A Mini Project Report on

MedCheck

(a machine learning based mobile application)

Submitted in partial fulfilment for the requirements of

B. E. (CSE) V Semester Mini Project

in

COMPUTER SCIENCE AND ENGINEERING

by

Manideep Kalyanam (160118733102)

Sairaj Gurram (160118733113)

Shanmukha Manoj Bharadwaj Korrapati (160118733115)

Under the guidance of

Mr. B Sateesh

Assistant Professor, Department of CSE, CBIT.

Department of Computer Science and Engineering

Chaitanya Bharathi Institute of Technology (Autonomous),

(Affiliated to Osmania University, Hyderabad)

Hyderabad, Telangana (India) – 500 075

CHAITANYA BHARATHI
INSTITUTE OF TECHNOLOGY (A)
Affliated to Osmania University

CERTIFICATE

This is to certify that the project entitled “**MedCheck (a machine learning based mobile application)”**, submitted to the Computer Science and Engineering Department, Chaitanya Bharathi Institute of Technology, in partial fulfilment of the requirement for the course Mini Project, is a bonafide record of work done by **Manideep Kalyanam (160118733102) , Sairaj Gurram (160118733113)** and **Shanmukha Manoj Bharadwaj Korrapati (160118733115)** , from August, 2020 to November 2020 under our guidance and supervision.

Mentors,

Ms. Ch Vijaya Lakshmi
Assistant Professor,
Department of CSE,
CBIT

Mr. B Sateesh
Assistant Professor,
Department of CSE,
CBIT

Ms. G Kavita
Assistant Professor,
Department of CSE,
CBIT

Supervisor,

Mr. B Sateesh
Assistant Professor,
Department of CSE,
CBIT

ABSTRACT

The main focus is on to use machine learning in healthcare to supplement patient care for better results. Machine learning has made easier to identify different diseases and diagnosis correctly. Predictive analysis with the help of efficient multiple machine learning algorithms helps to predict the disease more correctly and help treat patients.

The healthcare industry produces large amounts of healthcare data daily that can be used to extract information for predicting disease that can happen to a patient in future while using the treatment history and health data. This hidden information in the healthcare data will be later used for affective decision making for patient's health.

MedCheck, based on predictive modelling predicts the disease of the user on the basis of the symptoms that user provides as an input to the system. The system analyses the symptoms provided by the user as input and gives the predicted disease as output. Disease prediction is done by implementing various machine learning algorithms.

The expected outcome and scope of this project is that if disease can be predicted then early treatment can be given to the patients which can reduce the risk of life and save life of patients and cost to get treatment of diseases can be reduced up to some extent by early recognition.

ACKNOWLEDGEMENT

Firstly, we are grateful to Chaitanya Bharathi Institute of Technology for giving us this opportunity to do a Mini project. We acknowledge our sincere gratitude to our project guide Mr. B Sateesh (Assistant Professor, Dept. of CSE, CBIT) for providing us his valuable assistance for doing this project. We are grateful to our mentors Ms. Ch Vijaya Lakshmi (Assistant Professor, Dept. of CSE, CBIT), Mr. B Sateesh (Assistant Professor, Dept. of CSE, CBIT) and Ms. G Kavita (Assistant Professor, Dept. of CSE, CBIT) for their valuable guidance and support even in their very busy schedule. Finally, we thank each of our teammates for working hard for all the days and completing this project successfully.

INTRODUCTION

At present, when one suffers from particular disease, then the person has to visit to doctor which is time consuming and costly too. Also, if the user is out of reach of doctor and hospitals it may be difficult for the user as the disease cannot be identified. So, if the above process can be completed using an automated program which can save time as well as money, it could be easier to the patient which can make the process easier.

Machine learning is programming computers to optimize a performance using example data or past data. Machine learning is study of computer systems that learn from data and experience. Machine learning algorithm has two passes: Training, Testing. Prediction of a disease by using patient's symptoms is one of the areas machine learning technology is struggling from past decades. Machine Learning technology gives a good platform in medical field, so that healthcare issues can be solved efficiently.

Thanks to mobile technology, we have everything in the palm of our hand these days. Smartphones in combination with app are providing people with extensive information in less than a second. Today we have apps that can help us shop, order food, order a cab and much more in a blink of an eye. These apps are not just available but are being used by millions all over the globe. In this context, a healthcare app like disease predictor seems to be the most logical and convenient development.

MedCheck is an android application which focuses on the healthcare of the user. It provides the user with medication, diagnosis, health tips and suggesting the best hospitals in various specializations. It also has the facility of calling emergency services of that particular geographical location. The other salient feature of this application is to predict the disease of the user based on the symptoms given by the user.

Technologies used

Machine learning:

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

Some machine learning methods:

Machine learning algorithms are often categorized as supervised or unsupervised.

- **Supervised machine learning algorithms** can apply what has been learned in the past to new data using labelled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

- **Unsupervised machine learning algorithms** are used when the information used to train is neither classified nor labelled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabelled data. The system doesn’t figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabelled data.

- **Semi-supervised machine learning algorithms** fall somewhere in between supervised and unsupervised learning, since they use both labelled and unlabelled data for training – typically a small amount of labelled data and a large amount of unlabelled data. The systems that use this method are able to considerably improve learning accuracy. Usually, semi-supervised learning is chosen when the acquired labelled data requires skilled and relevant resources in order to train it / learn from it. Otherwise, acquiring unlabelled data generally doesn’t require additional resources.

Neural Networks:

A neural network is a type of machine learning which models itself after the human brain, creating an artificial neural network that via an algorithm allows the computer to learn by incorporating new data.

While there are plenty of artificial intelligence algorithms these days, neural networks are able to perform what has been termed deep learning. While the basic unit of the brain is the neuron, the essential building block of an artificial neural network is a perceptron which accomplishes simple signal processing, and these are then connected into a large mesh network.

The computer with the neural network is taught to do a task by having it analyse training examples, which have been previously labelled in advance. A common example of a task for a neural network using deep learning is an object recognition task, where the neural network is presented with a large number of objects of a certain type, such as a cat, or a street sign, and the computer, by analysing the recurring patterns in the presented images, learns to categorize new images.

