PRAESCRIPTIO

A Report

Submitted by

Ayushi Das
Roll: 191001112047
Regn. No. 1001911538 of 2019-2022

In partial fulfillment of the award of the Degree of

Bachelor of Science
in
Cloud Technology & Information Security

Under the guidance of

Arpita Sen,

Senior Faculty, Department of I.T., iNurture Education Solutions Ltd.

DEPARTMENT OF CLOUD TECHNOLOGY & INFORMATION SECURITY
Techno India University, EM-4, EM Block, Sector V, Bidhannagar,
Kolkata, West Bengal 700091

March 2022

CERTIFICATE

The report of the Project titled Praescriptio is submitted by Ayushi Das (Roll No.: 191001112047 of B. Sc (IT-CTIS) 6th Semester of 2019--2022 has been prepared under our supervision for the partial fulfillment of the requirements for B.Sc. degree in Techno India University.

Signature of the Guide

DECLARATION

This report has been prepared on the basis of my own work. Where other published and unpublished source materials have been used, these have been acknowledged.

Word Count: 7,630

Student Name: Ayushi Das

Date of Submission: 10th June, 2022.

Signature of the student

ACKNOWLEDGEMENT

I express my sincere gratitude to Ma'am, Arpita Sen, Senior Faculty from Department of I.T., TIU and for extending their valuable times for me to take up this problem as a Project.

I am also indebted to Sir, Deepjyoti Choudhury, Senior Faculty, Department of I.T., TIU For their unconditional help andinspiration.

Last but not the least I would like to express my gratitude to one of my supportive senior, Aviral Sharma, one of my senior who helped me in his own way whenever needed.

Date: 10th June, 2022.

Ayushi Das
Reg. No.: 1001911538
Roll No.: 191001112047
B. Sc (IT-CTIS) – 6th Semester, 2022, TIU

TABLE OF CONTENTS

Abstract

The adoption of artificial intelligence (AI) is reshaping the Indian healthcare market significantly. AI-enabled healthcare services like automated analysis of medical tests, predictive healthcare diagnosis, automation of healthcare diagnosis with the help of monitoring equipment, and wearable sensor-based medical devices, are expected to revolutionize medical treatment processes in the country. According to the recent market insights, it is predicted that the applications of artificial intelligence in the healthcare space will be worth INR ~431.97 billion by 2021, expanding at a rate of ~40%. Based on this growth of AI applications in healthcare, the doctor-patient ratio in India is expected to reach ~6.9:1,000 by 2023, from its 2017 ratio of ~4.8:1000. The capability of AI applications to improve doctors' efficiency will help in tackling challenges like uneven doctor-patient ratio, by providing rural populations high-quality healthcare, and training doctors and nurses to handle complex medical procedures.

Artificial intelligence is used in six healthcare segments: hospitals, pharmaceuticals, diagnostics, medical equipment and supplies, medical insurance, and telemedicine. Artivatic Data Labs Private Limited, Analytix Private Limited, IBM India Limited, Niramai Health Sigtuple Technologies Private Limited, and Tricog Health Services Private Limited are the major players operating in this sector. There are also some of the government initiatives in this domain. The Information Technology Act, 2000, and the Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules, 2011, mandate that service providers and patients exchange information constantly by using the latest technologies. National eHealth Authority (NeHA) – An authority which is responsible for the expansion of the integrated health information system within India. The United States–India Science and Technology Endowment Fund, is aimed at helping teams of innovators and entrepreneurs from both countries, whose products will improve the quality of healthcare, by harnessing the power of artificial intelligence. Let's talk about few benefits of AI in the healthcare sector. The patient doctor ratio in India is as low as 1,700:1. With the use of artificial intelligence applications, doctors can offer their services to more patients and reduce the existing gap in demand and supply of medical services in the country.AI-enabled healthcare services can be delivered at lower costs with increased efficiency and an emphasis on diagnostics. Moreover, artificial intelligence enables hospitals to implement patient centric plans and eliminate unnecessary hospital procedures, making delivery of healthcare services faster and efficient in India.

Key deterrents to the growth of the market can be understood by stating that India lacks standardized guidelines for designing AI applications that can be used in healthcare systems. Lack of clarity deters the use of artificial intelligence in the Indian healthcare industry. Also, most AI companies which aid the healthcare sector, are start-ups. Medical practitioners are not quick to trust start-ups whose products are not nationally or internationally certified. As a result, start-ups' sales get hampered, resulting in the limited implementation of AI in the Indian healthcare industry [1]. This papers aims at shedding light on developing one of the business models which will work inside the existing healthcare model till it gains trust to get completely automated and scalable. Keeping in mind the healthcare and technology gaps it propose an AI based solution to be deployed in existing hospitals for quick and automatic results.

Introduction

This section describes the unique opportunities and challenges for artificial intelligence (AI) and digital health technology in India, describes some success stories, and brings up some current trends. With its vast inequalities in healthcare distribution, glaring lack of trained healthcare clinicians and infrastructure, and low government spending on healthcare, India is one of the countries in the world with the most room for innovative, sustainable and scalable healthcare technology to improve lives. Yet, in a country with 1 billion people, many now equipped with internet connections and smartphones, it is still difficult to name more than a handful of examples of digital technology that have significantly impacted healthcare outcomes or been used widely. This piece describes the unique opportunities that the system offers, the challenges which prevent small initiatives from scaling up, describes some success stories, and brings up some worrying trends around artificial intelligence (AI) and Indian healthcare. For India, it is imperative to design and develop technology that takes into account local constraints, among them affordability. There are many local and behavioural challenges in the Indian healthcare sector, but cost is still a key driver. For it to succeed and make a difference at scale, new technology has to be priced for the country and developed to tackle its constraints. The good news is that this is exactly what AI promises. If implemented correctly, AI boils down to redistributing scarce expert knowledge to a large number of beneficiaries by training algorithms machines to replicate this knowledge.

Unique opportunities

There is probably no better place than India to find a problem in need of a creative solution. Around every corner is a new gap to bridge - a skill gap, geographical gap, an infrastructure gap, an urban-rural divide or a spending gap. The diversity and potential scale of the Indian healthcare system affords an opportunity and incentives like no other to pilot and operationalize innovations. Each one of the systems that is a failing for the Indian healthcare system is a unique opportunity for AI or even for simple digital technology. The country has witnessed the rapid penetration of internet and smartphones over the last decade and now meets the requirements for efficient delivery of digital solutions. Government enthusiasm for innovation and locally made technology is at an all-time high - both at the central policy level, as well as at local level, with individual states seeking to outdo each other at the adoption of new technology that can help solve old problems. Support for public-private partnerships is high. While this must be tempered by a healthy dose of scepticism about real ground conditions, this is an encouraging sign for innovators.

There are a plethora of healthcare issues that are still 'virgin territory' to technology, each with millions of potential beneficiaries in India. Areas such as antibiotic resistance, health insurance, communicable diseases like malaria and tuberculosis, as well as emerging ones like diabetes are a few of the many worth looking at through the lens of the technology, connectivity and artificial intelligence available today.

Challenges

The upside for an innovator in India is that she is very likely to find a use-case and a chance to pilot new technology. However, in a country that is often called 'the land of pilots', the challenge is usually with scaling and distributing technology - even technology that has been proven to be cost-effective and useful. Several pilots of public-private partnerships have been successful. However, none of them has been scaled up to meet India's health challenges.

AI can potentially leapfrog some other technologies, but for AI to be used at any scale, digitalisation is a pre-requisite. Considering that, in many Indian health centres, medical records are still paper, and radiology still uses films (although this is changing rapidly). The pace of this change is rapid, but statistics on digitalisation of records, prescriptions, and radiology are hard to come by.

Healthcare systems everywhere are slower to adopt change than their counterparts in other industries, often with good reason. But in India, it is not only regulation which stifles innovation. Most healthcare services are provided by the private sector and paid for out-of-pocket. This means that to be broadly adopted, technology has to provide a clear short-to-medium term incentive to the private sector, rather than directly aligning with health outcomes. The lack of government spending on healthcare means that public health programmes are still largely funded from outside the country. This sometimes results in importing technology rather than fostering the development of indigenously developed locally appropriate inventions. Medical education in India does not place enough emphasis on research and on keeping up with new developments. Combined with an overburdened system, this results in generations of practicing clinicians with little motivation to innovate or to understand and adopt technology.

Pioneering examples

There are many examples of digital innovation that have demonstrated success in screening, prevention and treatment in India. For each of these, there are likely dozens of comparable projects ongoing. Most of the country's healthcare is pre-digital, and paper medical records and film-based radiology are still more common than their electronic counterparts. In this setting, even seemingly simple systems such as an online appointment-booking system at the country's largest public hospitals in New Delhi can have a large impact by sparing patients long waits and saving numerous trips to the hospital for those who can ill-afford to take a day off. The last decade has also seen some great examples of dedicated hardware and technology that is engineered for the unique challenges of Indian health ecosystem. These include products for tuberculosis medication adherence monitoring (one of India's most significant public health issues), low-cost vital parameter monitors for use in the primary healthcare setting, and telemedicine programs that provide clinical expertise to areas without doctors. These are more mature than the artificial intelligence applications, which have begun to emerge over the last 5 years. Primarily used for screening, monitoring, and diagnostic assistance, AI applications include algorithms that analyse chest X-rays and other radiology images, read ECGs and spot abnormal patterns, automatically scan pathology slides and even assess fundus images for signs of retinopathy.

