BRAC University

Identification of Childhood Leukemia Using Deep Learning

By

Farana Naz Tultul

ID: 13101235

Supervised By

Mr. Moin Mostakim

Lecturer

Department of Computer Science & Engineering

BRAC University

DECLARATION

This is to certify that the research work titled “Identification of Childhood Leukemia Using Deep Learning” is submitted by Farana Naz Tultul to the Department of Computer Science and Engineering, BRAC University, in partial fulfillment of the requirement of the degree of Bachelor of Science in Computer Science and Engineering. I hereby declare that, this thesis is based on the results found from our own work. Materials of work by other researchers have been mentioned by references. This thesis, neither in whole nor in part, has been submitted previously for any degree. I carried out my research under the Supervision of Mr. Moin Mostakim, Lecturer, Department of Computer Science and Engineering, BRAC University.

Signature of Supervisor **Signature of Author**

ABSTRACT

Although cancer in children is rare, it is the leading cause of death past infancy amongst children. According to Afshar, Abdolrahmani, Tanha, Seif, Taheri(2010), Leukemia or blood cancer is one of the most common cancers in children, comprising of more than a third of all childhood cancers. Despite the advances of technology and research and overall decrease in mortality, nearly 2000 children die of cancer each year in the United States according to www.cancer.gov(2017). The website also tells us that if Leukemia cases are identified late or proper treatment isn't applied, then it can be mortal. For this reason, we have decided to use deep learning for the rapid identification of leukemia in the absence of doctors, which can be done in clinics by present nurses and lab workers. We are going to use ID3 and C4.5 (extension of ID3) classifiers, Naïve Bayes and Multi-layer Perceptron (MLP) Neural network on the data I have gathered of the 78 cases and check which one gives the most accurate result.

ACKNOWLEDGEMENT

My deepest thanks to my supervisor Mr. Moin Mostakim, without whose constant support and flexibility towards deadlines, completing my research work would have become really hard. He helped me through my thick and thin and always encouraged me when I would break down due to all the pressure as I did it all by myself.

I am ever grateful to my mother and father, who are a physician and an orthopedic surgeon respectively, for helping me out in learning about this malignant disease as much as they could!

I would also like to thank my parent's colleagues who have helped me for months in order to grant me the official permission of collecting data from National Institute of Cancer Research and Hospital, Dhaka.

Finally, a big thanks to my University for writing an official application to the National Institute of Cancer Research and Hospital on my behalf, so that they would let me do my research work in their institute.

Table of Contents

Declaration

Abstract

Acknowledgement

List of Figures

Abbreviations

Introduction

Literature Review

Methodology and Result

Conclusion

References

LIST OF FIGURES

Figure No. and Name **Page No.**

Fig 1: Input Layer 11

Fig 2: Hidden Layer 1 12

Fig 3: Hidden Layer 2 12

Fig 4: Output Layer 13

Fig 5: Application for permission to collect data 15

Fig 6: Application with approved signature from the hospital 16

Fig 7: The permission form 17

Fig 8: Certificate of Ethical Approval 19

Fig 9: Dataset from the journal Clinical Characteristics and Treatment Outcome of Childhood Acute Lymphoblastic Leukemia With the t(4;ll) (q21;q23): A Collaborative Study of 40 Cases (1991) 20

Fig 10: CSV file 21

Fig 11: CSV file(Continued) 22

Fig 12: CSV file(Continued) 23

Fig 13: ID3 & C4.5 algorithm result 24

Fig 14: Naïve Bayes algorithm result 25

Fig 15: Multi-layer Perceptron (MLP) Neural network algorithm result 26

ABBREVIATIONS

ALL - Acute Lymphocytic Leukemia

AML - Acute Myelogenous Leukemia

CML - Chronic Myelogenous Leukemia

CLL - Chronic Lymphocytic Leukemia

CNS - Central Nervous System

FAB - French-American-British group

CSV - Comma Separated Values

MLP - Multi-layer Perceptron

INTRODUCTION

Childhood should be the happiest and most carefree phase in every human's life. Every child is precious and deserves a normal and healthy childhood in order to have proper mental and physical growth, to grow up into a proper human being. However, there are a lot of unfortunate little souls who end up missing the essence of childhood completely. When a child should be running around in playgrounds, they have to run from hospitals to hospitals. Instead of being energetic and enthusiastic like a child should be, they feel malaise and suffer a lot in an age so tender. These unfortunate children are the victims of malignant diseases. One such malignant disease that drains out all the life from a child is Childhood Leukemia.

According to Canadian Cancer Society (2017), Leukemia is a cancer that starts in blood stem cells, which are basic cells that develop into different types of cells that have different jobs. In leukemia, there is an overproduction of blast cells as the stem cells of the blood become blast cells when they develop. Blast cells are immature blood cells which overtime, crowd out normal blood cells and as a result they can't do their jobs. Their presence alone is strong enough evidence to diagnose leukemia.

Generally there are 4 types of Childhood Leukemia according to Dr. Molla(April,2009). Amongst those I am going to work with Acute Lymphocytic or Lymphoblastic Leukemia (ALL).

In this research, we will be collecting real data of patients at first and then will be running some deep learning algorithms like ID3, C4.5, Naïve Bayes and Multi-layer Perceptron (MLP) Neural network on those data and figure out the accuracy of each algorithm. According to LeCun, Bengio and Hinton (2015), Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction.

LITERATURE REVIEW

According to the study from www.cancer.org(2017),Childhood leukemia is a type of leukemia, usually Acute Lymphocytic Leukemia (ALL), and a type of childhood cancer. It is the most common cancer in children and teens, accounting for almost 1 out of 3 cancers. Leukemia is a cancer that starts in early blood forming cells found in the bone marrow, the soft inner parts of certain bones. Most often, leukemia is a cancer of the white blood cells, but some leukemia start in other blood cell types. It is either acute (fast growing) or chronic (slow growing). Almost all childhood leukemia is acute.

As we have mentioned previously that generally there are 4 types of Childhood Leukemia according to Dr. Molla(April,2009). Two of them are acute and the other two are chronic.

The main types of acute leukemia are:
Acute Lymphocytic (Lymphoblastic) Leukemia (ALL): about 3 out of 4 childhood leukemia are ALL. This leukemia starts from early forms of lymphocytes in the bone marrow.

Acute Myelogenous Leukemia (AML): This type of leukemia, also called Acute Myeloid Leukemia, Acute Myelocytic Leukemia or Acute non-lymphocytic Leukemia, accounts for most of the remaining cases. AML starts from the myeloid cells that form white blood cells (other than lymphocytes), red blood cells or platelets.

Hybrid or Mixed Lineage Leukemia: In these rare leukemia, the cells have both features of ALL and AML. In children, they are generally treated like ALL and usually respond to treatment like ALL.

Chronic leukemia is much more common in adults than in children. They tend to grow more slowly than acute leukemia. They can be divided into two types:

Chronic Myelogenous Leukemia (CML): This rarely occurs in children. Treatment is similar to that used for adults.

Chronic Lymphocytic Leukemia (CLL): This leukemia is extremely rare in children.

