

MOBILE APP FOR DISEASE PREDICTION AND MANAGEMENT

Sreeparvathy S Menon

1st June 2023

Abstract

This report serves as a comprehensive guide for the development and deployment of a disease prediction and management mobile app. India's digital transformation and notoriously cumbersome and stressed healthcare system this app is the need of the hour. The report examines various customer needs and emphasizes the various challenges in the development and deployment of the app. It highlights the significance of a huge dataset for accurate prediction, data privacy, regulatory compliance, collaboration with healthcare professionals, and effective deployment strategies. It discusses the chances of acquiring a patent for this technology. The report compares other competitors that are already present in the market space. It also touches upon the various steps required in the product development process. A sample model using a Kaggle dataset was created to check the feasibility of the product and accuracy was determined.

1.0 Introduction

Today, India is the most populated country in the world. With the rising population, the responsibility of creating a new generation of healthy Indians lies with the health sector. India has made unprecedented progress in the health care sector but challenges remain. The rise in obesity and lifestyle risk factors has led to an increase in non-communicable diseases (NCDs) like diabetes, cardiovascular diseases, and cancer. Communicable diseases like tuberculosis, malaria, and dengue though decreased, still pose a threat and with new health scares like COVID 19 India's healthcare system is in dire need of a boost.

Accessibility is also a major issue. 30% of Indians don't have health insurance and have to pay out of pocket. Most doctors are present in urban areas and are unwilling to serve in rural areas. Only 31.5% of hospitals and 16% of hospital beds are situated in rural areas where 75% of the total population resides. Due to the above-mentioned factors, healthcare is not very accessible to the poor.

COVID-19 has sped up digitalization by several years. People are more and more reliant on digital technologies day by day. Digitalization of healthcare has already started with government apps like Arogya Setu which controlled the COVID vaccination. It is just a matter of time for this digitalization to spread to other areas of healthcare.

This product aims to conduct the initial diagnosis for any patient that has access to an internet connection. A mobile development app that could predict the disease with symptoms and general information about the patient. This would decrease outpatient doctor consultation costs. It would also decrease the stress for doctors and cases with healthy diagnoses need not come to the hospital. As no fee is required, the poorest of Indians could use this app. People also can skip traveling for a diagnosis. People tend to be negligent of earlier symptoms and would visit the hospital only after symptoms get more dire. An app would help to bridge this problem and provide the diagnosis in the earlier stages of the disease.

After diagnosis, the app would work as a health app, having medicine reminders, exercise trackers, and food trackers. Medical records could be inputted and different health goals could be set accordingly to stop the progression of disease and hopefully eliminate it. The product proposes to only provide an initial diagnosis, work with doctors and patients, and complement our healthcare system and not replace it.

2.0 Customer Needs Assessment

This section describes the various needs of the customers. It is important to understand the perspective of the customers to sustain a business in a competitive market like India.

- Accuracy: Prediction should be accurate
- Affordable: The product is also meant for people who do not have the time or money to spend.
- Easy to use: Account creation should be easy and usage of the app should be intuitive and user friendly. Uploading different data and symptoms should be efficient, taking less amount of time.
- Security and Privacy: Data, especially medical records should be protected against threats. Multilayer protection with encryption is required. Accounts shouldn't be hackable. Data leak needs to be avoided at all costs.
- Localization: The app must be accessible and user-friendly for people who may not be proficient in English. Health Plans should be created keeping the Indian audience in mind.
- Collaboration with healthcare professionals
- Clean UI UX: The app shouldn't be very cluttered and needs to be easy to read.
- App maintenance: The app should be maintained and updated regularly

3.0 External Search

Researchers have devised various AI-based techniques such as machine and deep learning models to detect the diseases in skin, liver, heart, etc. Numerous studies were able to predict the risk certain individual possess to diseases like diabetes, heart disease, Alzheimer's and cancer. Researchers through image recognition techniques were able to identify the progression of cancer. Techniques like Naive Bayes Algorithm, K nearest neighbour (kNN), support vector machine (SVM), decision tree, random forest, logistic regression, and artificial neural network were used to diagnose the diseases with great accuracy.

Kaggle data set is used to make a sample model

Kaggle data set link: <https://www.kaggle.com/datasets/kaushil268/disease-prediction-using-machine-learning>

Two CSV files are available: Training and Testing. Each CSV file has 133 columns. 132 of these columns are symptoms that a person experiences and the last column is the prognosis. These symptoms are mapped to 42 diseases you can classify these set of symptoms to. The first five rows of training and testing files are shown.