Certain branches of medicine in India have been more successful than others at fostering the development of innovation and adopting it. Ophthalmology is a clear leader on both these counts, with a relatively broad range of innovative technologies - high-quality imaging of both retina and cornea using smartphone-coupled devices, artificial intelligence for the screening of diabetic retinopathy - being developed and tested, and then brought into clinical use. This has largely been due to private-sector efforts and would not have been possible without the foresight shown by a set of well-organised large private eye care centres in the South of India that facilitated the data collection and piloted the new technologies.

Worrying trends

As the official use of healthcare electronic medical records and secure electronic means of aggregating and transferring medical data is gradually being adopted, an alternative is quickly taking its place. Indians are prolific users of the free messaging application ‘Whatsapp’ and doctor-patient communication and even inter-clinician communication by Whatsapp is frequent, with the medium used regularly by groups of clinicians to share and consult on cases.

A second, perhaps even more concerning trend for AI, in particular, is that India is starting to be seen as a ‘data source’ for radiology and ophthalmology images, and perhaps more. A combination of factors, including English-language reporting, privately owned healthcare systems and lax privacy, data protection and ownership laws. The data sharing itself may or may not be a cause for worry - but the lack of transparency and regulation around it certainly should raise red flags.

Current and future

The last 5 years in India have seen consumer-facing ‘health tech’ being talked about and embraced by investors, by the government and gradually by the public. Technology aimed at the urban, educated consumer is gaining traction, mostly in the form of online health service aggregators, telemedicine, e-commerce for home delivery of pharmaceuticals and a wave of fitness apps. Existing methods are also being used to reinvent healthcare delivery in the form of online consults or chat-based basic healthcare service apps.

More recently, physician-facing digital healthcare tech has begun to make its appearance - such as technology that performs or assists with core healthcare or medical tasks like analysing radiology, pathology or ophthalmology images.

Conclusion

The significant need for technology to bridge resource gaps in India, and the potential of AI to offer affordable solutions at scale means that India may soon be poised to realise the benefits of these technologies on health outcomes.

Key Points

1. Innovative, sustainable and scalable artificial intelligence technology has the potential to greatly improve healthcare outcomes in India
2. AI applications being developed and deployed in India include algorithms that analyze chest x-rays and other radiology images, read ECGs and spot abnormal patterns, automatically scan pathology slides and assess fundus photographs for signs of retinopathy.
3. Scaling up and distributing technology in India is challenging. [2]

Problem Analysis

Background

India still struggles with rural doctor shortages, with hospitals made available in considerably distant places. Although the number of health facilities has risen in the past decade, workforce shortages are substantial Even in health facilities where doctors, specialists, and paramedic staff have been posted, their availability remains in question because of high rates of absenteeism. Also in India, at least nine out of 10 adults suffer from low health literacy. It is reported that even in the US and the UK, more than 50 per cent of the people have low health literacy; the 'cost of ignorance' is about $200 billion in the US alone. As of March 31, 2015, more than 8% of 25 300 primary health centers in the country were without a doctor, 38% were without a laboratory technician, and 22% had no pharmacist. Nearly 50% of posts for female health assistants and 61% for male health assistants remain vacant. In community health centers, the shortfall is huge—surgeons (83%), obstetricians and gynecologists (76%), physicians (83%), and pediatricians (82%). [3]

Challenges

Most important challenges faced by the healthcare AI industry starts with the huge pressure on healthcare systems and equipment. Exponential growth of healthcare data. Producing perfect insights at the point of decision making. Augmented intelligence for the clinicians. Integration and legal challenges. Interoperability and integration is one of the key factors that separates academic research from practical applications of AI and that's what all the hospitals are looking at right now. They are setting up departments to deal with companies. We need to connect the dots along the continuum of care. For instance, linking hospitals to home, primary care etc. Also medical equipment vendors have to start enabling the creation of applications by third parties like innovative start-ups or academic clinical centres through publishing application programming interfaces (API).
We need APIs to consume data from the systems and that's also a bit of a challenge for the industry right now. Lastly, the legal challenges needless to say, safeguarding patient privacy is very important. And without doing that the FDA doesn't allow us to release any AI algorithms. [4]

Integration

With intellectual arbitrage, we can use the experience of other industries that have adopted AI and deep learning to apply it to healthcare. Arbitrage exploits incomplete distribution of information or experience. Healthcare's future is the past or present of other industries. When we look at other business verticals, without exception they bend over backwards to build human-machine cyber optimized IT collaboration workflow.

The whole idea of human-machine cybernetic collaboration is that it takes over what humans are bad at, such as remembering things. That leaves the human to do what they do best, and that is to understand what needs to be done for the patient, prioritize what is important, gain insight and help patients. That's not a threat to a medical experts, that's being complementary and other industries do it very well. People who are threatened by AI need to understand that this is a natural evolution of human machine collaboration, even predating computers. Being ten years behind other industries on AI and deep learning is not a bad thing. However, now that we have shared risk and capitation, we're competing in earnest, and it has to be more strategic, and our managers have to be more aggressive in looking at how we can achieve a differential competitive advantage by leveraging and optimizing human-machine cybernetic collaboration. To understand how best to achieve human-machine cybernetic balance all I need do is look at other industries. It's the best way to mitigate risk. The problem is we don't do it. We tend to recapitulate the errors of early adopters, including buying into hype. Some are even boldly claiming that human radiologists will soon be obsolete, replaced by these powerful creations. However, building the best deep learning system in the lab is not enough. Even if you assume that it is possible to create a deep learning system that will be a better radiologist then humans (a very strong and I believe suspect assumption), it is like building the best race car in a world without gasoline or roads. Even the best race car needs gasoline and roads. [5]

Recommendation

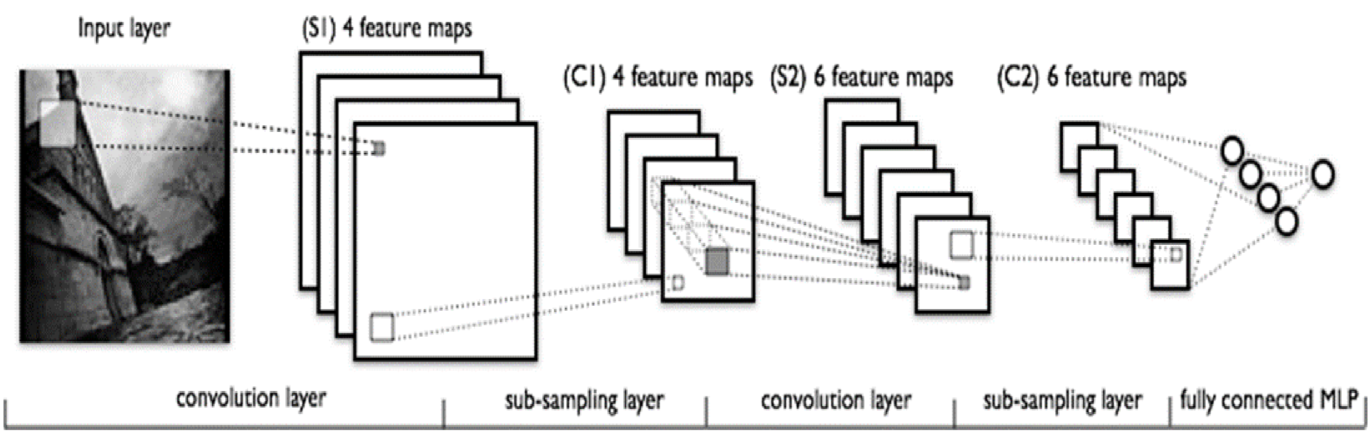
The medical departments in radiology aim to minimise exposure to the patients and get the scans right in the first try. For this, there are AI-related algorithms that get initiated by healthcare companies along with equipment that ensure high-quality images in a shorter span.

Secondly, these equipment are too sensitive and if a patient is moving a lot, the scanner can't take the image properly. So the staff need to control the motion. To do that, we need to think of software which can work even without making patients wear or use some physical equipment such as belts. Additionally, the biggest challenge today for radiology is making patients run from one department to another collecting reports. This also can be addressed by using AI-enabled workflow solutions. Additionally, chest X-rays, accurate precision diagnostics, and providing the right measurements for ultrasound are some standard areas where AI is being leveraged. Praescriptio will answer some of the concerned issues in the later part of this report.