There's another type of childhood leukemia which is, Juvenile Myelomonocytic Leukemia (JMML). This rare type of leukemia is neither chronic, nor acute. It begins from myeloid cells, but it usually doesn't grow as fast as AML or as slow as CML. It occurs most often in young children (under age 4). Symptoms can include pale skin, fever, cough, easy bruising or bleeding, trouble breathing (from too many white blood cells in the lungs) and an enlarged spleen or lymph nodes.

From this we can figure out that, there's a vast variety of symptoms for all the types of leukemia mentioned above and each one of them can be identified and therefore, treated differently.

However, in this research we will only be focusing on Acute Lymphocytic (Lymphoblastic) Leukemia (ALL) as the patients data we have collected are all ALL patients.

Deep learning is a fascinating field. According to DeepLearning.TV in the video "What is a Neural Network - Ep. 2 (Deep Learning SIMPLIFIED)" (2015), Deep learning is about neural networks. Artificial neural networks have been around for a long time, but something special has happened in recent years. The mixture of new faster hardware, new techniques and highly optimized open source libraries allow very large networks to be created with frightening ease

Jason Brownlee in his book 'Deep Learning With Python' said, "I have used many of the top deep learning platforms and libraries and I chose what I think is the best-of-breed platform for getting started and very quickly developing powerful and even state of the art deep learning models in the Keras deep learning library for Python".

From DeepLearning.TV's video(2015) we also get to know that, The structure of the neural network is like any other kind of network. There's an interconnected group of nodes which is called neurons and there are edges to join them together. A neural network's main function is to receive a set of inputs, perform progressively complex calculations and then use the output to solve a problem.

A neural network is highly structured and comes in layers. The first layer is the input layer, the final layer is the output layer and all the layers in between are referred to as hidden layers. Forward propagation is neural network's way of classifying. It is shown with pictures below.

First, nodes take input.

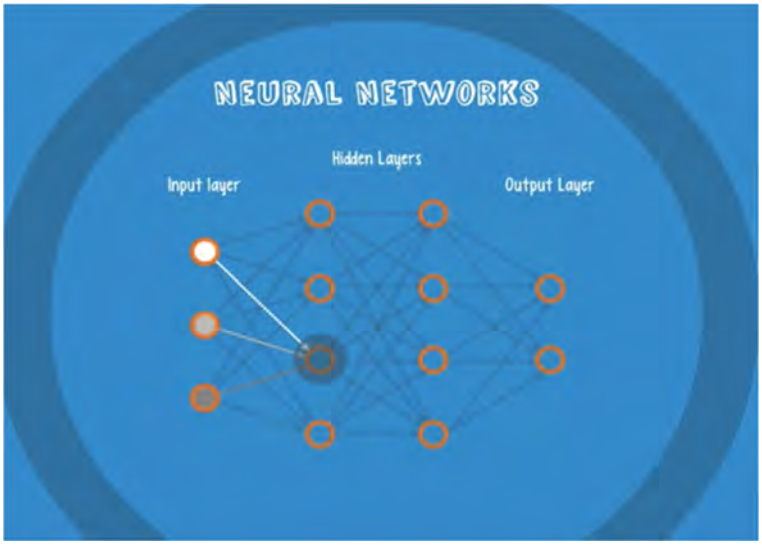


Fig 1: Input Layer

Then the result is passed on to the next layer.

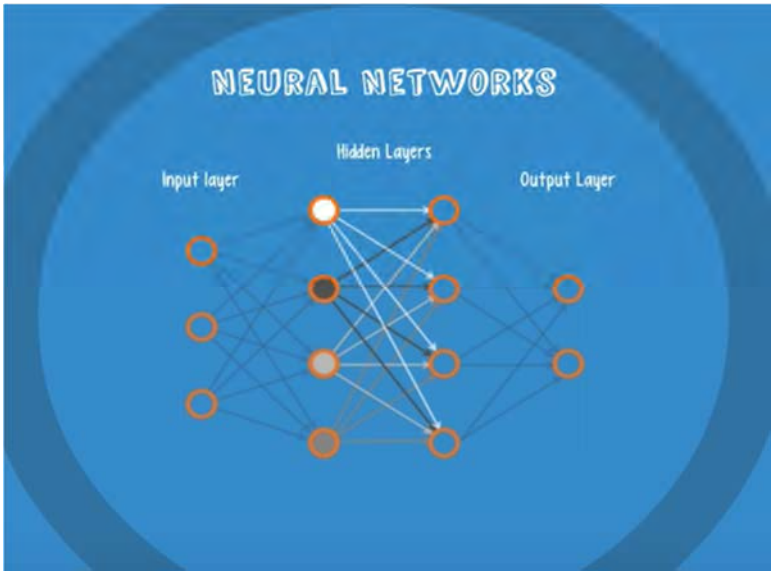


Fig 2: Hidden Layer 1

Then that result is passed on to another following layer.

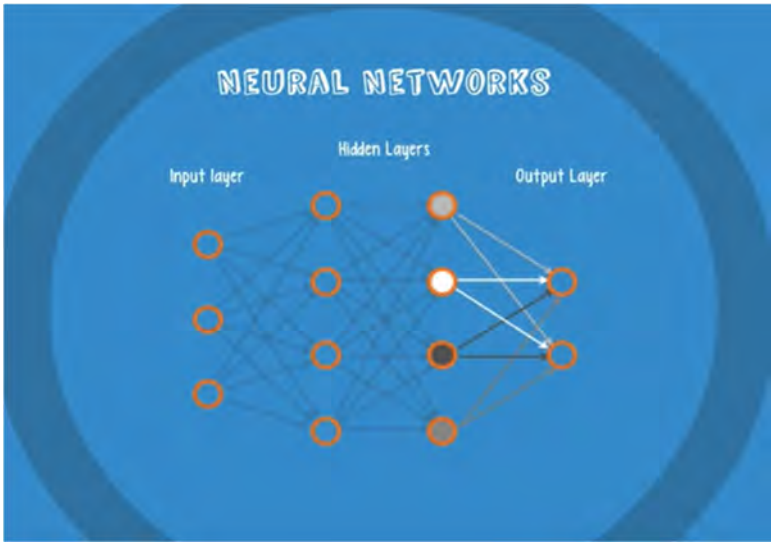


Fig 3: Hidden Layer 2

This goes on until we reach the output layer.

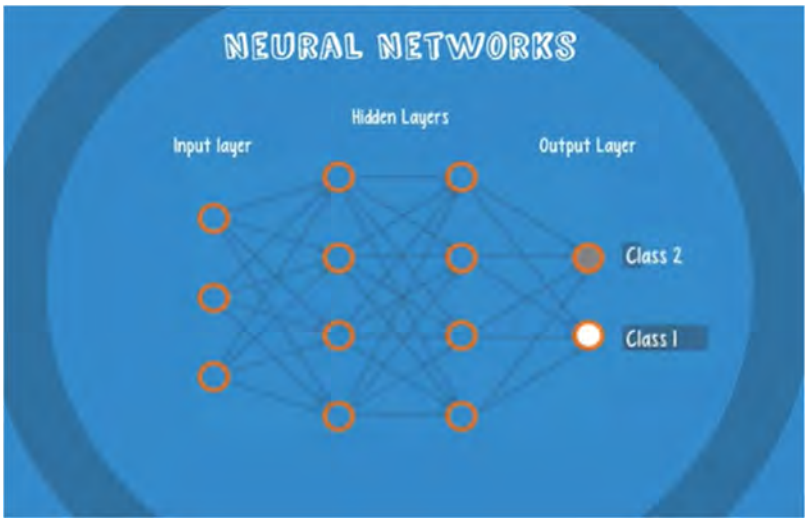


Fig 4: Output Layer

According to scikit-learn.org (2017), ID3 (Iterative Dichotomiser 3) was developed in 1986 by Ross Quinlan. The algorithm creates a multi way tree, finding for each node (i.e. in a greedy manner) the categorical feature that will yield the largest information gain for categorical targets. Trees are grown to their maximum size and then a pruning step is usually applied to improve the ability of the tree to generalize to unseen data.

