heavily dependent on the used imagery and supplementary data. It also is a highly subjective task, inevitably leading to inaccuracies and ambiguities depending on the priorities of the operator. The ground truth dataset can be complemented by existing property parcel information that could be indistinguishable based on the skin imagery alone. Furthermore, a single field parcel polygon can potentially show several subfields due to unreported land use changes within the growing season. Given these constraints, even a very powerful automated image processing algorithm can only become as good as the ground truth data but will never be in perfect alignment with reality. Although the model can generalize within certain limits (local generalization), its transferability is mainly limited by the variability of the training data. Furthermore, higher spatial resolution would certainly improve the exact delineation of the field boundaries, especially for small objects. At the time of writing, presents the highest resolution of freely available satellite imagery. Instead, the methodology focuses on a simple input-output mapping. Additional pre-processing or post-processing techniques could certainly add value later on, but would distract from the current goal of establishing baseline result. To exploit the pre-trained CNN for skin cancer detection of the three different datasets, a top model is to stacked on the deep learning. We explore three different top models in this thesis to maximally make use of the labeled, supervised data. A first and obvious choice is a mere fully-connected layer, applied on the flattened CNN-output with added dropout. In light of the high parametric demand this top-model places on the learning process, we opted to also investigate performance with residual blocks as top model. Not only are these residual blocks hypothesized to facilitate training by learning the residual mapping, they also decrease the number of parameters while allowing a multi-layered top model.

6.2 FUTURE RECOMMENDATIONS

In future, we want to add up the research line of deep learning with big data, training on RNN with Apache Spark engine for combinations of different feature from different pre-trained hybrid models leads to an enhanced classification performance compared to stand-alone features for skin cancer detection.

A more advanced use of fine-tuning however proved to be adapting the pre-trained model, to the task at hand and fine-tuning up to a certain depth in the RNN model.

This scheme will give rise to the optimal results achieved in this study for skin cancer detection. Finally, these results were compared developed using RNN framework in future for pre-training on any skin image dataset.

The best performance results will be achieved with high precision entered the same ballpark as the higher-scoring RNN classifiers on multiple pre-trained model framework, but could compete with the any other skin cancer detection work in previous studies.

We note however that in light of the trends observed in the RNN framework, a more thorough exploration (of architecture, training time and hyper-parameter settings) could potentially improve these results and open a window of opportunity to thoroughly using three labeled datasets for kick-starting the supervised training phase with limited three labeled datasets.

REFERENCES

[1] Albayark, A.; Bilgin, G. Mitosis detection using convolutional neural network based features. In Proceedings of the IEEE Seventeenth International Symposium on Computational Intelligence and Informatics (CINTI), Budapest, Hungary, 17–19 November 2016; pp. 335–340.

[2] Swiderski, B.; Kurek, J.; Osowski, S.; Kruk, M.; Barhoumi, W. Deep learning and non-negative matrix factorization in recognition of mammograms. In Proceedings of the Eighth International Conference on Graphic and Image Processing, International Society of Optics and Photonics, Tokyo, Japan, 8 February 2017; Volume 10225, p. 102250B.

[3] Suzuki, S.; Zhang, X.; Homma, N.; Ichiji, K.; Sugita, N.; Kawasumi, Y.; Ishibashi, T.; Yoshizawa, M. Mass detection using deep convolutional neural networks for mammoghraphic computer-aided diagnosis. In Proceedings of the 55th Annual Conference of the Society of Intruments and Control Engineers of Japan (SICE), Tsukuba, Japan, 20–23 September 2016; pp. 1382–1386.

[4] Wang, C.; Elazab, A.; Wu, J.; Hu, Q. Lung nodule classification using deep feature fusion in chest radiography. Comput. Med. Imaging Graph. 2017, 57, 10–18.

[5] Dou, Q.; Chen, H.; Yu, L.; Qin, J.; Heng, P.-A. Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection. IEEE Trans. Biomed. Eng. 2017, 64, 1558–1567.

[6] Shen, W.; Zhou, M.; Yang, F.; Yu, D.; Dong, D.; Yang, C.; Tian, J. Multicrop convolutional neural networks for lung nodule malignancy suspiciousness classification. Pattern Recognit. 2017, 61, 663–673.

[7] Available Online: https://www.healthline.com/health/layers-of-skin

[8] Hiryama, K.; Tan, J.K.; Kim, H. Extraction of GGO candidate regions from the LIDC database using deep learning. In Proceedings of the Sixteenth International Conference on Control, Automation and Systems (ICCAS), Gyeongju, Korea, 16–19 October 2016; pp. 724–727.

[9] Hosny KM, Kassem MA, Foaud MM (2019) Classification of skin lesions using transfer learning and augmentation with Alex-net. PLoS ONE 14(5): e0217293. https://doi.org/10.1371/journal.pone.0217293.

