SKIN CANCER DETECTION USING NEURAL NETWORKS

WIDER TOPICS IN DATA SCIENCE

- MEENAKSHI DATTU PATIL -

- APRIL 2022 -

Disclaimer:
This report contains images that might not be comfortable to look at, you have been warned.

ABSTRACT:

Skin cancer is an alarming disease for mankind. The necessity of early diagnosis of the skin cancer have been increased because of the rapid growth rate of Melanoma skin cancer, its high treatment costs, and death rate. The most common human malignancy is primarily diagnosed visually, beginning with an initial clinical screening, and followed potentially by dermoscopic analysis, a biopsy and histopathological examination. Automated classification of skin lesions using images is a challenging task owing to the fine-grained variability in the appearance of skin lesions. This paper will talk about how we went from visual detection of skin cancer growth to artificial skin cancer detection system using image processing and machine learning methods. The features of the affected skin cells are extracted after the segmentation of the dermoscopic images using feature extraction techniques. A deep learning-based method convolutional neural network classifier is used for the stratification of the extracted features.

1. INTRODUCTION

When the cells of a specific part of the body grow and reproduce uncontrollably, it is called cancer. Cancer arises when the normal cell transforms into a tumor cell in a multi-stage process that generally progresses from a pre-cancerous lesion to a malignant tumor. It is the leading cause of death worldwide accounting for nearly 10 million deaths in 2020, which means that 1 in 6 deaths is caused by cancer. It is extremely dangerous with the survival rate of 50%, according to estimates in the UK. There are over 200 different types of cancer, and each is diagnosed in a different way. Between 30 – 50% of cancers are preventable just by avoiding risk factors and employing prevention strategies. Furthermore, the burden of cancer can be reduced through early detection, and suitable treatment and care for patients who develop cancer.

Skin cancer is the most common cancer of all the other types, and each year more people are being diagnosed with this type of cancer each year than all the other cancers combined. Melanoma is the deadliest form of skin cancer with an estimated half-a-million cases projected by 2040. There is a need for remote and automated diagnosis solutions to reduce the impact of skin cancer cases on the healthcare services. Especially in poor countries where patients don't have access to latest medical services for accurate diagnosis. Datasets such as the International Skin Imaging Collaboration (ISIC) datasets and HAM10000(Human against Machine) are open source and contain images with a wide usage space. Most of the tasks focus on classification and segmentation, with the most common research involving binary classification. Some other uses of the ISIC dataset are effects of color consistency and data augmentation using generative adversarial networks (GAN's). The idea of having a model that is trained on the image dataset, is to be able to distinguish between cancerous and non-cancerous skin lesions will not only help in early detection of skin cancer, along with an early diagnosis, but also will reduce the impact on the health system. This paper will explore some techniques, that are used to provide information on whether the skin lesion is a skin cancer patch or just a normal patch. There are a plethora of methods, formats, and techniques, and although this paper is not sufficient to get into the details of all the methods and techniques used, it will provide some brief information as to the some of the techniques that are commonly used worldwide.

2. WHAT IS SKIN CANCER?

Skin cancer is one of the most common cancers worldwide and come in different types. Basal Cell Carcinoma (BCC) and Squamous Cell Carcinoma (SCC) are the most dominant forms, and they start in the epidermis (the outermost of the three layers of the skin) and are caused by sun-exposure. Hence, they typically develop in areas exposed to the sun such as the face, ears, neck, arms, and hands. Although this paper will not go into much detail regarding types of skin cancer, but a brief description of a few types of melanoma cancer will be described, and it will also have information regarding the rare types of melanomas that appear in areas that aren't exposed to the sun, and the types of cancer that are found in people of dark skin color (which is a rare occurrence).

