

PREDICTING PATIENT'S DISEASE BASED ON THEIR SYMPTOMS USING MACHINE LEARNING

By

Ahmed Nahian
14301120

Md. Jawad Bin Yusuf
18301296

Sajid Hasan Mahmud
19241032

A thesis submitted to the Department of Computer Science and Engineering in partial
fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
October 2020

© 2020. Brac University
All rights reserved.

Declaration

It is hereby declared that

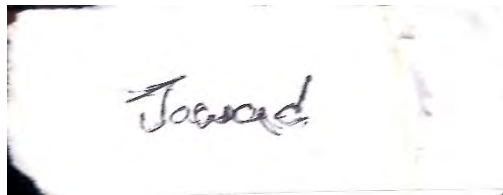
1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I/We have acknowledged all main sources of help.

Student's Full Name & Signature:



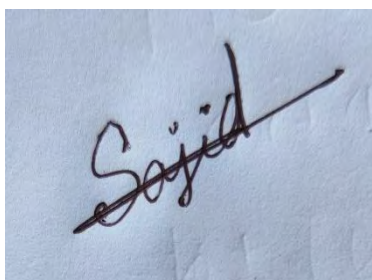
Ahmed Nahian

14301120



Md. Jawad Bin Yusuf

18301296



Sajid Hasan Mahmud

19241032

Approval

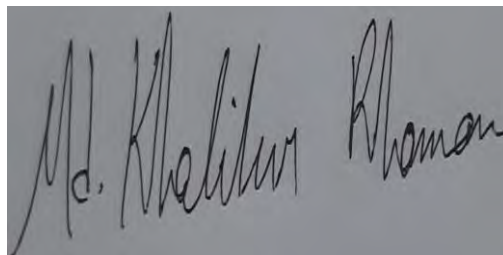
The thesis/project titled “PREDICTING PATIENT’S DISEASE BASED ON THEIR SYMPTOMS USING MACHINE LEARNING” submitted by

1. Ahmed Nahian (14301120)
2. Jawad Bin Yusuf (18301296)
3. Sajid Hasan Mahmud (19241032)

of Summer 2020 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering on October 09, 2020.

Examining Committee:

Supervisor:



Dr. Md. Khalilur Rahman

Associate Professor

Department of Computer Science and Engineering
BRAC University



Program Coordinator:

Dr. Md. Golam Rabiul Alam

Associate Professor

Department of Computer Science and Engineering
BRAC University

Prof. Mahbub Majumdar
Chairperson
Dept. of Computer Science & Engineering
BRAC University

Departmental Head:

DR. Mahbubul Alam Majumdar

Professor

Department of Computer Science and Engineering
BRAC University

Abstract

This paper analyzes how the patient can easily figure out the appropriate doctor based on the physical problem or the health issues which they could not figure out in previous because of lack of knowledge related to the disease. The methods we are using is machine learning which will figure out the machine readable details of the information that are provided by the patient. The patient will justify their problems by answering a list of questionnaire that helps them to figure out their problem details. The questionnaire will narrow down the problem and justify the appropriate disease or health issues that the patient is facing. Our result will show the details of the problem that the patient is facing and suggest them their preferred doctors. The program will also suggest the tests related to the issue. It will also figure out the preference of the patient. From a safety perspective this study will emphasize on the digital solution to the patient whom took hassle to figure out their preferred doctors.

Keywords: You may include up to six keywords or phrases. Keywords should be separated with semicolons.

Dedication

Dedicated to our beloved family members who were with us with their trust, love and support...

Acknowledgement

We would like to show our humble gratitude towards Dr. Khalilur Rahman sir, our thesis supervisor for assisting us with all the necessary facilities throughout our thesis journey. We would also like to thank Dr. Khalilur Rahman sir for his sincere and valuable guidance and encouragement which we were in need most. Undoubtedly, we will forever be indebted to

Dr. Khalilur Rahman sir for providing us the insights regarding the technical aspects of this work. Moreover, we would like to thank all the Faculty and Staff members of BRAC University for their guidance throughout our study period in BRAC University. Finally, we would like to thank our loving parents, siblings, friends for being our support and guide. They were always helping us directly or indirectly into making this project a success.

Table of Contents

Declaration	ii
Approval.....	iii
Abstract	v
Dedication.....	vi
Acknowledgement	vii
Table of Contents	viii
List of Tables	ix
List of Acronyms	x
Chapter 1 Getting Started	11
1.1 Introduction	11
1.2 Methods.....	13
1.3 Objectives.....	13
Chapter 2 Literature Review	14
2.1 Diseases.....	23
2.1.1 Stroke	23
2.1.2 Hypertension.....	30
Treatment and Recovery	31
2.1.3 Diabetes.....	32
Diagnosis.....	33
Chapter 3 Algorithm Discussion.....	36

3.1 Support Vector Machine (SVM)	37
Chapter 4 Prediction Model Using Neural Network.....	39
4.1 Importing Libraries	39
Chapter 5.....	40
Result	40
Chapter 6.....	43
Future Plan	43
References.....	43

List of Tables

Table 1 Patient's Information.....	23
------------------------------------	----

List of Acronyms

SVM	Support Vector Machine
	International Centre for Diarrhoea Disease Research,
ICDDR	Bangladesh
WHO	World Health Organization
PCP	Primary Care Physician
ADA	American Diabetes Association
ML	Machine Learning