Training.csv

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain	acidity	ulcers_on_tongue	...
0	1	1	1	0	0	0	0	0	0	0	...
1	0	1	1	0	0	0	0	0	0	0	...
2	1	0	1	0	0	0	0	0	0	0	...
3	1	1	0	0	0	0	0	0	0	0	...
4	1	1	1	0	0	0	0	0	0	0	...

...	scurrying	skin_peeling	silver_like_dusting	small_dents_in_nails	inflammatory_nails	blister	red_sore_around_nose	yellow_crust_ooze	prognosis	Unnamed: 133
...	0	0	0	0	0	0	0	0	Fungal infection	NaN
...	0	0	0	0	0	0	0	0	Fungal infection	NaN
...	0	0	0	0	0	0	0	0	Fungal infection	NaN
...	0	0	0	0	0	0	0	0	Fungal infection	NaN
...	0	0	0	0	0	0	0	0	Fungal infection	NaN

Testing.csv

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain	acidity	ulcers_on_tongue	...
0	1	1	1	0	0	0	0	0	0	0	...
1	0	0	0	1	1	1	0	0	0	0	...
2	0	0	0	0	0	0	0	1	1	1	...
3	1	0	0	0	0	0	0	0	0	0	...
4	1	1	0	0	0	0	0	1	0	0	...

...	blackheads	scurrying	skin_peeling	silver_like_dusting	small_dents_in_nails	inflammatory_nails	blister	red_sore_around_nose	yellow_crust_ooze	prognosis
...	0	0	0	0	0	0	0	0	0	Fungal infection
...	0	0	0	0	0	0	0	0	0	Allergy
...	0	0	0	0	0	0	0	0	0	GERD
...	0	0	0	0	0	0	0	0	0	Chronic cholestasis
...	0	0	0	0	0	0	0	0	0	Drug Reaction