Review of Literature

In the latter years, there has been a lot of work done to diagnose the discrete diseases in thyroid. Many authors have used various kinds of data mining technique. The authors proved to obtain an adequate approach and certainty to find out the diseases analogous to the thyroid by the work that includes various datasets and algorithms linked with the work that is to be done in the future perspective to accomplish effective and better results. The article from geeksforgeeks has inspired me in developing the common Disease Prediction Model using Machine Learning. They have portrayed the Confusion Matrix in Machine Learning and have implemented various other algorithms and compared them with respect to their accuracy. Some of them are Liner Regression and Logistic Regression. [6]

There are several methods and tests which can be used for malaria detection and diagnosis. The original paper on which our data and analysis is based on, ‘ Pre-trained convolutional neural networks as feature extractors toward improved Malaria parasite detection in thin blood smear images’ introduces us briefly to some of these methods. These include but are not limited to, thick and thin blood smear examinations, polymerase chain reaction (PCR) and rapid diagnostic tests (RDT). While we won’t cover all the methods here in detail, an important point to remember is that the latter two tests are alternative methods typically used an alternative particularly where good quality microscopy services cannot be readily provided. Deep Learning models, or to be more specific, Convolutional Neural Networks (CNNs) have proven to be really effective in a wide variety of computer vision tasks. While we assume that you have some knowledge on CNNs. Briefly, the key layers in a CNN model include convolution and pooling layers as depicted in the following figure. The paper by Rajaraman, leverages a total of six pre-trained models on the data mentioned in their paper to obtain an impressive accuracy of 95.9% in detecting malaria vs. non-infected samples. Our focus would be on VGG-19 model using transfer learning to see the kind of result we get on the same dataset! We will be using open-source tools and frameworks which include Python and TensorFlow to build our models. [7]



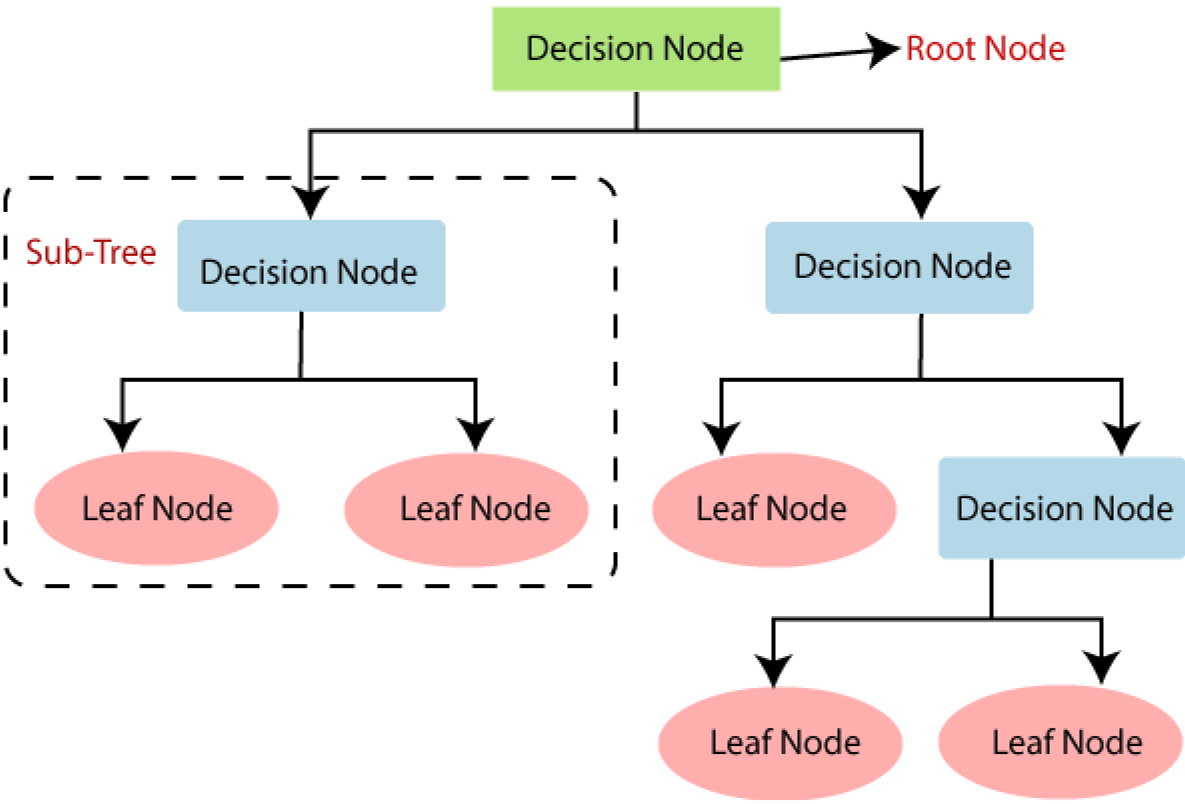
A typical CNN architecture (Source: deeplearning.net)

Various methods are available for brain tumour classification. Among various methods, neural network-based methods, convolutional neural network (CNN) methods, and DL methods are widely used.
The study has investigated the use of the deep features extracted from the CNNs that are pre-trained in the prediction of survival time. It provides further evidence for domain-specific fine-tuning to improve its performance. Standard dataset is available in the internet. It has an accuracy of approximately 81% in the case of leave-one-out-based cross-validation. [12] Some study has proposed a hybrid method for classifying the tissues of the brain tumour image. In this technique, the system employs a genetic algorithm (GA) for feature extraction and a support vector machine (SVM) for classification purposes. The features are further compared with the stored features, and the method is used to capture the images and their visual contents. This method signifies the raw image to simplify decision-making in a much focused form. The choice of the features composed is very difficult in the classification methods that are duly solved using the GA. The features with the SVM have been used to classify the tumour as normal or abnormal. If the tumour is detected based on mean, mode, and median, then the tumour is classified as either meningioma or pituitary tumour. The performance of the algorithm is assessed based on the images containing the brain tumour. [13]
Several methods have been introduced to describe a brief process in pneumonia detection using chest X-ray images in recent years, especially some deep learning methods. Deep Learning has been successfully applied to improve the performance of computer aided diagnosis technology (CAD), especially in the field of medical imaging [14], image segmentation [15] and image reconstruction. In 2017, Rajpurkar proposed a classical deep learning network named DenseNet-121 [16], which was a 121-layer CNN model to accelerate the diagnosis for pneumonia. In contrast to experienced doctors, the framework obtained a higher F1 score. Besides, in order to alleviate the effect of imbalanced classes, the team introduced Weighted Binary Cross-Entropy loss, whose difference between the Binary Cross Entropy losses was the different weights of imbalanced classes according to the number of each class. [17]

Formulation / Algorithm

Disease Prediction Using Machine Learning

Decision Tree is a supervised learning technique that can be used for both Classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches. The decisions or the test are performed on the basis of features of the given dataset.
It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions. It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure. In order to build a tree, we use the CART algorithm**,** which stands for Classification and Regression Tree algorithm. A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into sub trees. Below diagram explains the general structure of a decision tree:



The general structure of a decision tree (Source: tutorialspoint.com)

Decision Tree Terminologies

1. **Root Node:** Root node is from where the decision tree starts. It represents the entire dataset, which further gets divided into two or more homogeneous sets.
2. **Leaf Node:** Leaf nodes are the final output node, and the tree cannot be segregated further after getting a leaf node.
3. **Splitting:** Splitting is the process of dividing the decision node/root node into sub-nodes according to the given conditions.
4. **Branch/Sub Tree:** A tree formed by splitting the tree.
5. **Parent/Child node:** The root node of the tree is called the parent node, and other nodes are called the child nodes.

In a decision tree, for predicting the class of the given dataset, the algorithm starts from the root node of the tree. This algorithm compares the values of root attribute with the record (real dataset) attribute and, based on the comparison, follows the branch and jumps to the next node.

For the next node, the algorithm again compares the attribute value with the other sub-nodes and move further. It continues the process until it reaches the leaf node of the tree. The complete process can be better understood using the below algorithm:

Step-1: Begin the tree with the root node, says S, which contains the complete dataset.

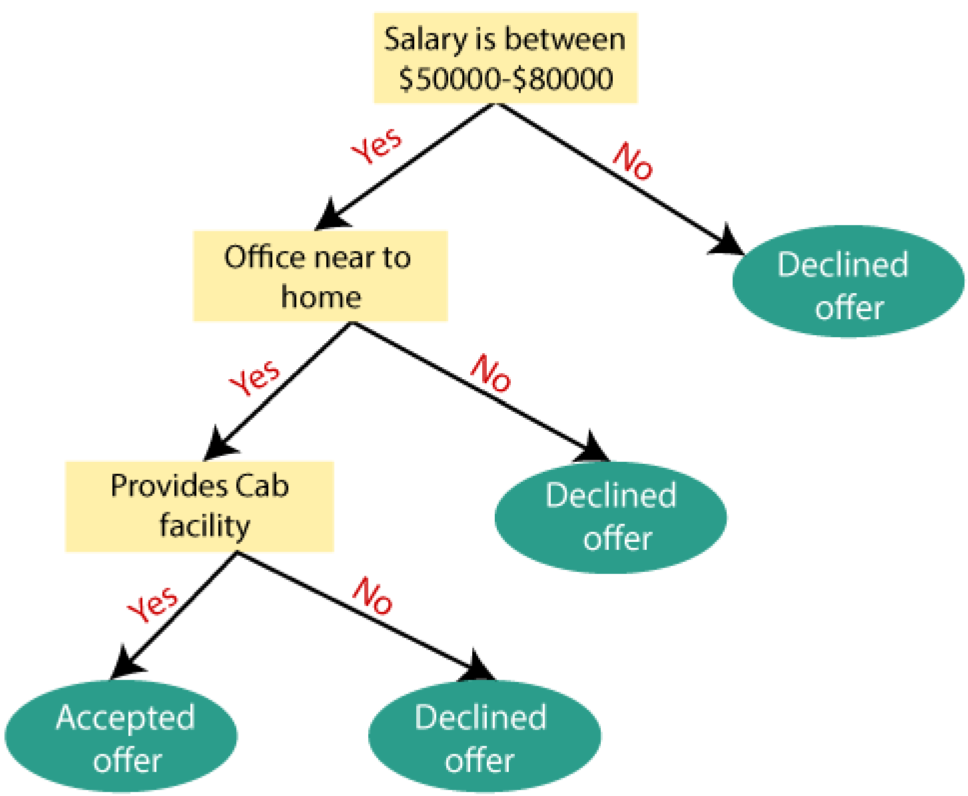
Step-2: Find the best attribute in the dataset using Attribute Selection Measure (ASM).

