**Real world case study:**

**Design a Heart Disease Detection through an AI based System**

|  |
| --- |
| Any disease diagnosis system that is developed using machine learning or |
| artificial intelligence needs to have the same capability as that of a human |
| doctor. This is a narrow application of technology and is also a quick win. |

A quote from Dr. Devi Shetty, the Chairman of Narayana Health in Bangalore, from the conference at XIME Bangalore:

*“Smart Software is smarter than the doctors. There are only* ***6,000 diseases,*** *out of which there are*

*1,000 that are common. There are* ***about 50,000 signs and symptoms of these diseases****. There are* ***100,000 types of lab reports****. Matching these to diagnose a disease can be done by* ***any software very easily, which takes a doctor year to practice.****”*

|  |  |
| --- | --- |
| In addition to what Dr. Devi Shetty said, any robotics software will need to | |
| understand how doctors work. It could also involve studying the surgeon’s | |
| logbook, etc., in order to gain such expertise | . |

Key reasons why we need disease diagnosis to be done by machines are:

1. **Every year 195,000 patients in the US die of medical diagnostic error**

1. **Inability of a human being to process huge amounts of information, analyse it, and apply it to a particular disease diagnosis**
2. **Cloning of an expert robot is cheap and fast compared to a doctor, where cloning of expert human doctors is forbidden by law in many countries**
3. **Accuracy in diagnosis is critical, as an incorrect diagnosis increases the cost of healthcare for the patient.**
4. **Using machine learning, we can track, measure, optimize, learn, and improve based on feedback on the accuracy of diagnosis by a robot or a machine.**
5. **To increase the value of healthcare to its patients, it is an absolute must to decrease the cost of diagnosis given to a patient.**
6. **Value = Patient health outcome per dollar spent on the treatment.**

If the patient If the patient health outcome is bad due to bad diagnosis, then the value of healthcare goes down. If the patient health outcome is good, then the value of healthcare provided goes up.

So it is extremely important that we use machine learning to increase the value of healthcare for the patient.

**Context:**

In the not so distant future in the year 2025, one fine morning an old lady receives an alert on her personal home management device that she is going to develop cancer in the near future. This report has been sent by her robot doctor, after her visit last week for a check-up. She is mildly shocked to hear such news. She then decides to get a second opinion from a human doctor. The human doctors are very few in numbers now in her city and are more expensive than the robot doctors. So she decides to visit the human doctor nearest to her home. She visits the doctor and shows him her report, which was sent to her by the robot doctor this morning. The human doctor carefully looks at the report and finds that the robot had mentioned a clinical study that was done in the year 2019 where it was proven that people with a sleeping disorder lasting more than 3 weeks in a row had a 90 percent chance of getting a certain type of cancer. Using its probe sensors installed in the patient’s house, the robot doctor had detected that she had experienced a disturbed sleeping pattern for more than 6 weeks in continuation. Based on this fact, the robot doctor had looked at her vital statistics data, such as her heart rate, blood pressure, breathing patterns, etc., and had reached the conclusion that she was on the path to get cancer. The human doctor, on the other hand, checks her vital statistics again and asks her to conduct some blood tests and other required tests for determining her current medical condition. After a few days, when her medical reports arrive, the human doctor declares that she does not have any risk.

Does this sound far-fetched and something too distant?

|  |  |
| --- | --- |
| This is not an unlikely scenario but something that we may witness once the | |
| robot doctors become a reality. |  |

|  |  |
| --- | --- |
| There is a robot in China that has already successfully passed the medical | |
| examination and has attained the medical degree of a doctor | . |

1. What questions arise in your mind once you read the situation?
2. What would you do if something like this happened to you?
3. Would you trust the robot doctor?
4. Would you trust the human doctor more?
5. Would you dismiss the report by the robot doctor as false and ignore it after the human doctor gave you a clean chit on your current medical condition?

These are some of the questions that the future society is going to have to deal with once we accept robots as specialists in the healthcare industry. This is a scenario where the human expert does not have the ability to prescribe any medicine based on the patterns that it is observing in a human being. In this case, the robot doctor is better prepared to predict and prescribe course-corrective medication to a human being based on the data that it gets from its connected probes or sensors.

The healthcare industry in particular deals with human beings and their lives. This is one of those industries where a simple judgmental error could cause death to a patient.

However, when we talk about building prediction models based on machine learning (ML), which is the brain behind any robot, we know that no matter what algorithm is selected for predicting the outcome from any data set, there is going to be a percentage of errors in the final prediction by the model. In the case of human beings, a human being or a human doctor or a healthcare professional is also prone to errors.This is something that we know as human error.