Android Studio:

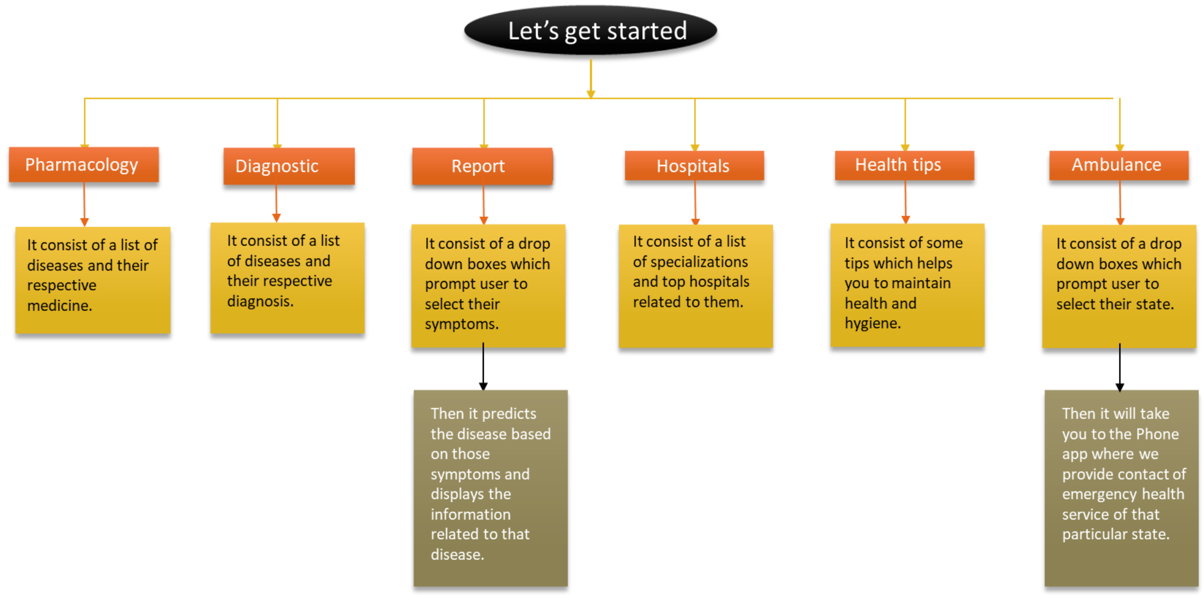
Android Studio is the official integrated development environment (IDE) for Android application development. It is based on the IntelliJ IDEA, a Java integrated development environment for software, and incorporates its code editing and developer tools.

To support application development within the Android operating system, Android Studio uses a Gradle-based build system, emulator, code templates, and GitHub integration. Every project in Android Studio has one or more modalities with source code and resource files. These modalities include Android app modules, Library modules, and Google App Engine modules.

Android Studio uses an Instant Push feature to push code and resource changes to a running application. A code editor assists the developer with writing code and offering code completion, refraction, and analysis. Applications built in Android Studio are then compiled into the APK format for submission to the Google Play Store.

The software was first announced at Google I/O in May 2013, and the first stable build was released in December 2014. Android Studio is available for Mac, Windows, and Linux desktop platforms. It replaced Eclipse Android Development Tools (ADT) as the primary IDE for Android application development. Android Studio and the Software Development Kit can be downloaded directly from Google.

BLOCK DIAGRAM


Let's get started
Pharmacology
Diagnostic
Report
Hospitals
Health tips
Ambulance
It consist of a list of diseases and their respective medicine.
It consist of a list of diseases and their respective diagnosis.
It consist of a drop down boxes which prompt user to select their symptoms.
It consist of a list of specializations and top hospitals related to them.
It consist of some tips which helps you to maintain health and hygiene.
It consist of a drop down boxes which prompt user to select their state.
Then it predicts the disease based on those symptoms and displays the information related to that disease.
Then it will take you to the Phone app where we provide contact of emergency health service of that particular state.

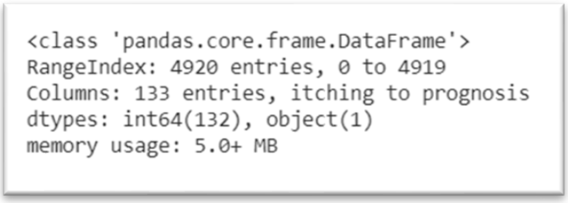
This block diagram shows workflow of the mobile application. On the home page, we have a button, Let's get started which direct us to the next screen on clicking it. In the next screen, we have six image buttons- Pharmacology, Diagnostic, Report, Hospitals, Health tips and Ambulance. Description for these buttons are mentioned in the block diagram. The application of machine learning can be observed in Report, where it asks the user to give their symptoms and predicts the disease based on them.

Results and Discussions

Dataset Description:

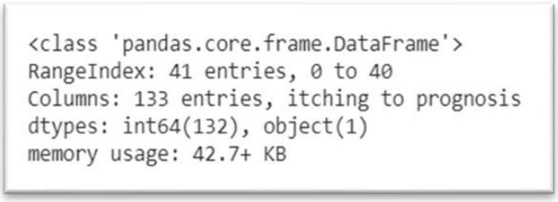
Link for data: https://www.kaggle.com/rabisingh/symptom-checker

We will get two different datasets from the source. One is the Train_set and another one is Test_set. The information of the two datasets are given below:



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4920 entries, 0 to 4919
Columns: 133 entries, itching to prognosis
dtypes: int64(132), object(1)
memory usage: 5.0+ MB
```

(i) Train_set



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 41 entries, 0 to 40
Columns: 133 entries, itching to prognosis
dtypes: int64(132), object(1)
memory usage: 42.7+ KB
```

(ii) Test_set

The datasets consist of 132 symptoms and one prognosis as shown in the above figure. Train_set consists of 4920 rows and 133 columns. Test_set consists of 41 rows and 133 columns. The dataset consists of no null values.

We trained our model using some classification Machine learning algorithms. They are Logistic regression, Support vector machine and Neural network.

