



Chemotherapy Efficiency Detector



Haydey, Ibrahim, and Eman
Supervisor: Dr. sabah

Abstract

Breast cancer is one of the most dangerous diseases that women face, with the greatest mortality rate.

Cancer classifications are intended to provide an accurate diagnosis of disease and predict tumor behavior to facilitate oncologist decision-making.

Our project is a website working in cancer classification with different stages of treatment. And the aim of this project is to use image processing and machine learning techniques in analyze a CT scan. This analysis study the effect of chemotherapy on a patient. Whether positive or negative, to assess their progress and build a personalized treatment using data from patients diagnosed with breast cancer.

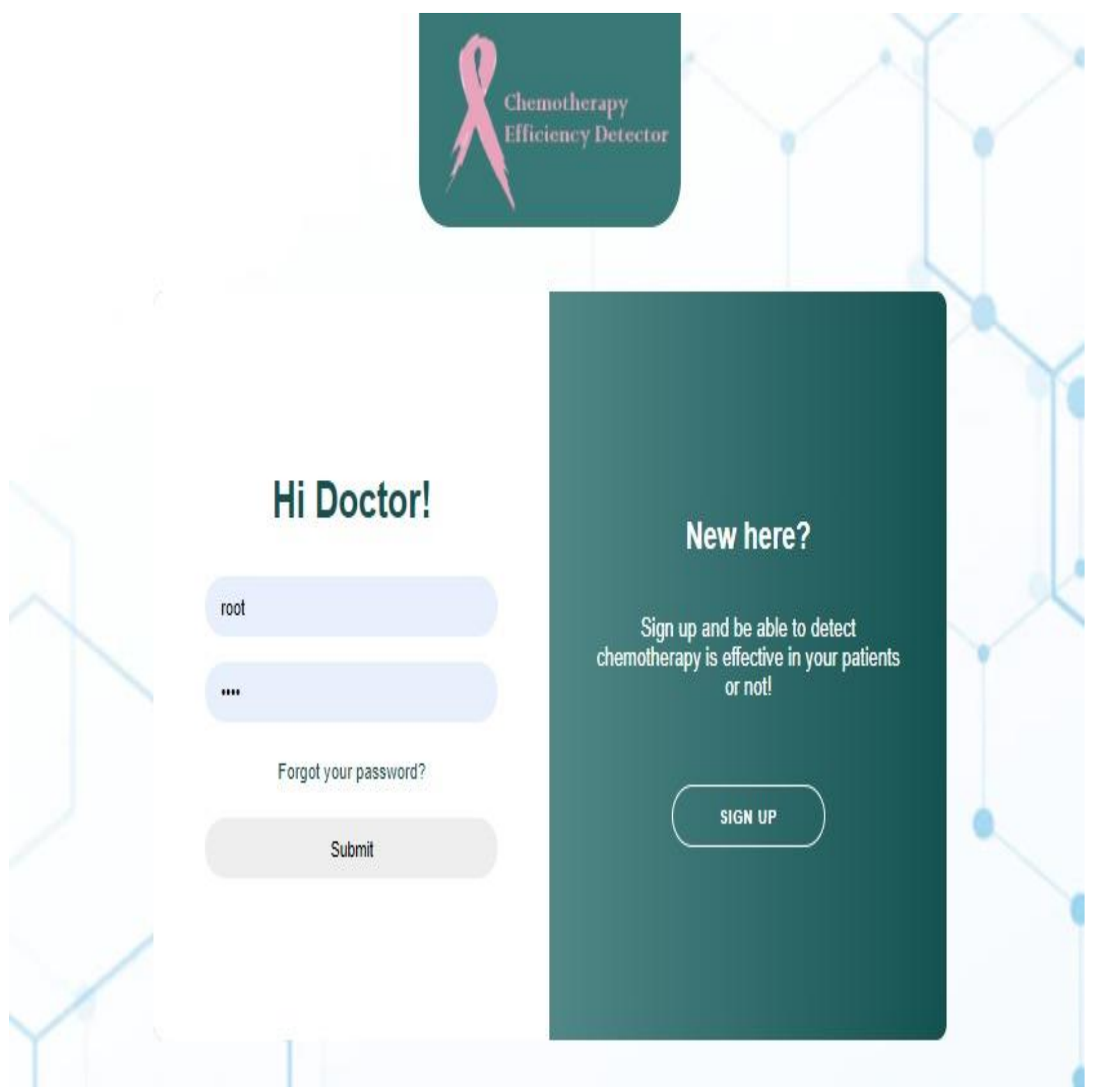
The website takes the data in the form of CT scans. The result will be whether chemotherapy is effective in this case or not.