Chapter 1

Getting Started

1.1 Introduction

This paper analyzes how the patient can easily figure out the appropriate doctor based on the physical problems or the health issues which they could not figure out in previous because of lack of knowledge related to the disease. The methods we are using is machine learning which will figure out the machine readable details of the information that are provided by the patient. The patient will justify their problems by answering a list of questionnaire that helps them to figure out their problem details. The questionnaire will narrow down the problem and justify the appropriate disease or health issues that the patient is facing. Our result will show the details of the problem that the patient is facing and suggest them their preferred doctors. The program will also suggest the test related to the issue. It will also figure out the preference of the patient for the doctors and also suggest them their preferable location for getting their treatment. It will also show the details about the public and private hospitals and the cost of their treatment. From a safety perspective this study will emphasizes on the digital solution to the patient whom took hassle to figure out their preferred doctors at budget cost on a preferable place. Medical Service is one of the crucial things needed for humanity. According to WHO, health is one of the fundamental rights of human being. In order to do that, people should get proper medical care. Here, we are going to discuss three of the most significant health issues that we usually find our country.

1.1.1 Stroke

This is such a disease that affects the veins in the brain. It is one of the major reasons of demise of and certainly a leading cause of paralysis as well. It happens when the vein that carries oxygen and nutrients is either blocked by a clot or bursts.

The time when this happens, the vein doesn't really get enough oxygen or supplements, and synapses begin to pass on. Stroke is a cerebrovascular infection. This implies it influences the veins that feed the mind oxygen. In the event that the mind doesn't get enough oxygen, harm may begin to happen. This is a health related crisis. Albeit numerous strokes are treatable, some can prompt handicap or demise.

1.1.2 Hypertension

It is such a condition when force of blood in artery of the heart is high and it may end up by causing diseases like heart disease. It is basically determined both by the rate of blood pumped by our heart as well as the rate of how it flows within the veins or arteries. A person may have this problem for several years and he may not get any symptoms whatsoever and it may be detected when a severe damage is done. So, it's truly necessary that we measure our blood pressure from time to time and keep it under control. Due to our weather, it can also be seen that our blood pressure remains high.

1.1.3 Diabetes

The full name of this disease is Diabetes Melitus and it occurs when there is a significant amount of glucose detected in our body. Here, sugar is carried by hormone insulin from blood to cells. For this, our body won't be able to make enough insulin that its need or unable to use the insulin

that its making. An expected 10 million individuals in Bangladesh have diabetes. A comparative report uncovers an all the more stunning reality, right around one of every ten grown-ups in Bangladesh was found to have diabetes. WHO expressed 83 % populace old enough gathering 25-65 never checks for diabetes. An alternate report says practically comparative thing. Around 17% of men and 23% of ladies were distinguished to have hindered fasting glucose or debilitated glucose resistance, on the whole named transitional hyperglycaemia. Bangladesh is ranked 57th due to this balanced death rate of 40.08 out of 100,000 persons.

1.2 Methods

Data mining uses different techniques to mine data and provides necessary decision making scenario by which those data come into use. In our thesis, we have used SVM to predict the disease out of symptoms. It is such a guided ML algorithm that uses by differentiating algorithms for two section classification subjects. Our goal will be to tune the way we use to get more accuracy and obviously better results.

1.3 Objectives

- ☐ Allocating doctors list to the patient to figure the suitable doctors for him/her
- ☐ Reducing hassle for the patient to figure out the problems
- ☐ Allocating doctors for the patient to their preferred areas
- ☐ Using machine learning to provide proper guideline to the program
- ☐ Reducing excess cost for the patient
- ☐ Making the whole process understandable towards the patient

Chapter 2

Literature Review

We are trying to help the people who cannot have instant doctor while they are suffering from some disease. It is being a major problem to our country nowadays. That's why we are trying to invent something which can help them to find a choice to have doctor instantly or at least they can get an idea what kind of symptoms their body is showing. According to our constitution, the health care service is one of the basic needs of the citizens. But, the medical service in our country is still not up to mark. Health care service is basically limited within major cities but not in the rural areas. In addition to the lack of basic medical facilities throughout the country, there is a major concern in Bangladesh regarding communicable diseases. Due to the geo location, we eventually face numerous amount of disease such as typhoid, cholera, malaria, dengue etc. This is a notable problem and should be considered in order to serve a better healthcare service. When a patient is found in severe condition it is either referred to the capital or outside of our country by which we can easily assume the medical service still needs to get a lot better. These medical referrals can be expensive to later on can be a burden to a person as we don't have medical insurances. According to ICDDR, 59% people died in non-communicable disease (2018) and according to (bdnews24.com), 32% people died in communicable diseases. It is major issues for Bangladesh. People cannot get instant treatment or they do not capable of paying in private hospital. So, they stuck with decisions and some, who can get instant treatment but they do not know which physician will be better for that disease. People suffer everyday from different diseases. So they always want something which can create an easy way to prevent the diseases. In our country, people more or less suffer from various diseases like Stroke, Hypertension, Diabetes Mellitus, Rheumatoid Arthritis, Blood Cancer (Leukemia, Lymphoma), Anemia and so on. Basically, most of the people of our country don't really care that much about their health, they

only go to the physician after reaching a severe condition of some diseases just because most of our people lives below the poverty line. For that reason, it is hard for them to maintain a good or regular health checkup. According to World Data Bank and Financial Express Bangladesh has a doctor for 1847 people as per 2018 which is significantly low.

