Proceedings of the 34th Accra Bespoke Multidisciplinary Innovations Conference & the Africa AI Stakeholders Summit
Academic City University College, Accra Ghana
19th – 21st December, 2022
www.isteams.net/accrabespoke2022

The Design and Implementation Of A Machine Learning Based Breast Cancer Diagnostic System

[1]Danso, J & Agbo, D-P
Faculty of Computational Sciences & Informatics
Academic City University College
Accra, Ghana
E-mails joseph.danso@acity.edu.gh, paul-david.agbo@acity.edu.gh
Phones: [1]+233268903634

ABSTRACT

The adoption of information technology in the Ghanian health sector remains for the most part limited to the Greater Accra region. The technologies adopted are mainly used for communication, telemedicine, record keeping and database systems. Advanced analytical technologies like machine learning and artificial intelligence can greatly improve the Ghanian health. But despite their great potential, they are still unpopular. The current project is on the use of classification-based machine learning technology for breast cancer disease diagnosis. This project is designed and implemented as an online based platform to be used mainly by laboratories agents and researchers. This system was built to simplify the diagnostic process for the disease and increase its accuracy level.

Keywords: Machine Learning, K-Nearest Neighbor and eHealth, Breast Cancer, Diagnostics

Proceedings Reference Format
Danso, J. & Agbo, D-P. (2022): The Design And Implementation Of A Machine Learning Based Breast Cancer Diagnostic System. Proceedings of the 34th Accra Bespoke Multidisciplinary Innovations Conference & The Africa AI Stakeholders Summit. Academic City University College, Accra Ghana, 2022. Pp 85-102. www.isteams.net/accrabespoke2022
https://doi.org/10.22624/AIMS/ACCRABESPOKE2022/V34P8

1. INTRODUCTION

The goal of this work is to employ the use of machine learning to efficiently recommend selected crops to grow in selected fields. It will seek to glean insight from the experiential knowledge employed by farmers in selecting crops they choose to plant. The data collected would be used to build a machine learning model to efficiently recommend selected crops to plant in the future.

2021 marks the year when the Ghanaian government started doubling down on the new digital national identification program called Ghana Card. According to the National Identification Authority of Ghana, the technology that is being put in place will facilitate the social, economic and political development of Ghana. With this project, the government hopes to encourage the digitization of all activity sectors. The health sector is key to this program; The digitalization of the health sector would be beneficial to all patients, medical staff, health institutions and the government. Various technologies can be used to improve the overall efficiency, ease of operation and convenience of all healthcare related activities.

Problem Statement
The adoption of information technology in the Ghanian health sector remains for the most part limited to the Greater Accra region. The technologies adopted are mainly used for communication, telemedicine, record keeping and database systems. Advanced analytical technologies like machine learning and artificial intelligence can greatly improve the Ghanian health medical system. But despite their great potential, they are still unpopular.

Aim Of Study
The aim of the work is to produce a machine learning solution to analyze laboratory result data.

1.4 Specific Objective Of The Study
The specific objective of the study is the design and implementation of a graphical user interface that automates the analysis of breast cancer data points and the diagnosis prediction process.

1.5 Deliverable Product
A system (online application) that applies a machine learning model to process breast cancer data measurements in order to produce a prediction based on previous training with real data. The prediction will be performed with the data that the user uploads.

The following diagram shows the work breakdown structure for the design and implementation phase.

2. LITERATURE REVIEW

Numerous studies have been done in this area in the past with the goal of developing crop predicting systems all with their merits and demerits. Devdatta A. Bondre and Santosh alluded to the fact that a farmer always has an interest in knowing how much yield to expect. Previously, this was done with experiential knowledge on a particular field and crop. However, this is an issue that can be made better by using available data since machine learning techniques can provide more optimal solutions (Bondre & Mahagaonkar, 2019). Sayed Mazhar Ali et al highlighted that crop cultivation in any part of the world depended on the climate which could also represent seasons and depended on soil properties as well. The paper however stressed on the fact that the enhancement of the production of crops was barred on factors related to temperature (Ali, Das, & Kumar, 2021).

There is no singular factor that is the be-all-end-all to giving clear insight into what a farmer should plant for suitable crop yield. However, there are factors, if not one, that significantly increase the chances of one getting a favorable yield. Collecting all necessary data from the soil type of a plant to the season it is planted contributes to making it yield as anticipated.