Logistic Regression:

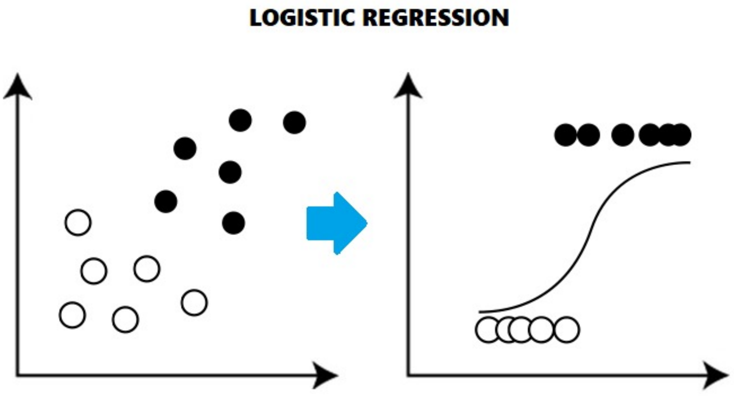
Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.

Logistic regression predicts the output of a categorical dependent variable. Therefore, the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, True or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.

In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).

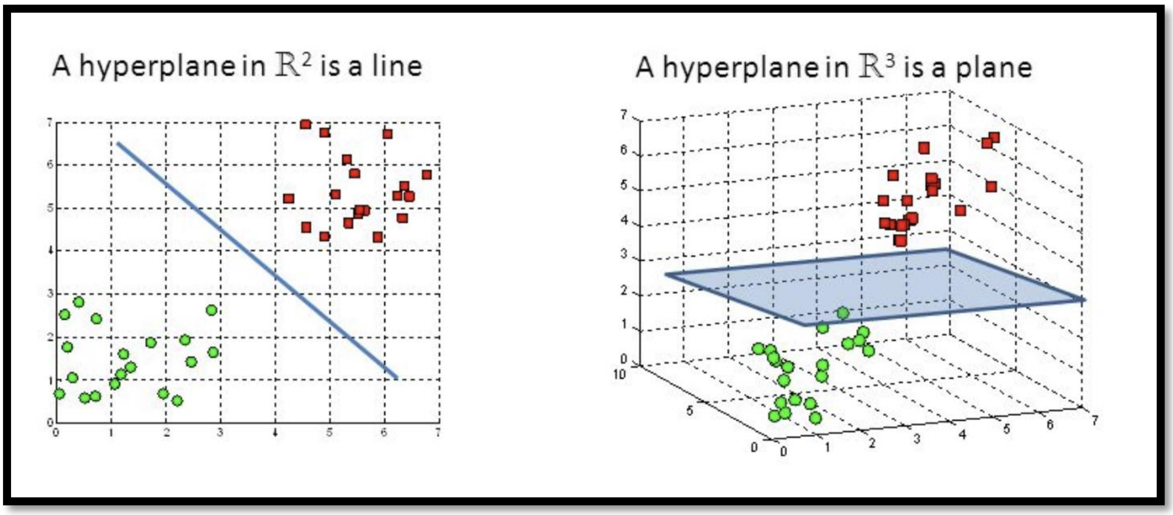
Logistic Regression is a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets.



Support Vector Machine:

The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N — the number of features) that distinctly classifies the data points.

To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e., the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence.



Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane depends upon the number of features. If the number of input features is 2, then the hyperplane is just a line. If the number of input features is 3, then the hyperplane becomes a two-dimensional plane. It becomes difficult to imagine when the number of features exceeds three.

Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyperplane. These are the points that help us build our SVM.

Neural network:

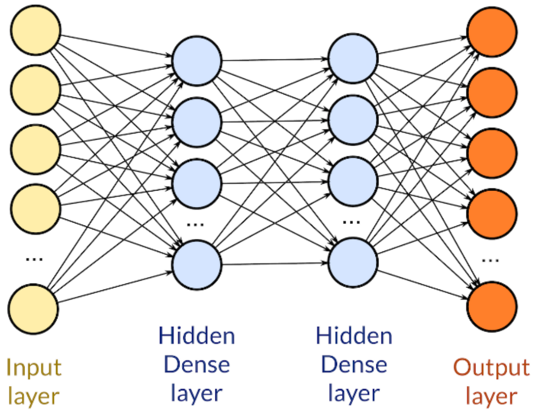
Deep learning is an increasingly popular subset of machine learning. Deep learning models are built using neural networks. A neural network takes in inputs, which are then processed in hidden layers using weights that are adjusted during training. Then the model spits out a prediction. The weights are adjusted to find patterns in order to make better predictions. The user does not need to specify what patterns to look for — the neural network learns on its own.

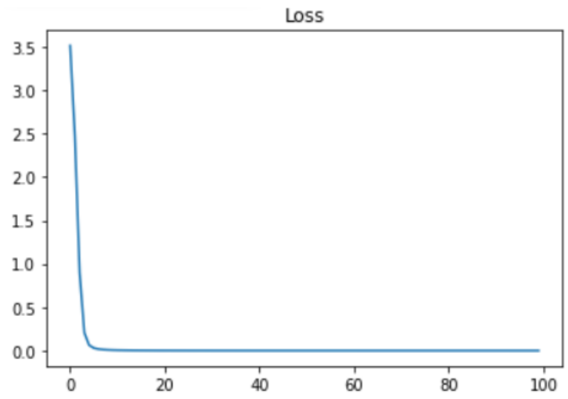
The model type that we will be using is Sequential. Sequential is the easiest way to build a model in Keras. Keras is a user-friendly neural network library written in Python. It allows you to build a model layer by layer. Each layer has weights that correspond to the layer the follows it.

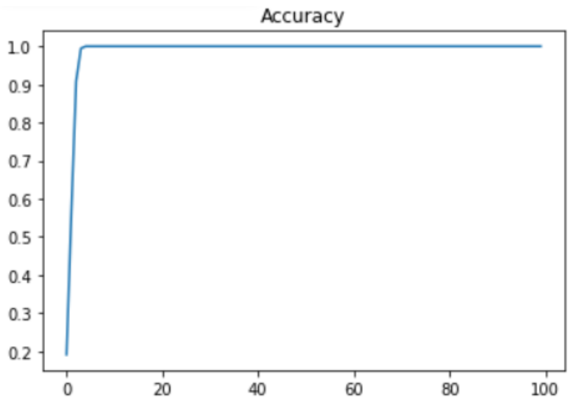
We use the ‘add()’ function to add layers to our model. We will add two layers and an output layer.

‘Dense’ is the layer type. Dense is a standard layer type that works for most cases. In a dense layer, all nodes in the previous layer connect to the nodes in the current layer.