Here is some patients overview that were gathered from patients medical report by visiting them in the hospital assisted by duty doctors. (All of data are given here uses a different name than the original one to protect the privacy of these patients).

Patient's Name	Age	Name of Disease	Symptoms	Treatment
1. Faruk	52	Ischemic stroke	Left sided weakness for 5 hours Unable to speak for same duration belching	NG tube feeding 200mL 2 hourly Inf.N/S 1500 mL iv @15 d/min Inj.ceftriaxone 2g 1 vial iv 12hourly Inj omeprazole 40mg 1vial iv 12 hourly Tab.Aspirin

				75mg 0+1+0 Tab Atorvastatin 10mg 0+0+1 Tab Bislol 2.5mg 1+0+0
2.Umar	55	Ischemic stroke	Right sided weakness for 1 day	Inf.N/S 1L iv @10d/min Inj.ceftriaxone 1gm 1 vial iv 12 hourly Inj omeprazole 1 vial iv 12hourly Injroxadex 1 amp iv 8 hourly Tab Risdon 4mg ½+1/2+1 Tab Ecosprin 75mg 0+1+0 Tab ATV 10mg 0+0+1 SypLivwel 200 mL 2 tsf 12 hourly Sypneurolep 100mL 1 tsf 8 hourly

				Tab Arotril 2mg ½+0+1
--	--	--	--	--------------------------

4. A.Hannan	95	Ischemic stroke	Left sided weakness from yesterday Unconsciousness from yesterday night	NG feeding 200mL 2hourly Inj ceftriaxone 1gm 1 vial iv 12 hourly Inj Omeprazole 1 vial iv 12 hourly InjRoxadex 1 amp iv 8 hourly Injansulin 100IU 10+0+6 s/c ½ hour before meal Tab Ecosprin 75mg 0+1+0 Tab ATV 10mg 0+0+1 Tab Amloten 50mg 1+0+0
5. Tarik	55	Ischemic stroke	Right sided weakness for 1 day	Inf N/S 1L iv @10d/min Inj Ceftriaxone 1gm 1 vial iv 12 hourly Inj Omeprazole

				40mg 1 vial iv 12 hourly Inj Roxadex 1 amp iv 8 hourly Tab clindipin plus 0+0+1 Tab Ecosprin 75mg 0+1+0 Tab ATV 10mg 0+0+1
6. A Latif	56	Hemorrhagic stroke	Right sided weakness of the body Restlessness for 9 hours	Inj. Ceftriaxone 1gm 1 vial iv 12 hourly Inj Omeprazole 40mg

				1 vial iv 12 hourly InjRoxadex 1 amp iv 8 hourly InjHalopid 1 amp i/m stat Injperkinil 1 amp iv stat Tab calnor 5/20 0+0+1 Change posture 2 hourly
5. Masum	60	Ischemic stroke	Left sided weakness of the body for 1 day Slurring of speech for 1 day	Inf N/S 1L iv@30d/min Inj Ceftriaxone 1gm 1 vial iv 12 hourly Inj Omeprazole 1 vial iv 12 hourly InjRoxadex 1 amp iv 8 hourly Tab Ecosprin 75mg 0+1+0 Tab ATV 10mg 0+0+1
7.Nurul	70	Ischemic stroke	Left sided weakness of the body for 1 day	Inf.N/S 1L iv @10d/min Inj.ceftriaxone

				1gm 1 vial iv 12 hourly Inj omeprazole 1 vial iv 12hourly Injroxadex 1 amp iv 8 hourly Tab Risdon 4mg ½+1/2+1 Tab Ecosprin 75mg 0+1+0 Tab ATV 10mg 0+0+1 SypLivwel 200 mL 2 tsf 12 hourly
--	--	--	--	---

				Sypneurolep 100mL 1 tsf 8 hourly Tab Arotril 2mg ½+0+1
8. Md Saidur Rahman	60	Ischemic Stroke	Left sided weakness of the body Unconsciousness for 2 days	NG tube feeding 200mL 2 hourly Inf.N/S 1500 mL iv @15 d/min Inj.ceftriaxone 2g 1 vial iv 12hourly Inj omeprazole 40mg 1vial iv 12 hourly InjRoxadex 1 amp iv 8 hourly Tab.Aspirin 75mg 0+1+0 Tab Atorvastatin 10mg 0+0+1 Tab Osartil 50mg 0+0+1 Tab cerevas 5mg 1+1+1 SypLivwel

				1 tsf 3 times daily
--	--	--	--	------------------------

Table 1Patient's Information

2.1 Diseases

As we have chosen three of the most common diseases in our country, we are going to discuss about how the physician diagnose the disease alongside their symptoms. These are the things we have got from registered and prominent physicians from public hospitals.

2.1.1 Stroke

- **Ischemic stroke**

It happens when the brain don't get enough oxygen due to the narrowing down of the veins or by a blood clot. This causes ischemia, or a seriously diminished blood stream, which harms synapses. Blood clusters frequently cause ischemic stroke. Clusters can shape in the cerebrum veins and other veins in the body. The circulatory system conveys these into smaller veins in the cerebrum. Greasy plaque stores inside the corridors can likewise cause clusters that bring about ischemia.