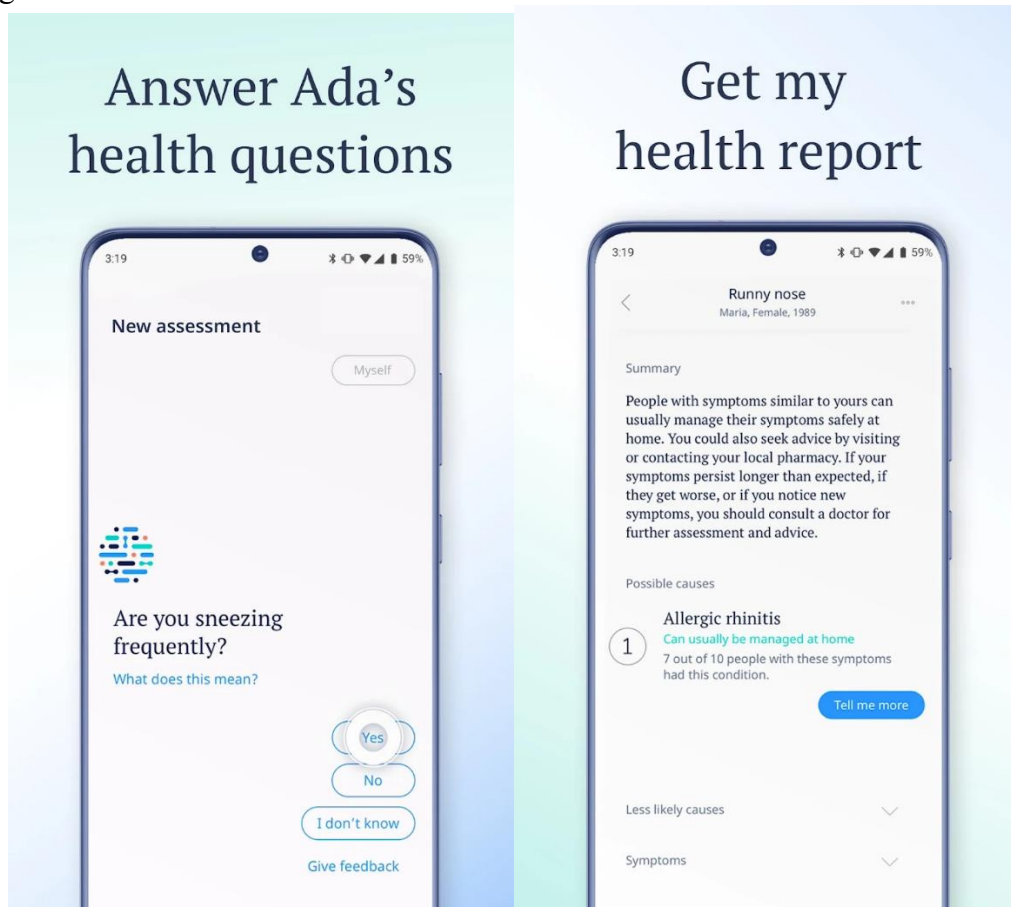
3.1 Benchmarking

Currently, there are not many companies that provide all the features our product is addressing. This gives us a major first-mover advantage as the market space is free.

Few commercially available products slightly entertain these ideas but not have fully materialized. This section identifies products that have few features and partially implemented our product. 3 products are compared: Ada Health predicts diseases based on symptoms entered by the user while SickHealth collects data from wearables but only predicts weakness but not particular diseases. HealthifyMe although not a disease prediction app, tracks health goals and curate diet plans by taking into account different health conditions of the user. Ada Health and SickHealth haven't found a huge user base in India, but HealthifyMe is very popular in India.

1. Ada Health:

Ada Health is a free mobile application based in Berlin. Ada collects data using a health assessment. The app analyzes the data received and generates a report on what can be causing these symptoms as well as gives a suggestion on what action you should take. It also tracks the severity of the symptoms. Ada Health has 10 million users worldwide and 25 million completed assessments. Ada Health relies only on data acquired through question air for diagnosis.



2. SickPredict App:

SickPredict App calculates personalized SickNumber, based on the specific health metrics from different wearables like apple watch, fit bit, etc., SickNumber helps to anticipate any

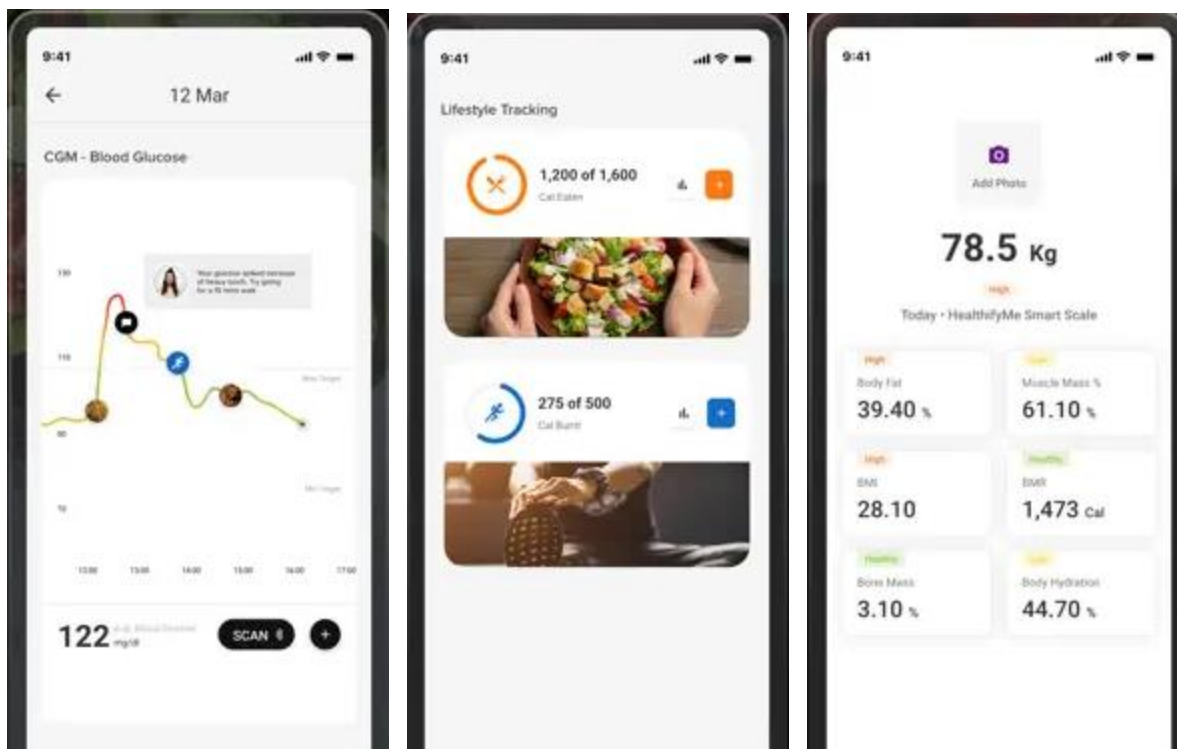
upcoming changes in your health in advance of experiencing symptoms. The app is only available on Appstore which is a disadvantage as Android phones are mostly used in India. Also, it only predicts the upcoming changes in health, not a particular disease.



