Detection of Pancreatic Cancer Using ML

MINOR PROJECT

BACHELOR OF TECHNOLOGY

Computer Science & Engineering

SUBMITTED BY

Parija Anupreet	Nitin Pundir	Niranjan	Nitish Kumar
196301065	196301062	196301060	196301063

UNDER THE SUPERVISION OF

Dr. Suyash Bhardwaj

Assistant Professor

Faculty of Engineering and Technology

Gurukul Kangri(Deemed to be University)

Certificate by student

This is to certify that the project in this file are done by all the members of the group and the code used are not copied from anywhere . This work has not been presented anywhere else.

Parija Anupreet	Nitin Pundir	Niranjan	Nitish Kumar
196301065	196301062	196301060	196301063

Certificate by Teacher

This is to certify that the programs in this file are done under my supervision and are not being copied from anywhere or presented anywhere for any sort of benefit.

Dr. Suyash Bhardwaj

Assistant Professor

Department of Computer Science & Engineering

Faculty of Engineering and Technology

Index

S.No	Title	Page No
1	Introduction	6
2	Aim/ Objective	7-10
3	Methodology	11-12
4	Facilities Required	13-22
5	Result and Conclusion	23

1.INTRODUCTION

Pancreatic cancer is an extremely deadly type of cancer. Once diagnosed, the five-year survival rate is less than 10%. However, if pancreatic cancer is caught early, the odds of surviving are much better. Unfortunately, many cases of pancreatic cancer show no symptoms until the cancer has spread throughout the body. A diagnostic test to identify people with pancreatic cancer could be enormously helpful.

1.1 Pancreatic Cancer

Pancreatic cancer is a disease in which malignant (cancer) cells form in the tissues of the pancreas. Pancreatic cancer is difficult to diagnose early.

The pancreas has two main jobs in the body:

- To make juices that help digest (break down) food.
- To make hormones, such as insulin and glucagon that help control blood sugar levels. Both of these hormones help the body use and store the energy it gets from food.

The digestive juices are made by exocrine pancreas cells and the hormones are made by endocrine pancreas cells. About 95% of pancreatic cancers begin in exocrine cells.

.

1.2 Causes of Pancreatic Cancer

Smoking causes the majority of pancreatic cancers — both in smokers and in people exposed to secondhand smoke. But pancreatic cancer also occurs in people who never smoked and in those who never had prolonged exposure to secondhand smoke.

Some major causes are:

- Smoking.
- Being very overweight.

- Having a personal history of diabetes or chronic pancreatitis.
- Having a family history of pancreatic cancer or pancreatitis.
- Having certain hereditary conditions, such as:
- Multiple endocrine neoplasia type 1 (MEN1) syndrome.
- Hereditary nonpolyposis colon cancer (HNPCC; Lynch syndrome).
- von Hippel-Lindau syndrome.
- Peutz-Jeghers syndrome.
- Hereditary breast and ovarian cancer syndrome.
- Familial atypical multiple mole melanoma (FAMMM) syndrome.
- Ataxia-telangiectasia.

1.3 Symptoms of Pancreatic cancer:

Pancreatic cancer may not cause early signs or symptoms . Signs and symptoms may be caused by pancreatic cancer or by other conditions. Check with your doctor if you have any of the following:

- Jaundice (yellowing of the skin and whites of the eyes).
- Light-colored stools .
- Dark urine .
- Pain in the upper or middle abdomen and back.
- Weight loss for no known reason.
- Loss of appetite .
- Feeling very tired.
- Pancreatic cancer is difficult to diagnose early.

Pancreatic cancer is difficult to detect and diagnose for the following reasons:

- There aren’t any noticeable signs or symptoms in the early stages of pancreatic cancer.

- The signs and symptoms of pancreatic cancer, when present, are like the signs and symptoms of many other illnesses.

- The pancreas is hidden behind other organs such as the stomach, small intesine , liver, gallbladder , spleen, and bite ducts .