‘Activation’ is the activation function for the layer. An activation function allows models to take into account nonlinear relationships. The activation function we will be using is ReLU or Rectified Linear Activation. Although it is two linear pieces, it has been proven to work well in neural networks.

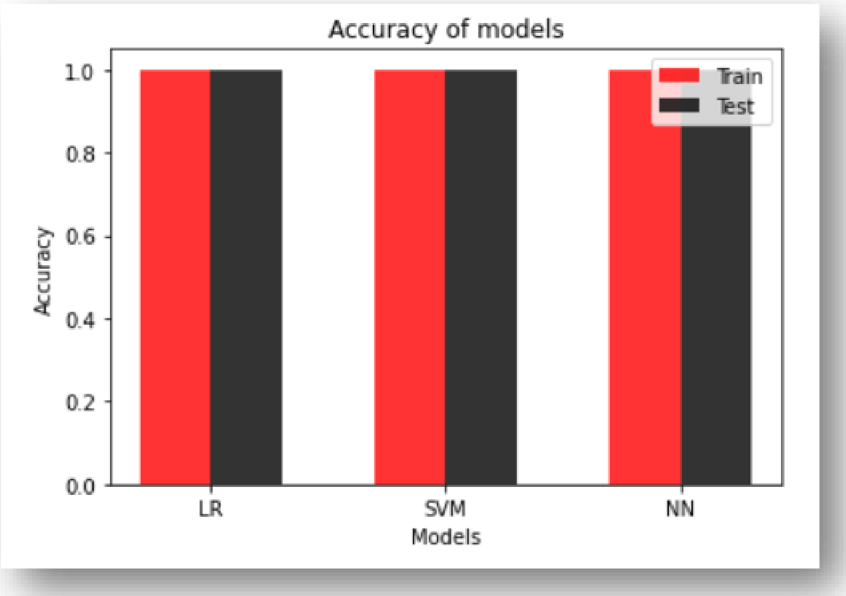






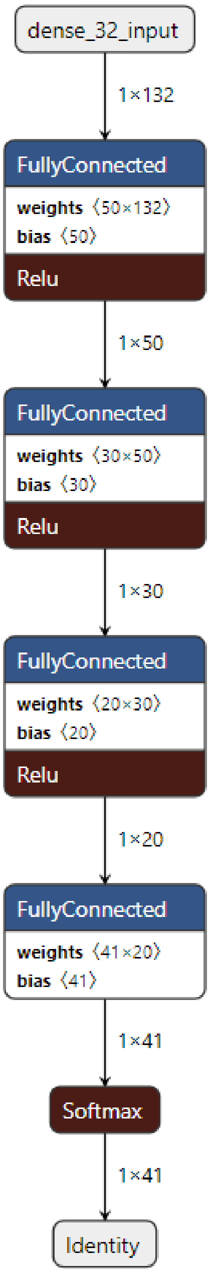
The loss and accuracy while training the neural network model can be observed from the above graphs.

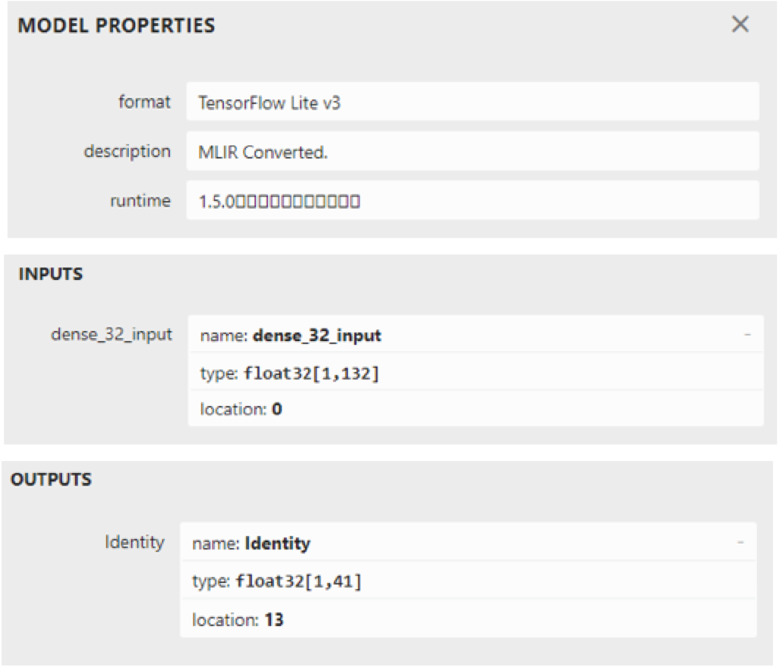
Comparison of models:



All the models are giving 100 percent accuracy, but we have to use neural network in order deploy it in android application. We have generated a **tflite** file of the model to use it in android studio.

Tflite file:





MODEL PROPERTIES

format TensorFlow Lite v3

description MLIR Converted.

runtime 1.5.0□□□□□□□□□□□

INPUTS

dense_32_input name: **dense_32_input**

type: `float32[1,132]`

location: **0**

OUTPUTS

Identity name: **Identity**

type: `float32[1,41]`

location: **13**

From this picture, we can observe the tflite file (i.e., neural network). It takes a float type array containing 132 entries, there are 3 hidden layers using Relu activation function and finally output is a float type array containing 41 entries (i.e., they are the probabilities of 41 different diseases that are available in the dataset). Based on the probabilities, we display the most probable disease as the output to the user.

We add this tflite file to assets folder in android studio and access it by adding required dependencies in the build.

Android application:

We built the app in a user-friendly manner where we provided six different categories as mentioned in the block diagram. Medical report is the area where the application of machine learning come into picture. The app takes four unique symptoms from the user and provides information related to the predicted disease.

CONCLUSION

In this project, we have practiced different machine learning techniques and different models for data training, attempting to achieve the highest accuracy of predicting the disease. Thus, this study settled on classifying disease based on given user's symptoms using three different algorithms and consequently testing its accuracy.

Comparing all the classification models, we conclude Neural network is the preferred choice in terms of its high accuracy and computational efficiency and for the deployment of the model in android studio. However, there is no single classifier that works best on all given problems. The accuracies of different algorithms are compared by plotting a graph which was shown in previous section.