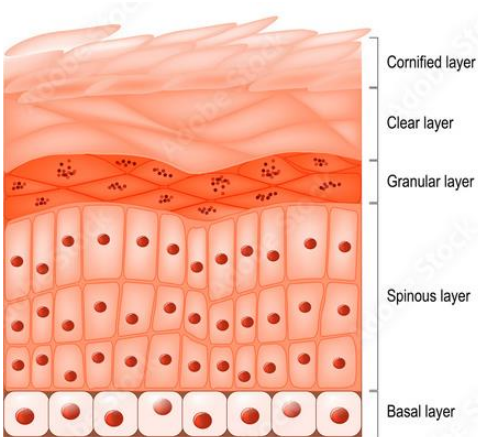


Fig1: Layers of Epidermis

While BCC very rarely spreads to other parts of the body, SCC may spread to nearby organs or lymph nodes. Both BCC and SCC are sometimes called keratinocyte cancers as they both originate in keratinocytes (the most common type of skin cell). Melanoma is another form of skin cancer that develops in melanocytes (a type of skin cell in the epidermis that produces melanin (a dark pigment responsible for skin color)). With around 16,000 cases registered yearly, melanoma is the 5th most common cancer in the UK. It is known to take over 2,300 lives every year in the UK alone.

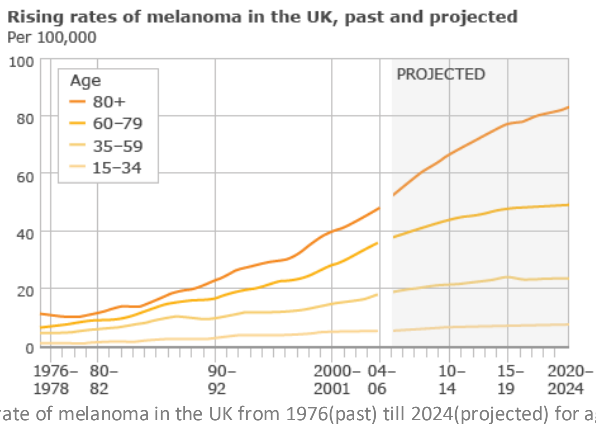


Fig2: Rising rate of melanoma in the UK from 1976(past) till 2024(projected) for ages 15 to 80+

In most of the cases, the melanomas have an irregular shape and are more than one color, with the mole mostly being itchy or may ooze blood and be larger in size. These are all visible externally on the surface of the skin and with the recent advancements in hardware and software technologies, we can use various machine learning tools for feature engineering to determine the severity or type of pigmentation.

2.1 TYPES OF MELANOMAS

2.1.1 SUPERFICIAL SPREADING MELANOMAS:

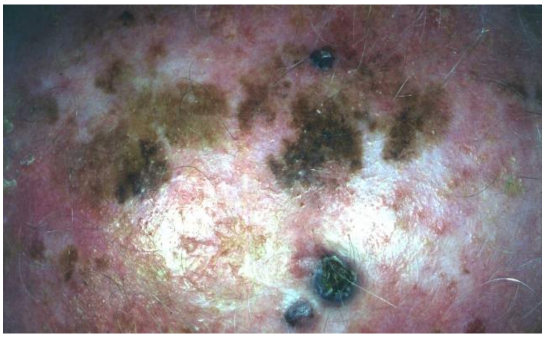
This is the most common type of melanoma in the UK. People with pale skin and freckles are most likely to get it, while darker skinned people are not very common. The melanoma tends to grow outward, rather than growing downward, in most cases, and so they do not pose a problem. But they will spread to other parts of the body if it begins growing downward.

2.1.2 NODULAR MELANOMA:

Nodular melanoma's grow rapidly and if they aren't removed, they have a tendency to grow downwards in the layers of skin. They are generally found on previously typical skin, in areas near the chest or back and head or neck. They appear as a changing lump on the skin and can go from black to red in color. This type of melanoma also causes bleeding and oozing.

2.1.3 LENTIGO MALIGNA MELANOMA:

This melanoma mostly affects people of age, especially those who've spent a lot of time outdoors. It is a slow developing melanoma that appears in the areas exposed to the sun, and takes years to show up, such as the face. Lentigo Maligna Melanoma appears as freckles that are darker, larger and they can be easily differentiated from normal freckles.



2.1.4 ACRAL LENTIGINOUS MELANOMA

This melanoma is also very aggressive. It is rare, generally grows on the palm of the hands, soles of the feet and sometimes around the nails and isn't caused due to sunlight. Compared to other skin colors, dark skin people usually tend to have Acral Lentiginous Melanoma, but they do appear in other skin types as well.