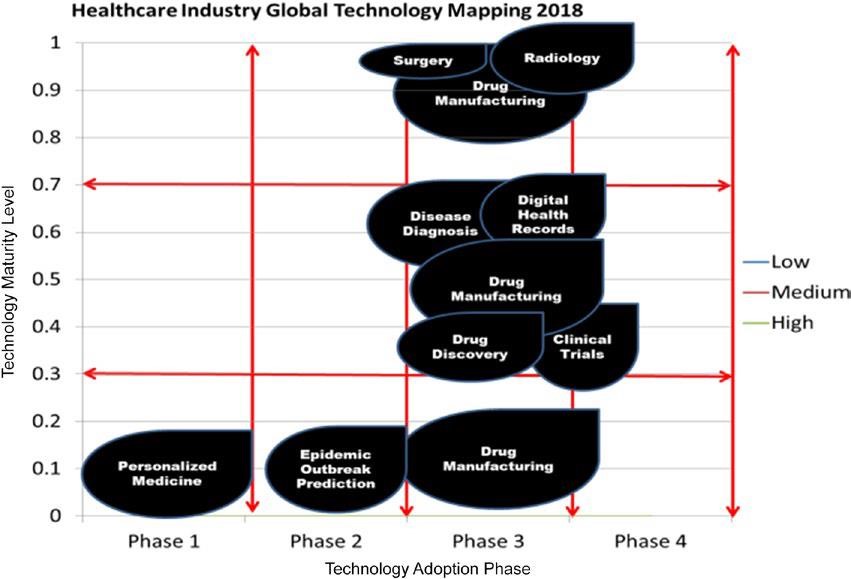
*A recent research by Hopkins Medical Organization or the Johns Hopkins Medical Organization shows that 10 percent of all the U.S. states happened due to medical errors by the doctor and it is the third highest cause of death in the US*

So if we were to build and create a replacement or a competitor for

|  |  |
| --- | --- |
| a human doctor, we know that it would have to do better than this error rate. It | |
| can only survive if it gives predictive diagnosis at a lower error rate than that of | |
| the human doctor. |  |

Since we are dealing with human life in the healthcare industry, we need a gradual and careful way of adopting technology, as a lot is at stake. The requirement is to build robust algorithms with prediction models with higher accuracy levels.

# Healthcare industry technology adoption phases



There are certain areas in the graph that lie in Phase 1 and are low in technological maturity level. Although these hold potential, they do not give us a huge area of opportunity, as the technology is not currently supporting developments in these areas. For example, personalized medicine is very new and there is huge amount of research that must happen, including use of AI, to

|  |  |
| --- | --- |
| enable it to move to the next phase. Such research must be linked to the | |
| healthcare industry very closely so that the adoption happens faster. |  |

Next is the Phase 2 area of epidemic outbreak prediction, which has a few hits and misses and needs to address privacy issues in order to move to Phase 3. The real potential lies in the Phase 3 column of areas, where the technology has moved into the assisted applications stage.

# Areas of Healthcare Research where there is Huge Potential

1. Digital Health Records
2. Disease Diagnosis
3. Radiology

# Therefore the focus of this project is Disease Diagnosis through AI

56

percent of experts feel that Disease diagnosis has a medium level of maturity

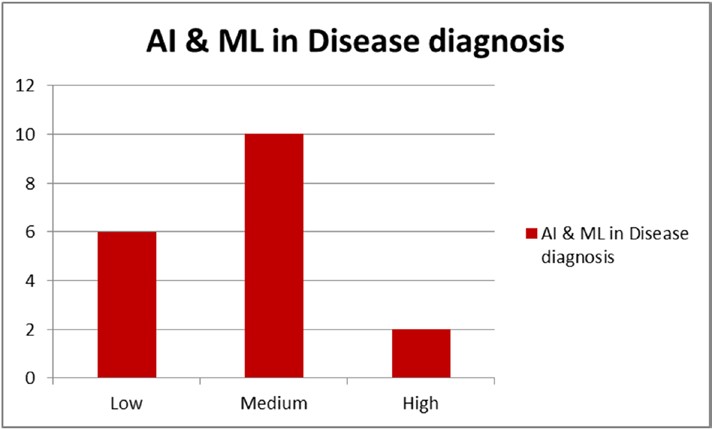
which means that technology has been implemented in the area of disease