C4.5 is the successor to ID3 and removed the restriction that features must be categorical by dynamically defining a discrete attribute (based on numerical variables) that partitions the continuous attribute value into a discrete set of intervals. C4.5 converts the trained trees (i.e. the output of the ID3 algorithm) into sets of if-then rules. These accuracy of each rule is then evaluated to determine the order in which they should be applied. Pruning is done by removing a rule's precondition if the accuracy of the rule improves without it.

METHODOLOGY & RESULT

At first we had to collect data for my research. For that reason we needed permission from National Institute of Cancer Research and Hospital. However, as we all know that trying to get our hands on the datasets of any government institution is very hard and requires some formalities, we also had to go through all of them.

At first we needed our University authority to verify our identities to the Cancer Institute. For that we needed an application written directly to the Director of National Institute of Cancer Research and Hospital by our University Registrar which we eventually got after writing a few applications which included my supervisor, Mr. Moin Mostakim's signature.

BRAC UNIVERSITY

BRAC UNIVERSITY
Inspiring Excellence

11 July 2017

The Director
National Institute of Cancer Research & Hospital
Mohakhali, Dhaka - 1212
Bangladesh

Dear Sir,

This is to certify that Farana Naz Tultul (BRACU ID # 13101235) is a student of the Bachelor of Science in Computer Science and Engineering program at BRAC University.

Currently, she is doing thesis under supervision of Mr. Moin Mostakim (Lecturer, Department of Computer Science and Engineering). Her thesis topic is "*Identification of Childhood Leukemia Using Deep Learning*". For completion of her thesis she needs data of the cancer patients of your institution.

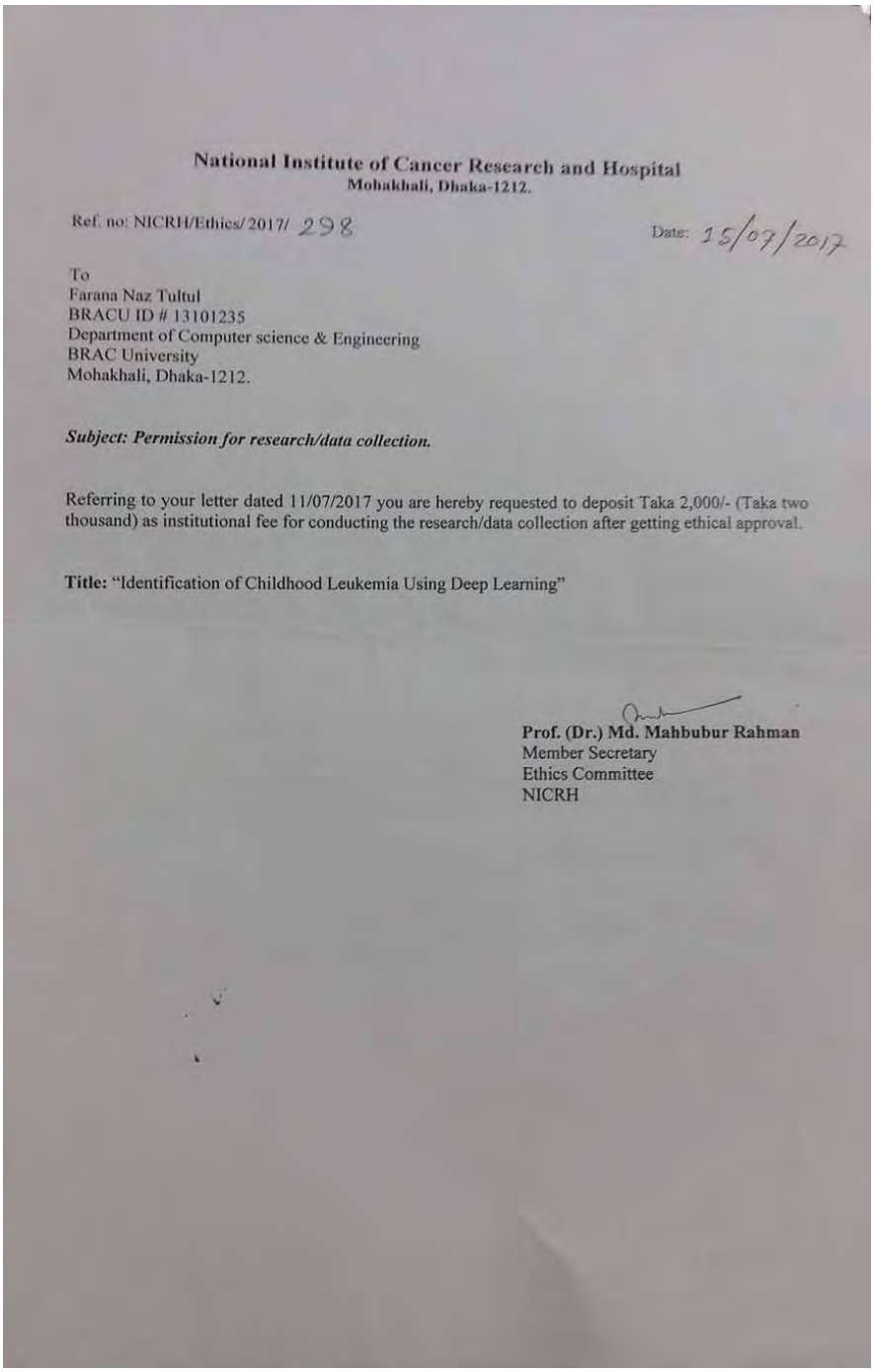
Therefore, I would like to request you to allow her to conduct her research work in your institution.

Muhammad Sahgol Afzal
Major General (Retired)
Registrar
BRAC University

66 Mohakhali, Dhaka-1212, Bangladesh Tel: +880 2 882 4051-4 Fax: +880 2 5881 0383 Email: info@bracu.ac.bd www.bracu.ac.bd

Fig 5: Application for permission to collect data

After that we had to go to the hospital for many days to make them approve of my request to collect data. It took us around two months to finally get the permission.



National Institute of Cancer Research and Hospital
Mohakhali, Dhaka-1212.

Ref. no: NICRH/Ethics/2017/ 298 Date: 15/07/2017

To
Farana Naz Tultul
BRACU ID # 13101235
Department of Computer science & Engineering
BRAC University
Mohakhali, Dhaka-1212.