[10] Hussein, S.; Gillies, R.; Cao, K.; Song, Q.; Bagci, U. TumorNet: Lung Nodule Characterization Using Multi-View Convolution Neural Network with Gaussian Process. In Proceedings of the IEEE 14th International Symposium on Biomedical Imaging (ISBI), Melbourne, Australia, 18–21 April 2017; pp. 1007–1010.

[11] Balagurunthan, Y.; Gu, Y.; Wang, H.; Kumar, V.; Grove, O.; Hawkins, S.H.; Kim, J.; Goldgof, D.B.; Hall, L.O.; Gatenby, R.A. Reproducibility and prognosis of quantitative features extracted from CT images. Transl. Oncol. 2015, 7, 72–87.

[12] Han, F.; Wang, H.; Zhang, G.; Han, H.; Song, B.; Li, L.; Moore, W.; Lu, H.; Zhao, H.; Liang, Z. Texture feature analysis for computer-aided diagnosis on pulmonary nodules. J. Digit. Imaging 2015, 28, 99–115.

[13] Barata, C.; Marquees, J.S.; Rozeira, J. A system for the detection of pigment network in dermoscopy images using directional filters. IEEE Trans. Biomed. Eng. 2017, 59, 2744–2754.

[14] Barata, C.; Marquees, J.S.; Celebi, M.E. Improving dermoscopy image analysis using color constancy. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2016; pp. 3527–3531.

[15] Barata, C.; Ruela, M.; Mendonca, T.; Marquees, J.S. A bag-of-features approach for the classification of melanomas in dermoscopy images: The role of color and texture descriptors. In Computer Vision Techniques for the Diagnosis of Skin Cancer; Springer: Berlin/Heidelberg, Germany, 2016; pp. 49–69.

[16] Sadeghi, M.; Lee, T.K.; McLean, D.; Lui, H.; Atkins, M.S. Detection and analysis of irregular streaks in dermoscopic images of skin lesions. IEEE Trans. Med. Imaging 2016, 32, 849–861.

[17] Ahmed, Saadaldeen. (2019). Breast Cancer Detection and Image Evaluation using Augmented Deep Convolution Neural Networks.

[18] Meier, R.; Bauer, S.; Slotnoom, J.; Wiest, R.; Reyes, M. A hybrid model for multi-modal brain tumor segmentation. In Proceedings of the MICCAI Challenge on MultimodalBrain Tumor Image Segmentation, NCI-MICCAI BRATS, Nagoya, Japan, 22 September 2018; pp. 31–37.

[19] Pinto, A.; Pereira, S.; Correia, H.; Oliveira, J.; Rasteito, D.M.; Silva, C.A. Brain tumour segmentation based on extremely randomized forest with high-level features. In Proceedings

of the 37th Annual International Conference on IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 3037–3040.

[20] Tustison, N.J.; Shrinidhi, K.; Wintermark, M.; Durst, C.R.; Kandel, B.M.; Gee, J.C.; Grossman, M.C.; Avants, B.B. Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation(simplified) with ANTsR. Neuroinformatics 2015, 13, 209–225.

[21] Bengio, Y.; Courville, A.; Vinvent, P. Representation learning: A review and new prespectives. IEEE Trans. Pattern Anal. Mach. Intell. 2013, 35, 1798–1828.

[22] LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436–444.

[23] Xu, B.; Wang, N.; Chen, T.; Li, M. Empirical evaluation of rectified activations in convolutional network. arXiv 2015, arXiv:1505.00853.

[24] Messadi, M.; Bessaid, A.; Taleb-Ahmed, A. Extraction of specific parameters for skin tumour classification. J. Med. Eng. Technol. 2019, 33, 288–295.

[25] Reddy, B.V.; Reddy, P.B.; Kumar, P.S.; Reddy, S.S. Developing an approach to brain MRI image preprocessing for tumor detection. Int. J. Res. 2017, 1, 2348–6848.

[26] Zacharaki, E.I.; Wang, S.; Chawla, S.; Soo Yoo, D.; Wolf, R.; Melhem, E.R.; Davatzikos, C. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. Magn. Reson. Med. 2019, 2, 1609–1618.

[27] Miah, M.B.A.; Yousuf, M.A. Detection of lung cancer from CT image using image processing and neural network. In Proceedings of the 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), Jahangirnagar University, Dhaka, Bangladesh, 21–23 May 2015.

[28] Ponraj, D.N.; Jenifer, M.E.; Poongodi, P.; Manoharan, J.S. A survey on the preprocessing techniques of mammogram for the detection of breast cancer. J. Emerg. Trends Comput. Inf. Sci. 2016, 2, 656–664.

[29] Zhang, Y.; Sankar, R.; Qian, W. Boundary delineation in transrectal ultrasound image for prostate cancer. Comput. Biol. Med. 2017, 37, 1591–1599.

