

Artificial Intelligence and Machine Learning

(6CS012)

BLOOD CANCER DETECTION

Student Id : 1928584

Student Name : Sunil Ghimire

Group : C3G4

Submitted to : Prakash Gautam

Cohort : 3

Submitted on : 29th Apr 2020

**Table of Contents**

[1. Finding a research topic 1](#_Toc42261423)

[2. Professional Activities 2](#_Toc42261424)

[2.1. Gantt Chart 2](#_Toc42261425)

[3. Literature Review 4](#_Toc42261426)

[4. Reflection of Research Project 8](#_Toc42261427)

# Finding a research topic

Just over a month ago, I did what seemed like a small research at first: a mine Convolutional Neural Network (CNN) revealed something I did not expect. Yet the more I researched, the more I noticed that I had witnessed the beginnings of a break in the base of CNN architecture and a first hint of a brand-new form of CNN.

The project i.e. Detection of Blood Cancer is the result of approximately a month of research I have done to build the latest form of CNN architecture. I have never imagined it to take this long, but I've learned vastly more than I ever felt imaginable, and in fact what I have done now touches almost existing types of blood cells and quite a bit besides.

According to the Leukemia & Lymphoma Community, one adult in the U.S. is infected with blood cancer around every 3 minutes and an approximate number of 174,250 individuals in the U.S. are predicted to have leukemia, lymphoma, or myeloma in 2018. The projected reported diagnoses in 2019 was about 61,780 and the percentage of all current cancer cases is 3.5 percent, according to the National Cancer Institute. As in the case of acute leukemia, the individual died within a few months if the treatment is not performed in a specific period. And identifying cancer in the early stages is very important for managing this type of cancer or any other type of cancer.  It takes more time and effort to do the detection process by technicians manually and it costs more with the help of the instrument.

The aim of this research is to learn our capacity to analyze various CNN architecture and operating processes focused on the identification of blood cancer and acquire information about the specific parameters and hyperparameters that make up a complete system.

The research objectives are:

1. Getting the information by proper research
2. Develop a system that can automatically detect cancer from blood cell image
3. Able to visualize the model summary of the CNN architecture
4. To configure and reconfigure the model made by the aggregation of the blood cell images

It has certainly taken me a month to come to terms with the conclusions I have reached. And while I hope I have put into my research and project will make it easier for others. Some of the research strategy guide for finding quality research are listed below:

1. Consider the scope of my topic
2. Locate background information
3. Identify information needs
4. Analyze and adjust research strategy

To maximize my knowledge regarding blood cancer detection is one of the most rewarding benefits of eLearning provided by IEEE which helps me to find most desirable knowledge about detecting cancer using CNN architecture also provides best knowledge on model summary along with weights, bias, back propagation, input, hidden and output layer, flattening layer, pooling layer, activation function and fully connected layer.

For me what has always been most important is the actual process of research. For I know of nothing exciting as to glimpse for the first time some new and basic truth.

# Professional Activities

## Gantt Chart

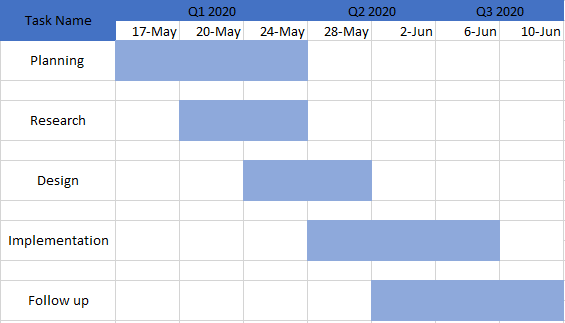


Figure : Gantt Chart

When talking about the research and professional activities, day zero is a day when I know that I am interested about the machine learning outside the scope of the medical, books, and Wikipedia posts that are more easily available or originally attracted my attention. It helps to be ferociously enthusiastic about the topic, as this is the initial stage that gets me through stressful moments.

Day One is expected to run for a couple of years. Day One describes the related authorities, scholar, and subject-specific scientific publications and include the sort of cutting-edge and in-depth work that I am hoping to read more about.