Subject: Permission for research/data collection.

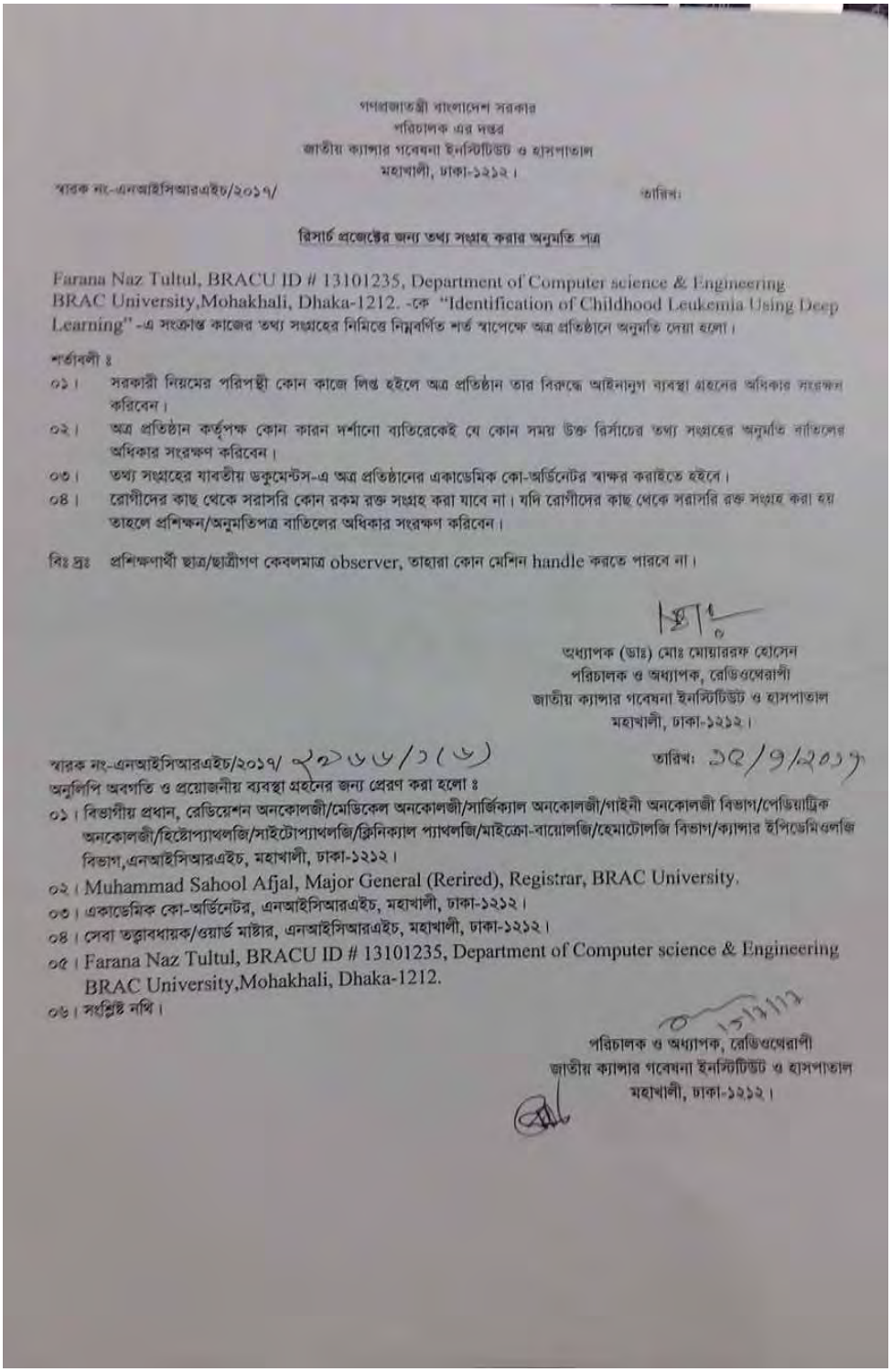
Referring to your letter dated 11/07/2017 you are hereby requested to deposit Taka 2,000/- (Taka two thousand) as institutional fee for conducting the research/data collection after getting ethical approval.

Title: "Identification of Childhood Leukemia Using Deep Learning"

Prof. (Dr.) Md. Mahbubur Rahman
Member Secretary
Ethics Committee
NICRH

Fig 6: Application with approved signature from the hospital

Then we finally got the form, showing which we could collect data from the hospital.



গণপ্রজাতন্ত্রী বাংলাদেশ সরকার
পরিচালক এর দপ্তর
জাতীয় ক্যান্সার গবেষনা ইনস্টিটিউট ও হাসপাতাল
মহাখালী, ঢাকা-১২১২।

স্মারক নং-এনআইসিআরএইচ/২০১৭/ তারিখ:

রিসার্চ প্রজেক্টের জন্য তথ্য সংগ্রহ করার অনুমতি পত্র

Farana Naz Tultul, BRACU ID # 13101235, Department of Computer science & Engineering BRAC University,Mohakhali, Dhaka-1212. -কে "Identification of Childhood Leukemia Using Deep Learning" -এ সংক্রান্ত কাজের তথ্য সংগ্রহের নিমিত্তে নিম্নবর্ণিত শর্ত স্বাপেক্ষে অত্র প্রতিষ্ঠানে অনুমতি দেয়া হলো।

শর্তাবলী ঃ

০১। সরকারী নিয়মের পরিপন্থী কোন কাজে লিপ্ত হইলে অত্র প্রতিষ্ঠান তার বিরুদ্ধে আইনানুগ ব্যবস্থা গ্রহনের অধিকার সংরক্ষন করিবেন।

০২। অত্র প্রতিষ্ঠান কর্তৃপক্ষ কোন কারন দর্শানো ব্যতিরেকেই যে কোন সময় উক্ত রির্সাচের তথ্য সংগ্রহের অনুমতি বাতিলের অধিকার সংরক্ষণ করিবেন।

০৩। তথ্য সংগ্রহের যাবতীয় ডকুমেন্টস-এ অত্র প্রতিষ্ঠানের একাডেমিক কো-অর্ডিনেটর স্বাক্ষর করাইতে হইবে।

০৪। রোগীদের কাছ থেকে সরাসরি কোন রকম রক্ত সংগ্রহ করা যাবে না। যদি রোগীদের কাছ থেকে সরাসরি রক্ত সংগ্রহ করা হয় তাহলে প্রশিক্ষন/অনুমতিপত্র বাতিলের অধিকার সংরক্ষণ করিবেন।

বিঃ দ্রঃ প্রশিক্ষণার্থী ছাত্র/ছাত্রীগণ কেবলমাত্র observer, তাহারা কোন মেশিন handle করতে পারবে না।

অধ্যাপক (ডাঃ) মোঃ মোয়াররফ হোসেন
পরিচালক ও অধ্যাপক, রেডিওথেরাপী
জাতীয় ক্যান্সার গবেষনা ইনস্টিটিউট ও হাসপাতাল
মহাখালী, ঢাকা-১২১২।

স্মারক নং-এনআইসিআরএইচ/২০১৭/ ২০৬৫/১(৬) তারিখ: ১২/৯/২০১৭

অনুলিপি অবগতি ও প্রয়োজনীয় ব্যবস্থা গ্রহনের জন্য প্রেরণ করা হলো ঃ

০১। বিভাগীয় প্রধান, রেডিয়েশন অনকোলজী/মেডিকেল অনকোলজী/সার্জিক্যাল অনকোলজী/গাইনী অনকোলজী বিভাগ/পেডিয়াট্রিক অনকোলজী/হিষ্টোপ্যাথলজি/সাইটোপ্যাথলজি/ক্লিনিক্যাল প্যাথলজি/মাইক্রো-বায়োলজি/হেমাটোলজি বিভাগ/ক্যান্সার ইপিডেমিওলজি বিভাগ,এনআইসিআরএইচ, মহাখালী, ঢাকা-১২১২।

০২। Muhammad Sahool Afjal, Major General (Rerired), Registrar, BRAC University.