2.1.5 AMELANOTIC MELANOMA

This type of melanoma is a poorly differentiated subtype of a typical melanoma and are tough to detect because they have little to no color, but occasionally is pink or red, with light brown or grey edges. Excessive UV radiation is a common cause of this type of cancer.

3. THE CONCEPT OF NEURAL NETWORKS

Deep leaning is a part of machine learning in which the methods used are inspired or based on artificial neural networks (ANN). In this method, learning can be supervised, semi-supervised or unsupervised. The concept of neural networks is usually compared to the neurons in the brain, as these neural networks also function in a similar manner.

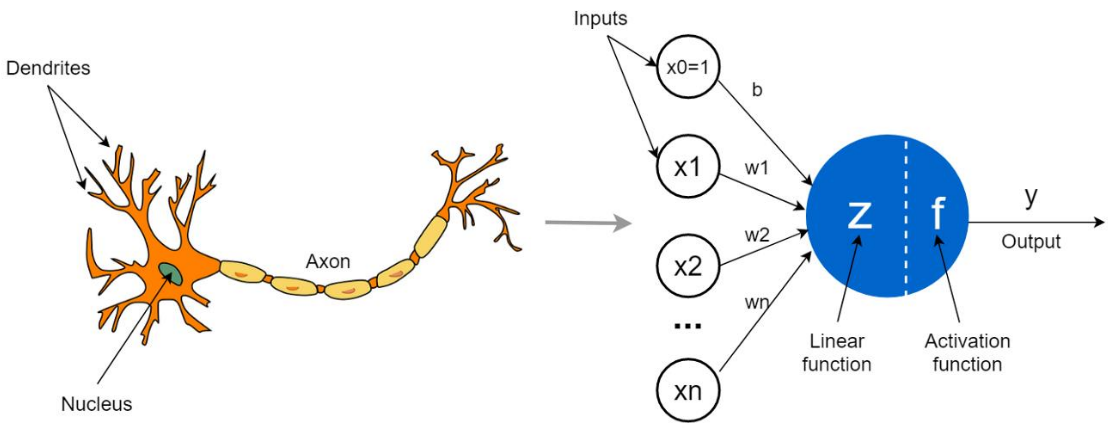


Fig3: Comparing a neuron of the brain to the structure of a perceptron

There are many types of neural networks such as Convolutional Neural Networks (CNN), Stochastic Neural Network (SNN) and Recurrent Neural Networks (RNN), they use Network Science and different neural networks are used for different purposes. Although the initial idea for neural networks was derived from biological neurons, they were soon shifted towards getting better empirical results, and remaining true to the biological ancestor wasn't a priority anymore. Each artificial neuron has inputs and produces one output that can be sent to multiple other neurons.

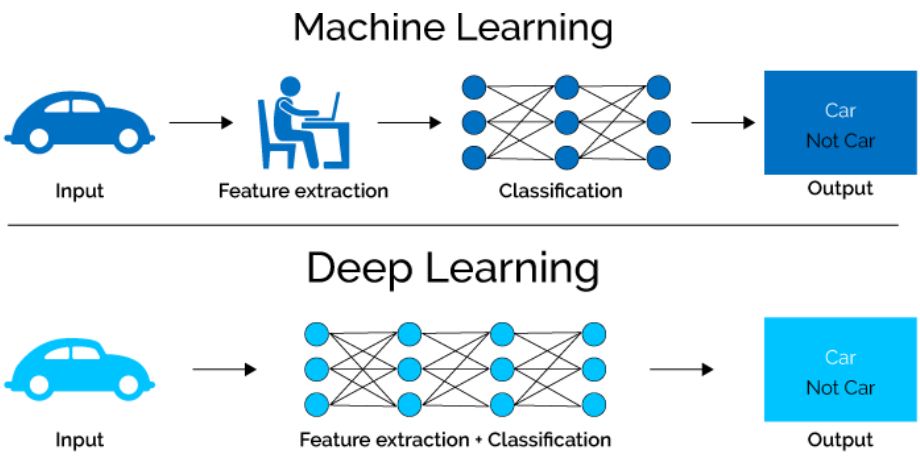


Fig4: Comparison of Machine Learning and Deep Learning

Although the image on top explains how machine learning varies from deep learning, the concepts are way more complex than the picture leads one to believe. For starters, each line in the image has its own separate weights that must be accounted for, and these values are computed by the values of the previous layer along with the weights, and this is called the activation function. The concept of deep learning and neural networks is extremely complex and a complete specialized topic, which this paper will not cover as it is out of the scope of this study.