As widely said "Prevention is better than cure", prediction of diseases and epidemic outbreak would lead to an early prevention of an occurrence of a disease. So, with this project, we successfully implemented the disease prediction system.

FUTURE WORK

This project has not implemented accessing locations. So, we can access location of the user and recommend nearby hospitals and doctors. We can also associate with doctors and can provide the users with the option of contacting a doctor through chat. Various specialists are available with whom patients can interact and get a diagnosis and treatment plan or even just an opinion. We can even add a chatbot which can behave as a health assist to the user.

Links for code

Dataset: https://www.kaggle.com/rabisingh/symptom-checker

Google colaboratory (for Machine learning part):
https://colab.research.google.com/drive/15IXfzz00avmWR_oVN3VbUDJuDLOXvY5X?usp=sharing

MedCheck (Android app): https://github.com/shanmukha-manoj/MedCheck

SYSTEM MANUAL

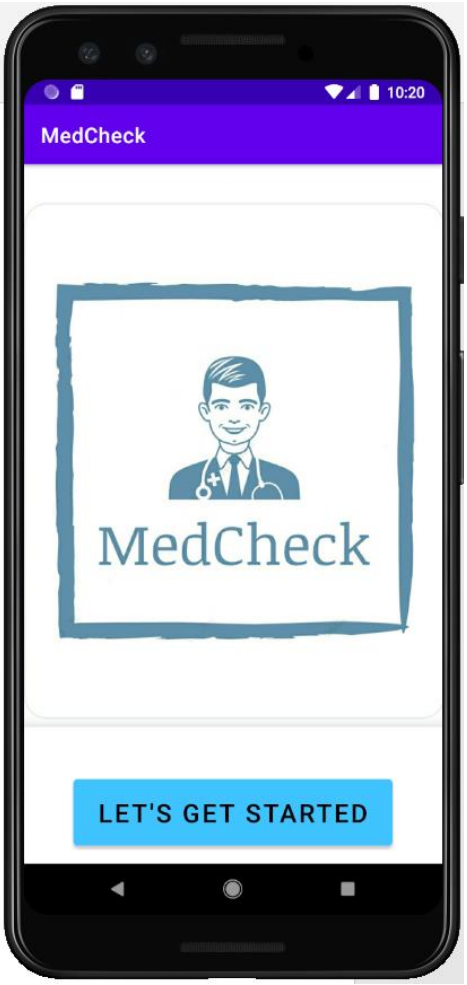


Fig. (1)

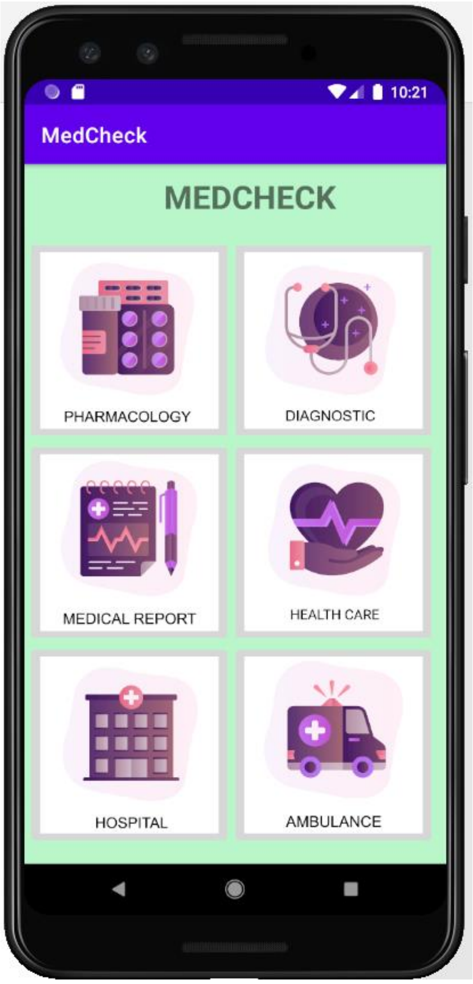
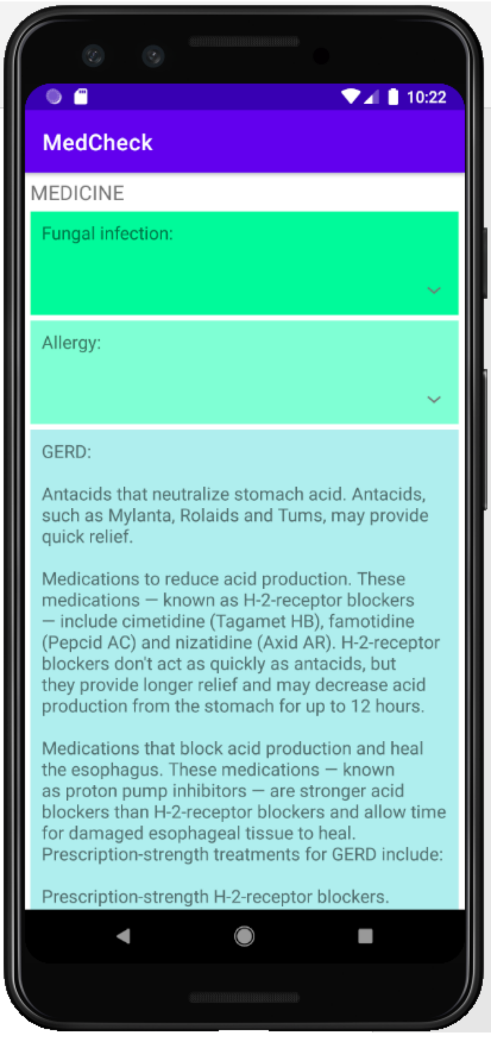


Fig. (2)

Fig. (1) shows the home page of the application. It has a button **Let's get started,** by clicking on it, you will be directed to next screen i.e., Fig.(ii).

Fig. (2) is like navigator, it has six image buttons namely Pharmacology, Diagnostic, Medical report, Health care, Hospital and Ambulance. Each image button will take you to that particular field.