**Day One: Identifying Researchers**

* Academia.edu and ResearchGate are the perfect for discovering individuals who think for the same stuff I Do. And, depending on the topic, that’s Quora and Reddit
* I can often begin to know experts after I have served on variety of related articles.

**Day One: Identifying Relevant Journals**

This one is challenging, and it takes a lot of time and effort in my experience. The working mechanism of my topic typically occurs in journals and, least in my area (Mathematics and Algorithms), will take couple of years before it occurs common books on the topic.

Several resources that will enable me to classify the relevant journals are:

* Google Scholar
* The catalog search engine of the university library
* Share problem with module leader because they are amazing to figure out how to find stuff.

When I have any relevant research papers which support project, I need to sit down and read the paper for couple of hours and bibliographies of that research paper helps me to find additional resources. Some of them are listed below,

* Which authors are frequently cited? This could boost the selection criteria above
* What journals are giving me the kind of knowledge that I am searching for?

**Day Two: Be visible**

Once I have spent a lot of time to gain more experience in my profession, it is time to focus on getting more noticeable in the field. If I want people to take my work seriously, I need to walk very fine line. I have got to be clear researcher, so I have got to be patient not to mention something that I have studied in a lot of detail. I am still studying, because it takes me couple of times to recognize what a reasonably well-researched, well-supported hypothesis looks like.

Some of the points are listed below which helps to be visible:

* Build a profile and follow the people who research the same thing you do.
* Answer a ton of technical questions regarding project at Q&A sites like stack overflow and Quora
* Attend conferences in field of interest
* Submit articles/project to peer-reviewed papers in field of interest

# Literature Review

Making a mind map is a practice that makes me understand and gives my mind ideas when I see connection between certain things with my project. So, that is what it is for sometimes, support for thinking and encouraging innovation. I often use mind maps to organize a project, and then to control the project while it progresses. I come back to it periodically to study it in the light of changes incidents and new results. And how I “USE mind maps” is the subject of project supervision. For the subject of project supervision, I always use the visualization software like Microsoft Visio and Balsamiq to avoid too much re-drawing. Many times, I use mind maps to coordinate information when researching a topic. This needs to be accomplished with apps, as I add notes and site kinks or web archive files to the mind map- it allows me a live connection and all observations are available from the mind map, while I focus on the investigation. Finally, I also use a mind maps to go back to one I did a long time ago to give me a head start to current idea close to a previous one.

There are several published research papers to detect blood cancer. A paper published on “Leukemia Blood Cell Image Classification Using Convolutional Neural Network”. The main purpose of this paper is to introduce a system focused on the Convolutional Neural Network (CNN) to differentiate between normal and abnormal representations of blood cells which is based on computer vision in recent years. The algorithm for this this approach consists of several rigid steps: image pre-processing, clustering, morphological filtering, segmentation, feature selection or extraction, classification, and evaluation. For the detection and classification of blood cells in images, most of authors in the literature have adopted machine learning techniques such as K-means clustering with some disadvantage like time-consuming and mainly to determine the optimize the precision of classification. While deep learning will automatically learn and remove high-level attributes and at the same time to perform classification. The experiment in this research paper with 1118 images were performed on Matlab where 70% i.e. 831 images is used for training and the remaining 30% i.e. 357 images is used for testing set. Therefore, proposed CNN model with the size of the input volume to increase the precision of leukemia detection achieved the accuracy of 96.6%. Finally, from the 162 normal blood cell images, 152 were predicted as the normal cell images and from the 192 abnormal cell images, 2 were the normal cell images and 193 were abnormal cell images (Vununu, et al., 2018).

Another research paper published on “Automatic detection and classification of leukocytes using convolutional neural networks”. The key goal of this paper is to build an automatic system for detection and classification for superficial blood representations of WBCs. Based on the relation of color R, B, and morphological operation to remove noise and full nucleus this paper first proposes an algorithm for detecting WBCs for microscope images from peripheral blood images. PRICoLBP and SVM feature were first used with the detected WBCs to first distinguish eosinophils and basophils from other three types of WBCs they are neutrophil, monocyte, and lymphocyte where CNN is used to remove high-level features, and a random forest is added to separate them. Some detection experiments on the Cellavison database and ALL-IDB database show that proposed method of detection has a stronger and impact almost than iterative threshold method with a lower cost period, and some classification experiments show that proposed method of classification is more reliable than almost every other process (Zhao, et al., 2016).

