

Detection of Skin Cancer using Machine learning Algorithm such as Convolutional Neural Network (CNN).

Semester 1 - 2019

Understanding Research –
IFN600

Vinay Huchanahalli Nagaraju – n10180893

Content

1. Problem Statement -----	02
2. Approach -----	03
3. Literature Review -----	04
4. Conclusion -----	06
5. References -----	06
6. Reflective Statement -----	07
7. Appendix -----	07

Abbreviations:

CNN – Convolutional Neural Network

ANN – Artificial Neural Network

DT – Decision Tree

SVM – Support Vector Machine

QUT – Queensland University of Technology

CPU – Central Processing Unit

GPU – Graphical Processing Unit

1. Problem Statement:

1.1. General Overview:

The research problem is to personalize the skin cancer diagnosis using machine learning algorithms such as Artificial Neural Networks. It will save the cost as well as time for diagnosis. Machine Learning algorithm will learn the classification based on the analysis of previous cases. It can be called as "Supervised Learning". It will mainly benefit to the rural area people and people from developing countries. Rural people from developing countries do not have proper skin cancer Diagnosis center. People do not have enough money for medical checks. The normal clinical lab test will need one or two days to diagnose skin cancer. Hence, personalizing the skin cancer will help them to diagnose skin cancer with cheaper/no cost. Diagnosis centers will take at least 2 or 3 days to detect cancer. But Machine will learn the Cancer pattern from previous cases and detects cancer quickly.

Detection of skin cancer using Machine learning Algorithms such as Convolutional Neural Network (CNN). CNN will generally learn based on an image of a single mixed resolution. CNN is a class of Artificial Neural Network (Krizhevsky, 2012)ⁱ. This method helps to detect cancer in the early stage. If skin cancer is detected in early stage then we can save millions of lives but our Machine Learning Algorithms will take very less time to detect cancer. We know that Machine has more computational power than humans. Hence, it detects skin cancer in a quicker manner. Melanoma is the most deathly of all skin cancers but early diagnosis can ensure a high degree of survival. In some emergency cases, Diagnosis centers want to detect the cancer in the early stage with less time. Accuracy of the CNN is roughly around 76%. However, it will be not that much useful for medical field because it needs more accuracy than that (Hinton, 2012). Hinton achieved 76% accuracy using CNN and which is highest among all the other machine learning algorithms. Diagnosis centers will take at least 2 or 3 days to detect cancer. But Machine will learn the Cancer pattern from previous cases and detects cancer quickly. New researcher can download the data set that he used for the skin cancer prediction is present in the website called "Kaggle".

1.2. Background:

Overall Trends and Conflicts:

Artificial Intelligence is in core trend in 21st century. It is used to find the patterns in the voice to autonomous cars. It can also use to detect the face and face recognition using the technique called CNN (Abdel-Zaher, 2016). There are very less research articles available on ANN (Artificial Neural Network) and CNN. Moreover, all research articles are latest and they are not cited by many people. I found many Trustworthy articles in Google Scholar and QUT Quick Search.

Previously, people used to detect skin cancer based on the base Machine Learning Algorithms such as SVM, Naïve Bayes, and Decision Trees. These Algorithms are weak learners and their accuracy to predict the cancer is very poor. CNN learns the pattern from the image and predict the occurrence of skin cancer based on the image of the patient and it will also consider the previous cancer patient cases. CNN will also learn from the patterns from the image and classify whether it is skin cancer or not.

2. Approach

2.1. Strategy:

I wanted to search articles which are related to personalizing skin cancer diagnosis using the Machine Learning algorithm such as CNN. I mainly used 2 techniques they are “Google Scholar” and “QUT Quick Search”. They helped me to get the relevant articles based on Language, Time (Latest), and Trustworthy. Google Scholar is effective because I got all the information and articles that are needed to understand my research problem. Information is retrieved based on the number of citations which also includes all the version of the article. I found all the relevant references. I also found Author information such as number of citation he/she got for all his research articles and what he/she contributed towards his domain.

I typed the below search terms in Google Scholar order to find the relevant articles.

The main search terms are

- a. Artificial Neural Networks
- b. Machine Learning Algorithms
- c. Supervised Learning
- d. Skin cancer

Main Phrases are

- a. Personalizing skin cancer
- b. Skin cancer diagnosis
- c. Skin cancer classification based on Machine Learning Algorithms
- d. Skin Cancer detection based on Neural Network Learning

QUT Quick Find is very helpful because we can search the relevant articles and apply filters based on

- a. Date
- b. Peer reviewed
- c. Articles/Books/News Papers etc.
- d. Keywords (keywords are mentioned in question 2)
- e. Language such as English, French, Chinese, etc.

Peer reviewed articles are trustworthy and cited by the professionals in that domain. QUT Quick Find is helpful to retrieve the information/articles quickly and accurately.