- **Hemorrhagic stroke**

Broken or burst corridors in the cerebrum can offer ascent to hemorrhagic strokes. Spilled blood squeezes synapses and harms them. It likewise decreases the blood gracefully that can arrive at the mind tissue after the discharge. Veins can get cracked an spell blood in the brain. This may likewise send blood into the space between the brain and the skull. Having

hypertension, encountering physical injury, taking blood-diminishing drugs, and having an aneurysm would all be able to make a vein break or burst.

Intracerebral drain is the most well-known sort of hemorrhagic stroke. This happens when cerebrum tissue floods with blood after a course blasts. Subarachnoid drain is another sort of hemorrhagic stroke. These are more uncommon. In a subarachnoid discharge, draining happens in the zone between the mind and the slender tissues that spread it.

Symptoms

The indications of a stroke regularly show up abruptly. A portion of the principle indications include:

- Disarray, including trouble talking and getting discourse
- A cerebral pain, perhaps with modified cognizance or spewing
- Deadness or a powerlessness to move portions of the body like face deviation, numbness of arm or legs, or even it can affect a finger.
- Blurriness in one or both eyes.

Stroke can prompt long haul medical issues. Contingent upon the speed of the determination and treatment, an individual can encounter brief or perpetual incapacities after a stroke.

A few people may likewise act:

- Bladder or inside control issues
- Despondency
- Loss of body balance
- Trouble expressing their feelings

- Side effects change and may go in seriousness.

Learning the abbreviation "Quick" is a decent method to recall the manifestations of stroke.

This can enable an individual to look for brief treatment. Quick represents:

- Face hanging: It like the deviation of mouth.
- Hand shortcoming: If the individual attempts to use both their arms they should be able to do that.
- Discourse trouble: If the individual attempts to rehash a basic expression, it shouldn't be discourse absurd or bizarre.
- Time to act: If the above mentioned things happen we should contact the emergency medical service right away.

The result relies upon how rapidly somebody gets treatment. Brief consideration likewise implies that they would be less inclined to encounter perpetual cerebrum harm or demise.

Diagnosis

Stroke starts quickly. For the best result, an individual ought to get treatment at an emergency clinic inside 3 hours of their manifestations first showing up.

There are a few distinctive analytic tests a specialist can use to decide the sort of stroke. These include:

- Physical assessment: A specialist will get some information about the individual's side effects and clinical history. They will check muscle quality, reflexes, sensation, vision, and coordination.

- Blood tests: A physician may give blood tests to check if there any contamination or the blood is thickened by which there might be a clot that results in this kind of diseases.
- CT Scan: A certain X-Ray can be performed to see if there is any ischemic or hemorrhagic thing happened.
- MRI Scan: These utilization radio waves and magnets to make a picture of the mind, which a specialist can use to identify harmed cerebrum tissue.
- Cerebral angiogram: A physician may infuse a color into the brain's veins to make them noticeable under X-beam or MRI. This gives a definite perspective on the veins in the mind and neck.
- Echocardiogram: This makes a nitty gritty picture of the heart, which specialists can use to check for any wellsprings of clusters that could have gone to the cerebrum.

It is just conceivable to affirm the sort of stroke utilizing a mind check in a medical clinic condition.

Treatment and Recovery

Ischemic stroke

To treat stroke, blood flow must be restored as it was. For that a physician may suggest these necessary things.

- **Emergency IV medication.** Drugs or necessary medicines can certainly break up a clot inside a brain that restrains the blood flow and it should be done as soon as possible.

An IV injection of tissue plasminogen activator (tPA) — also called alteplase (Activase) is the most used treatment for ischemic stroke now a days . This type of injections can be given within the very few hours like 3.5 – 4 to save the patients.

This medicine is really helpful in order to restore blood flow by dissolving the clot that stops the blood flow within the brain by which the stroke has occurred. A physician first need to identify whether this types of tPA should be given or not by several tests we said earlier.

- **Emergency endovascular procedures.** Physicians performs this procedure ASAP by giving the medication directly in the veins.
 - o **Medications delivered directly to the brain.** Physicians insert a catheter (a long thin tube) into the brain unlike a tPA that directly inserted through veins. This type of medication is less practiced now a days.
 - o **Removing the clot with a stent retriever.** Basically, a neurosurgeon performs this type of services to their patients to remove a bigger blood clot from the brain that cannot be dissolved only by tPA. So, they uses a stent retriever to remove the clot from the brain.

The time period is expanding to treat someone for such activities due to the recent technology (image tests) developments. Now, physicians get more time to treat their patients with procedures like endovascular tests.

Other procedures

To reduce the risk of strokes in near future a doctor can suggest these type of procedures to open up a vein that is narrowed by any plaque.

- **Carotid endarterectomy.** These arteries are the main two arteries in our neck that supplies oxygen in the brain. This procedure evades the plaque that stopped a carotid artery. This can results in heart diseases any other side effects.
- **Angioplasty and stents.** Most of the patients basically treated with these type of stents that enlarge the affected vein that is narrowed down for any reason like blood clot. Here, a surgeon basically uses a balloon to inflame the artery and plant a stent to stable the artery to maintain the blood flow.