Accuracy comparison of proposed system of classification with Sayed method and HSVM on the mixed sample of Databases Cellavision , ALL-IDB and Jiashan is show below

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Methods | Basophil (%) | Eosinophil (%) | Lymphocyte (%) | Monocyte (%) | Neutrophil (%) | Classification accuracy (%) |
| Proposed Method | 100 | 70 | 74.8 | 85.3 | 97.1 | 92.8 |
| Sayed | 53.0 | 63 | 85.0 | 39.0 | 50.8 | 76.8 |
| HSVM | 43.8 | 0 | 66.8 | 0 | 7.5 | 76.6 |

Another research paper published on “A Deep Learning Based Framework for Accurate Segmentation of Cervical Cytoplasm and Nuclei”. The key goal of this paper is to propose a system of segmentation which focused on superpixel and convolution neural network (CNN) for the segmentation of cervical cancer cells. Since, the contrast between background and cytoplasm is not relatively, cytoplasm segmentation is done first which is based on experience and observation. CNN which is based on deep learning is investigated for field identification of interest. A course of fine nucleus segmentation is also being established for segmentation of cervical cancer cells and further refining. The result after the experiment that for nucleus area identification an accuracy of 94.5 percent and recall of 0.8726±0.0008 and and a precision of 0.9143±0.0202 are achieved. The experiment in this research paper performed with 1400 of dataset where 1200 image are used for training set and 200 images are used for testing test. Different algorithms like CNN, backward propagation neural network (BPNN), probabilities neural network (PNN), support vector network (SVM), and learning vector quantization (LVQ) are used to detect performance evaluation based on accuracy, sensitivity, specificity, and f1-score where CNN has best result (Song, et al., 2014).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algorithms | Accuracy | Sensitivity | Specificity | F1 measure |
| BPNN | 0.8900 | 0.8578 | 0.9286 | 0.8947 |
| PNN | 0.8775 | 0.8912 | 0.8647 | 0.8753 |
| SVM | 0.8975 | 0.8841 | 0.9919 | 0.8993 |
| LVQ | 0.9000 | 0.9124 | 0.8883 | 0.8985 |
| CNN | 0.9450 | 0.9406 | 0.9495 | 0.9453 |

Another research paper published on “Unsupervised Blood Microscopic Image Segmentation and Leukemia Detection using Color based Clustering”. The key goal of this paper is color-based clustering of stained blood smear images of WBC nucleus segmentation accompanied by appropriate attribute extraction for leukemia detection. Some standard clustering technique like K-means, K-Medoid, Fuzzy C-Means (FCM), Gustafson Kessel (GK), and unsupervised blood microscopic image segmentation and Leukemia identification using color based clustering 483 Fuzzy Possibilistic C Means (FCM) were used for color-based segmentation and their output was compared. The paper also proposes two innovative methods for calculating boundary defects of nuclei using the aspect of the hausdorff and the signature of contours. Satisfactory results were obtained for leukemia detection identification using SVM classifier with the proposed features. The proposed approach is validated with microscopic images gather in blood and adequate findings were obtained.

Reviewing the relevant literature is the researcher’s most important work to complete before writing an effective proposal or starting a project that will help the scholar to give final shape of goals. And lets the researcher delineate the nature of his/her research and establish a sound approach, the scholar must complete a pilot survey and offer the idea tuning to make it more effective. A study of the literature indicates that have carried out extensive research on topic and are aware of past research which can offer a brief analysis of evidence for chosen topic. The key component which I have forget is novelty when writing a review. A review should not regurgitate information a reader might obtain from other reviews. Some of the points that I have choose while writing a review:

1. I have addressed a current subject by bringing together references that were never addressed within each other’s context. This would usually include identifying a similar theme in 50-100 source as well as thinking on how to exploit the link between them to discover new patterns.
2. I have addressed a novel issue or placing a fresh proposal or focus on a current category of publications already addressed in other articles. Many articles are published with a review portion outing a possible usage of the results. However, as several more similar articles are written in an area, other emerging technologies are more obvious depending on certain relevant or different innovations, and these new points of debate may be written in a summary.