2.1 Review of Machine Learning Technologies
Machine learning (ML) is a sort of artificial intelligence (AI) that allows software applications to improve their prediction accuracy without being expressly designed to do so. In order to forecast new output values, machine learning algorithms use historical data as input. Fraud detection, Health systems, spam filtering, malware threat detection, business process automation (BPA), and predictive maintenance are all common applications where machine learning is used.

Machine learning is significant because it allows businesses to see trends in customer behavior and business operating patterns while also assisting in the development of new goods. Machine learning is at the heart of many of today's most successful businesses, like Facebook, Google, and Uber. For many businesses, machine learning has become a crucial competitive differentiation. Machine Learning is often classified in four groups:

Supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning are the four basic methodologies. The algorithm that data scientists use is determined on the type of data they work on and the type of data they are trying to predict. Supervised machine learning is an artificial intelligence and machine learning subcategory. It is the use of labeled datasets to train algorithms that accurately classify data or predict outcomes. As input data is fed into the model, the weights are adjusted until the model is properly fitted, which happens during the cross-validation phase. Organizations can use supervised learning to tackle a range of real-world problems at scale, such as spam classification in a distinct folder from your email. Unsupervised machine learning analyzes and clusters unlabeled datasets using machine learning methods. This type of algorithm uncovers hidden patterns or data groupings without the need for human intervention. It is the best solution for exploratory data analysis, cross-selling techniques, consumer segmentation, and image identification because of its capacity to detect similarities and differences in information. Semi-supervised machine learning involves a small number of labeled examples and many unlabeled examples.

Reinforcement machine learning is based on rewarding desired behaviors and/or punishing undesired ones. A reinforcement learning agent can perceive and interpret its environment, act, and learn through trial and error.

3. SPECIFICATION, DESIGN AND DEVELOPMENT METHODOLOGY

The Waterfall methodology—also known as the Waterfall model—is a sequential development process that flows like a waterfall through all phases of a project (analysis, design, development, and testing), with each phase completely wrapping up before the next phase begins. The figure bellow shows the sequential nature of the waterfall model; each phase is completed before going to the next.

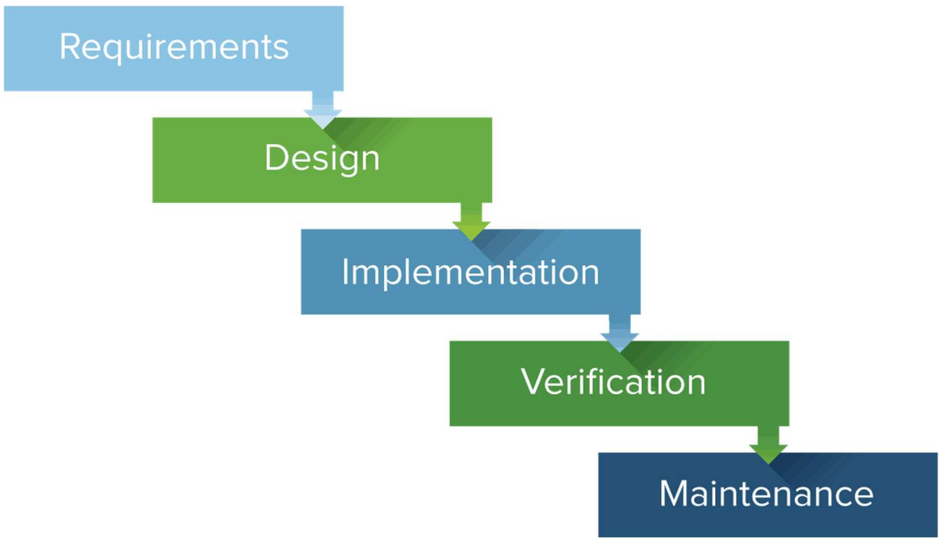


Fig 1: The Waterfall Model

3.1 Machine Learning Algorithm

A machine learning algorithm is the method by which the AI system conducts its task, generally predicting output values from given input data. The field of Machine learning offers a variety of algorithms to choose from. The two main processes of machine learning algorithms are classification and regression. Classification recognizes specific entities within the dataset and attempts to draw some conclusions on how those entities should be labelled or defined. Regression is a technique for investigating the relationship between independent variables or features and a dependent variable or outcome. It's used as a method for predictive modelling in machine learning, in which an algorithm is used to predict continuous outcomes. Looking at this project, a classification model, with the K-Nearest Neighbor (KNN) Algorithm is best.