4. SKIN CANCER IMAGE DATASETS

4.1 HAM10000

The HAM10000 (“Human Against Machine with 10000 training images”) dataset is a collection of dermatoscopic images that are collected from multiple sources in Austria and Australia. The actual dataset consists of 10015 images of benign and malignant types of skin cancer of different types, such as melanocytic nevi, melanoma, keratosis etc. The images in this dataset are in batches of 32, randomly cropped and resized to 224 x 224 pixels. The dataset is public and because of this, a lot of research work is performed using this dataset. The cases included in this dataset are a collection of pigmented lesions. More than half of the images in this dataset are all confirmed through histopathology (the study of changes in the tissue, caused due to a disease), while the remaining data is procured after either a follow-up examination, in-vivo confocal microscopy, or expert consensus. This dataset is a popular benchmark that can be used to test or compare machine learning results with the human expert counterparts.

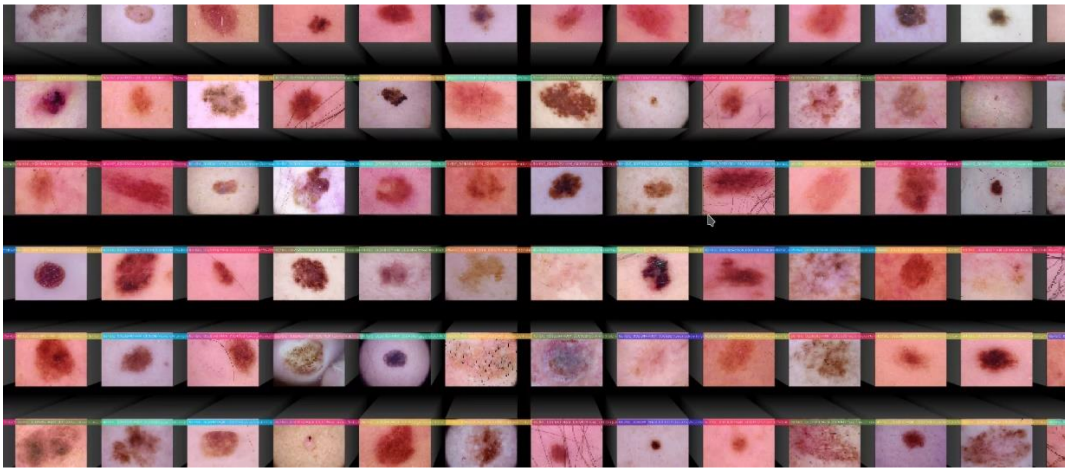


Fig5: Representative images from the HAM10000 dataset

4.2. INTERNATIONAL SKIN IMAGING COLLABORATION

The ISIC dataset consists of images of malignant and benign oncological diseases. These images were collected from the International Skin Imaging Collaboration and the subsets were divided into the same number of images, except for melanomas and moles, because those images were slightly dominant. The data is collected from Hospital Clinic de Barcelona, Medical University of Vienna, Memorial Sloan Kettering Cancer Centre, Melanoma Institute Australia, The University of Queensland, and the University of Athens Medical School. This dataset has garnered quite a lot of interest in researchers because of the yearly updates provided. The issue with the dataset is that there are duplicates embedded, and they must be eliminated at the initial stages. Even the HAM10000 dataset is constructed using some data from the ISIC dataset.