Hemorrhagic stroke

Treatment of Hemorrhagic stroke means to solve the case and restore the normal blood flow of patients and remove the blood that caused by the erupted veins in the brain. The treatment are as follows”

- **Emergency measures.** A physician more or less give medicines to thin the blood so that the blood or the clot in the brain can be removed and blood flow can be restored to counter react this thing they also suggest some drugs so that the blood don't really become too much thinner as well as blood transfusions. Physician may also give medications to lower the blood pressure (intracranial pressure) inside the brain, reducing the blood pressure will prevent spasms in blood veins and will reduce the chance of any forthcoming seizure.

- **Surgery.** If a physician found the area where the stroke occurred is large and flowed with blood, he may suggest or perform surgery to remove the blood from the brain. This includes repairing blood veins that is erupted and caused by the hemorrhagic stroke. Usually, a Neurosurgeon suggest these types of treatment.
- **Surgery Clamps.** A Neurosurgeon may place tiny little clamps to close the damaged veins and protect it from future seizure. They usually place it when the hemorrhagic stroke occurs.
- **Coiling (endovascular embolization).** With the help of a catheter a surgeon uses heart artery of the patients that guided to his brain, and he places small detachable coils that prevents the bleeding in the brain. This thing prevents the blood clot.
- **Surgical AVM removal.** Neurosurgeons may perform AVM removal procedure if that is places within the area that can be accessible. If not or it is within the deep part of the brain they don't really follow this thing cause the surgery may affect brain functions.
- **Stereotactic radiosurgery.** It is such a process where ray beams are used to repair any damaged blood vessel due to the hemorrhagic stroke. It may not be performed everywhere but where the up to the mark technology is used to treat patients.

Recovery from Stroke and back to the Normal Life

After the proper treatment, patient will be thoroughly checked in a daily basis. From that point onward, rehabilitation centers around helping him recoup however much capacity as could reasonably be expected and re-visitation of free living. The effect of your stroke relies upon the region of the mind in question and the measure of tissue harmed.

Following the stroke, patient's body part or either side of the body may be affected and seen numbness due to the stroke. Patient's PCP will ensure the most thorough treatment program he can deal with depending on his age and endurance, in general wellbeing and level of incapacity from stroke. Primary care physician may contemplate his way of life, interests and needs, and the accessibility of relatives or different parental figures.

Recovery may start before patient's leave the emergency clinic. After release, he may proceed with his program in a restoration unit of a similar emergency clinic.

2.1.2 Hypertension

It is such a condition where the blood flow of the heart is significantly high and can cause any damage to an organ uncertainly. BP means the blood pressure of one's body that flows within the brain and the rate of flow is usually measured. The pulse is the thing that rate of blood pumped by the heart and it flows within the arteries. Patients may have this disease for quite a long time and it might be happening that they find no symptoms. Any sort of heart diseases can be seen due to this hypertension and more or less heart attacks or cardiac failure are most commonly seen results for such issue. If it is not kept under controlled then it may results into severe heart illness. This hypertension always affect patients life sooner or later. Thanks to the medical science that there are numerous ways to keep this thing under control with proper guidance of a registered physician.

We can distinguish Hypertension into two types.

- **Primary (basic) hypertension**

Most grown up people suffer from hypertension due to several reasons. It basically form over numerous years.

- **Secondary hypertension**

There are very few people who are seen to have this type of hypertension where they fall into this every now and then and it can be controlled by proper medications.

- Kidney diseases
- Obstructive sleep apnea
- Thyroid problems
- Adrenal tumors

Symptoms

Most of the people or patient shows no signs of hypertension till they reach in a severe stage.

There might be a couple of patients who deal with nose bleeding, hand sweating, dizziness and other issues.

Treatment and Recovery

Lifestyle change can save a person from hypertension. PCP may suggest a patients to undergo the following advices.

- Eating solid food that contains less salts
- Physical exercises
- Keep the normal BMI in case of overweight or underweight
- Consuming liquor is forbidden

Still the changing lifestyle may not effect that much when the physician give the patients the following instructions.

- Depending on how sound he is, his pulse treatment relies.
- Pulse should be kept between the margin 130/80 mm Hg if:
- Patient's should be 65 or more years age.

Medication

Diuretics, more or less called water pills, are prescribed to dispose the sodium and water through kidneys and lessening the blood volume eventually. This is the most used medicine in home or abroad to treat such kind of patients. Thiazide diuretics basically contains chlorthalidone, hydrochlorothiazide (Microzide). When the patient is not taking any kind of medicine or taking but the condition is still not improving he should consult with PCP to get more this type of drug or something else.

2.1.3 Diabetes

It is such a disease that referred to a group of illness where the body deals with glucose which is basically the main source of body energy. It can be also called the primary source of brain's energy supply. Nonetheless, what types of diabetes a patient's have it increases the level of glucose in his blood and can cause serious health issues. Diabetes include type 1 and 2. Pre diabetes stage is called when they find sugar in one's blood but not as significant as that can be called as diabetes. There's a case of pregnancy where its natural to find mothers to have pre diabetes.

Symptoms

- Become more thirsty every now and then
- Increasing of urination
- Become hungry frequently
- Uncertain weight loss
- Ketones seen in urine
- Overweight patients
- Slow healing process than natural
- Gum of skin infections

Diagnosis

There might be seen diabetes in a patient out of nowhere which is basically the type 1 diabetes and the American Diabetes Association (ADA) has declared some rules to determine it.

- A person over 25 years suffering from anxiety or hypertension or above mentioned symptoms may have diabetes and can't be neglected.
- Aging more than 45 years a person have diabetes and should be screened the glucose level frequently.