diagnosis in production , but there are hits and misses

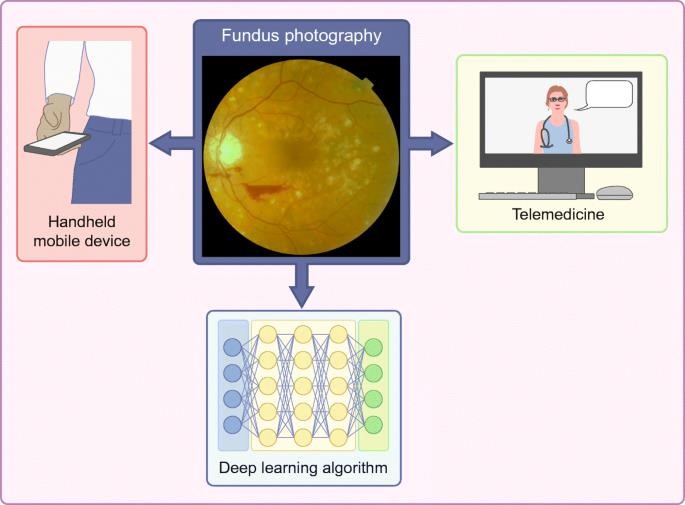
and therefore it needs

more research to move in

to the mainstream production



|  |  |
| --- | --- |
| A good example of this area would be Google’s deep learning for detection of | |
| Diabetic eye disease |  |



**Key References :**

Study Suggests Medical Errors Now Third Leading Cause of Death in the U.S., May 3 2016,

https://www.hopkinsmedicine.org/news/media/releases/

Google Flu Public Data:

https://www.google.com/publicdata/explore?ds=z3bsqef7ki44ac\_#!ctype=m&strail

=false&bcs=d&nselm=s&met\_s=flu\_index&scale\_s=lin&ind\_ s=false&ifdim=region&hl=en\_US&dl=en\_US&ind=false

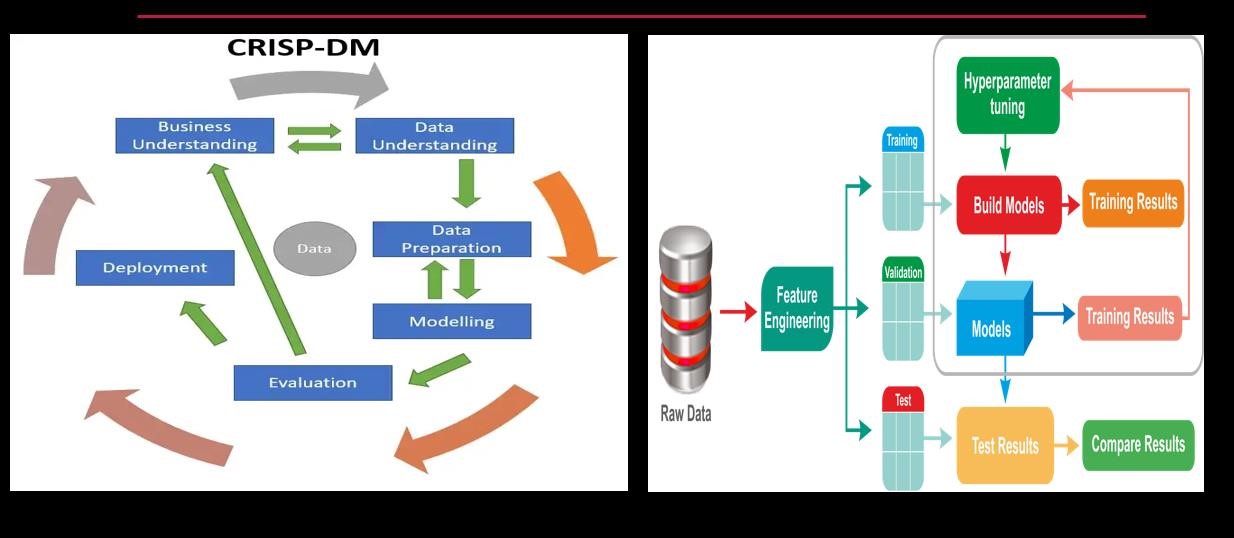
Siemens Healthineers: https://www.healthcare.siemens.com/about

Niramai: http://niramai.com/

# Challenges /Opportunity of Technology led efficiencies and effectiveness in the Healthcare sector in India

1. Low penetration of quality healtcare services.
2. Absymal doctor to patient ratio.
3. Low outreach in Rural India.
4. Affordability .