Some of research methodology I have followed while doing this research are listed below:

1. Identification of research problem
2. Literature review
3. Specifying the purpose of research
4. Data collection
5. Analyzing and interpreting the data
6. Reporting and evaluating research

The step in this project usually reflect the overall method but should be treated as a continues shift rather than a defined collection of steps. Most of the research in this project stats with a general statement of the problem, or rather, the purpose of the study. The review of literature identifies flaws or holes in previous research which justifies the study. The research purpose identifies one hypothesis. So, while doing blood cancer detection I have collects the dataset from Kaggle to test the hypothesis and then analyzes and interprets the data via different mathematical techniques via graph. The effects of the data collection are then recorded and analyzed in affirmation or failure to refute the Null hypothesis. At the end of I can explore more study areas of blood cancer detection which makes me to describe the methods to plan to use to undertake different research project.

# Reflection of Research Project

Some of the key achievement while doing blood cancer detection I have found existing methods for diagnosis and they are listed below:

1. Medical history and physical examination: The record of present symptoms, and problems a person has had in the past. The medical history of a person’s family also helps in diagnose leukemia.
2. Complete blood count (CBC): Blood is taken and the number of RBC, WBCs abd platelets checked under a microscope.
3. Bone marrow aspiration: Bone marrow is extracted from breastbone with the help of a needle. Under a microscope the extracted sample is examined to search for irregular cells.
4. Cytogenetic analysis: To help identify individual chromosomes, cytogenetic testing takes the blood or bone marrow. It shows chromosome abnormalities while help diagnose and identify the leukemia type. Typically reports are accessible in 3 weeks’ time.
5. Immunohistochemistry: In immunohistochemistry, the blood sample of the cells is loaded with specific antibodies. The change in color can be seen under the microscope which helps to decide what cells types are present.

While doing this project I have found before CNN there are several algorithms that people used for image classification. People used to create images features, and the feed them into some classification algorithm such as SVM. Some algorithms also used image level pixel values as vector of features. For example, you may be able to train an SVM with 784 features in which each feature is the pixel value for an image of 28 × 28. The automated attribute extractors from the image can be thought of as CNNs. Although we lack a lot of spatial contact between pixels if we use an algorithm with a pixel vector, a CNN essentially utilizes neighboring pixel knowledge to analyze the picture by convolution first and only using a predictive layer at the end. Yann le cun first proposed this definition for digit classification in 1998, in which he used a single layer of convolution. It was later popularized in 2012 by Alex net which used multiple layers of convolution to achieve state of the art on the net of photos. So, henceforth make them am algorithm of choice for challenges of image classification.

Following are the expertise I gave gained as a researcher

1. Segmentation method has been proposed using color-based clustering to obtain nucleus region and cytoplasm area from stained images of blood smears. SVM classifiers are implemented with the appropriate features and the tests are satisfactory.
2. Automatically identify white blood cells (WBCs) from peripheral blood images and recognize five types of WBCs they are eosinophil, basophil, neutrophil, monocyte, and lymphocyte. Eosinophil and basophil from other WBCs are first identified with a granularity attribute by SVM. Other three types are then identified using CNN to extract features, and these features are used by random forest to identify certain WBC.

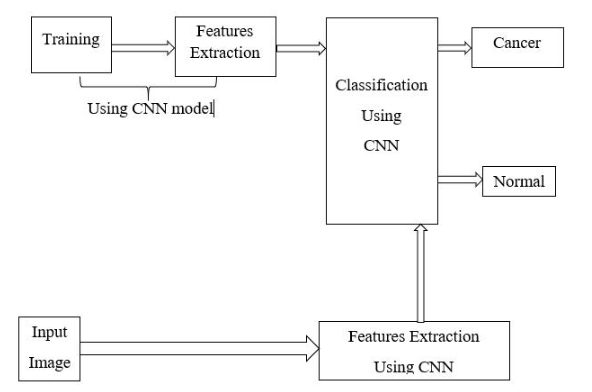


Figure : Block Diagram