2.2. Analysis:

Initially approach is to search the research topic based on random Machine learning algorithms such as “(Linear regression | Logistic Regression | Support Vector Machine | Artificial Neural Network) algorithm to diagnose skin cancer”. After many search in Google Scholar and QUT Quick Search later it seems that Convolutional Neural Network (one of the Machine Learning algorithm) gives accurate result than all other base Machine Learning Algorithms. CNN have more articles which are related to skin cancer diagnosis (Zhang, 2016)ⁱⁱ. Hence, Zhang found that CNN is best Machine Learning algorithm to predict skin cancer.

2.3. Scope:

2.3.1. In Scope:

In Understanding research I analyzed the articles which are related to the Machine Learning algorithms such as CNN, SVM, and ANN to personalize skin cancer diagnosis.

Comment	Literature Source
Comparing the accuracy achieved among all the Machine Learning Algorithms such as SVM, DT, and CNN.	Research Articles.
There are many types of skin cancer such as Melanoma, Basal, and Squamous. I'm giving general interpretation of skin cancer prediction without discussing the specific type of skin cancer.	Theoretical articles, Tutorials and Lecture Notes.
Skin Cancer prediction using CNN Model.	Case Studies and Research Articles.
Analyzing the Various Machine Learning Algorithms.	Literature review.

2.1.2. Out of Scope:

1. I'm not implementing the Machine learning Algorithms to personalize skin cancer diagnosis.
2. I didn't mentioned any coding part of Convolutional Neural Network or other Machine Learning algorithms.
3. I'm not discussing about the basics of Machine Learning Algorithms.
4. I'm not expert in skin cancer domain and hence I didn't mentioned more on skin cancer.
5. I didn't discussed more on types of skin cancer and type of diagnosis for each type of cancer.

3. Literature Analysis

3.1. Major Concerns:

a. Blur Image:

The Major concern about the detection of skin cancer is the accuracy. Because if the Image is blur or image of skin cancer is not taken properly then CNN will not accurately detect the presence of the skin cancer. Image must have good number of pixels. So, usage of better camera is mandatory.

b. Less Data set:

CNN will take large amount of data set and it needs to learn from those data set. But, there is not much data sets are available. Data set is already present in the site called "Kaggle" but it is not having enough data set for accurate skin cancer prediction (Krizhevsky , 2012). It is also not preprocessed. Many rows are blanks and filled with some typed errors. In order to increase the accuracy, data need to preprocess.

c. Computational speed:

It will also needs large computational speed to learn the patterns of skin cancer. Normal computer don't have that much computational speed. Normal computer will run CNN on central processing unit (CPU). Graphical processing unit (GPU) is best suitable processing unit to run CNN to predict skin cancer(Hinton, 2012).

3.2. Influential studies:

I found 3 beautiful research articles for skin cancer detection using Machine Learning Algorithms such as CNN.

1. Computer aided diagnosis for skin cancer is best article to understand about the techniques to identify skin cancer using machine learning algorithms (Masood, 2013)ⁱⁱⁱ. The article emphasize more on the machine learning algorithms and image detection techniques. This article also discussed more on types of skin cancer and its classification technique. This research article also discussed about computer vision and implementation of algorithms using python programming language.
2. Another Research article is about the identification of the cancer in early stage (Wulfkuhle, 2013)^{iv}. Article emphasized more on the nature and the behavior of the cancer. It also explained more on the medical part of the cancer rather than Machine Learning. He also mentioned few things about the prevention part. This research article explained the different types of cancer and the method to classify it.
3. Another Research article is about the Computer-aided diagnosis in medical imaging. The article gave the brief introduction to the computer vision and methods to identify the skin cancer using the Machine Learning Algorithms (Abdel-Zaher, 2016)^v. It also gives the brief introduction to the Deep Neural Network and Multi-Layer Perceptron. Deep Neural Network is the advanced version of the Neural Network.

3.3. Agreements, Disagreements, Tensions, and Contentious issues:

Agreement and Disagreement: All research articles agreed that Convolutional Neural Network is best suitable to detect the cancer accurately. Because, accuracy of CNN is more than all the other algorithms. Hence, it is easy to find more research articles which are related to CNN to detect skin cancer compare to other Machine Learning Algorithms. Few research laureate thinks that some mistakes from the machine learning algorithms will cost lives. Because, no machine learning algorithms achieved 100% accuracy. Hence, they are planning to keep Machine learning algorithms out from the medical field. But, some scientists are against it. Because machine learning is growing filed and they predict that it will take time to achieve high accuracy.

Tensions and Contentious issues: Few Researchers thought that Machine Learning will take the job of Doctors. But Machine require more time and more analytical power. The most contentious issues about the Machine Learning algorithms is the replacement of human power. Many people thinks that Artificial Intelligence will reduce the jobs and create the unemployment. It is true for some extent. Researchers are thinking that Machine can never replace human jobs. It is just helping humans to complete their task quickly and efficiently.