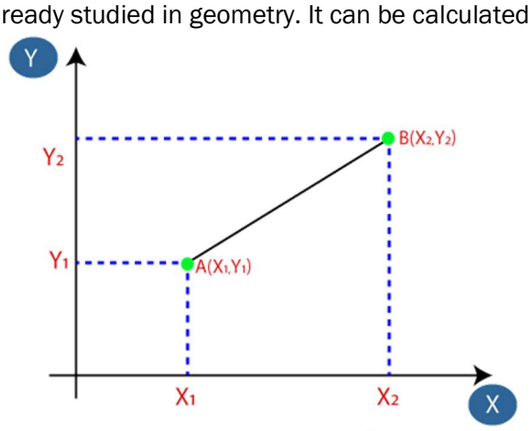
K-Nearest Neighbours Algorithm

The K-Nearest Neighbor algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. While it can be used for either regression or classification problems, it is typically used as a classification algorithm, working off the assumption that similar points can be found near one another.

How Does The K-Nearest Neighbours Algorithm Work

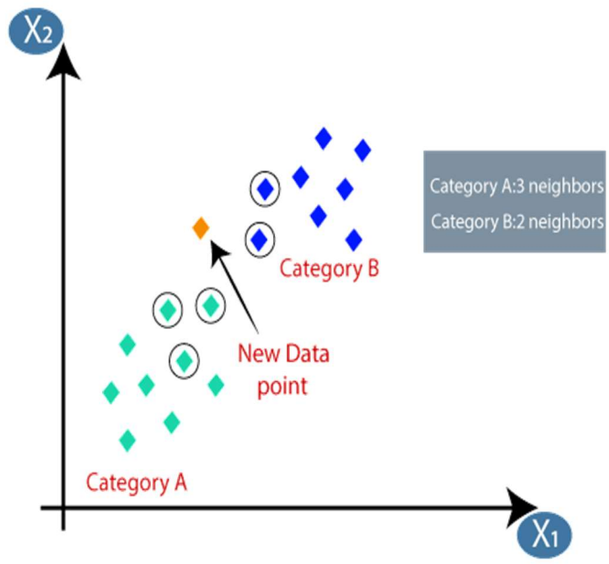
The KNN algorithm assumes that the new case/data and existing cases are similar and places the new case in the category that is most similar to the existing categories. The KNN method saves all the available data and classifies a new data point based on its similarity to the existing data. This means that new data can be quickly sorted into the most appropriate category.
Suppose we have a new data point, and we need to put it in the required category. Consider the below image:

- Firstly, we will choose the number of neighbors, so we will choose the k=5.
- Next, we will calculate the **Euclidean distance** between the data points. The Euclidean distance is the distance between two points,
  which we have already studied in geometry. It can be calculated as:



Euclidean Distance between A_1 and B_2 = $\sqrt{(X_2-X_1)^2+(Y_2-Y_1)^2}$

- By calculating the Euclidean distance, we got the nearest neighbors, as three nearest neighbors in category A and two nearest neighbors in
  category B. Consider the below image:



- As we can see the 3 nearest neighbors are from category A, hence this new data point must belong to category A.

Here is a pseudo code that explains how the KNN algorithm is implemented:

1. Load the training data.
2. Prepare data by scaling, missing value treatment, and dimensionality reduction as required.
3. Find the optimal value for K:
4. Predict a class value for new data:
5. Calculate distance (X, Xi) from i=1,2,3,....,n.
6. where X= new data point, Xi= training data, distance as per your chosen distance metric.
7. Sort these distances in increasing order with corresponding train data.
8. From this sorted list, select the top 'K' rows.
9. Find the most frequent class from these chosen 'K' rows. This will be your predicted class.

3.2 Development Tools and Data in Use

The technical tools and elements used for this project are : Breast cancer data points compiled into csv training sets, Google Collaboratory, Google Drive, Anvil Works, HTML, CSS and Python programming languages.

Type Of Data Being Used

The dataset used here includes 569 examples of cancer biopsies, each with 32 features. The features listed below are the identification number for each set of measurement, the diagnosis and 30 numeric-valued laboratory measurements,
all listed below:
radius_mean
texture_mean
perimeter_mean
area_mean
smoothness_mean
compactness_mean
concavity_mean
concave points_mean
symmetry_mean
fractal_dimension_mean
radius_se
texture_se
perimeter_se
area_se
smoothness_se
compactness_se
concavity_se
concave points_se
symmetry_se
fractal_dimension_se
radius_worst
texture_worst
perimeter_worst
area_worst
smoothness_worst
compactness_worst
concavity_worst
concave points_worst
symmetry_worst
fractal_dimension_worst
The diagnosis is coded as "M" to indicate malignant or "B" to indicate benign.