০৩। একাডেমিক কো-অর্ডিনেটর, এনআইসিআরএইচ, মহাখালী, ঢাকা-১২১২।

০৪। সেবা তত্ত্বাবধায়ক/ওয়ার্ড মাষ্টার, এনআইসিআরএইচ, মহাখালী, ঢাকা-১২১২।

০৫। Farana Naz Tultul, BRACU ID # 13101235, Department of Computer science & Engineering BRAC University,Mohakhali, Dhaka-1212.

০৬। সংশ্লিষ্ট নথি।

পরিচালক ও অধ্যাপক, রেডিওথেরাপী
জাতীয় ক্যান্সার গবেষনা ইনস্টিটিউট ও হাসপাতাল
মহাখালী, ঢাকা-১২১২।

Fig 7: The permission form

However, despite having completed the formalities, they still could not provide us with any data yet as they had prospective data only, not retrospective. That too of only 6 to 7 patients!

As our deadline was approaching and as it would take time to collect the desired amount of data from them, we decided to use online dataset of childhood leukemia patients till we get the actual data from the hospital. Here's a picture of our Ethical Certificate which actually certifies that we tried collecting data from them and we still can take it from them in near future to further implement the algorithms.

National Institute of Cancer Research and Hospital
Mohakhali, Dhaka-1212.

Ref. no: NICRH/Ethics/ 2017/ 299 Date: 15/07/2017

Certificate of Ethical Approval

Title: "Identification of Childhood Leukemia Using Deep Learning"

Investigator: Farana Naz Tultul, BRACU ID # 13101235, Department of Computer science & Engineering
BRAC University,Mohakhali, Dhaka-1212.

Recommendation: APPROVED/~~APPROVED AFTER CORRECTION~~/~~NOT APPROVED~~

Note: The research study to be carried out must comply with the national laws and regulations of the country and "WMA declaration of Helsinki-Ethical Principles for Medicai Research Involving Human Subjects, amended, October 2008'

15/7/17

Prof (Dr.) Md. Moarraf Hossen
Chairperson
Ethics Committee
NICRH

Prof. (Dr.) Md. Mahbubur Rahman
Member Secretary
Ethics Committee
NICRH

Fig 8: Certificate of Ethical Approval

We took a dataset of 36 patients, all of whom had ALL. We made another 36 patients data with anomalies as we needed data of people who do not have ALL in order to train my algorithm with both Leukemia and non-Leukemia symptoms. Below is the table from which we used 36 patients data, ignoring the ones which has unknown parameters.

Table 1. Clinical and Laboratory Findings in the 40 Cases With t(4;11)(q21;q23)

Patient No.	Age	Race/Sex	Leukocyte Count (10^9/L)	Hemoglobin (g/dL)	Platelet (10^9/L)	CNS Leukemia	FAB	Remission Duration (mo)	Type of Failure
1	1 mo	W/F	432	10.4	93	Yes	L1	45+	—
2	1 mo	B/F	2.7	12.9	60	Yes	L2	5	BM + CNS
3	2 mo	W/F	31	12.5	220	Yes	L1	1	BM + CNS
4	2 mo	W/F	368	5	52	No	L1	11	BM
5	2 mo	W/F	336	5	17	No	L1	13	BM
6	2 mo	H/M	448	8	45	Yes	L1	10	BM
7	2 mo	W/F	12.1	4.5	25	No	L2	5	BM + CNS
8	3 mo	W/F	104	5.8	43	No	L1	6+	—
9	3 mo	W/F	832	6	41	Yes	L1	6	CNS
10	3 mo	W/M	78	10.7	336	No	L1	18	BM
11	3 mo	W/M	336	4.5	14	No	L2	3	Death
12	4 mo	W/M	234	6.9	171	Yes	L1	15	BM
13	4 mo	H/F	22	6	153	No	L1	17+	—
14*	4 mo	W/F	201	5.8	15	Yes	L1	29	BM
15	4 mo	B/M	78.2	6.6	173	Yes	L1	1+	—
16	6 mo	H/M	149	5.4	25	No	L1	6	T
17	7 mo	W/F	218.4	3.1	51	No	L1	3+	—
18	8 mo	W/F	256	7	8	No	L1	0	IF
19	8 mo	B/F	8.3	8.2	345	No	L1	0	IF
20	9 mo	W/F	816	Unknown	10	No	L1	5	BM
21	9 mo	W/F	78	5	124	No	L1	3	BM
22	9 mo	W/M	255	8.8	22	No	L1	17+	—
23	10 mo	W/F	2.5	6	15	No	L1	40	BM + T
24	11 mo	B/F	46	5.8	38	No	L1	3+	—
25	1 y 1 mo	W/F	61	8.1	176	No	L1	6+	—
26	1 y 1 mo	O/F	650	4	28	No	L1	2+	—
27	1 y 9 mo	W/F	164	5.7	5	No	L1	31+	—
28	2 y 11 mo	W/M	97	7	26	No	L1	0	IF
29*	3 y 10 mo	W/F	27.3	9.5	114	Unknown	L1	88+	—
30*	7 y 1 mo	W/M	12.1	10.2	159	No	L1	69+	—
31	8 y 6 mo	W/F	216	7.5	104	No	L1	4+	—
32	8 y 6 mo	W/F	102	10.5	61	No	L1	37+	—
33	10 y	H/F	320	5	25	No	L1	16+	—
34	10 y 2 mo	H/ M	49	9.7	26	No	L1	8	BM
35*	13 y 6 mo	B/M	345	6	110	No	L2	8	BM
36	13 y 7 mo	W/M	656	4.8	23	No	L1	0	IF
37	13 y 8 mo	B/M	581	11.3	313	No	ND	3+	—
38*	13 y 9 mo	W/F	346	6.6	19	No	L1	22	BM
39	13 y 9 mo	W/M	1.6	7.4	25	No	L2	6	BM
40	14 y 8 mo	H/ M	107	Unknown	45	No	ND	8	BM