- A lady with gestation diabetes should be kept under observation and checked the diabetes time to time.
- Any subject diagnosed with pre-diabetes should be checked every year.

Suggested tests:

- **Glycated hemoglobin (A1C) test:** It's a blood test that shows normal glucose level of a person. It measure the glucose level of red platelet of the blood that contains hemoglobin.
- **Random blood sugar test:** A blood test will be taken before and after the patients have eaten to see whether their glucose level have increased or not. If yes then it should not be increased more than 11.1 gm/L.
- **Fasting blood sugar test:** Physicians often suggest the patients to take a blood test after fasting to determine whether he's in pre-diabetes stage or anything serious.
- **Oral glucose tolerance test:** On this type of tests, patients are asked to consume any sweet food and their glucose level is measured several times for two hours while fasting.

Treatment and Medication

Medicines for a wide range of diabetes

- **Good Diet Plan:** A patient should maintain a regular routine while eating. There's no fixed diet plan for diabetes. Patient should maintain his blood sugar under control for which the physicians suggest to eat natural foods, less oily food, no sweet food whatsoever.

However, a dietician can suggest a proper diet plan according to the patient's lifestyle and can make his breakfast, lunch dinner routine. The patients should eat in a fixed time in a day to keep his glucose level in control.

- **Physical Activity:** Not only the patients but also we need a standard form of exercise to keep the body and mind fresh. The patient should maintain this. Too much of exercise also can hamper a person's health as well when he has this sort of diseases.
-
- **Monitoring Blood Sugar:** For type 1 diabetes a person should check his glucose level frequently or up to 4 times a day if he's taking insulin. But for the case of type 2 where the patient don't take insulin can check less significantly.

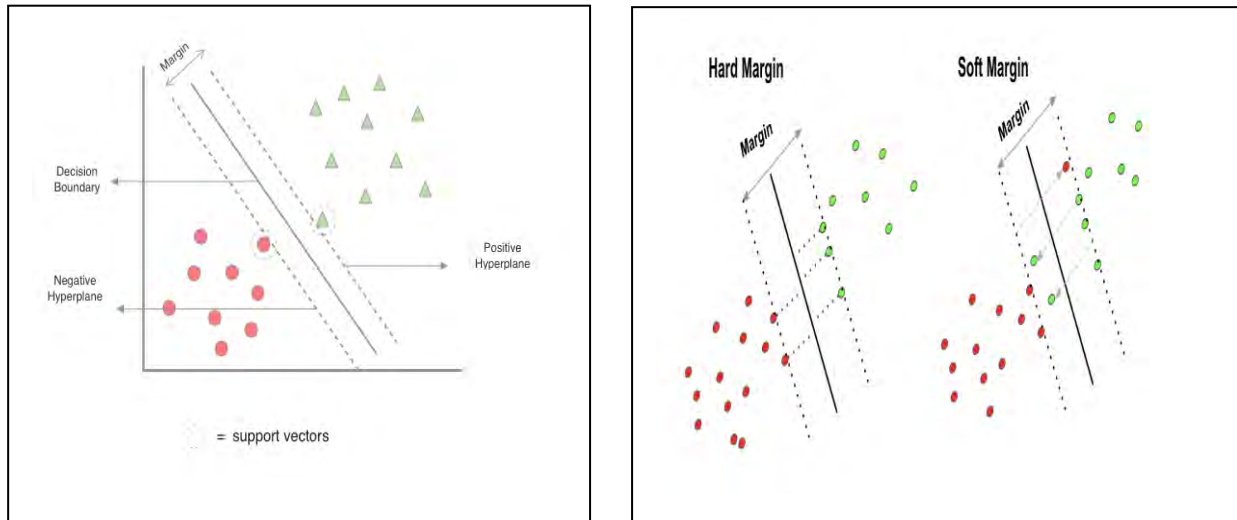
Patients who get insulin to keep their glucose level normal should check the blood sugar with the help of a diabetes test machine that is available everywhere now a days. Physicians will suggest A1C test for up to several months to check the blood sugar level. Increased glucose in this test may end up to change the diet plan as well the prescribed medications that the patient is having.

Chapter 3

Algorithm Discussion

In ML, we may use different classifiers to help us predict the standard result. In this project we are targeting to achieve a good score where the machine will predict the disease based on the symptoms of patient that stroke, hypertension or diabetes running SVM to our data set. Out of all the models we think this will help us determine a better and obviously dependable result with our dataset. The primary cause of using SVM that it can joint with kernel trick which will be very efficient in our thesis. That is why by undergoing and fine tuning kernels we will achieve appropriate feature spaces that will help to determine the result we need. Here, the linear classification is also available to classify data created by non-linear classification. Furthermore, the optimization problem solved in learning is a convex one, i.e. if it has a solution, i.e. it is feasible, the solutions global. At the time SVM became popular, they were introduced as neural networks with this additional nice property. When using a penalization approach, the problem is always feasible. This algorithm is used because it performs optimally in cases where there is a distinct margin of separation among classes. Also, it has been shown to perform well in highly dimensional data particularly when the number of dimensions exceeds the sample size. It is also memory efficient. Another important point is to try linear SVM. Text data is high dimensional, with linear SVM all that we do is to find a linear hyperplane that separates your classes. Linear SVM can have a short training time as well. In general, Naive bayes, Linear SVM, SVM work well with text data. Other classifiers may work well too as other answers point out; it always depends upon our data.