3.4. Limitations of CNN:

CNN follows the supervised learning methodology technique. It will learn skin cancer patterns from the previous data and detect the cancer in patients. But CNN will not support for unsupervised learning. Unsupervised Learning don't require any labelled data set to detect the skin cancer patterns. It will cluster the unlabeled dataset and detects the cancer. But there is very less research articles are available for it. We don't have enough data set to predict skin cancer using CNN in Machine Learning Algorithm (Krizhevsky, 2012).

Another major gap is CNN is a new field and it first introduced in the year 2012. Hence, we can't get any research article before that. There are very less research articles available on Convolutional Neural Network (Ali Al-Jumaily, 2013) and Image recognition. Accuracy of CNN is good when compare to all other algorithm but it is not enough for the medical field (Gutman D, 2016)^{vi}.

4. Conclusion:

Machine learning in skin cancer detection makes an impressive progress. Next this is to find the path to increase the accuracy in CNN. Researchers are also working in order to introduce another advanced algorithm to improve the skin cancer prediction. Big MNC such as Google, Microsoft and others needs to concentrate more on Machine Learning Methodologies by investing more on Research and Development.

CNN will help to predict skin cancer using the Supervised Learning approach. It will save the time as well as money. It does not require any diagnosing equipment like the traditional clinical approach. It will detect the cancer in early stage and helps to save many lives. Developing countries will get more benefit from this approach. Because, most people are under the poverty line and they cannot afford more money on cancer diagnosis.

Accuracy of CNN needs to improve from 76% to at least 85%. Model needs to train again and again with different data set. There are many procedures to improve the accuracy. First one is to recruit more human resource to collect skin cancer data from different hospitals and integrate them for Data Preprocessing. In data preprocessing, unnecessary information should be deleted. Model needs to train from it until the desired accuracy is achieved. Hybrid model can also use to improve the accuracy. The combination of hybrid models are like CNN + SVM or many other combinations. But There is no much articles related to it.

5. Reference:

ⁱ Krizhevsky, A., Sutskever, I., Hinton, G.E. (2012): ImageNet classification with deep convolutional neural networks. In: NIPS, pp. 1097–1105.

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

ⁱⁱⁱ Masood A, Ali Al-Jumaily A (2013): Computer aided diagnostic support system for skin cancer: a review of techniques and algorithms. International journal of biomedical imaging.

^{iv} Wulfschlegel JD, Liotta LA, Petricoin EF (2013): Early detection: proteomic applications for the early detection of cancer. Nature reviews cancer.

^v Abdel-Zaher AM, Eldeib AM(2016). Breast cancer classification using deep belief networks. Expert Systems with Applications. 2016.

^{vi} Gutman D, Codella NCF, Celebi E, Helba B, Marchetti M, Mishra N, Halpern A. arXiv.(2016): Skin lesion analysis toward melanoma detection: A challenge at the International Symposium on Biomedical Imaging (ISBI).

6. Reflective statement:

Week	Tutor	Student
1	Patrick explained about the unit overview and he gave us the sticky note to write our research questions. He explained about brainstorming. How to pose the good research questions. He formed team and asked us to discuss about our research questions.	I understood how to pose the qualitative research question. How to pose the research questions using Brain storming.
2	He explained about the problem statement. How to write problem statement and how to search the trustworthy articles. How to use filter in QUT Quick Search in order to get the relevant articles.	I understood how to search the peer reviewed articles from Google Scholar and QUT quick search. I understood about problem statement and what information needed to include in it.
3	He taught us about the Literature review. How to add novelty, what information need to include in the literature review. He took one example from pacemaker technology and explained about literature review.	I understood how to write the literature review and the information that are needed to include in it.
4 and 5	He again explained about the problem statement and literature review. He mentioned the triangle format for the literature review. In week 5, he gave more importance to the assignment. He mentioned what he is expecting in the assignment 1. He also explained about the marking criteria for assignment 1.	I understood about the problem statement and literature review clearly. I got clear picture about the assignment 1 and the assessment criteria.

7. Appendix: Review from Supervisor

Name: Vinay Huchanahalli Nagaraju

ID: N10180893

Supervisor: Patrick Delaney

Supervisor review	Comment
1. Problem statement is not clear	I changed the problem statement and made it clearer.
2. Add the supportive reference to the Analysis	Added the evidence literature to it.
3. First person should not include in problem statement and literature review.	Removed all the first person in this literature review.
4. Instead of using points, it is better to write in paragraph in conclusion.	I changed few points to the paragraph.
5. Conclusion must have 3 things, 1. What next?, summary and recommendation.	I included these 3 things in conclusion.