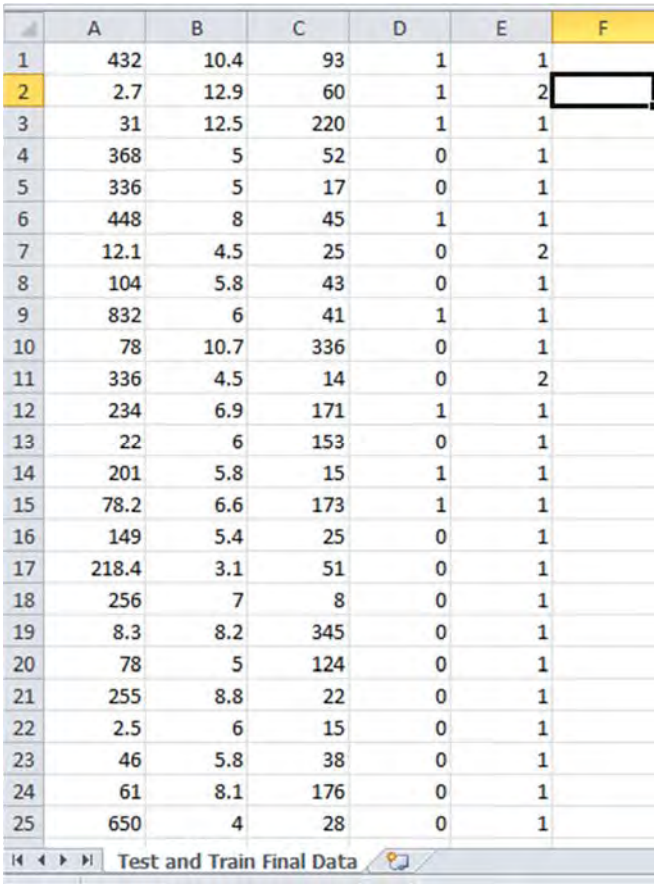
Abbreviations: B, black; W, white; H, Hispanic; O, Oriental; ND, not done; BM, bone marrow relapse; T, testicular relapse; IF, induction failure.
*Some clinical features and early treatment outcome have been reported previously.[23,52,53]

Fig 9: Dataset from the journal Clinical Characteristics and Treatment Outcome of Childhood Acute Lymphoblastic Leukemia With the t(4;ll) (q21;q23): A Collaborative Study of 40 Cases (1991)

We only took the Leukocyte count, Hemoglobin Count, Platelet Count and If the patient has CNS leukemia or not. CNS leukemia is central nervous system leukemia according to Molla (April, 2009). A patient with ALL can be with or without CNS. FAB is the outcome. L1 and L2 are types of ALL. L1 has better prognosis while L2 has more aggressive disease course.

In all the three algorithms we have used, we used Leukocyte count, Hemoglobin count, Platelet count and CNS leukemia as input data and FAB as output data. FAB is classification of ALL by the French-American-British group according to Molla (April, 2009).

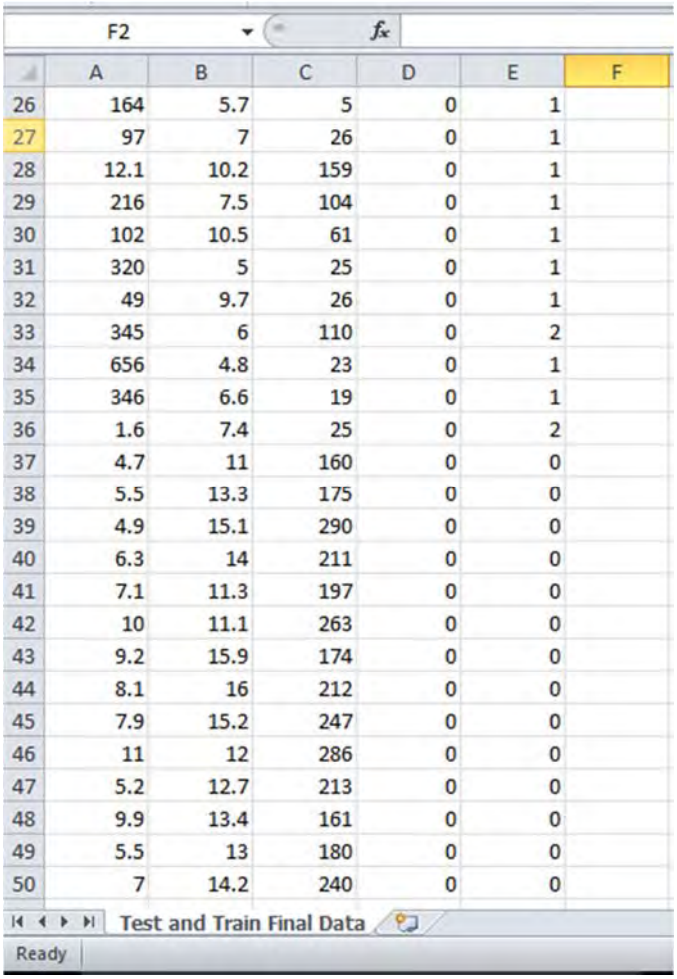
We made a CSV file, which is Comma Separated Data, with the 72 dataset we had. We used half of it for training the classification model and the other half for testing the model.



	A	B	C	D	E	F
1	432	10.4	93	1	1	
2	2.7	12.9	60	1	2	
3	31	12.5	220	1	1	
4	368	5	52	0	1	
5	336	5	17	0	1	
6	448	8	45	1	1	
7	12.1	4.5	25	0	2	
8	104	5.8	43	0	1	
9	832	6	41	1	1	
10	78	10.7	336	0	1	
11	336	4.5	14	0	2	
12	234	6.9	171	1	1	
13	22	6	153	0	1	
14	201	5.8	15	1	1	
15	78.2	6.6	173	1	1	
16	149	5.4	25	0	1	
17	218.4	3.1	51	0	1	
18	256	7	8	0	1	
19	8.3	8.2	345	0	1	
20	78	5	124	0	1	
21	255	8.8	22	0	1	
22	2.5	6	15	0	1	
23	46	5.8	38	0	1	
24	61	8.1	176	0	1	
25	650	4	28	0	1	