[30] Emre Celebi, M.; Kingravi, H.A.; Iyatomi, H.; Alp Aslandogan, Y.; Stoecker, W.V.; Moss, R.H.; Malters, J.M.; Grichnik, J.M.; Marghoob, A.A.; Rabinovitz, H.S.; et al. Border detection in dermoscopy images using statistical region merging. Skin Res. Technol. 2018, 14, 347–353.

[31] Tong, N.; Lu, H.; Ruan, X.; Yang, M.H. Salient object detection via bootstrap learning. In Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1884–1892.

[32] Bozorgtabar, B.; Abedini, M.; Garnavi, R. Sparse coding based skin lesion segmentation using dynamic rule-based refinement. In International Workshop on Machine Learning in Medical Imaging; Springer: Cham, Switzerland, 2016; pp. 254–261.

[33] Li, X.; Li, Y.; Shen, C.; Dick, A.; Van Den Hengel, A. Contextual hypergraph modeling for salient object detection. In Proceedings of the 2013 IEEE International Conference on the Computer Vision (ICCV), Sydney, Australia, 1–8 December 2017; pp. 3328–3335.

[34] Yuan, X.; Situ, N.; Zouridakis, G. A narrow band graph partitioning method for skin lesion segmentation. Pattern Recogn. 2019, 42, 1017–1028.

[35] Sikorski, J. Identification of malignant melanoma by wavelet analysis. In Proceedings of the Student/Faculty Research Day, New York, NY, USA, 7 May 2014.

[36] Chiem, A.; Al-Jumaily, A.; Khushaba, N.R. A novel hybrid system for skin lesion detection. In Proceedings of the 3rd International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIPâTM07), Melbourne, Australia, 3–6 December 2017; pp. 567–572.

[37] Maglogiannis, I.; Zafiropoulos, E.; Kyranoudis, C. Intelligent segmentation and classification of pigmented skin lesions in dermatological images. In Advances in Artificial Intelligence; Springer: Berlin/Heidelberg, Germany, 2016; pp. 214–223.

[38] Tanaka, T.; Torii, S.; Kabuta, I.; Shimizu, K.; Tanaka, M.; Oka, H. Pattern classification of nevus with texture analysis. In Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBCâTM04), San Francisco, CA, USA, 1–5 September 2016; pp. 1459–1462.

[39] Zhou, H.; Chen, M.; Rehg, J.M. Dermoscopic interest point detector and descriptor. In Proceedings of the 6th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Boston, MA, USA, 28 June–1 July 2019; pp. 1318–1321.

[40] Lee, C.; Landgrebe, D.A. Feature extraction based on decision boundaries. IEEE Trans. Pattern Anal. Mach. Intell. 2019, 15, 388–400.

[41] Lee, C.Y.; Gallagher, P.W.; Tu, Z. Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree. In Proceedings of the 19th International Conference on

Artificial Intelligence and Statistics (AISTATS), Cadiz, Spain, 9–11 May 2016.

[42] Nwankpa, C.; Ijomah, W.; Gachagan, A.; Marshall, S. Activation functions: Comparison of trends in practice and research for deep learning. arXiv 2018, arXiv:1811.03378.

[43] Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICLM), Haifa, Israel, 21–24 June 2018; pp. 807–814.

[44] Kingma, D.; Ba, J. Adam: A method of stochastic optimmization. arXiv 2014, arXiv:1412.6980.

[45] Albarqouni, S.; Baur, C.; Achilles, F.; Belagiannis, V.; Demirci, S.; Navab, N. AggNet: Deep learning from crowds for mitosis detection in breast cancer histology images. IEEE Trans. Med. Imaging 2016, 35, 1313–1321.

[46] Spanhol, F.A.; Oliveira, L.S.; Petitjean, C.; Heutte, L. Breast cancer histopathological image classification using convolutional neural networks. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 2560–2567.

[47] Wichakam, I.; Vateekul, P. Combining deep convolutional networks and SVMs for mass detection on digital mammograms. In Proceedings of the 8th International Conference on Knowledge and Smart Technology (KST), Bangkok, Thailand 2016; pp. 239–244.

[48] Ertosun, M.G.; Rubin, D.L. Probabilistic visual search for masses within mammography images using deep learning. In Proceedings of the 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Washington, DC, USA, 9–12 November 2015; pp. 1310–1315.

[49] Demyanov, S.; Chakravorty, R.; Abedini, M.; Halpern, A.; Garnavi, R. Classification of dermoscopy patterns using deep convolutional neural networks. In Proceedings of the 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 13–16 April 2016; pp. 364–368.

[50] Giotis, I.; Molders, N.; Land, S.; Biehl, M.; Jonkman, M.F.; Petkov, N. MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images. Expert Syst. Appl. 2015, 42, 6578–6585.