A21 | 8510426

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	id	diagnosis	radius_me	texture_m	perimeter	area_mea	smoothne	compactn	concavity_	concave p	symmetry	fractal_dir	radius_se	texture_se	perimeter	area_se	smoothne	compactn	concavity_	concave p	symmetry	fractal_dir	radius_wo te
2	842302	M	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001	0.1471	0.2419	0.07871	1.095	0.9053	8.589	153.4	0.006399	0.04904	0.05373	0.01587	0.03003	0.006193	25.38
3	842517	M	20.57	17.77	132.9	1326	0.08474	0.07864	0.0869	0.07017	0.1812	0.05667	0.5435	0.7339	3.398	74.08	0.005225	0.01308	0.0186	0.0134	0.01389	0.003532	24.99
4	84300903	M	19.69	21.25	130	1203	0.1096	0.1599	0.1974	0.1279	0.2069	0.05999	0.7456	0.7869	4.585	94.03	0.00615	0.04006	0.03832	0.02058	0.0225	0.004571	23.57
5	84348301	M	11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	0.1052	0.2597	0.09744	0.4956	1.156	3.445	27.23	0.00911	0.07458	0.05661	0.01867	0.05963	0.009208	14.91
6	84358402	M	20.29	14.34	135.1	1297	0.1003	0.1328	0.198	0.1043	0.1809	0.05883	0.7572	0.7813	5.438	94.44	0.01149	0.02461	0.05688	0.01885	0.01756	0.005115	22.54
7	843786	M	12.45	15.7	82.57	477.1	0.1278	0.17	0.1578	0.08089	0.2087	0.07613	0.3345	0.8902	2.217	27.19	0.00751	0.03345	0.03672	0.01137	0.02165	0.005082	15.47
8	844359	M	18.25	19.98	119.6	1040	0.09463	0.109	0.1127	0.074	0.1794	0.05742	0.4467	0.7732	3.18	53.91	0.004314	0.01382	0.02254	0.01039	0.01369	0.002179	22.88
9	84458202	M	13.71	20.83	90.2	577.9	0.1189	0.1645	0.09366	0.05985	0.2196	0.07451	0.5835	1.377	3.856	50.96	0.008805	0.03029	0.02488	0.01448	0.01486	0.005412	17.06
10	844981	M	13	21.82	87.5	519.8	0.1273	0.1932	0.1859	0.09353	0.235	0.07389	0.3063	1.002	2.406	24.32	0.005731	0.03502	0.03553	0.01226	0.02143	0.003749	15.49
11	84501001	M	12.46	24.04	83.97	475.9	0.1186	0.2396	0.2273	0.08543	0.203	0.08243	0.2976	1.599	2.039	23.94	0.007149	0.07217	0.07743	0.01432	0.01789	0.01008	15.09
12	845636	M	16.02	23.24	102.7	797.8	0.08206	0.06669	0.03299	0.03323	0.1528	0.05697	0.3795	1.187	2.466	40.51	0.004029	0.009269	0.01101	0.007591	0.0146	0.003042	19.19
13	84610002	M	15.78	17.89	103.6	781	0.0971	0.1292	0.09954	0.06606	0.1842	0.06082	0.5058	0.9849	3.564	54.16	0.005771	0.04061	0.02791	0.01282	0.02008	0.004144	20.42
14	846226	M	19.17	24.8	132.4	1123	0.0974	0.2458	0.2065	0.1118	0.2397	0.078	0.9555	3.568	11.07	116.2	0.003139	0.08297	0.0889	0.0409	0.04484	0.01284	20.96
15	846381	M	15.85	23.95	103.7	782.7	0.08401	0.1002	0.09938	0.05364	0.1847	0.05338	0.4033	1.078	2.903	36.58	0.009769	0.03126	0.05051	0.01992	0.02981	0.003002	16.84
16	84667401	M	13.73	22.61	93.6	578.3	0.1131	0.2293	0.2128	0.08025	0.2069	0.07682	0.2121	1.169	2.061	19.21	0.006429	0.05936	0.05501	0.01628	0.01961	0.008093	15.03
17	84799002	M	14.54	27.54	96.73	658.8	0.1139	0.1595	0.1639	0.07364	0.2303	0.07077	0.37	1.033	2.879	32.55	0.005607	0.0424	0.04741	0.0109	0.01857	0.005466	17.46
18	848406	M	14.68	20.13	94.74	684.5	0.09867	0.072	0.07395	0.05259	0.1586	0.05922	0.4727	1.24	3.195	45.4	0.005718	0.01162	0.01998	0.01109	0.0141	0.002085	19.07
19	84862001	M	16.13	20.68	108.1	798.8	0.117	0.2022	0.1722	0.1028	0.2164	0.07356	0.5692	1.073	3.854	54.18	0.007026	0.02501	0.03188	0.01297	0.01689	0.004142	20.96