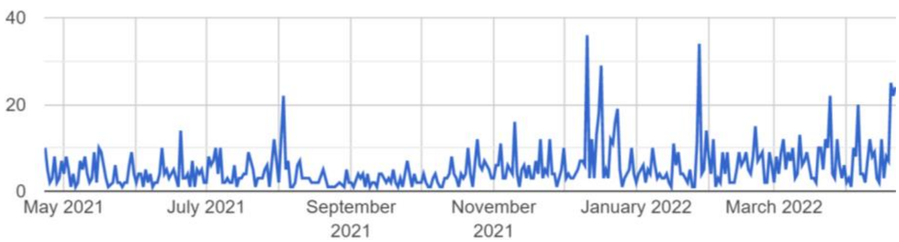


Fig6: Number of downloads for the ISIC dataset from Kaggle (to depict the interest in the dataset)

5. PRACTICAL APPLICATIONS

5.1 HUMAN – COMPUTER COLLABORATION

One such research uses a 34-layer residual network(ResNet34), which is a type of convolution Neural Network (CNN) on the training dataset from one publicly available pigmented lesions image benchmark, with seven diagnostic categories including malignant (Melanoma's (MEL's), Basal Cell Carcinomas (BCC's), Actinic Keratoses and Intraepithelial Carcinomas (AKIEC's)), benign (Melanocytic Nevi (NVs), Benign Keratinocytic Lesions (BKLs), Dermatofibromas (DFs) and vascular lesions (VASCs)) proliferations from the HAM10000 dataset. When this was tested on a publicly available dataset, an accuracy of 80.3% was achieved, and when compared to the results of other published reader studies, this CNN model outperforms most humans, and it ranks in the top quartile of machine-learning that were tested and developed with the same dataset. A possible elucidation for this is that the pigmented and non-pigmented variants of keratinocyte cancer share common measures. But this cannot be definite in other settings, like, for example, the International Skin Imaging Collaboration (ISIC) 2019 challenge established that AI doesn't work consistently on out-of-distribution images. Adding to this, within the realm of pigmented skin lesions, AI-based support provides backing for the less skilled raters to improve the expertise in telemedicine. This study is neither simple nor concrete, it needs more cognitive engagement of time and decision-making. It was also observed that the dermatologists were in some cases uncertain with the decision made and choose to change their decision after looking at the result of the AI. This study suggests the shift from human computer competition to human computer collaboration and making testing possible in the hands of the intended user, and not as a stand-alone device.

5.2 DERMATOLOGIST-LEVEL CLASSIFICATION WITH DNNs

Skin cancer is analysed clinically by a dermatologist visually, commencing with a clinical screening followed by a dermoscopic analysis, biopsy, and histopathological examination. Automating this process is a challenging task because of the fine-grained variability in the appearance of the skin lesions. In order to achieve this task, GoogleNet Inception v3 CNN architecture was used. The GNet architecture is pre-trained on roughly 1.28 million images from the 2014 ImageNet Large Scale Visual Recognition Challenge, and it is trained using transfer learning. The model is validated using nine-fold cross-validation.

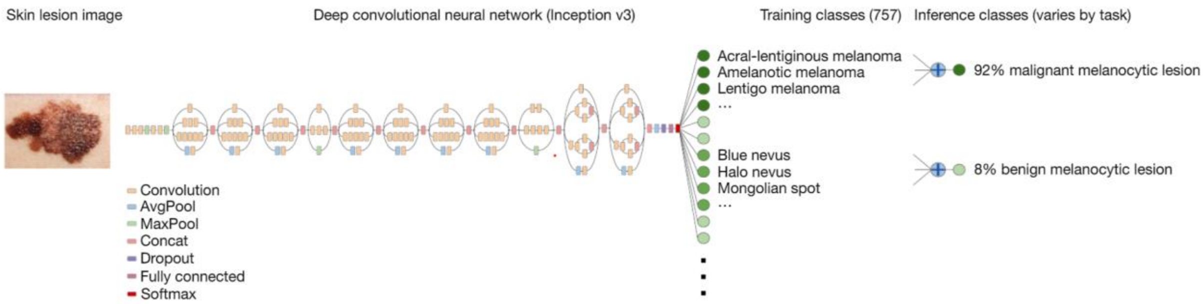


Fig7: Deep CNN layout

First, a three- class disease partition- the first level nodes of the taxonomy, which represents benign lesions, malignant lesions, and non-neo-plastic lesions. In this task, the CNN achieved an overall accuracy of 72.1% while 2 dermatologists who were part of the experiment attained 65.56% and 66.0% accuracy in the validation set from the ISIC dermoscopic archive. The second validation was performed using a nine-class disease partition, so that the diseases of each class have similar medical treatment plans. In this validation model, an accuracy of 55.4% was achieved, while the dermatologists attained 53.3% and 55.0%. The images of the validation set are labelled by dermatologists, but not necessarily confirmed by biopsy, which makes this metric inconclusive and shows that the CNN is learning the required relevant information.