Test and Train Final Data

Fig 10: CSV file



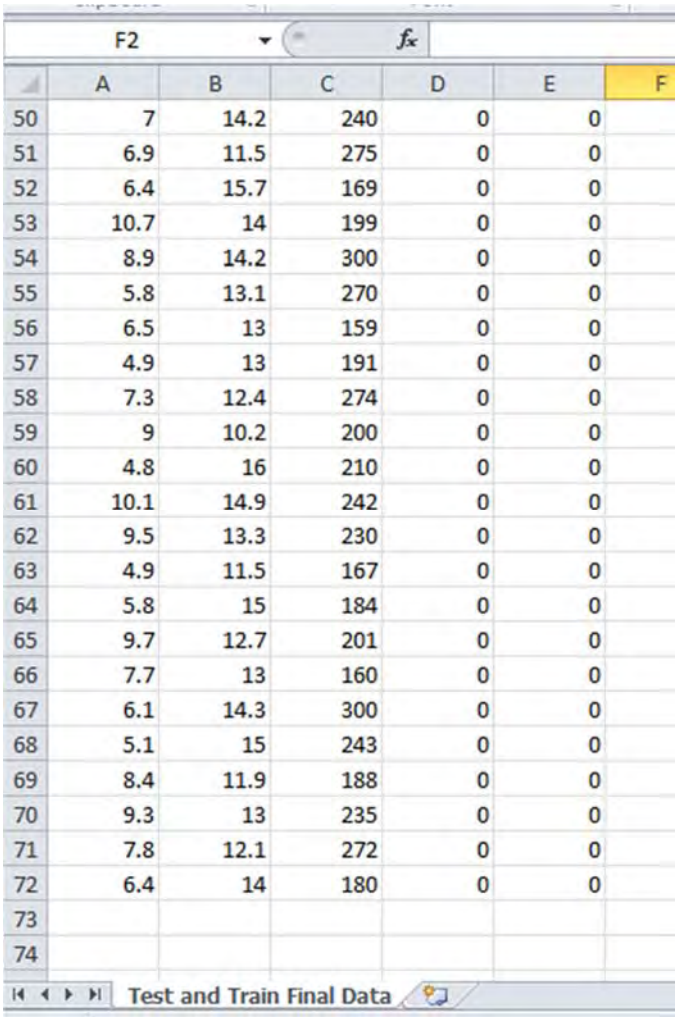
F2

	A	B	C	D	E	F
26	164	5.7	5	0	1	
27	97	7	26	0	1	
28	12.1	10.2	159	0	1	
29	216	7.5	104	0	1	
30	102	10.5	61	0	1	
31	320	5	25	0	1	
32	49	9.7	26	0	1	
33	345	6	110	0	2	
34	656	4.8	23	0	1	
35	346	6.6	19	0	1	
36	1.6	7.4	25	0	2	
37	4.7	11	160	0	0	
38	5.5	13.3	175	0	0	
39	4.9	15.1	290	0	0	
40	6.3	14	211	0	0	
41	7.1	11.3	197	0	0	
42	10	11.1	263	0	0	
43	9.2	15.9	174	0	0	
44	8.1	16	212	0	0	
45	7.9	15.2	247	0	0	
46	11	12	286	0	0	
47	5.2	12.7	213	0	0	
48	9.9	13.4	161	0	0	
49	5.5	13	180	0	0	
50	7	14.2	240	0	0	

Test and Train Final Data

Ready

Fig 11: CSV file(Continued)



F2

	A	B	C	D	E	F
50	7	14.2	240	0	0	
51	6.9	11.5	275	0	0	
52	6.4	15.7	169	0	0	
53	10.7	14	199	0	0	
54	8.9	14.2	300	0	0	
55	5.8	13.1	270	0	0	
56	6.5	13	159	0	0	
57	4.9	13	191	0	0	
58	7.3	12.4	274	0	0	
59	9	10.2	200	0	0	
60	4.8	16	210	0	0	
61	10.1	14.9	242	0	0	
62	9.5	13.3	230	0	0	
63	4.9	11.5	167	0	0	
64	5.8	15	184	0	0	
65	9.7	12.7	201	0	0	
66	7.7	13	160	0	0	
67	6.1	14.3	300	0	0	
68	5.1	15	243	0	0	
69	8.4	11.9	188	0	0	
70	9.3	13	235	0	0	
71	7.8	12.1	272	0	0	
72	6.4	14	180	0	0	
73						
74						

Test and Train Final Data

Fig 12: CSV file(Continued)

First we used the ID3 with C4.5 extension as we not only have discreet data but continuous as well. Then we used Naïve Bayes and afterwards recurrent neural network on the dataset.

We learned about the algorithm and how to implement it on our data from the existing examples of scikit.org. We used Spyder to make the model, **Spyder** is an IDE which integrates NumPy, SciPy, Matplotlib and IPython, as well as other open source software. For doing Python code in Windows we need NumPy, SciPy and these open source software. First we tried doing this in the command prompt by installing ony NumPy and SciPy but it required installing other software

too. Afterwards we found the IDE Spyder, which included all of it so we had no further problems.

By running ID3 and C4.5 on the CSV file we got an accuracy of 88.8889%

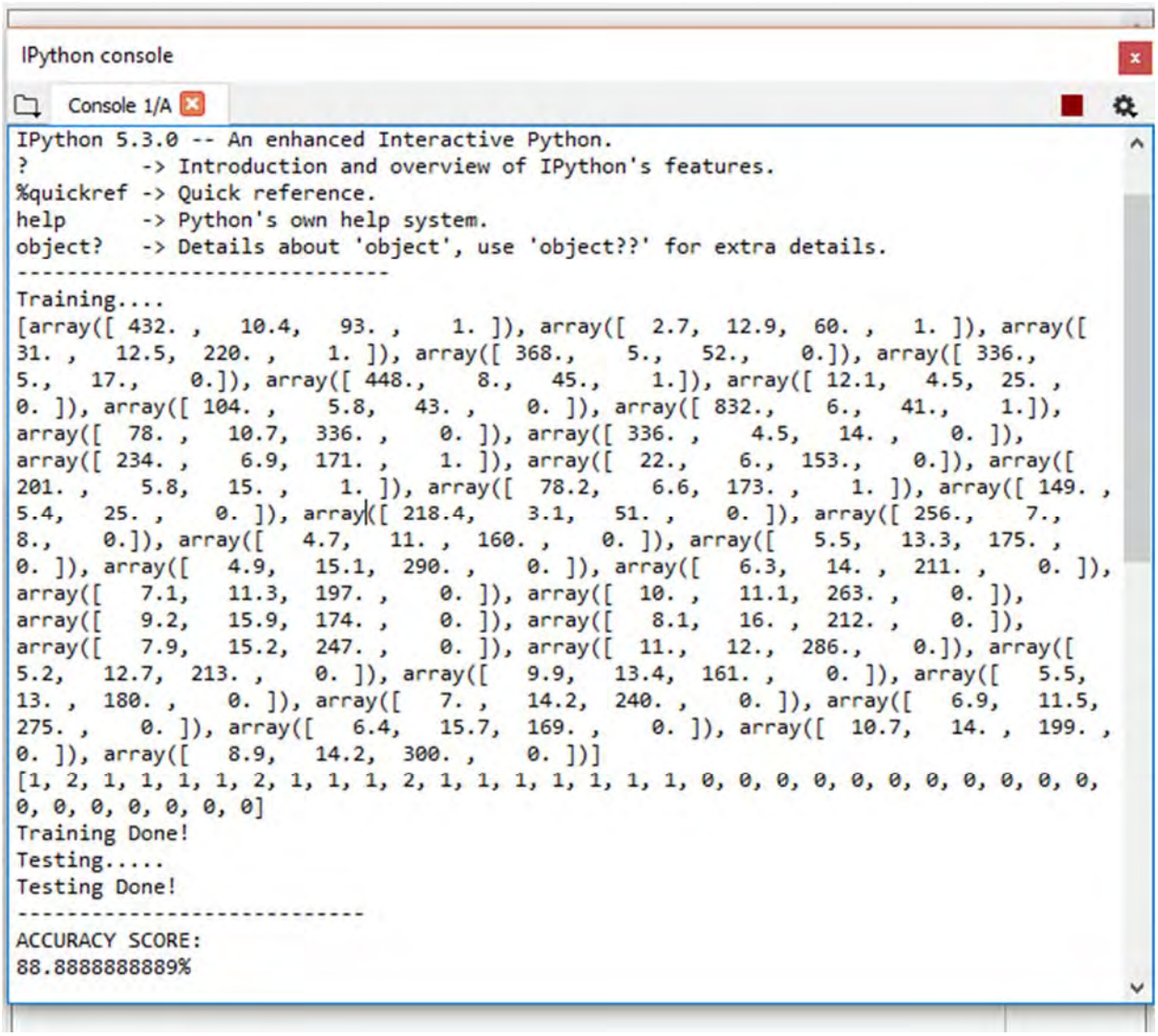


Fig 13: ID3 & C4.5 algorithm result

By running Naïve Bayes algorithm on the CSV file we got and accuracy of 77.7778%

Fig 14: Naïve Bayes algorithm result

By running Recurrent Neural Network on the CSV file we got an accuracy of 86.1112%