20	849014	M	19.81	22.15	130	1260	0.09831	0.1027	0.1479	0.09498	0.1582	0.05395	0.7582	1.017	5.865	112.4	0.006494	0.01893	0.03391	0.01521	0.01356	0.001997	27.32
21	8510426	B	13.54	14.36	87.46	566.3	0.09779	0.08129	0.06664	0.04781	0.1885	0.05766	0.2699	0.7886	2.058	23.56	0.008462	0.0146	0.02387	0.01315	0.0198	0.0023	15.11
22	8510653	B	13.08	15.71	85.63	520	0.1075	0.127	0.04568	0.0311	0.1967	0.06811	0.1852	0.7477	1.383	14.67	0.004097	0.01898	0.01698	0.00649	0.01678	0.002425	14.5
23	8510824	B	9.504	12.44	60.34	273.9	0.1024	0.06492	0.02956	0.02076	0.1815	0.06905	0.2773	0.9768	1.909	15.7	0.009606	0.01432	0.01985	0.01421	0.02027	0.002968	10.23
24	8511133	M	15.34	14.26	102.5	704.4	0.1073	0.2135	0.2077	0.09756	0.2521	0.07032	0.4388	0.7096	3.384	44.91	0.006789	0.05328	0.06446	0.02252	0.03672	0.004394	18.07
25	851509	M	21.16	23.04	137.2	1404	0.09428	0.1022	0.1097	0.08632	0.1769	0.05278	0.6917	1.127	4.303	93.99	0.004728	0.01259	0.01715	0.01038	0.01083	0.001987	29.17
26	852552	M	16.65	21.38	110	904.6	0.1121	0.1457	0.1525	0.0917	0.1995	0.0633	0.8068	0.9017	5.455	102.6	0.006048	0.01882	0.02741	0.0113	0.01468	0.002801	26.46
27	852631	M	17.14	16.4	116	912.7	0.1186	0.2276	0.2229	0.1401	0.304	0.07413	1.046	0.976	7.276	111.4	0.008029	0.03799	0.03732	0.02397	0.02308	0.007444	22.25
28	852763	M	14.58	21.53	97.41	644.8	0.1054	0.1868	0.1425	0.08783	0.2252	0.06924	0.2545	0.9832	2.11	21.05	0.004452	0.03055	0.02681	0.01352	0.01454	0.003711	17.62
29	852781	M	18.61	20.25	122.1	1094	0.0944	0.1066	0.149	0.07731	0.1697	0.05699	0.8529	1.849	5.632	93.54	0.01075	0.02722	0.05081	0.01911	0.02293	0.004217	21.31
30	852973	M	15.3	25.27	102.4	732.4	0.1082	0.1697	0.1683	0.08751	0.1926	0.0654	0.439	1.012	3.498	43.5	0.005233	0.03057	0.03576	0.01083	0.01768	0.002967	20.27
31	853201	M	17.57	15.05	115	955.1	0.09847	0.1157	0.09875	0.07953	0.1739	0.06149	0.6003	0.8225	4.655	61.1	0.005627	0.03033	0.03407	0.01354	0.01925	0.003742	20.01
32	853401	M	18.63	25.11	124.8	1088	0.1064	0.1887	0.2319	0.1244	0.2183	0.06197	0.8307	1.466	5.574	105	0.006248	0.03374	0.05196	0.01158	0.02007	0.00456	23.15
33	853612	M	11.84	18.7	77.93	440.6	0.1109	0.1516	0.1218	0.05182	0.2301	0.07799	0.4825	1.03	3.475	41	0.005551	0.03414	0.04205	0.01044	0.02273	0.005667	16.82
34	85382601	M	17.02	23.98	112.8	899.3	0.1197	0.1496	0.2417	0.1203	0.2248	0.06382	0.6009	1.398	3.999	67.78	0.008268	0.03082	0.05042	0.01112	0.02102	0.003854	20.88
35	854002	M	19.27	26.47	127.9	1162	0.09401	0.1719	0.1657	0.07593	0.1853	0.06261	0.5558	0.6062	3.528	68.17	0.005015	0.03318	0.03497	0.009643	0.01543	0.003896	24.15
36	854039	M	16.13	17.88	107	807.2	0.104	0.1559	0.1354	0.07752	0.1998	0.06515	0.334	0.6857	2.183	35.03	0.004185	0.02868	0.02664	0.009067	0.01703	0.003817	20.21
37	854253	M	16.74	21.59	110.1	869.5	0.0961	0.1336	0.1348	0.06018	0.1896	0.05656	0.4615	0.9197	3.008	45.19	0.005776	0.02499	0.03695	0.01195	0.02789	0.002665	20.01
38	854268	M	14.25	21.72	93.63	633	0.09823	0.1098	0.1319	0.05598	0.1885	0.06125	0.286	1.019	2.657	24.91	0.005878	0.02995	0.04815	0.01161	0.02028	0.004022	15.89
39	854941	B	13.03	18.42	82.61	523.8	0.08983	0.03766	0.02562	0.02923	0.1467	0.05863	0.1839	2.342	1.17	14.16	0.004352	0.004899	0.01343	0.01164	0.02671	0.001777	13.3