Step-3: Divide the S into subsets that contains possible values for the best attributes.

Step-4: Generate the decision tree node, which contains the best attribute.

Step-5: Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

Example: Suppose there is a candidate who has a job offer and wants to decide whether he should accept the offer or Not. So, to solve this problem, the decision tree starts with the root node (Salary attribute by ASM). The root node splits further into the next decision node (distance from the office) and one leaf node based on the corresponding labels. The next decision node further gets split into one decision node (Cab facility) and one leaf node. Finally, the decision node splits into two leaf nodes (Accepted offers and Declined offer). Consider the below diagram



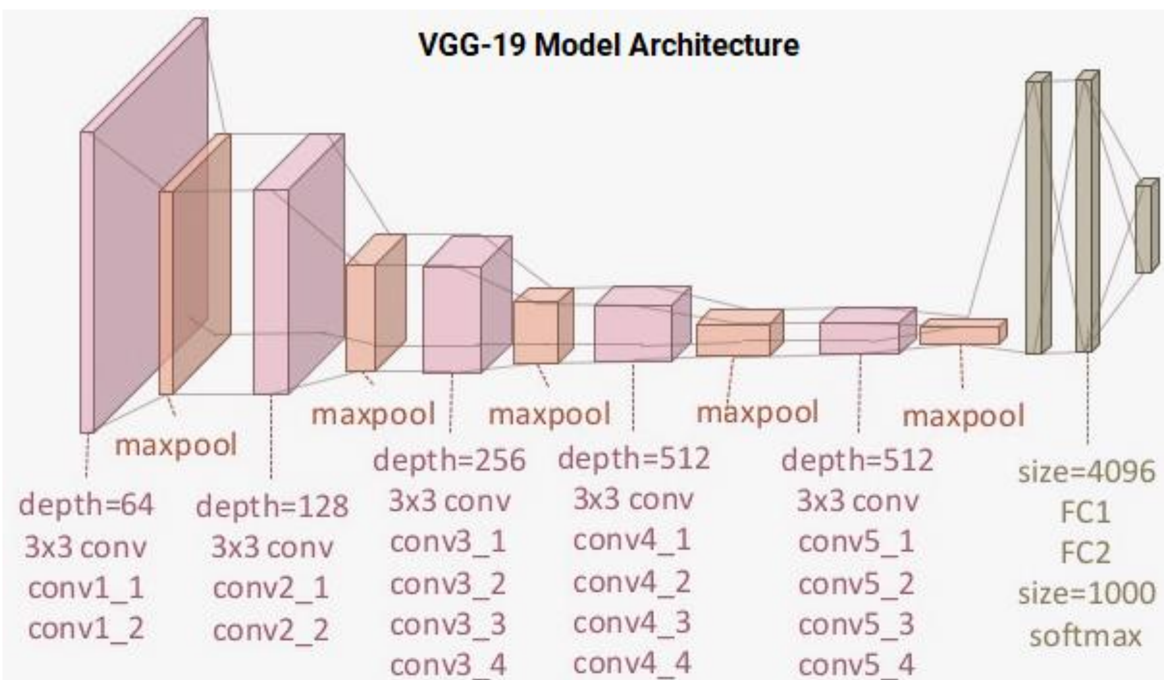
A flowchart of Decision Tress example (Source: Javapoint.com)

Malaria Detection Using Deep Learning and Brain Tumor Classification and Segmentation Using Deep Learning

We will be using the pre-trained VGG-19 deep learning model, developed by the Visual Geometry Group (VGG) at the University of Oxford, for our experiments. A pre-trained model like the VGG-19 is an already pre-trained model on a huge dataset (Image Net) with a lot of diverse image categories. Considering this fact, the model should have learned a robust hierarchy of features, which are spatial, rotation, and translation invariant with regard to features learned by CNN models. Hence, the model, having learned a good representation of features for over a million images, can act as a good feature extractor for new images suitable for computer vision problems just like malaria detection! Let's briefly discuss the VGG-19 model architecture before unleashing the power of transfer learning on our problem.

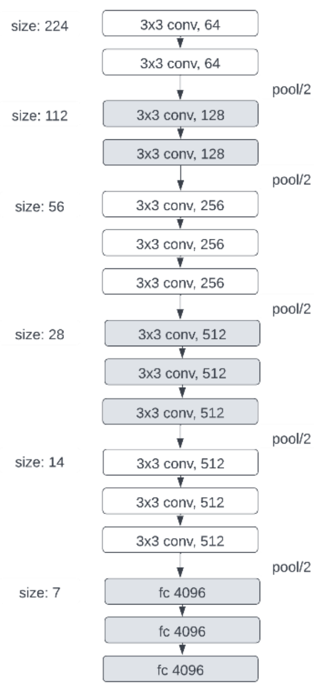
Understanding the VGG-19 model

The VGG-19 model is a 19-layer (convolution and fully connected) deep learning network built on the ImageNet database, which is built for the purpose of image recognition and classification. This model was built by Karen Simonyan and Andrew Zisserman and is mentioned in their paper titled 'Very Deep Convolutional Networks for Large-Scale Image Recognition'. I recommend all interested readers to go and read up on the excellent literature in this paper. The architecture of the VGG-19 model is depicted in the following figure. [8]



VGG-19 Model Architecture, Source (Medium.com)

We can clearly see that we have a total 19 convolution layers which uses 3X3 convolution filters along with max pooling layers for down sampling and a total of two fully connected hidden layers, 4096 units in each layer followed by a dense layer, 1000 units, where each unit represents one of the image categories in the ImageNet database. We do not need the last three layers since we will be using our own fully connected dense layers to predict malaria. We are more concerned with the first five blocks, so that we can leverage the VGG model as an effective feature extractor. We will use it as a simple feature extractor by freezing all the five convolution blocks to make sure their weights don't get updated after each epoch as we train our own model. [9]



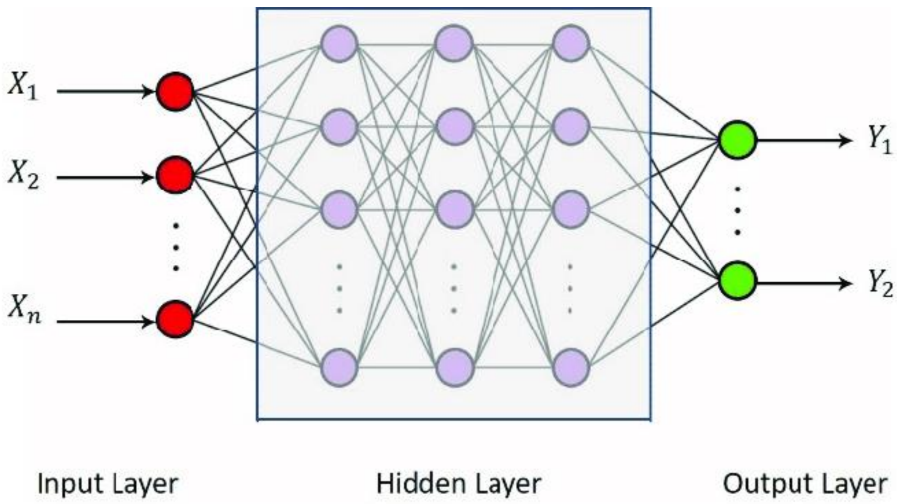
Flow Diagram of VGG-19 Architecture

Number	Convolution	Output Dimension	Pooling	Output Dimension
Layer 1&2	Convolution Layer of 64 channel of 3x3 kernel with padding 1, stride 1	224x224x64	Max Pool Stride=2, Size 2x2	112x112x64
Layer 3&4	Convolution Layer of 128 channel of 3x3 kernel	112x112x128	Max Pool Stride=2, Size 2x2	56x56x128
Layer 5,6,7	Convolution Layer of 256 channel of 3x3 kernel	56x56x256	Max Pool Stride=2, Size 2x2	28x28x256
Layer 8,9,10	Convolution Layer of 512 channel of 3x3 kernel	28x28x512	Max Pool Stride=2, Size 2x2	14x14x512
Layer 11,12,13	Convolution Layer of 512 channel of 3x3 kernel	14x14x512	Max Pool Stride=2, Size 2x2	7x7x512

Table to explain VGG-19 Architecture

Covid-19 and Pneumonia Detection Using Deep Learning

Artificial Neural Networks is one of the most common and frequently tackled problems in the machine learning domain. In its simplest form the user tries to classify an entity into one of the two possible categories. Through the effective use of Neural Networks (Deep Learning Models), problems can be solved to a fairly high degree. We will create a Neural Network that has three layers. There are two layers of 16 nodes each and one output node. The last node uses the sigmoid activation function that will squeeze all the values between 0 and 1 into the form of a sigmoid curve. The other two layers use ReLU (Rectified Linear Units) as the activation function. ReLU is a half rectified function; that is, for all the inputs less than 0 (e.g. -120,-6.7, -0.0344, 0) the value is 0 while for anything positive (e.g. 10, 15, 34) the value is retained. One output unit is used since for each record values in X, a probability will be predicted. If it is high (>0.9) than the molecule is definitely active. If it is less (<0.2) then it is definitely not active. [18]



ANN with 3hidden layers for Binary Classification, Source (Researchgate.com)

Problem Discussion

Disease Prediction Using Machine Learning

There are various algorithms in Machine learning, so choosing the best algorithm for the given dataset and problem is the main point to remember while creating a machine learning model. There are some notable reasons for me choosing Decision Tress Algorithm. Firstly, it is simple to understand as it follows the same process which a human follow while making any decision in real-life which in turn can be very useful for solving decision-related problems. It helps to think about all the possible outcomes for a problem and there is less requirement of data cleaning compared to other algorithms. The logic behind the decision tree also can be easily understood because it shows a tree-like structure.