Introduction

If we search about breast cancer, we can find a lot of projects detected the cancer if it malignant or benign. Or we find projects about chemotherapy, but we will not find project can detect if the chemotherapy is effective with patient or not.

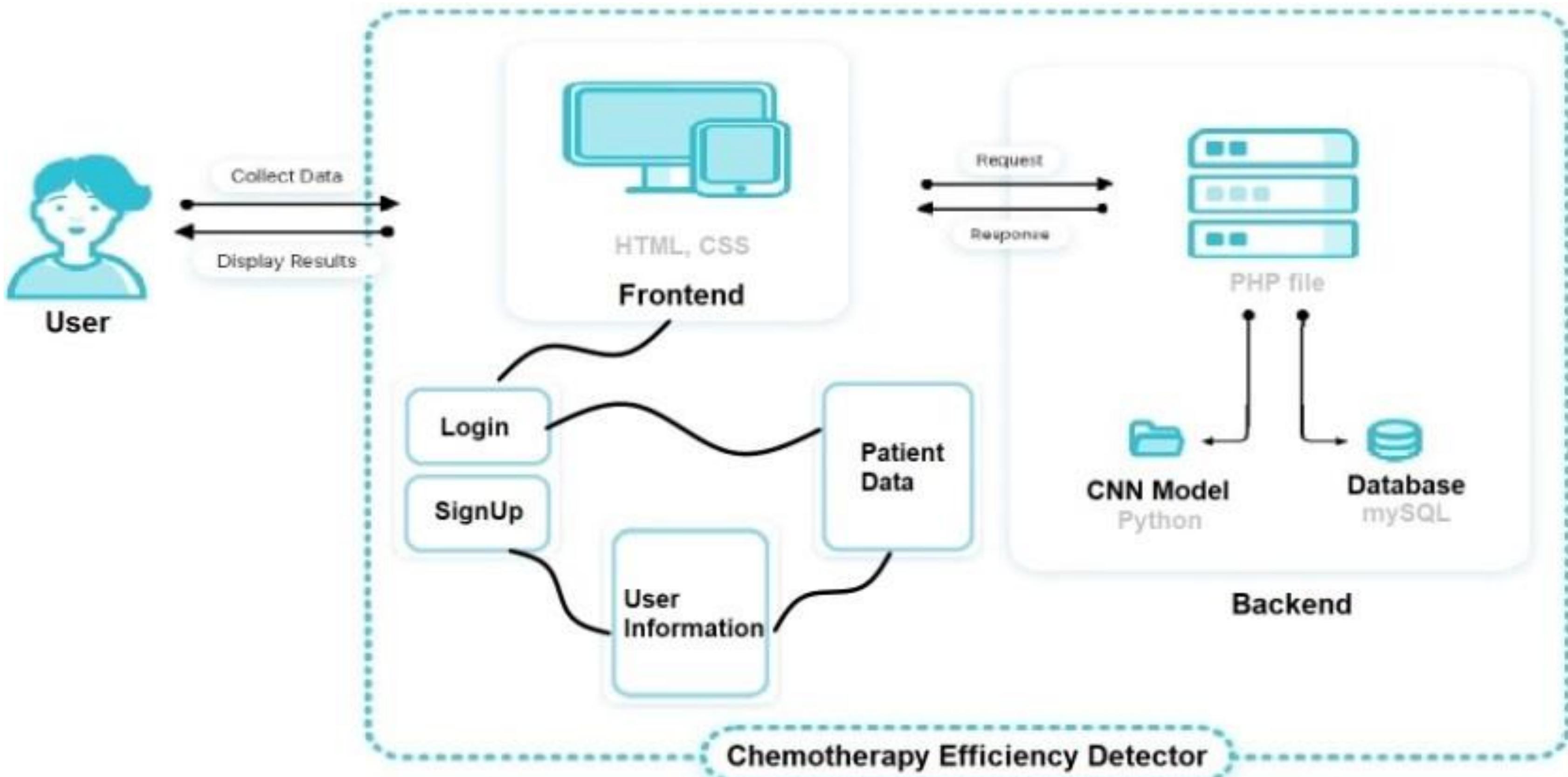
In this project we will try to build a model that help to know the efficiency of drug.