5.3 DEEP LEARNING ENSEMBLES FOR MELANOMA RECOGNITION

IBM has also shown interest in this field and have developed a smart computer system that is more effective at identifying a form of skin cancer than most dermatologists (at least in the early research stages). The technology is very young and has a long way to go, and a lot of more real-world tests to pass. The system was provided with hundreds of images from the International Skin Imaging Collaboration 2016 dataset that were labelled as skin cancer, so that it could understand the appearance of a "dangerous skin lesions". This process of exposing the computer to pictures helps the computer suss out the underlying traits of what patch is cancerous and what's harmless. The process of visual recognition uses segmentation and classification. With segmentation, skin lesion is identified and distinguished from healthy skin. This allows the system to subsequently perform analysis within two contexts. It is unsure as to when this tech will be available for the public to use.

Fig8: IBM's Image Processing model, marking, and prioritizing pigments of a patient in a primary-care setting

5.4 USING WIDE-FIELD IMAGES OF PATIENT'S SKIN FOR BETTER EFFICIENCY

The challenge is to quickly find, identify and prioritize suspicious pigmented lesions (SPLs) from a high volume of pigmented lesions which are usually assessed for potential biopsies. This method of using Deep Convolutional Neural Networks (DCNNs) to investigate SPLs via wide-range photography, which is mostly found in smartphones and personal cameras is being researched upon by researchers in MIT and elsewhere by using a new AI pipeline.

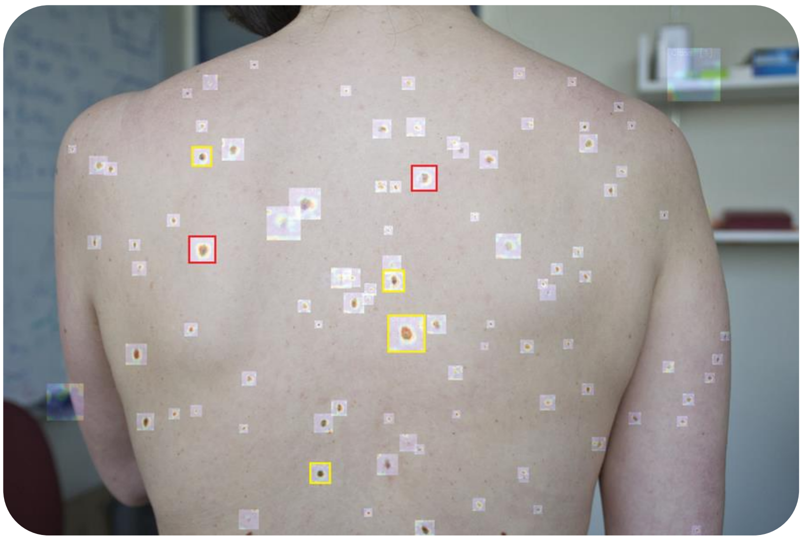


Fig8: Image Processing and prioritizing of pigments using a smartphone camera of a patient in a primary-care setting

The image captured by the smartphone shows a large skin section that is sent through an automated system that detects, extracts, and analyses all the pigments. Then the suspiciousness of each individual pigment is determined using the pre-trained DCNN, marking them (red = requires further inspection or referral to dermatologist, yellow = consider further inspection, and any other colour meaning that no inspection is required). In this case, the researched used 20,388 wide-field images from 133 patients at the Hospital Gregorio Maranon, Madrid to train the model, along with other publicly available images. The images were captured using a wide variety of ordinary cameras that are in the consumer market, readily available to purchase in the market. The sensitivity of the system in distinguishing the SPLs from non-suspicious lesions, while avoiding the need for separate lesion imaging sessions, which are cumbersome and time consuming is 90.3%.