Fungal infection:

Allergy:

GERD:

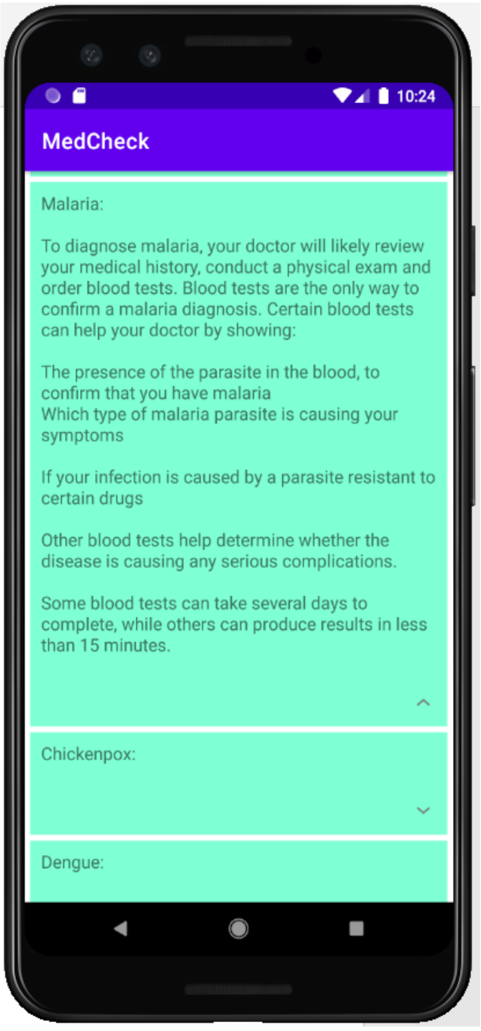
Antacids that neutralize stomach acid. Antacids, such as Mylanta, Rolaids and Tums, may provide quick relief.

Medications to reduce acid production. These medications — known as H-2-receptor blockers — include cimetidine (Tagamet HB), famotidine (Pepcid AC) and nizatidine (Axid AR). H-2-receptor blockers don't act as quickly as antacids, but they provide longer relief and may decrease acid production from the stomach for up to 12 hours.

Medications that block acid production and heal the esophagus. These medications — known as proton pump inhibitors — are stronger acid blockers than H-2-receptor blockers and allow time for damaged esophageal tissue to heal.
Prescription-strength treatments for GERD include:

Prescription-strength H-2-receptor blockers.

Fig. (3)



Malaria:

To diagnose malaria, your doctor will likely review your medical history, conduct a physical exam and order blood tests. Blood tests are the only way to confirm a malaria diagnosis. Certain blood tests can help your doctor by showing:

The presence of the parasite in the blood, to confirm that you have malaria
Which type of malaria parasite is causing your symptoms

If your infection is caused by a parasite resistant to certain drugs

Other blood tests help determine whether the disease is causing any serious complications.

Some blood tests can take several days to complete, while others can produce results in less than 15 minutes.

Chickenpox:

Dengue:

Fig. (4)

You will reach Fig. (3) when you click **Pharmacology** in Fig. (2). Pharmacology consist of a list of diseases and their corresponding medication.

You will reach Fig. (4) when you click **Diagnostic** in Fig. (2). Diagnostic consist of a list of diseases and their corresponding diagnosis.

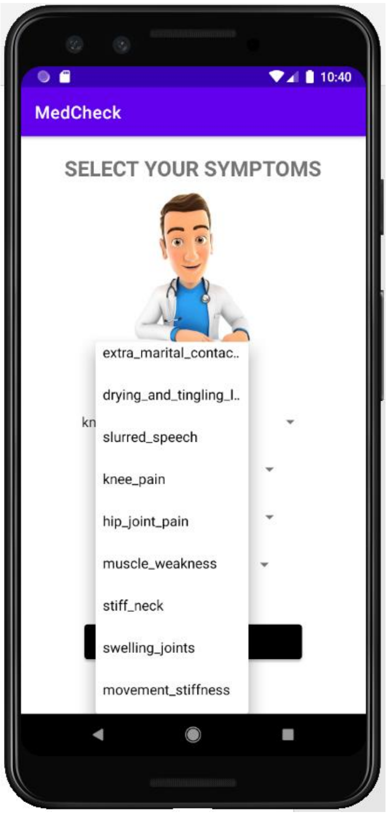


Fig. (5)

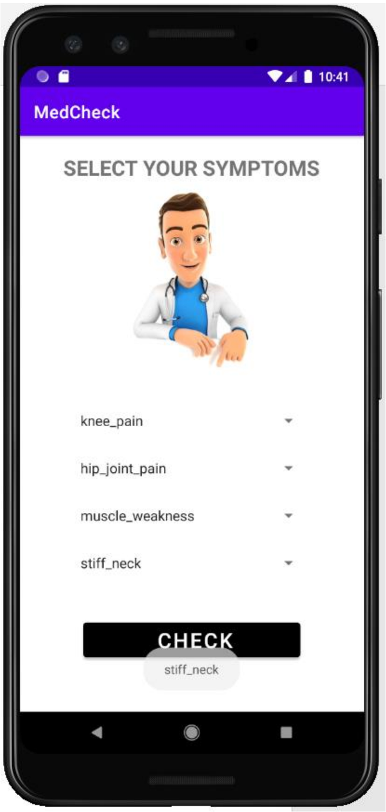
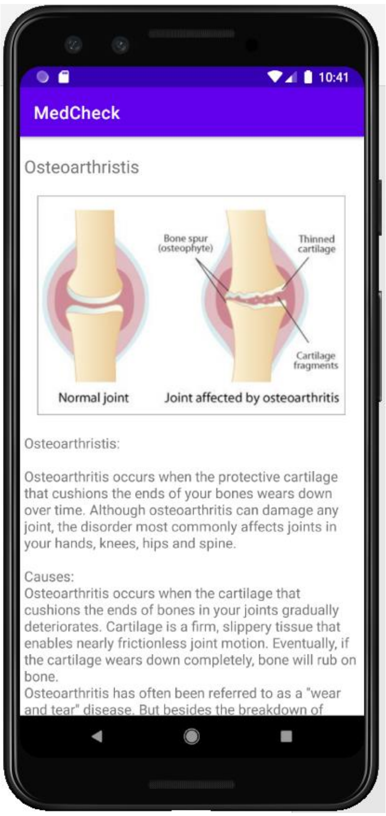


Fig. (6)



Osteoarthristis

Osteoarthristis:

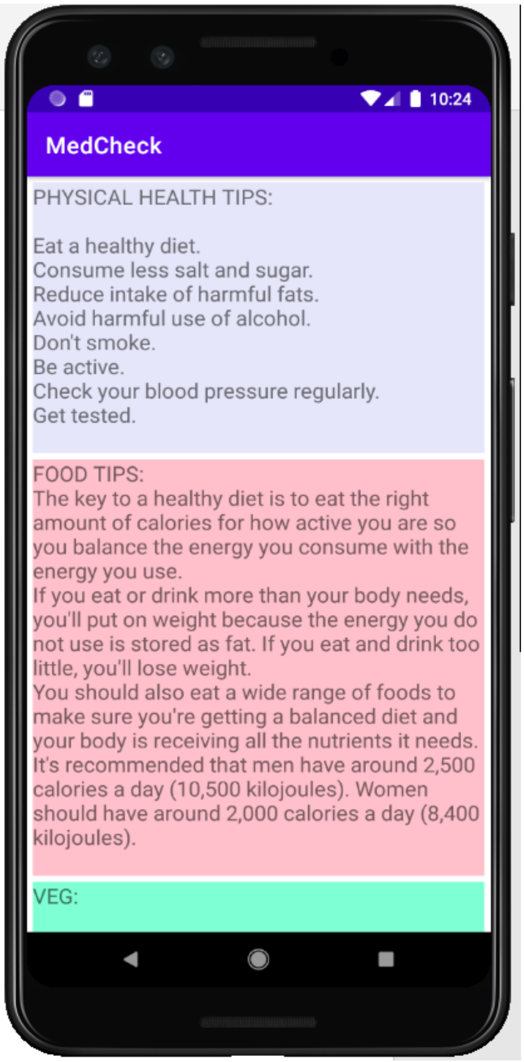
Osteoarthritis occurs when the protective cartilage that cushions the ends of your bones wears down over time. Although osteoarthritis can damage any joint, the disorder most commonly affects joints in your hands, knees, hips and spine.

Causes:
Osteoarthritis occurs when the cartilage that cushions the ends of bones in your joints gradually deteriorates. Cartilage is a firm, slippery tissue that enables nearly frictionless joint motion. Eventually, if the cartilage wears down completely, bone will rub on bone.
Osteoarthritis has often been referred to as a "wear and tear" disease. But besides the breakdown of

Fig. (7)

On clicking **Medical report** on Fig. (2), you will reach Fig. (5) where there are four drop-down menus which shows a list of symptoms. The user should select four unique symptoms from the list as shown in Fig. (6) and click on **Check** button.

After clicking Check button, the system analyses the symptoms and predicts the disease. So, you will reach Fig. (7), where the information like causes, symptoms and treatment related to predicted disease is provided.



MedCheck

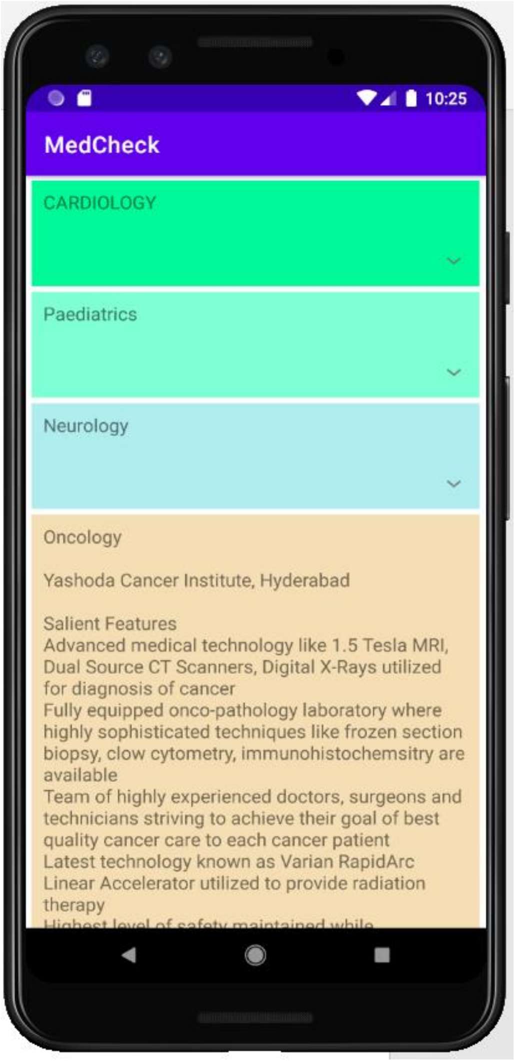
PHYSICAL HEALTH TIPS:

Eat a healthy diet.
Consume less salt and sugar.
Reduce intake of harmful fats.
Avoid harmful use of alcohol.
Don't smoke.
Be active.
Check your blood pressure regularly.
Get tested.

FOOD TIPS:
The key to a healthy diet is to eat the right amount of calories for how active you are so you balance the energy you consume with the energy you use.
If you eat or drink more than your body needs, you'll put on weight because the energy you do not use is stored as fat. If you eat and drink too little, you'll lose weight.
You should also eat a wide range of foods to make sure you're getting a balanced diet and your body is receiving all the nutrients it needs.
It's recommended that men have around 2,500 calories a day (10,500 kilojoules). Women should have around 2,000 calories a day (8,400 kilojoules).

VEG:

Fig. (8)



MedCheck

CARDIOLOGY

Paediatrics

Neurology

Oncology

Yashoda Cancer Institute, Hyderabad

Salient Features
Advanced medical technology like 1.5 Tesla MRI, Dual Source CT Scanners, Digital X-Rays utilized for diagnosis of cancer
Fully equipped onco-pathology laboratory where highly sophisticated techniques like frozen section biopsy, clow cytometry, immunohistochemsitry are available
Team of highly experienced doctors, surgeons and technicians striving to achieve their goal of best quality cancer care to each cancer patient
Latest technology known as Varian RapidArc Linear Accelerator utilized to provide radiation therapy
Highest level of safety maintained while

Fig. (9)

You will reach Fig. (8) when you click **Health care** in Fig. (2). Health care consist of physical health tips, food tips and required diet for vegetarians and non-vegetarians.

You will reach Fig. (9) when you click **Hospital** in Fig. (2). Hospital consist of a list of specialisations in medical field and five best hospitals in India corresponding to that field.