The main application in our project is a website response input data from user as computed tomography CT images and send result to user as PCR or NONPCR Which shows whether the patient has been completely cured or not



Methods

In a deep learning technique, we use a framework of 6-layer convolutional neural networks (CNNs), Containing one input layer with input_shape (200*200*1), 3 sets of dual convolutional and pooling layers (2×2), 3 fully connected (dense) layers, Activation function Relu/Sigmoid and one output layer and loss function categorical_crossentropy. And we use data augmentation method with rotation_range= 40, width_shift_range = 0.08, height_shift_range = 0.08, shear_range = 0.08, zoom_range = 0.08, horizontal_flip = True, fill_mode = 'reflect' and batch_size = 10. And last method is SMOTE from Imbalanced-learn library and use it with balanced weights. This model accuracy is 83.26%. In web application we use PHP programming language to implement the model python file and MySQL database. Whih was the best method for us to implement the model.



Primarily Design

Our project separates into two main components:

- First one is front end which represent the user interface in our website.

We use HTML and CSS to implement the web application pages.

When user visit our website first page appears is the Signup page, in this page user can choose if he wants to create a new account or he already has an account and want to log in.

If user clicked on register, then he will move to the next page which takes the user information and make him create a unique email, password and user id. When user click on submit, the last page will appear which takes user id, a unique patient id which user create, other information and path of the folder that contains the patient images. All these data will insert into database by clicking on submit button. User can get the result of his patients any time by write patient and user id then click show result button, the result will appear as PCR or nonPCR and the accuracy of detection.

But if user write his email and password and clicked on log in button, he will move to the last page directly.

- Second one is back end which represented in connecting all HTML files with database and CNN model.

In user page all values that existed at input boxes will insert into user table at MySQL database by PHP. The same thing happened in the patient page.

In PHP file of patient page, the path will send to python file which takes the path and call function predict to preprocessing the images before input to the CNN model. CNN model predicts the patient belonging to which class, then function predict handles the output and return it to PHP file.

PHP file takes the output and insert it to the database.

Conclusion

After take best sample from dataset where we did it twice, first sample has clearest 10 images from each stage in patient folder, second sample has clearest 5 images from each stage in patient folder.

We then preprocessing datasets and started to implement different models with different approaches, we compared and analyzed the results, nearly we tried a hundred of deep learning models.

A lot of problems faced us because the amount of our dataset. The biggest problem was overfitting, so we used augmented data and it gave us acceptable results.

One of the problems that the minority class is too small compared with majority class, we used balanced weights and it gave us a good result.

But the best result we got when we use SMOTE to equals the classes then we increase one of them by add a random sample to it, then we used balanced weights, that model gave us 83.26% accuracy.

We built a web application depend on this model as a back end. We built it using HTML, CSS and PHP. This web application has a database to save all information of user and patients result.

And the final results were very satisfiable to us.

Contact with us:

ia91142@gmail.com

haydeyalikhalil@gmail.com

eman010038@gmail.com