3.1 Support Vector Machine (SVM)



Prior to our start with SVM we should be acquainted with linear regression and logistic regression. We should at least have some idea before moving on with SVM. It is such an essential algorithm that every LM user or data scientist should have a knowledge on. Particularly, in this scenario vector machine is used to have better accuracy with less power consumption. It can be used both in regression and classification aspects. In classification objects it is commonly used. Let us discuss a bit more about what support vector machine is. Getting a hyper plane in N-dimension space which specifically defines the data points is the primary goal of the SVM. There is significant amount of hyper planes that can be chosen to differentiate between two aspects of data points. As per the requirement or to get a better accuracy we need to find out maximum distance between two datasets of both aspects. We will get better accuracy by getting maximized marginal distance. To define the data points hyper planes are known as decision boundaries. Data points falling on either aspect of the hyper plane will be attributed to completely different categories. Number of used Hyper planes basically depends on the number of features we are going to use in our project. In the case of using 2 features the hyper plane will just be a line. This scenario changes if we use 3

features and it will be two dimensional. Now, we can easily assume that it is difficult to understand if we use more than 3 features. We should always remember that support vectors are data points that are close to the hyper plane and influence the position and orientation of the hyper plane. The data points which is basically the SVM is close to the hyper plane and get affected by the position. Margin of classifier is increased by using the SVM.

Position of hyper plane will be changes if we delete the vector machine. We can build the SVM with the assistance of these data points. Now, we are going to enlighten a bit more about our purpose. Using sigmoid function our output will fall within the range (0-1) we will start. We will label it as 1 if the value is greater than 0.5 which is basically our threshold or else it's going to be a 0. Here we identify an output with one class if that is greater than 1 and another if that is below 1. Since, the value is $[-1, 1]$ and the values should stay within these range basically which is the threshold. Then, here comes the cost function.

To start, in the SVM algorithm,

```
save_model = 'SVM_classifier.sav'  
pickle.dump(model, open(save_model, 'wb'))
```

Our target will be to greater the margin of the hyper plane. Operation loss is the hinge when we try to maximize the margin. If predicted and original value becomes same the cost is 0. Else we try to calculate the loss. Then we add regularization parameter to operate the value. Regulation parameter basically balances the marginal maximization and loss. When we get the loss function, using partial derivatives from the weights we try to find the gradients. Then we can update the weights with the help of acquitted gradients. If there is found no misclassification, it can be said that the model is correctly finding the class of the data points, then all we need to do is to update the gradient using regularization parameter. If there's a case where we find misclassification as for, the model gives wrong prediction from the data points, we will insert the loss along with the regularization parameter to get gradient update. There will only two features that we will use which are Sepal length and Petal length.

There is another advantage of using it. Thanks to the SVM formula and its implementation. There is Sklearn library is available to use and simply decision to connect functions to implement this model. The thousands of line of code comes shorter with the use of this. To recapitulate, we can say that SVM is a efficient and largely used algorithm. We can use it in a lot of aspects and get viable results in ML world and get optimized performances.

Chapter 4

Prediction Model Using Neural Network

Neural network is best option for predicting and accurate percentage. And here, in our topic, is all based on predicting from symptoms and answering all diseases according to their symptoms. Linear regression is also used for these types of predicting but with neural networks, hidden layers makes it more accurate and less overfitting. It uses kernel and c parameter also.

4.1 Importing Libraries

We have used pandas to load the data as a data frame. We have used sklearn to load a simple dataset. We have used model_selection as sub library, and used train_test_split so I can split training and test the data. Os used to import the files from anywhere. Matplotlib used for graphs . We used metrics with classification for the quality of prediction from the disease in class which defines the more accurate output in code. We use the pickle for saving and loading the binary in the model. It is used perfectly.

```
In [1]: import numpy as np
        from numpy import array
        import pandas as pd
        import os
        import matplotlib.pyplot as plt
        import pickle #for saving and loading the model
        from sklearn.model_selection import train_test_split
        #from sklearn.preprocessing import OneHotEncoder
        from sklearn import svm
        from sklearn.metrics import classification_report
```

In this model, we are using useful libraries to load the data and convert the data frame to pandas to load the excel file.

```
df = pd.read_excel("Dataset Final(1).xlsx", sheet_name='All together ') # can also index sheet by name or fetch all sheets
df.head()
```

4.2 Kernel Discussion

After fitting the model, we input kernel and c parameter to more accurately predict the diseases. In neural network, kernel sigmoid is used and the equation is

$$k(x, y) = \tanh(\alpha x^T y + c)$$

And for c, it is a constant which is clarified mistakes with same classification.

```
In [10]: #using sigmoid kernel and gamma calculation automatically
        model = svm.SVC(kernel = 'sigmoid', gamma = 'scale', C = 2.0)

In [11]: model.fit(X_train, np.ravel(y_train, order='C')) #using ravel to convert to 1d array
Out[11]: SVC(C=2.0, kernel='sigmoid')
```

SVC(Support Vector Classification) is used to define the kernel and make it useful properly.