**Heart Disease Detection through an AI based System :**



**Project Framework**

# CRISP-DM FRAMEWORK

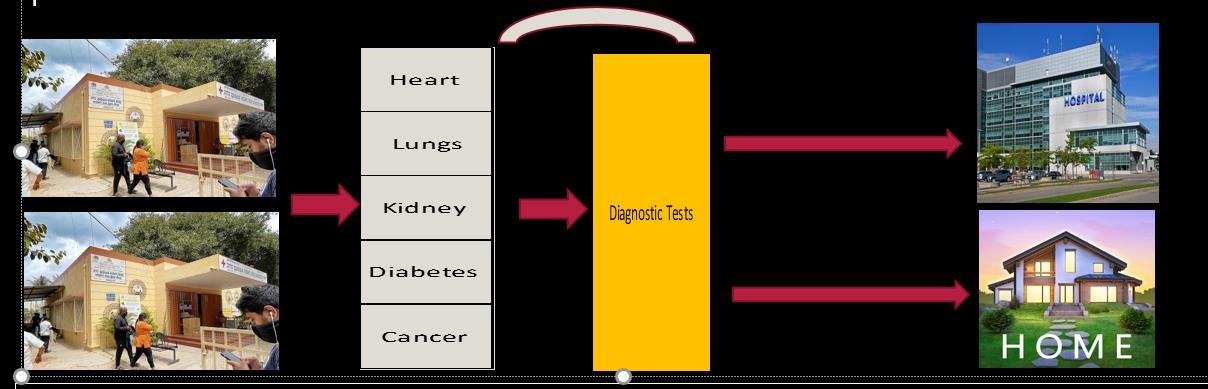
**Business set up :**

|  |  |
| --- | --- |
| Towards Mission Health for India , how can we leverage AI for the benefit of | |
| larger population towards timely medical intervention . |  |

|  |  |
| --- | --- |
| Design an AI system for a Multi specialty Hospital which will screen the patients | |
| at multiple local clinics for possible disease and refer them for further treatment | |
| to the multispeciality hospital . |  |

|  |  |
| --- | --- |
| The first part will be AI driven( at the clinics) and the Second part will be Human | |
| driven (Specialists at Multispeciality Hospitals) | . |

Rough design of the AI system which screens possible patients using an automated AI system ( designed using UI/UX, Design Principles, but at the core will be Predictive Analytics based on Data collected at these clinics .



Reference Data :

|  |
| --- |
| Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D. |
| University Hospital, Zurich, Switzerland: William Steinbrunn, M.D. |

14 Independent variables and 1 Dependent variable .

|  |  |  |
| --- | --- | --- |
| Dependent Variable is labelled as AHD (0 OR 1) ie Angiographic disease status . | | |
| Independent variables captures Demography, Lifestyle parameters , and other | |  |
| associated variables |  |

Data File : **heart.csv** Data Dictionary :

Features:

1.(age)

2.(sex) (0 = female, 1 = male)

3.(cp) cp: chest pain type -- 1: typical angina -- 2: atypical angina -- 3: non-anginal pain -- 4: asymptomatic

4.(trestbps) trestbps: resting blood pressure (in mm Hg on admission to the hospital)

5.(chol) chol: serum cholesterol in mg/dl

6.(fbs) fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)

7.(restecg) restecg: resting electrocardiographic results -- Value 0: normal -- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) -- Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria

8.(MaxHr): maximum heart rate achieved

9.(exang) exercise induced angina (1 = yes; 0 = no)

10.(oldpeak) oldpeak = ST depression induced by exercise relative to rest

11.(slope) slope: the slope of the peak exercise ST segment -- Value 1: upsloping -- Value 2: flat -- Value 3: downsloping

12.(ca) ca: number of major vessels (0-3) colored by flourosopy

13.(thal) thaldur: duration of exercise test in minutes

14.(AHD) (the predicted attribute): diagnosis of heart disease (angiographic disease status) --

15.No: < 50% diameter narrowin

16.Yes: > 50% diameter narrowing (in any major vessel: attributes 59 through 68 are vessels)

Key Objectives :

* Business objective :
  + **Screen maximum** possible walkins with respect to disease disorder , with minimal errors, before directing them to either hospital or home.
  + **Revenue maximization** through maximum possible walkins to the various Multispeciality hospitals located around the city .

* Data objective :
  + The classification system should be accurate enough to detect **False Negatives** .
  + Therefore **Recall** is the key metrics to be maximized for the candidate model/s for prediction.
  + **Appropriate thresholding** with scenario testing .

**Evaluation Rubrics :**

|  |  |
| --- | --- |
| Topic | Marks |
|  |  |
| Business Understanding | 20 |
|  |  |
| Data Understanding & Insighting | 20 |
|  |  |
| Pre -Modelling | 10 |
| Modelling | 30 |
|  |  |
| Interpretations and Presentation | 20 |
|  |  |
| Total | 100 |