1.4 Available methods

- Physical exam and health history: An exam of the body to check general signs of health, including checking for signs of disease, such as lumps or anything else that seems unusual.
- Blood chemistry studies: A procedure in which a blood sample is checked to measure the amounts of certain substances, such as bilirubin, released into the blood by organs and tissues in the body. An unusual (higher or lower than normal) amount of a substance can be a sign of disease.
- Tumor marker test: A procedure in which a sample of blood, urine, or tissue is checked to measure the amounts of certain substances, such as CA 19-9, and carcinoembryonic antigen (CEA), made by organs, tissues, or tumor cells in the body
- MRI (magnetic resonance imaging): A procedure that uses a magnet, radio waves, and a computer to make a series of detailed pictures of areas inside the body. This procedure is also called nuclear magnetic resonance imaging (NMRI).
- CT scan (CAT scan): A procedure that makes a series of detailed pictures of areas inside the body, taken from different angles. The pictures are made by a computer linked to an xray machine. A dye may be injected into a vein or swallowed to help the organs or tissues show up more clearly.

2.Aim/Objective

The main objective of this project is to make a User interface which provides uploading of CT scan images and identifying whether the person has Pancreatic cancer or not.

This project Uses concepts of Machine learning and Deep learning to perform the predictions about Pancreatic cancer.

3. Methodology/Planning of work

1. Collection of dataset: In this project we will use a dataset from Kaggle .First for checking the consistency of the data we will do exploratory data analysis. The data set is downloaded from the kaggle website. This dataset included 119 CP cases, 54 gallbladder diseases, 20 cystic lesions of the pancreas, and 15 cases with abdominal pain and gastrointestinal symptoms suggestive of pancreatic origin. Of the 590 samples, 332 (81 control, 89 benign, and 162 PDAC) have been previously analysed. The demographic details of the samples are described in . Centre of origin, details of diagnosis, and sample overlap are reported in . Matched plasma specimens were available for 350 samples

2.Analyse the dataset: Then we will analyse the dataset and make further decisions.

	sample_id	patient_cohort	sample_origin	age	sex	diagnosis	stage	benign_sample_diagnosis	plasma_CA19_9	creatinine	LYVE1	REG1B	TFF1	REG1A
0	S1	Cohort1	BPTB	33	F	1	NaN	NaN	11.7	1.83222	0.893219	52.94884	654.282174	1262.000
1	S10	Cohort1	BPTB	81	F	1	NaN	NaN	NaN	0.97266	2.037585	94.46703	209.488250	228.407
2	S100	Cohort2	BPTB	51	M	1	NaN	NaN	7.0	0.78039	0.145589	102.36600	461.141000	NaN
3	S101	Cohort2	BPTB	61	M	1	NaN	NaN	8.0	0.70122	0.002805	60.57900	142.950000	NaN
4	S102	Cohort2	BPTB	62	M	1	NaN	NaN	9.0	0.21489	0.000860	65.54000	41.088000	NaN
5	S103	Cohort2	BPTB	53	M	1	NaN	NaN	NaN	0.84825	0.003393	62.12600	59.793000	NaN
6	S104	Cohort2	BPTB	70	M	1	NaN	NaN	NaN	0.62205	0.174381	152.27700	117.516000	NaN
7	S105	Cohort2	BPTB	58	F	1	NaN	NaN	11.0	0.89349	0.003574	3.73000	40.294000	NaN
8	S106	Cohort2	BPTB	59	F	1	NaN	NaN	NaN	0.48633	0.001945	7.02100	26.782000	NaN
9	S107	Cohort2	BPTB	56	F	1	NaN	NaN	24.0	0.61074	0.278778	83.92800	19.185000	NaN

3.Feature Extraction: The process of feature extraction from the dataset is to obtain the features like modality, image size, pixel spacing etc.