3. HealthifyMe:

HealthifyMe is not a disease prediction app but it contains the health goal aspect of our product. It is a fitness-tracking app that tracks calorie intake and exercise to achieve different health goals. HealthifyMe provides features to help reach different health & fitness goals. It has a food tracker, workout tracker, water tracker, sleep tracker, weight loss tracker & handwash tracker.

HealthifyMe creates a diet chart and meal planner from health data and BMI. The specialized diet plan is carefully curated to easily manage dietary health conditions (diabetes, thyroid, PCOS, cholesterol, hypertension) & help the overall immune system fight viral & bacterial infections. HealthifyMe was specifically created for the Indian audience leading to its wide popularity in India with over 10 million downloads.



3.2 Applicable Patents

In India, inventions claiming diagnostic methods are not patentable. The act does not encompass detection tests, e.g., a screening test, wherein the primary purpose is early detection or risk factors of a disease which may never lead to a disease and are merely susceptibility markers in otherwise healthy individuals as well as prediction methods for progression rather than diagnosis in already diagnosed diseased individuals, since it has no bearing on developing methods for further course in therapy. Hence, detection tests, e.g. for screening, and prognostic methods for assessment of disease progression are fundamentally different from diagnostic methods of treatment.

Our product is a detection test under this definition and is patentable. A similar patent that is available is: Application No.202111026966 A: A HYBRID FEATURE SELECTION APPROACH FOR PARKINSONTMS DISEASE DIAGNOSIS BASED ON MUTUAL INFORMATION GAIN AND RECURSIVE FEATURE ELIMINATION

3.3 Applicable Standards

The product should comply with the following government regulations:

1. Drugs and Cosmetics Act, 1940: If the disease prediction app qualifies as a medical device under the definition provided by the Medical Device Rules, it may need to comply with the regulations specified under the Drugs and Cosmetics Act, 1940.
2. Telemedicine Guidelines: The app needs to follow the Telemedicine Practice Guidelines issued by the Ministry of Health and Family Welfare. It includes requirements for patient consent, privacy, and maintaining medical records.
3. Information Technology (IT) Act, 2000: The IT Act is a comprehensive legislation that governs electronic transactions, data protection, and cybersecurity in India.
4. Personal Data Protection Bill, 2022: The Bill will apply to the processing of digital personal data within India where such data is collected online, or collected offline and is digitized.
5. Electronic Health Records (EHR) Standards: By The Ministry of Health and Family Welfare.
6. Advertising Standards: Advertisements must comply with the ASCI (Advertising Standards Council of India) code.