Malaria Detection Using Deep Learning and Brain Tumor Classification and Segmentation Using Deep Learning

Both the models are trained using VGG-19 CNN Model. The VGG19 model achieves almost 92.7% top-5 test accuracy in ImageNet. ImageNet is a dataset consisting of more than 14 million images belonging to nearly 1000 classes. Moreover, it was one of the most popular models submitted to Large Scale Visual Recognition Challenge 2014 (ILSVRC2014) thereby making significant improvements over AlexNet. The VGG19 model was trained using Nvidia Titan Black GPUs for multiple weeks. VGG19 highly surpasses the previous versions of models in the ILSVRC-2012 and ILSVRC-2013 competitions. Moreover, the VGG19 result is competing for the classification task winner (GoogLeNet with 6.7% error) and considerably outperforms the ILSVRC-2013 winning submission Clarifai. It obtained 11.2% with external training data and around 11.7% without it. In terms of the single-net performance, the VGGNet-19 model achieves the best result with about 7.0% test error, thereby surpassing a single GoogLeNet by around 0.9%. [11]

Covid-19 and Pneumonia Detection Using Deep Learning

Some of the important reasons for choosing the ANN model are stated here. Problems in ANN are represented by attribute-value pairs. ANNs are used for problems having the target function, the output may be discrete-valued, real-valued, or a vector of several real or discrete-valued attributes. ANN learning methods are quite robust to noise in the training data. The training examples may contain errors, which do not affect the final output. It is used where the fast evaluation of the learned target function required. ANNs can bear long training times depending on factors such as the number of weights in the network, the number of training examples considered, and the settings of various learning algorithm parameters. [19]

Implementation Details

Disease Prediction Using Machine Learning

The dataset to predict the common type of diseases is collected from Kaggle, a subsidiary of Google LLC, is an online community of data scientists and machine learning practitioners. This dataset will help applying Machine Learning to great use. Applying Knowledge to field of Medical Science and making the task of Physician easy is the main purpose of this dataset. This dataset has 132 parameters on which 42 different types of diseases can be predicted. Complete Dataset consists of 2 files for Training and Testing. Each CSV file has 133 columns. 132 of these columns are symptoms that a person experiences and last column is the prognosis. These symptoms are mapped to 42 diseases we can classify these set of symptoms to. Training file has 1500 rows and Testing file has 42 rows. [20]

	DP	DQ	DR	DS	DT	DU	DV	DW	DX	DY	DZ	EA	EB	EC
1	prominent_veins_on_calf	palpitations	painful_walking	pus_filled_pimples	blackheads	scurring	skin_peeling	silver_like_dusting	small_dents_in_nails	inflammatory_nails	blister	red_sore_around_nose	yellow_crust_ooze	prognosis
2	0	0	0	0	0	0	0	0	0	0	0	0	0	Fungal infection
3	0	0	0	0	0	0	0	0	0	0	0	0	0	Allergy
4	0	0	0	0	0	0	0	0	0	0	0	0	0	GERD
5	0	0	0	0	0	0	0	0	0	0	0	0	0	Chronic cholestasis
6	0	0	0	0	0	0	0	0	0	0	0	0	0	Drug Reaction
7	0	0	0	0	0	0	0	0	0	0	0	0	0	Peptic ulcer diseae
8	0	0	0	0	0	0	0	0	0	0	0	0	0	AIDS
9	0	0	0	0	0	0	0	0	0	0	0	0	0	Diabetes
10	0	0	0	0	0	0	0	0	0	0	0	0	0	Gastroenteritis
11	0	0	0	0	0	0	0	0	0	0	0	0	0	Bronchial Asthma
12	0	0	0	0	0	0	0	0	0	0	0	0	0	Hypertension
13	0	0	0	0	0	0	0	0	0	0	0	0	0	Migraine
14	0	0	0	0	0	0	0	0	0	0	0	0	0	Cervical spondylosis
15	0	0	0	0	0	0	0	0	0	0	0	0	0	Paralysis (brain hemorrhage)
16	0	0	0	0	0	0	0	0	0	0	0	0	0	Jaundice
17	0	0	0	0	0	0	0	0	0	0	0	0	0	Malaria
18	0	0	0	0	0	0	0	0	0	0	0	0	0	Chicken pox
19	0	0	0	0	0	0	0	0	0	0	0	0	0	Dengue
20	0	0	0	0	0	0	0	0	0	0	0	0	0	Typhoid
21	0	0	0	0	0	0	0	0	0	0	0	0	0	hepatitis A
22	0	0	0	0	0	0	0	0	0	0	0	0	0	Hepatitis B
23	0	0	0	0	0	0	0	0	0	0	0	0	0	Hepatitis C
24	0	0	0	0	0	0	0	0	0	0	0	0	0	Hepatitis D
25	0	0	0	0	0	0	0	0	0	0	0	0	0	Hepatitis E
26	0	0	0	0	0	0	0	0	0	0	0	0	0	Alcoholic hepatitis
27	0	0	0	0	0	0	0	0	0	0	0	0	0	Tuberculosis
28	0	0	0	0	0	0	0	0	0	0	0	0	0	Common Cold
29	0	0	0	0	0	0	0	0	0	0	0	0	0	Pneumonia
30	0	0	0	0	0	0	0	0	0	0	0	0	0	Dimorphic hemmorhoids(piles)
31	0	0	0	0	0	0	0	0	0	0	0	0	0	Heart attack
32	1	0	0	0	0	0	0	0	0	0	0	0	0	Varicose veins
33	0	0	0	0	0	0	0	0	0	0	0	0	0	Hypothyroidism
34	0	0	0	0	0	0	0	0	0	0	0	0	0	Hyperthyroidism
35	0	1	0	0	0	0	0	0	0	0	0	0	0	Hypoglycemia

Testing

Testing Dataset for Disease Prediction Using Machine Learning

Malaria Detection Using Deep Learning

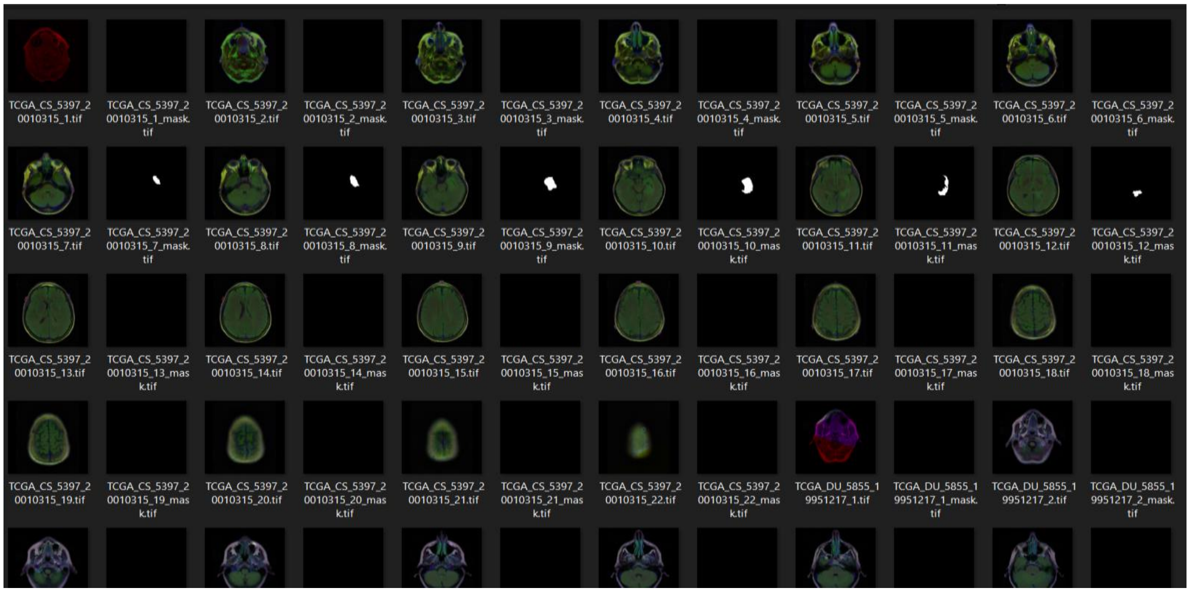
The dataset for Malaria detection has been collected again from Kaggle's website where the dataset contains two folder for cell images concerning stained Red Blood Cells. Infected Cell Images and Uninfected Cell Images are present respectively with the total of 27,558 images. [21]