wisc_bc_data

The dataset used for the training on the Machine Learning Model has been produced by researchers of the University of Wisconsin in Madison, USA. The dataset is made of multiple real fine-needle biopsy results combined in a csv format. A biopsy is a procedure to remove a piece of tissue or a sample of cells from your body so that it can be tested in a laboratory. Needle biopsy is a general term that's often used to describe inserting a special needle through the skin to collect cells from a suspicious area.

How The Data Flows Through The System

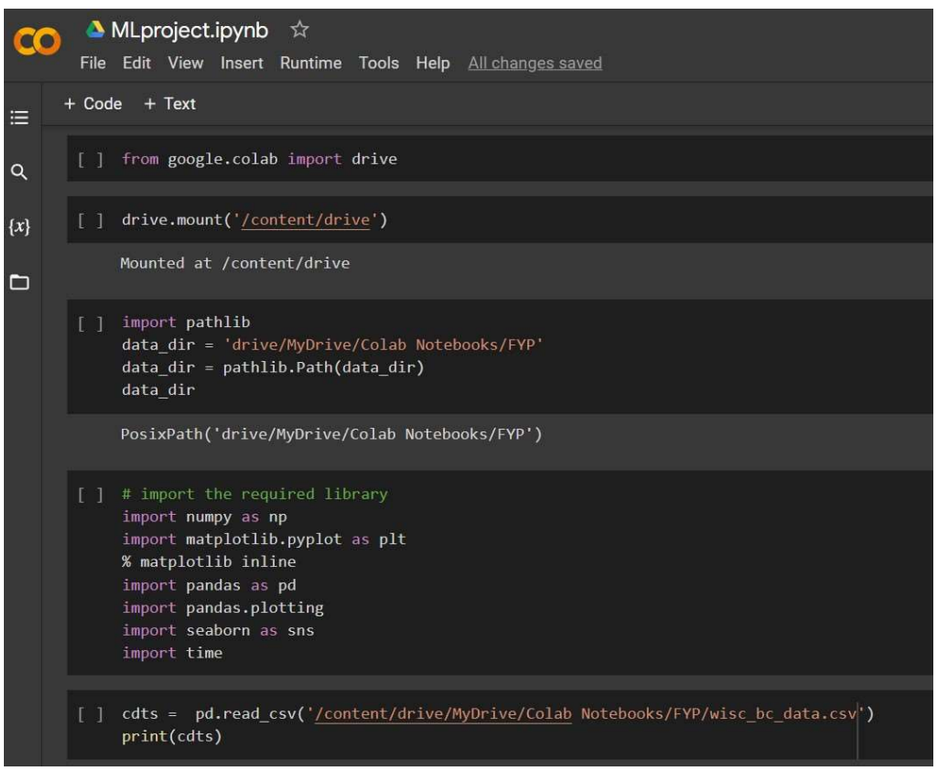
The system will consist of a front-end interface connected to an online back-end application that will process the data that has been uploaded. Here are the steps indicating how the data will flow through the system:

1. The data will be collected via from the Web User Interface
2. The data will be stored on a Cloud Storage Platform
3. The data will be fetched from the Cloud Storage Platform and sent to the back-end system
4. The system will process the data and produce a result
5. The data will be deleted from the Cloud Storage Platform