6. CONCLUSION

This essay has revealed a few of multiple ways and techniques that can be used to detect skin cancer, mainly using Deep Neural Networks along with its many features, advantages and short comings. Identifying diseases and performing diagnosis is now becoming a task that makes a use of both man and machine. But the fact remains that humans tend to make mistakes due to various reasons, and not all the humans perform in the same manner. With the onset of modern computing, availability of tech and downsizing of faster and more efficient processors, we enter a time in which everybody has access to smartphones that has better processing power and efficiency than the best of the best computers from just a decade ago. The advancement in technology has also brought in the advancement of medical science, where both go hand in hand to bring out new methods and techniques to tackle problems of the modern medical industry.

Soon it is certain that detecting a skin cancer will become a task that'll be performed only using a computer. New and amazing research work is being performed in the field of Cancer Research, and with the help of machine learning and image processing, the process of detecting such anomalies in the body is made less complex than it was earlier. Although there is a lot more work to be performed in this field, a lot of the work done seems very promising with the hope of this advanced technology not only becoming more efficient, but more accessible to the general public. With the dawn of faster and cheaper technology at hand, and continued research in this field, the next dermatologist appointment for a pigment on the skin can be done at home itself, without the need to travel and wait for an appointment.

BIBLIOGRAPHY

- Mahamudul Hasan, Surajit Das Barman, Samia Islam & Ahmed Wasif Reza (2019) ***Skin Cancer Detection Using Convolutional Neural Network*** - https://dl.acm.org/doi/abs/10.1145/3330482.3330525

- Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau & Sebastian Thrun (2017) ***Dermatologist-Level Classification Of Skin Cancer With Deep Neural Networks*** - https://pubmed.ncbi.nlm.nih.gov/28117445/

- NHS. ***Cancer*** - https://www.nhs.uk/conditions/cancer/

- Cancer Research UK. ***Melanoma Skin Cancer Statistics*** - https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/melanoma-skin-cancer

- World Health Organization. ***Cancer*** - https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/melanoma-skin-cancer

- Wikipedia. ***Deep learning*** - https://en.wikipedia.org/wiki/Deep_learning

- NHS. ***Skin Cancer (Melanoma)*** - https://www.nhs.uk/conditions/melanoma-skin-cancer/

- Ameri A (2020) ***A Deep Learning Approach to Skin Cancer Detection in Dermoscopy Images*** - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7753251/

- Bill Cassidya, Connah Kendricka, Andrzej Brodzickib, Joanna Jaworek-Korjakowskab, Moi HoonYap (2020) ***Analysis Of The ISIC Image Datasets: Usage, Benchmarks And Recommendations*** - https://www.sciencedirect.com/science/article/pii/S1361841521003509/

- Philipp Tschandl, Christoph Rinner, Zoe Apalla, Giuseppe Argenziano, Noel Codella, Allan Halpern, Monika Janda, Aimilios Lallas, Caterina Longo, Josep Malvehy, John Paoli, Susana Puig, Cliff Rosendahl, H. Peter Soyer, Iris Zalaudek & Harald Kittler (2020) ***Human–Computer Collaboration For Skin Cancer Recognition*** - https://www.nature.com/articles/s41591-020-0942-0

- Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun (2017) ***Dermatologist-Level Classification Of Skin Cancer With Deep Neural Networks*** - https://www.nature.com/articles/nature21056.epdf

- N. C. F. Codella, Q. B. Nguyen, S. Pankanti, D. Gutman, B. Helba, A. Halpern, J. R. Smith(2016) ***Deep Learning Ensembles For Melanoma Recognition In Dermoscopy Images*** - https://arxiv.org/ftp/arxiv/papers/1610/1610.04662.pdf

- Megan Lewis (2021) **An Artificial Intelligence Tool That Can Help Detect Melanoma** - https://news.mit.edu/2021/artificial-intelligence-tool-can-help-detect-melanoma-0402

- Ed Sperling (2018) ***Deep Learning Spreads*** - https://semiengineering.com/deep-learning-spreads/

- Rukshan Pramoditha (2021) ***The Concept of Artificial Neurons (Perceptrons) in Neural Networks*** - https://towardsdatascience.com/the-concept-of-artificial-neurons-perceptrons-in-neural-networks-fab22249cbfc