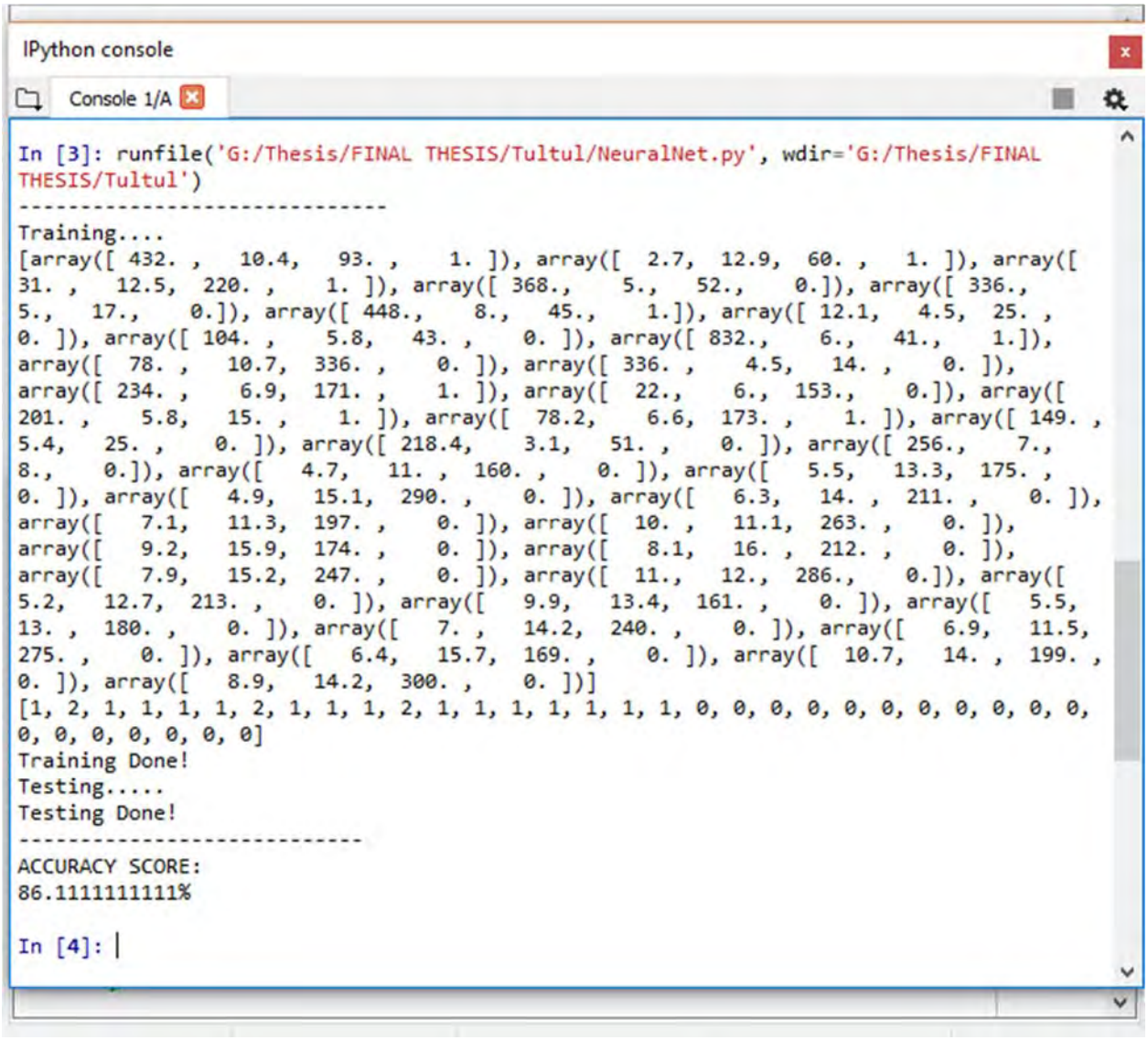


Fig 15: Recurrent Neural Network algorithm result

CONCLUSION

In many underdeveloped areas of our countries where there are no proper hospitals, and clinics are the only option for primary treatment, this model we have created will come in handy. In many rural clinics, doctors visit once weekly. As ALL is an acute disease, the earlier it is identified the better chances the patient has for getting better. The nurses and lab workers of a clinic can hence identify using this model after blood filming and other tests whether the person has ALL or not using this model. Based on the result they can send the patient away to a hospital for getting diagnosed and treated under a doctor.

We worked with only ALL so far and with very little data. In the future, the goal is to cover all types of leukemia and gather data from multiple hospitals throughout the country. If that works well, we can even use this model to identify other malignancies as well!

If this research work is able to save at least one child's life, it will be worthwhile.

REFERENCES

[1] Afshar, S., Abdolrahmani, F., Tanha, F.V., Seif, M.Z. & Taheri, K. (2011). Recognition and prediction of leukemia with Artificial Neural Network (ANN). *Medical Journal of Islamic Republic of Iran, 25*(1), 35-39. Retrieved from http://mjiri.iums.ac.ir/article-1-249-en.pdf

[2] Leukemia in Children. (2017, August 22). Retrieved from http://www.cancer.org/cancer/leukemiainchildren/

[3] Cancer in Children and Adolescents. (2017, August 22). Retrieved from http://www.cancer.gov/types/childhood-cancers/child-adolescent-cancers-fact-sheet

[4] Deep learning. (2017, August 21). In *Wikipedia, The Free Encyclopedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Deep_learning&oldid=796484501

[5] Brownlee, J. (2016). *Deep Learning with Python.* Melbourne, AU: (n.p.).

[6] Molla, M. R. (2009). *Concise Textbook of Pediatrics Revised Edition.* Dhaka, BD: Child Friendly Publications.

[7] DeepLearning.TV. (2015, December 8). *What is a Neural Network - Ep. 2 (Deep Learning SIMPLIFIED)* [Video file]. Retrieved from https://www.youtube.com/watch?v=P2HPcj8lRJE

[8] LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep Learning. *Nature, 521*(7553), 436-444. Retrieved from http://www.nature.com/nature/journal/v521/n7553/abs/nature14539.html?foxtrotcallback=true

[9] What is leukemia?. (2017, August 22). Retrieved from http://www.cancer.ca/en/cancer-information/cancer-type/leukemia/leukemia/?region=sk#ixzz4qJqQT1Jw

[10] scikit-learn Machine Learning in Python. (2017, August 22). Retrieved from http://scikit-learn.org