3.4 Applicable Constraints

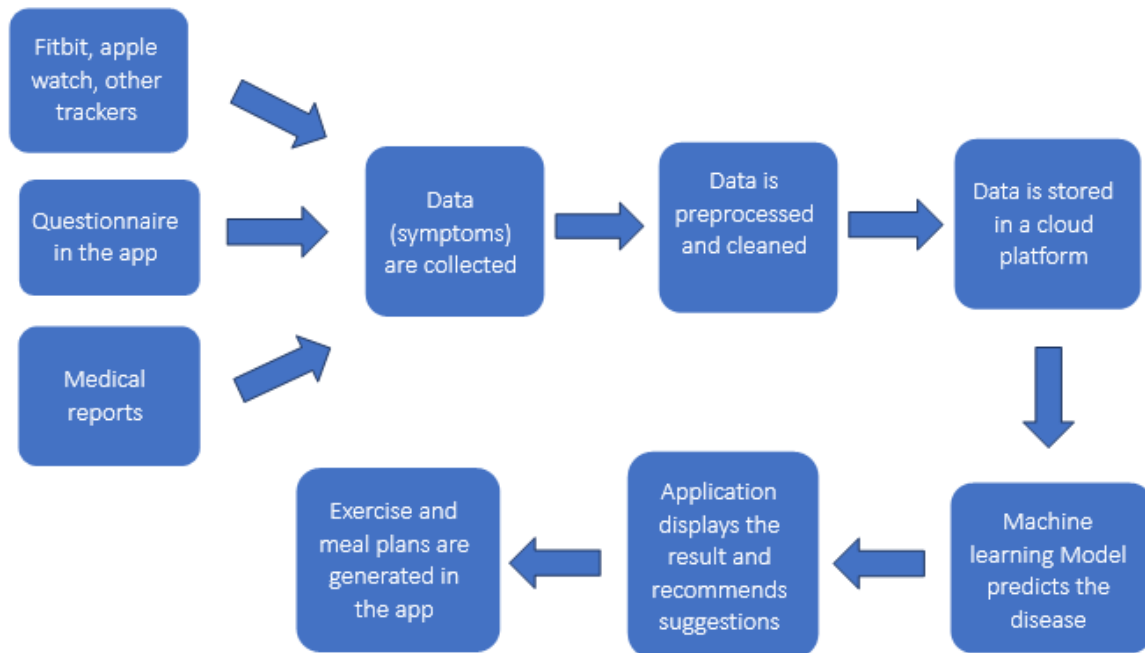
- The algorithm necessitates a huge amount of data for good accuracy. A Large scale study must be conducted to create a robust algorithm that can accurately predict diseases.
- App development and constant maintenance of the app require additional costs
- App deployment in Playstore and Appstore: Compliance with the specific guidelines and requirements provided by Google and Apple
- Cloud platforms to store the data: A reliable cloud service provider is required to store data. Some popular options include Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure.
- Compliance with Government laws and regulations

4.0 Concept Generation

This section describes the processes used to generate the app. A sample machine-learning algorithm is also given.

4.1 Problem Clarification

The various tasks to be done by the product are given below in terms of a flow chart:



Data is collected from three different sources: 1) Different wearables like apple watch, Fitbit, etc, 2) a questionnaire created by medical professionals, and 3) Medical reports like blood test results.

To check the feasibility of the system, a sample model is made using the Kaggle dataset. Random Forest Classifier is used to fit the data.

5.0 Sample Machine Learning Model

Using the Kaggle data set, a Random forest classifier is used to accurately predict the disease through symptoms. In the data set 132 different symptoms are used to predict 42 diseases. Discuss details of the design refinement process and the final detailed design.

Code Link: <https://github.com/Sreeparvathy-S-Menon/Disease-prediction>

5.1 Data Pre-processing and visualization

Data with any blank values need to be removed. As 'Unnamed: 133' column has many NaN values it is being dropped. Data is also rechecked for any other NaN values.

```
df.drop('Unnamed: 133', axis=1, inplace=True)
df.head()
```

...	blackheads	scurring	skin_peeling	silver_like_dusting	small_dents_in_nails	inflammatory_nails	blister	red_sore_around_nose	yellow_crust_ooze	prognosis
...	0	0	0	0	0	0	0	0	0	Fungal infection
...	0	0	0	0	0	0	0	0	0	Fungal infection
...	0	0	0	0	0	0	0	0	0	Fungal infection
...	0	0	0	0	0	0	0	0	0	Fungal infection
...	0	0	0	0	0	0	0	0	0	Fungal infection

```
df.isna().sum()
```

```
itching          0
skin_rash        0
nodal_skin_eruptions  0
continuous_sneezing  0
shivering        0
..
inflammatory_nails  0
blister           0
red_sore_around_nose  0
yellow_crust_ooze  0
prognosis         0
Length: 133, dtype: int64
```

There are no more NaN values. Data should have the same amount of samples for each disease-should be balanced.

```
5]: Fungal infection          120
    Hepatitis C              120
    Hepatitis E              120
    Alcoholic hepatitis      120
    Tuberculosis             120
    Common Cold              120
    Pneumonia                120
    Dimorphic hemorrhoids(piles) 120
    Heart attack              120
    Varicose veins           120
    Hypothyroidism           120
    Hyperthyroidism          120
    Hypoglycemia             120
    Osteoarthritis           120
    Arthritis                120
    (vertigo) Paroymsal  Positional Vertigo 120
    Acne                     120
    Urinary tract infection  120
    Psoriasis                120
    Hepatitis D              120
    Hepatitis B              120
    Allergy                  120
    hepatitis A              120
    GERD                     120
    Chronic cholestasis      120
    Drug Reaction            120
    Peptic ulcer disease     120
    AIDS                     120
    Diabetes                 120
    Gastroenteritis          120
    Bronchial Asthma         120
    Hypertension             120
    Migraine                 120
    Cervical spondylosis     120
    Paralysis (brain hemorrhage) 120
    Jaundice                 120
    Malaria                  120
    Chicken pox              120
    Dengue                   120
    Typhoid                  120
    Impetigo                 120
    Name: prognosis, dtype: int64
```

The correlation matrix is also obtained.