Testing Dataset for Malaria Detection Using Deep Learning

Brain Tumor Classification and Segmentation Using Deep Learning

The dataset for Brain Tumor Classification and Segmentation has been collected from Kaggle's website where there are two different dataset for classification and segmentation, that in turn also has a collection of various datasets put together to train a model. The classification dataset contains 7022 images of human brain MRI images which are classified into 4 classes: glioma - meningioma - no tumor and pituitary**.** The segmentation dataset contains brain MR images together with manual FLAIR abnormality segmentation masks. The images were obtained from The Cancer Imaging Archive (TCIA). They correspond to 110 patients included in The Cancer Genome Atlas (TCGA) lower-grade glioma collection with at least fluid-attenuated inversion recovery (FLAIR) sequence and genomic cluster data available. [22, 23]

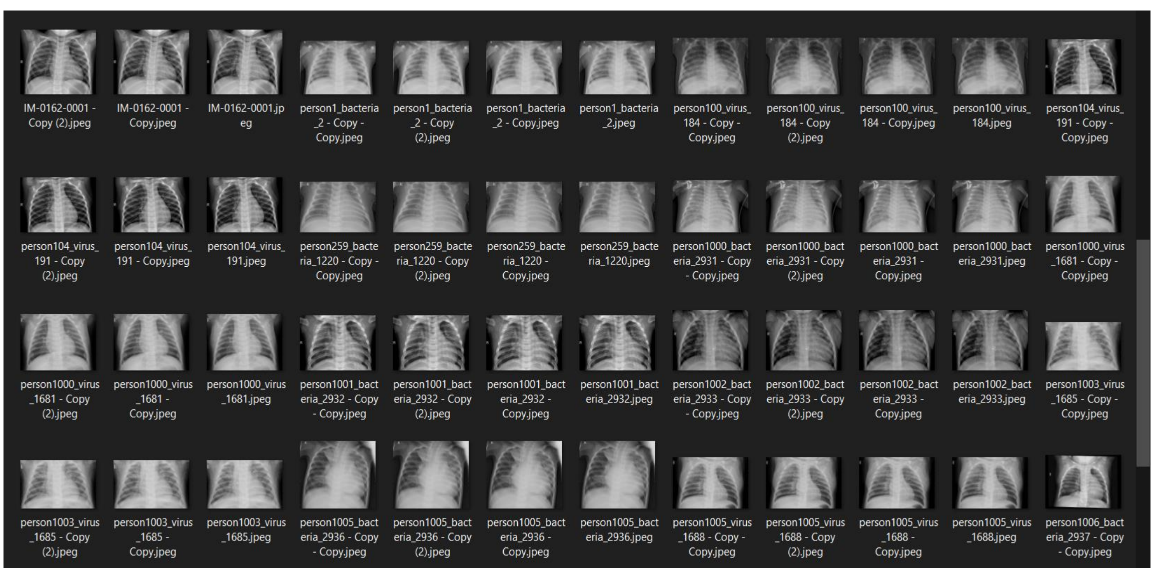
Testing Dataset for Brain Tumor Classification Using Deep Learning



Testing Dataset for Brain Tumor Segmentation Using Deep Learning

Covid-19 and Pneumonia Detection Using Deep Learning

The dataset for Covid-19 and Pneumonia Detection has been collected again from Kaggle's website where the dataset contains two folder for cell images concerning chest X-Ray. There is a collection and labeling of total 5,232 chest X-ray images from children, including 3,883 characterized as depicting pneumonia (2,538 bacterial and 1,345 viral) and 1,349 normal, from a total of 5,856 patients to train the AI system. The model was then tested with 234 normal images and 390 pneumonia images (242 bacterial and 148 viral) from 624 patients. [24]



Testing Dataset for Pneumonia Detection Using Deep Learning

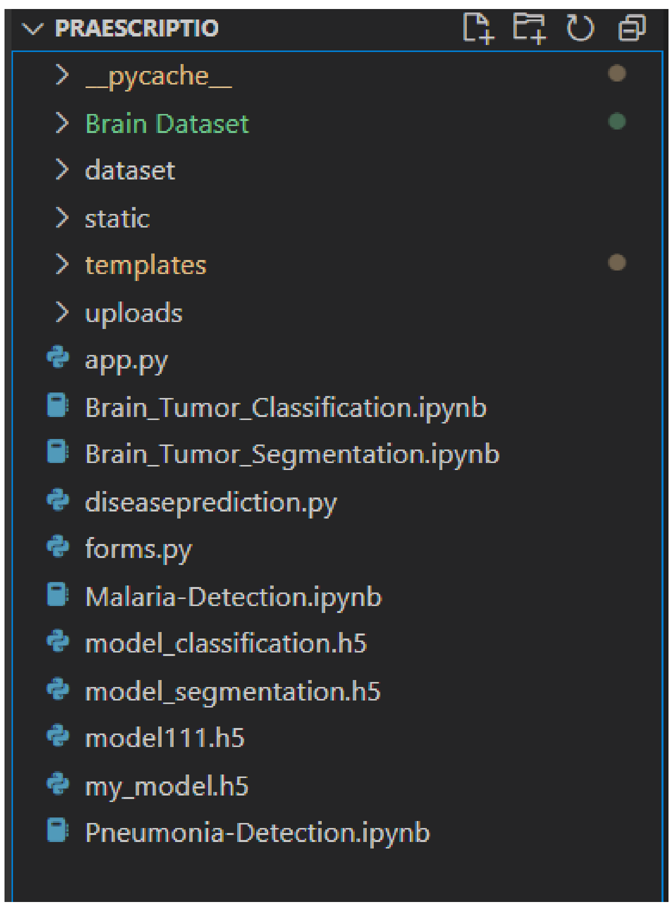
Implementation of Problem

System Specification used to develop the application

1. Laptop Configuration: Windows, core i5 10th generation, 8GB RAM, 1TB Hard disk.
2. Browser: Google Chrome
3. Code Editor: Visual Studio Code
4. Model Datasets: Kaggle Datasets, Excel Spreadsheets
5. Database: SQL Lite
6. Framework: Flask (Python)
7. Models training: Jupyter Notebook
8. Frontend: html, CSS, JavaScript, Bootstrap, Google Maps.

Application's Directory Structure

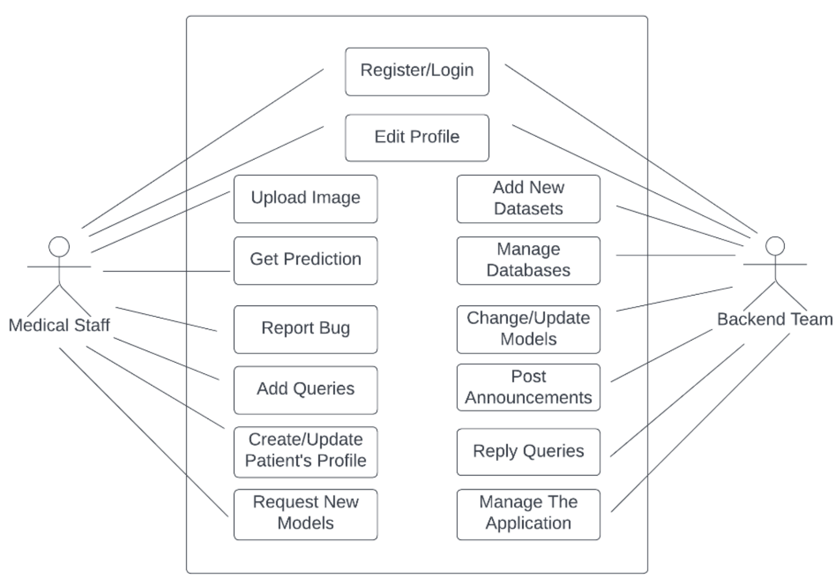
Below there is an attached screen shot of flask application's directory structure along with their respective details:



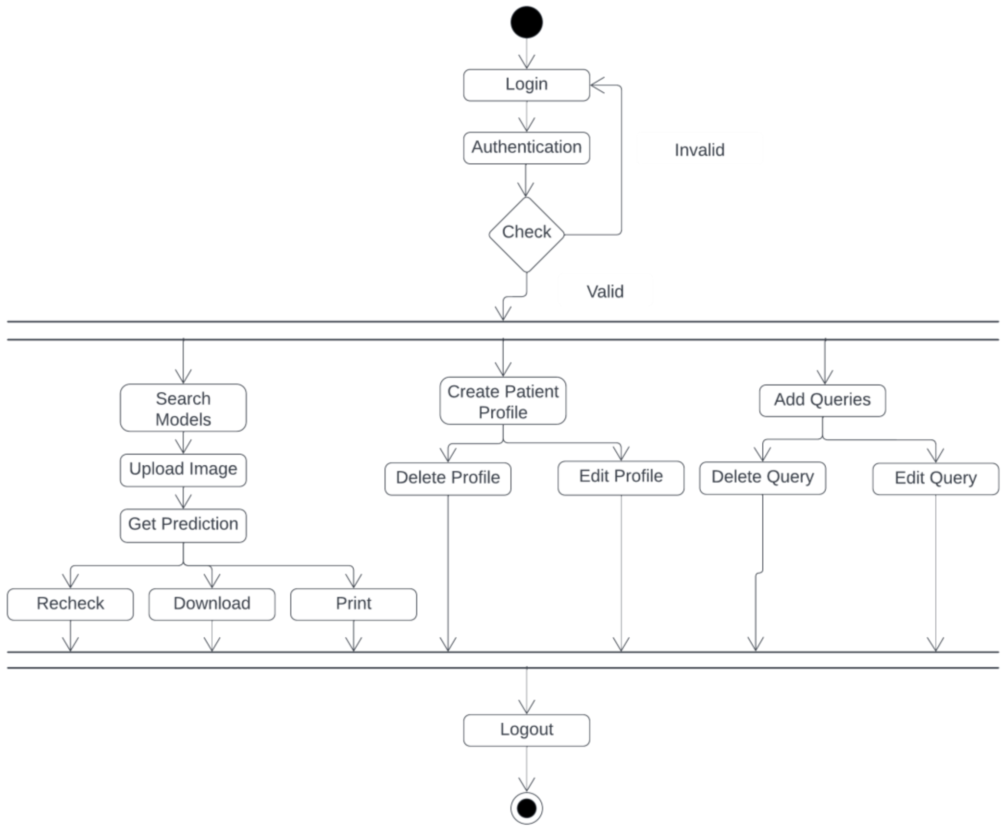
Directory Structure

1. app.py: main flask application where server runs.
2. Brain_Tumor_Classification.ipynb: Jupyter notebook to train Brain Tumor Model
3. Brain_Tumor_Segmentation.ipynb: Jupyter notebook to train Brain Tumor Model
4. Diseasepredictipn.py: Python file to train Common Diseases Model
5. Forms.py: Form for user registration, login and contact
6. Malaria-Detection.ipynb: Jupyter notebook to train Malaria Model
7. model_classifcation.h5: Saved weights of Brain MRI Model
8. model_segmentation.h5: Saved weights of Brain MRI Model
9. model111.h5: Saved weights of Pneumonia-Covid19 Model
10. my_model.h5: Saved weights of Malaria Model
11. Pneumonia-Detection.ipynb: Jupyter notebook to train Pneumonia-Covid19 Model
12. _pycache_: This is the folder where the interpreter compiles python code to byte code first (this is an oversimplification) and stores it.
13. Brain Dataset: Folder having both the classification and segmentation dataset of Brain Tumor.
14. dataset: Folder having the labeled dataset for common disease prediction.
15. static: Folder containing all the CSS, JS, Bootstrap components and Frontend Animations.
16. templates: Folder containing all the html files to run the application on web.
17. uploads: Folder containing all the uploaded images being given by end-users during testing the models.

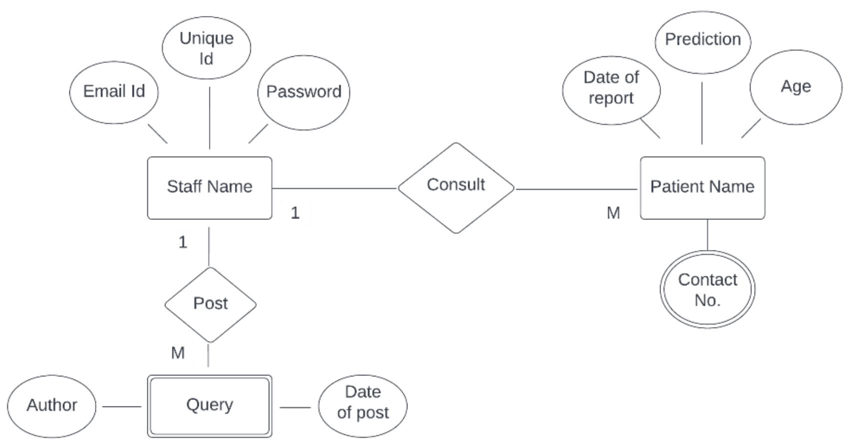
All the models that are trained in this application are mentioned in the Algorithm Section of this report. More information about the outcomes have been stated in the later part of the report.



Use Case Diagram of Praescriptio



Activity Diagram of Praescriptio



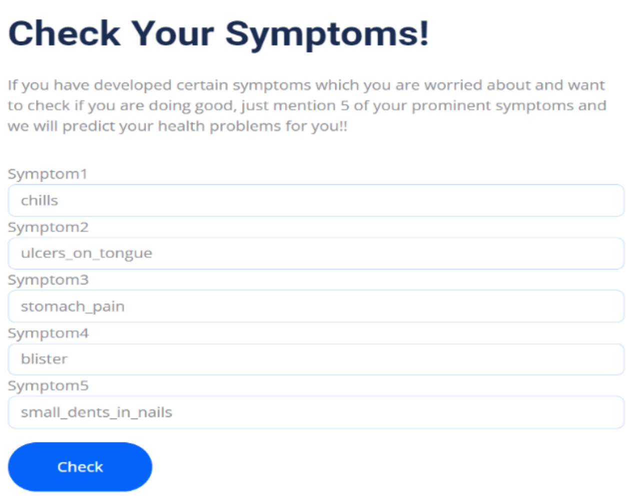
ER Diagram

Sample Output

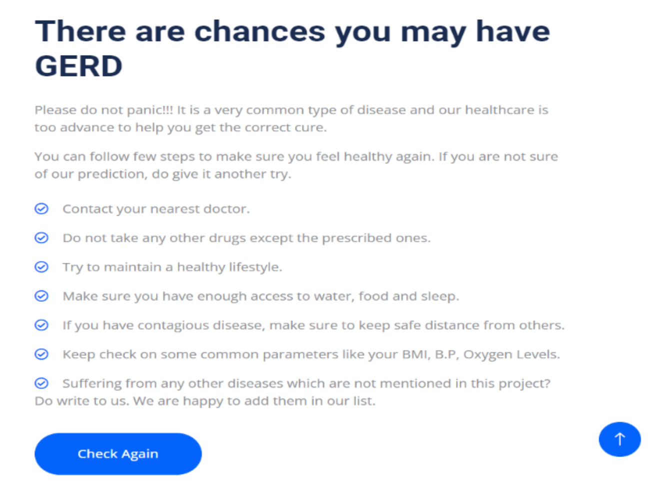
Implementing a research paper's findings and developing a project work out of it is fairly challenging. We need to keep in mind all the edge cases and probable errors one might encounter into along with the proper management of a project pipelines. The harmony of platforms, software and tools used should match with those of system configuration. Also not to forget a proper flow of data across the front-end and the back-end. Even if all these are possible with a disciplined timeframe, deploying such AI based projects inside an already existing business model, making a coexistence possible between humans and machine and generating revenue out of it still remains at the top of discussion. With all these kept in mind, a basic framework of the solution provided in this paper has been shown below as in how this project should look like after its completion.

Disease Prediction Using Machine Learning

Input:

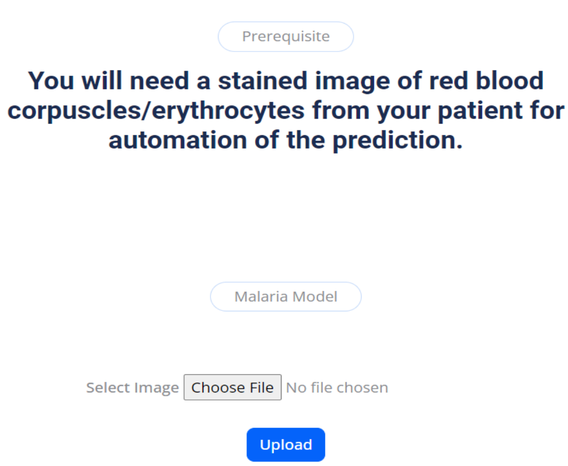


Output:

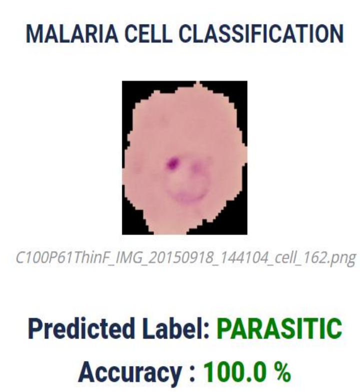


Malaria Detection Using Deep Learning

Input:

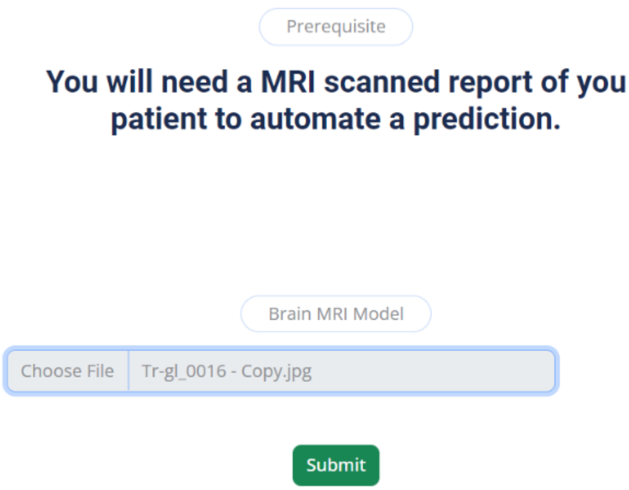


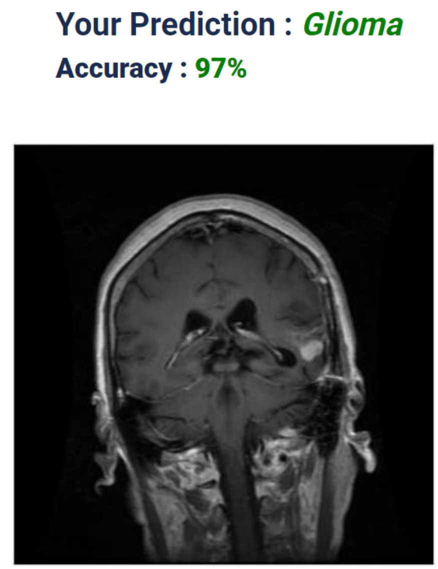
Output:



Brain Tumor Classification and Segmentation Using Deep Learning

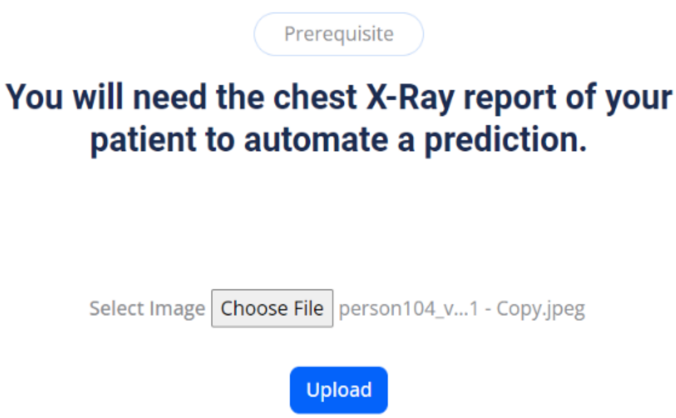
Input: **Output:**

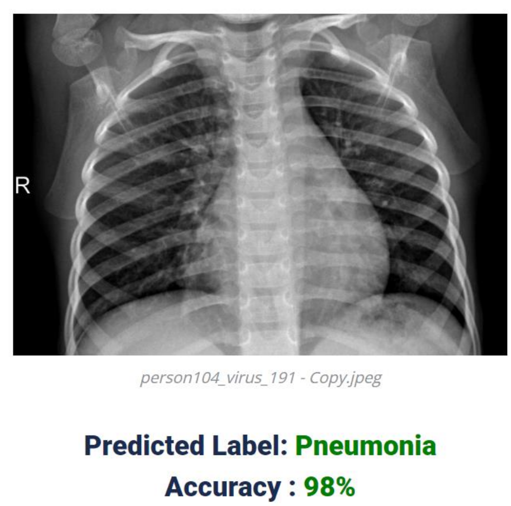




Covid-19 and Pneumonia Detection Using Deep Learning

Input: **Output:**





Conclusion / Future Scope of Work

There are a lot of work we can do in this paper but I have identified few of them which can be considered as a short term development work. Some of them are:

1. Load all the models into any cloud based platform and deploy it so that testers from the healthcare business can provide valuable feedback. More the data we generate after deployment, more will be the accuracy of the models and hence the probability of integrating it into other small businesses with reliable outcome will also increase. This means, the web based application can be made openly available for real hospitals to use as a part of testing phase. Due to the easy to use interface and availability of cloud resources, the initial set up should not be complicated.
2. Include more Medical Imaging Models which can read all types of X-Rays and MRI scans to automate human labor in reading the reports. Also we can include Predictive Machine Learning Models to help doctor better advice their patients and understand their health risks ahead of time.
3. Include Optical Character Recognition to read all the digitally printed or handwritten reports and prescriptions quickly. This will enable us to eliminate the human error which we might incur while generating outputs through this project.
4. Maintain a user management system for each patients inside a particular hospital along with their respective doctors they are consulting with. This will allow to track down all the historical medical data of a given patient and improve the Machine Learning Model's predictions drastically. Their doctors can easily identify when to start which type of treatments for that given patient and considerably reduce the death rate in the hospital.
5. Produce a backend team of Data Scientists who will constantly check on the outliers occurring and take required steps. Also they will ensure a legal and correct use of personal medical data of any given body.
6. Give the produced Big Data from the models and generate a dynamic trend reports of commonly occurring diseases to the pharm industry and help the decide when to release which drug into the market. This will ensure that there is a minimal shortage of any given drug or medical equipment and an all time availability of such basic means will elevate the healthcare industry considerably.

Reference

1. Netscribes (India) Private Limited. April 2009. Report References. [Online]. Available from: https://www.researchandmarkets.com/reports/4787271/artificial-intelligence-ai-in-healthcare-market [Accessed 1 May 2022].
2. Dr. Pooja Rao. May 2018. Article References. [Online]. Available from: https://healthmanagement.org/c/healthmanagement/issuearticle/ai-and-healthcare-technology-in-india-opportunities-challenges-and-emerging-trends [Accessed 1 May 2022].
3. Dinesh C Sharma. December 2015. Report References. [Online]. Available from: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(15)01231-3/fulltext [Accessed 1 May 2022].
4. Shanthi S. May 2021. Article References. [Online]. Available from: https://analyticsindiamag.com/challenges-and-future-of-ai-in-healthcare/ [Accessed 2 May 2022].
5. Professor Paul Chang. December 2015. Article References. [Online]. Available from: https://healthmanagement.org/c/healthmanagement/issuearticle/artificial-intelligence-and-radiology-

human-machine-collaboration-is-key [Accessed 2 May 2022].

6. Geeksforgeeks. February 2018. Article References. [Online]. Available from: https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm [Accessed 3 May 2022].
7. Sivaramakrishnan Rajaraman. January 2018. Article References. [Online]. Available from: https://lhncbc.nlm.nih.gov/LHC-publications/PDF/pub9752.pdf [Accessed 3 May 2022].
8. Karen Simonyan and Andrew Zisserman. April 2015. Conference Paper. [Online]. Available from: https://arxiv.org/pdf/1409.1556.pdf [Accessed 3 May 2022].
9. Dipanjan Sarkar. April 2015. Conference Paper. [Online]. Available from: https://towardsdatascience.com/detecting-malaria-with-deep-learning-9e45c1e34b60 [Accessed 3 May 2022].
10. Karen Simonyan and Andrew Zisserman. April 2015. Conference Paper. [Online]. Available from: https://arxiv.org/pdf/1409.1556.pdf [Accessed 1 May 2022].
11. Viso.ai April 2015. Article References. [Online]. Available from: https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/ [Accessed 1 May 2022].
12. Ahmed, K. B., Hall, L. O., Goldgof, D. B., Liu, R., and Gatenby, R. A. (2017). May 2017. Article References. [Online]. Available from: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.848784/full#B1 [Accessed 4 May 2022].
13. Badza, M. M., and Barjaktarovic. May 2017. Article References. [Online]. Available from: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.848784/full#B1 [Accessed 1 May 2022].
14. Le, W.T.; Maleki, F.; Romero, F.P.; Forghani, R.; Kadoury, S. June 2020. Article References. [Online]. Available from: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.848784/full#B1 [Accessed 4 May 2022].
15. Ronneberger, O.; Fischer, P. January 2015. Article References. [Online]. Available from: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.848784/full#B1 [Accessed 4 May 2022].
16. Milletari, F.; Navab, N.; Ahmadi, S.A. March 2016. Article References. [Online]. Available from: https://www.mdpi.com/2079-9292/10/13/1512/pdf [Accessed 4 May 2022].
17. Dejun Zhang, Fuquan Ren, and Yue Ma. September 2019. Article References. [Online]. Available from: https://www.mdpi.com/2079-9292/10/13/1512/pdf [Accessed 4 May 2022].
18. Vaibhav Sharma. May 2019. Article References. [Online]. Available from: https://www.pluralsight.com/guides/deep-learning-model-perform-binary-classification [Accessed 4 May 2022].
19. Asquiro. June 2020. Article References. [Online]. Available from: https://www.asquero.com/article/advantages-and-disadvantages-of-artificial-neural-networks/ [Accessed 4 May 2022].
20. https://www.asquero.com/article/advantages-and-disadvantages-of-artificial-neural-networks/

21. KAUSHIL268. December 2020. Kaggle Dataset. [Online]. Available from: https://www.kaggle.com/datasets/kaushil268/disease-prediction-using-machine-learning?select=Testing.csv [Accessed 2 May 2022].
22. ARUNAVA. December 2020. Kaggle Dataset. [Online]. Available from: https://www.kaggle.com/datasets/iarunava/cell-images-for-detecting-malaria [Accessed 2 May 2022].
23. SHOBHIT SHRIVASTAVA. November 2021. Kaggle Dataset. [Online]. Available from: https://www.kaggle.com/code/shobhit18th/keras-nn-x-ray-predict-pneumonia-86-54/data [Accessed 2 May 2022].
24. MASOUD NICKPARVAR. July 2021. Kaggle Dataset. [Online]. Available from: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset [Accessed 2 May 2022].
25. MATEUSZ BUDA. August 2019. Kaggle Dataset. [Online]. Available from: https://www.kaggle.com/datasets/mateuszbuda/lgg-mri-segmentation [Accessed 2 May 2022].