4.Test /Train the dataset: After the feature extraction we will split the data into tests and train set with test size as 01/10%. Then we will do networking and model building keras tensorflow and we will get the model summary. And we plot the graph of accuracies and loss using the matplotlib library. At last features have been extracted from the model that we builded.

```
[ ] target='diagnosis'

[ ] x = df.drop(target, axis=1)
    y = df[target]

[ ] #Spliting dataset into training set and testing test
    from sklearn.model_selection import train_test_split
    X_train,X_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0)

[ ] # Feature Scaling
    from sklearn.preprocessing import StandardScaler
    sc = StandardScaler()
    X_train = sc.fit_transform(X_train)
    X_test = sc.transform(X_test)

Using DecisionTreeRegressor

    # Fitting Decision Tree Regression to the dataset
    from sklearn.tree import DecisionTreeRegressor
    regressor = DecisionTreeRegressor()
    regressor.fit(X_train, y_train)

    DecisionTreeRegressor()

[ ] print('Ridge Regression Train Score is : ' , regressor.score(X_train, y_train))
    print('Ridge Regression Test Score is : ' , regressor.score(X_test, y_test))
```

5.Checking Accuracy of model: We will use the R-CNN algorithm of ML to get the best accuracy for predicting the outcome.

6.Deployment Process:

6.1 Creating user interface Using Flask: We will be creating a website using the Combination of HTML,CSS,JavaScript and Flask. As Flask is python framework it will help us more to integrate between our model and the website/

6.2 Deploying our model: Here we will be using heroku to deploy our model as it is a very good cloud platform for such purposes.

4.Facilities required for proposed work

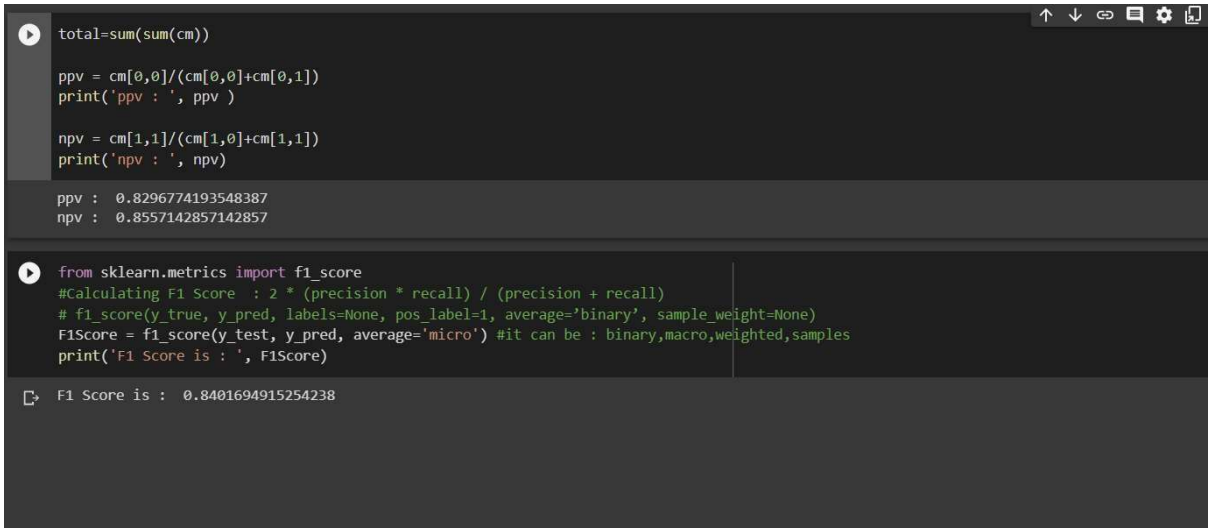
- Dataset
- Libraries
- Kaggle notebook
- Google colab

- Python with ML
- Frontend
- Backend- Flask/Django
- Cloud - Heroku for Deployment
- Hardware- A laptop with Intel i7 processor and 16GB Ram
- Software Required- Anaconda , Visual Studio Code(VS Code)

5. Result and Conclusion

After completion of this project we will be able to recognise pancreatic cancer only by uploading our CT scans to a website. The website will predict whether the person has lung cancer or not by analysing that uploaded image.

This Project will give an aspect in the field of medical science as detection of lung cancer becomes easier and faster after completion of this project.



6.Research/References

Dataset link:-

https://www.kaggle.com/code/mohamedbakrey/pancreatic -cancer-classification-by-ml-algo/data

How to deploy ML model: https://youtu.be/pMIwu5FwJ78

Deep learning concepts with Django:
https://www.youtube.com/watch?v=bFYzhTiJkuQ