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain	acidity	ulcers_on_tongue	muscle_wast
itching	1.00	0.32	0.33	-0.09	-0.06	-0.18	-0.16	0.20	-0.09	-0.06	-0.06
skin_rash	0.32	1.00	0.30	-0.09	-0.07	-0.03	0.17	0.16	-0.09	-0.07	-0.07
nodal_skin_eruptions	0.33	0.30	1.00	-0.03	-0.02	-0.07	-0.06	-0.03	-0.03	-0.02	-0.02
continuous_sneezing	-0.09	-0.09	-0.03	1.00	0.61	0.45	-0.09	-0.05	-0.05	-0.03	-0.03
shivering	-0.06	-0.07	-0.02	0.61	1.00	0.30	-0.06	-0.03	-0.03	-0.02	-0.02
chills	-0.18	-0.03	-0.07	0.45	0.30	1.00	-0.09	-0.10	-0.10	-0.07	-0.07
joint_pain	-0.16	0.17	-0.06	-0.09	-0.06	-0.09	1.00	-0.09	-0.09	-0.06	-0.06
stomach_pain	0.20	0.16	-0.03	-0.05	-0.03	-0.10	-0.09	1.00	0.43	0.65	0.65
acidity	-0.09	-0.09	-0.03	-0.05	-0.03	-0.10	-0.09	0.43	1.00	0.61	0.61
ulcers_on_tongue	-0.06	-0.07	-0.02	-0.03	-0.02	-0.07	-0.06	0.65	0.61	1.00	1.00
muscle_wast	-0.06	-0.07	-0.02	-0.03	-0.02	-0.07	-0.06	0.65	0.61	1.00	1.00

5.2 Creation of model

The Random Forest Classifier algorithm with Sickit Learn is used to create the model. A random forest fits several decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.

The following is the code to split the data into train and test sets, fit the data into the model, and find the accuracy of the model created.

```
x_train = df.drop('prognosis', axis = 1)
y_train = df['prognosis']
x_test = testset.drop('prognosis', axis = 1)
y_test = testset['prognosis']

rfc= RandomForestClassifier()
rfc.fit(x_train,y_train)
pred = rfc.predict(x_test)
acc = rfc.score(x_test, y_test)

print("Accuracy on test set: {:.2f}%".format(acc*100))
```

Accuracy on test set: 97.62%

The accuracy of the model was found out to be 97.62%.

6.0 Conclusions

With the Digitalization of healthcare and the emergence of machine learning, a good business environment has been created for healthcare apps. A disease prediction and management app would be a successful addition to the current market landscape.

Indian healthcare has been under stress due to the huge influx of patients. Any product that can decrease this stress would receive a warm welcome. A disease prediction app would be a huge time and money saver for the people of India. It would reduce the number of hospital visitations.

It also could lead to earlier detection of diseases, saving many lives. The disease management aspect of the app could complement the health system very well.

There are multiple challenges to the creation of this product. A huge dataset is required for accurate prognoses. Privacy and security should be the topmost priority. A data breach could taint the reputation of the app. Many government norms and regulations are there to be followed. As there is a large amount of data, usage of a cloud platform is necessary.

There are not many apps that have all the features of our app. The few competitors that have some of the features proposed have been examined and their pros and cons discussed.

Using a Kaggle data set, a sample machine-learning model was made with Random Forest Classifier and Scikit-Learn. The data was analyzed, cleaned, preprocessed, checked for balancing of different prognoses, and fit into the model. The model obtained a very high accuracy of 97.62%.

7.0 References

<https://www.economicdiscussion.net/articles/7-major-problems-of-health-services-in-india/2305>

<https://indianexpress.com/article/opinion/columns/healthcare-in-india-has-made-great-progress-but-challenges-remain-8524485/>

<https://www.ijraset.com/research-paper/disease-predictor-based-on-symptoms-using-machine-learning>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8754556/>

<https://ipindia.gov.in/>

https://ipindia.gov.in/writereaddata/Portal/IPOGuidelinesManuals/1_37_1_3-guidelines-for-examination-of-patent-applications-pharmaceutical.pdf