The result will be displayed on the Web User Interface

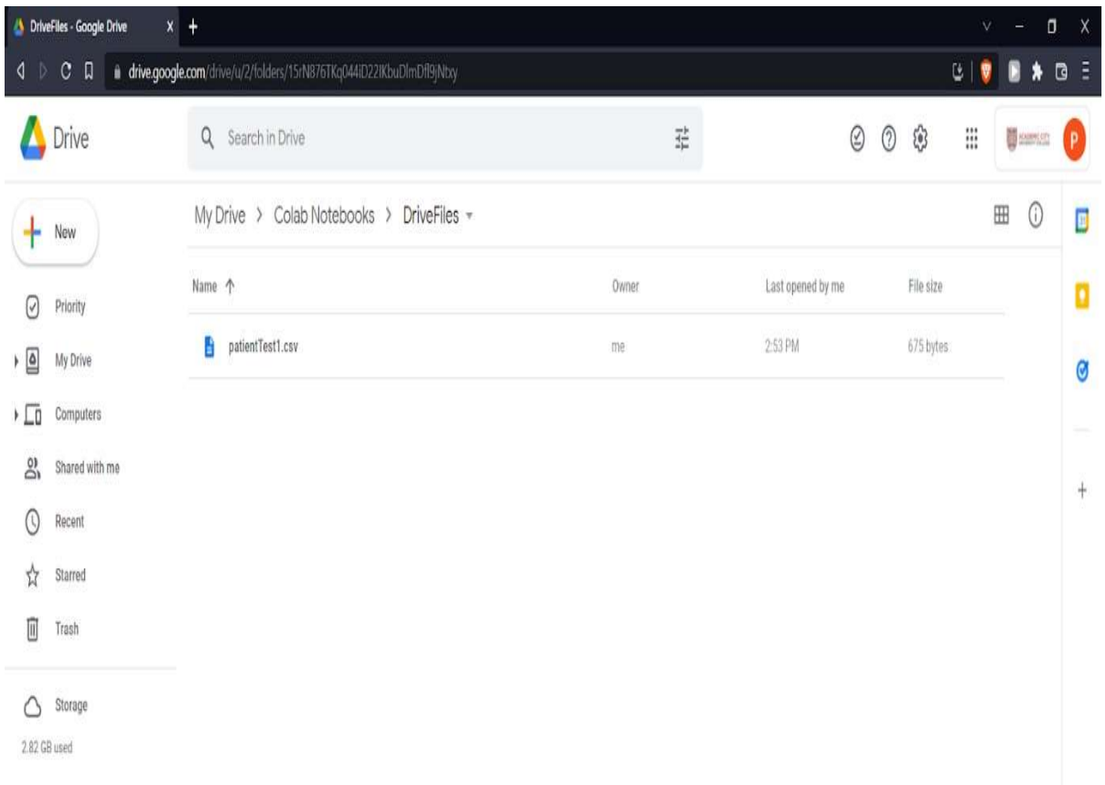
Google Collaboratory

Google Colaboratory, or "Google Colab", is a free to use Software product from Google Research. Google Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. It is the main tool used to implement the back-end system of the project.