Chapter 5

Result

After running kernel and all, we predict X and Y which are symptoms and binary numbers, we have gotten the output of Diseases.

```
In [12]: y_predict = model.predict(X_test)
print(y_predict)

['Stroke' 'Diabetes' 'Hypertension' 'Diabetes' 'Stroke' 'Hypertension'
'Stroke' 'Stroke' 'Diabetes' 'Hypertension' 'Hypertension' 'Diabetes'
'Stroke' 'Hypertension' 'Diabetes' 'Diabetes' 'Stroke' 'Hypertension'
'Diabetes' 'Diabetes' 'Stroke' 'Stroke' 'Hypertension' 'Hypertension'
'Hypertension' 'Hypertension' 'Stroke' 'Hypertension' 'Hypertension'
'Diabetes' 'Stroke' 'Hypertension' 'Hypertension' 'Diabetes' 'Diabetes'
'Stroke' 'Diabetes' 'Hypertension' 'Stroke' 'Diabetes' 'Stroke'
'Diabetes' 'Hypertension' 'Hypertension' 'Stroke' 'Diabetes'
'Hypertension' 'Hypertension' 'Stroke' 'Hypertension' 'Diabetes'
'Diabetes' 'Diabetes' 'Diabetes' 'Stroke' 'Diabetes' 'Diabetes'
'Diabetes' 'Diabetes' 'Diabetes' 'Diabetes' 'Hypertension' 'Stroke'
'Stroke' 'Hypertension' 'Diabetes' 'Hypertension' 'Diabetes' 'Stroke'
'Diabetes' 'Stroke' 'Hypertension' 'Stroke' 'Stroke' 'Hypertension'
'Stroke' 'Stroke' 'Diabetes' 'Hypertension' 'Stroke' 'Stroke' 'Diabetes'
'Stroke' 'Hypertension' 'Stroke' 'Diabetes' 'Stroke' 'Stroke' 'Stroke'
'Hypertension' 'Diabetes' 'Hypertension' 'Stroke' 'Diabetes' 'Diabetes'
'Diabetes' 'Diabetes' 'Hypertension' 'Diabetes' 'Hypertension' 'Diabetes'
'Diabetes' 'Diabetes' 'Diabetes' 'Diabetes' 'Diabetes' 'Stroke'
'Diabetes' 'Hypertension' 'Hypertension' 'Stroke' 'Diabetes' 'Diabetes'
'Diabetes' 'Hypertension' 'Diabetes' 'Stroke' 'Stroke' 'Diabetes'
'Diabetes' 'Hypertension' 'Hypertension' 'Stroke' 'Stroke' 'Hypertension'
'Stroke' 'Diabetes' 'Stroke' 'Hypertension' 'Diabetes' 'Stroke'
'Hypertension' 'Hypertension' 'Hypertension' 'Diabetes' 'Diabetes']
```

And we can easily divide it into individual diseases to change in X, and finally the percentage of fitting the result is

```
In [13]: model_score = model.score(X_test, y_test)
print(model_score)

0.8205980066445183
```

And within the model, previously which we did classification report on library for the model and it divides into 4 chapters and it gave us a number which realizes to be fitting I guess.

```
In [15]: model_report = classification_report(y_test, y_predict)
print(model_report)
```

	precision	recall	f1-score	support
Diabetes	0.72	0.80	0.75	104
Hypertension	1.00	0.89	0.94	96
Stroke	0.79	0.78	0.79	101
accuracy			0.82	301
macro avg	0.84	0.82	0.83	301
weighted avg	0.83	0.82	0.82	301

And we input it into SVM model as it is and it gave us final output of prediction. And it is fitted into model.

```
In [16]: loaded_model = pickle.load(open(save_model, 'rb'))
result = loaded_model.score(X_test, y_test)
print(result)
```

```
0.8205980066445183
```

Chapter 6

Future Plan

With this model and plan, we can establish a web application which can be accessed in online and anywhere in the world, they can use it for emergency asking for a doctor and suggested list which they need to do instantly. And we need support to have this going in manually and to make it more efficient, we make it easier to people to use it cheap.

References

- [1]“Hypertension.”WorldHealthOrganization,WorldHealthOrganization,www.who.int/news-room/fact-sheets/detail/hypertension.
- [2] Stern MP, Patterson JK, Haffner SM, Hazuda HP, Mitchell BD. Lack of Awareness and Treatment of Hyperlipidemia in Type II Diabetes in a Community Survey. JAMA. 1989;262(3):360–364. doi:10.1001/jama.1989.03430030048032
- [3] Mohiuddin, AK. “Diabetes Fact: Bangladesh Perspective.” International Journal of Diabetes Research, www.ghrnet.org/index.php/ijdr/article/view/2457/2836.
- [4] Gandhi, Rohith. “Support Vector Machine - Introduction to Machine Learning Algorithms.” Medium, Towards Data Science, 5 July 2018, towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47.
- [5] The Financial Express. “Bangladesh Has One Doctor for Every 1847 People.” The Financial Express, thefinancialexpress.com.bd/health/bangladesh-has-one-doctor-for-every-1847-people-1519053209.
- [6]“Physicians(per1,000People)Bangladesh.”Data,data.worldbank.org/indicator/SH.MED.PHYS.ZS?locations=BD.
- [7]“Stroke: Causes, Symptoms, Diagnosis, and Treatment.” Medical News Today,

MediLexicon International, www.medicalnewstoday.com/articles/7624.

[8] "https://towardsdatascience.com/train-test-split-and-cross-validation-in-python-80b61beca4b6"