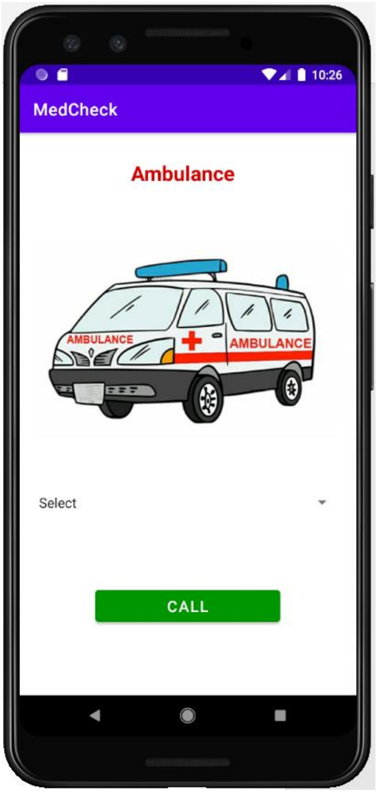


Fig. (10)

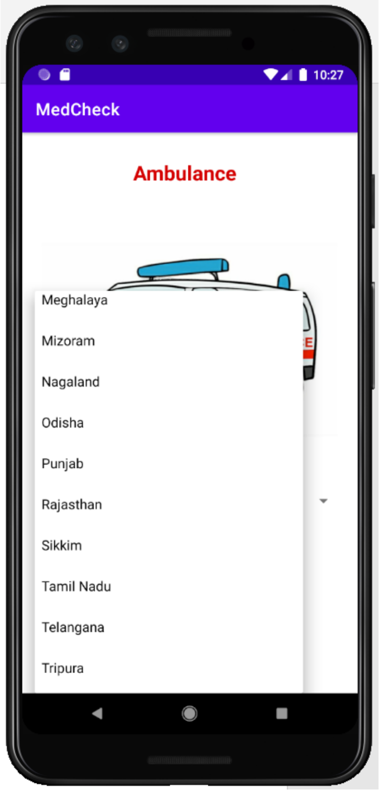


Fig. (11)

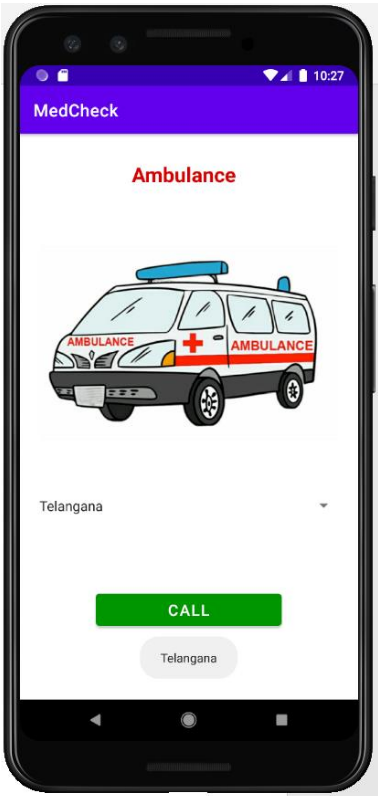


Fig. (12)

You will reach Fig. (10) when you click **Ambulance** in Fig. (2). Ambulance consist of a drop-down menu in which there is a list of states as shown in Fig. (11).

The user has to select their state of residence and click on **Call** button as shown in Fig. (12).

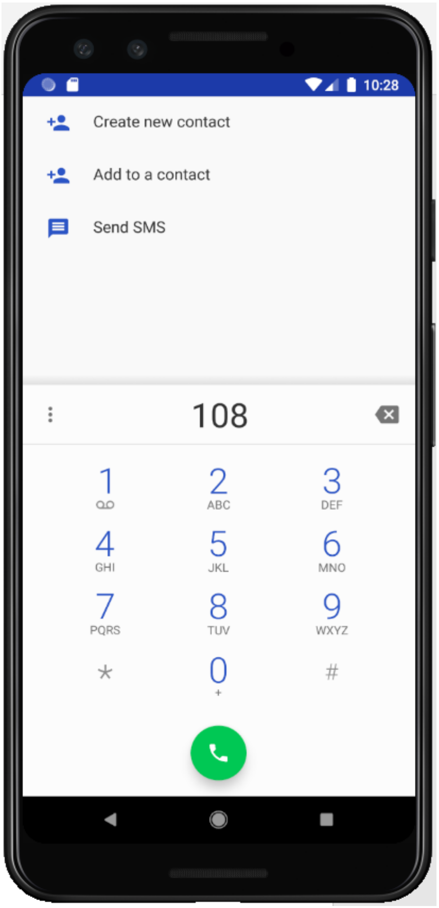


Fig. (13)

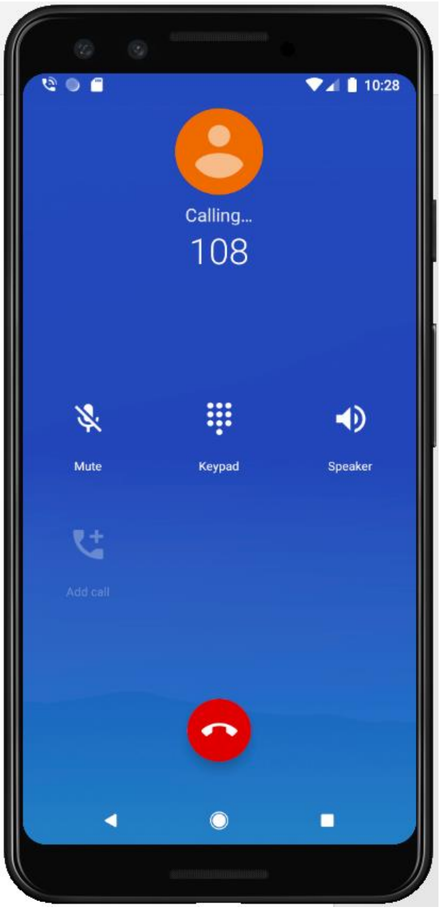


Fig. (14)

After clicking Call button in Fig. (12), you will be directed to Phone app with the respective state's emergency service number on the dial pad. You can call the emergency service as shown in Fig. (14).