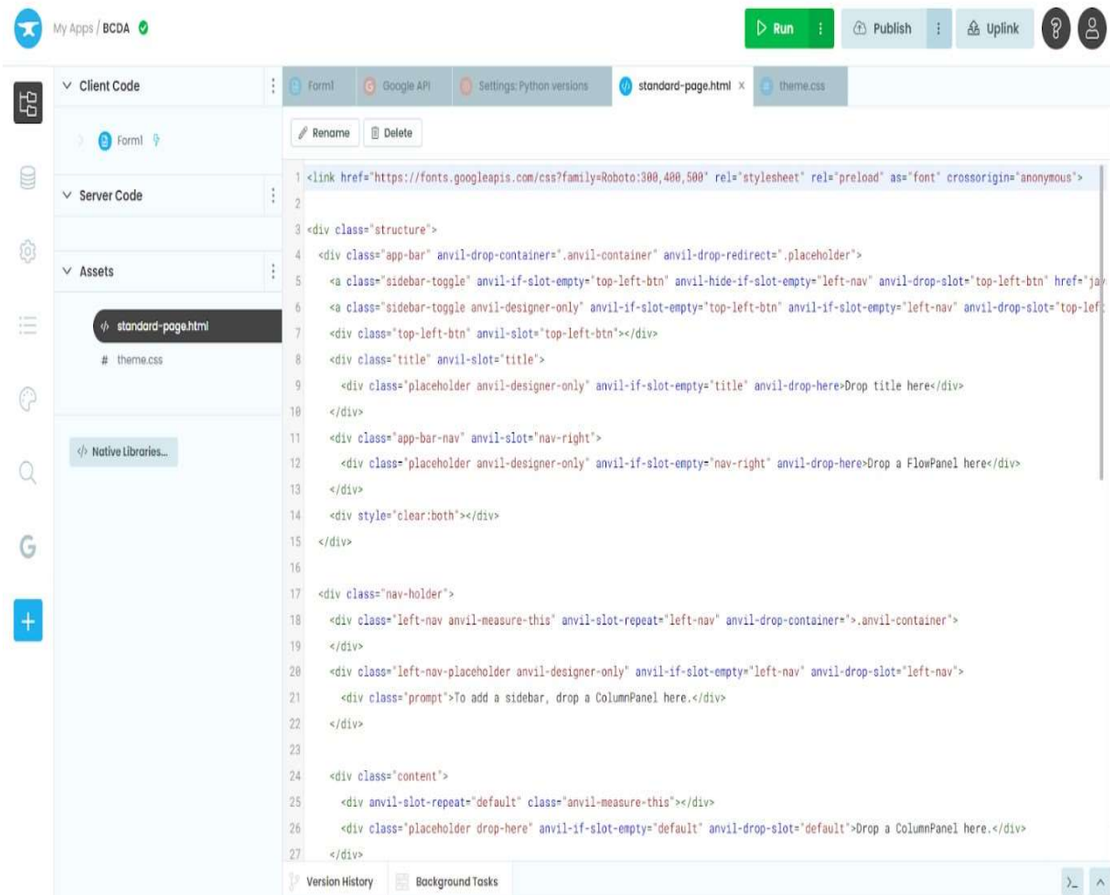
Google Drive

Google Drive is a free service from Google that allows users to store files online and access them anywhere using the cloud. It will be used as storage platform for our system.



ANVIL WORKS

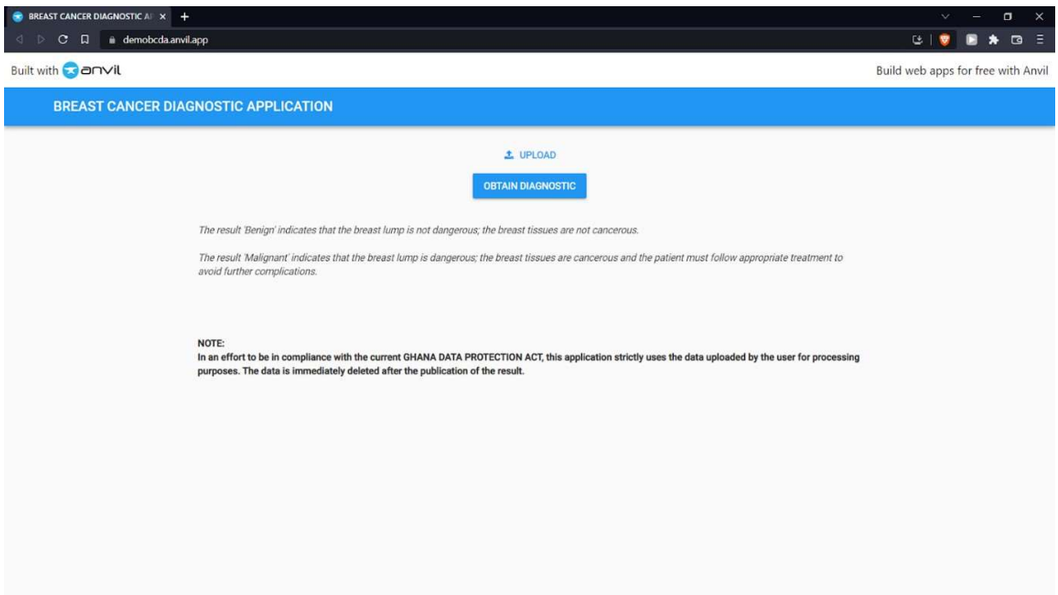
Anvil is a web framework that allows users to build and host python-based web application, with both front-end, and back-end support. It is used here to set-up the Graphical Interface, connect to the storage platform, and define functions that will be called on the back-end platform.



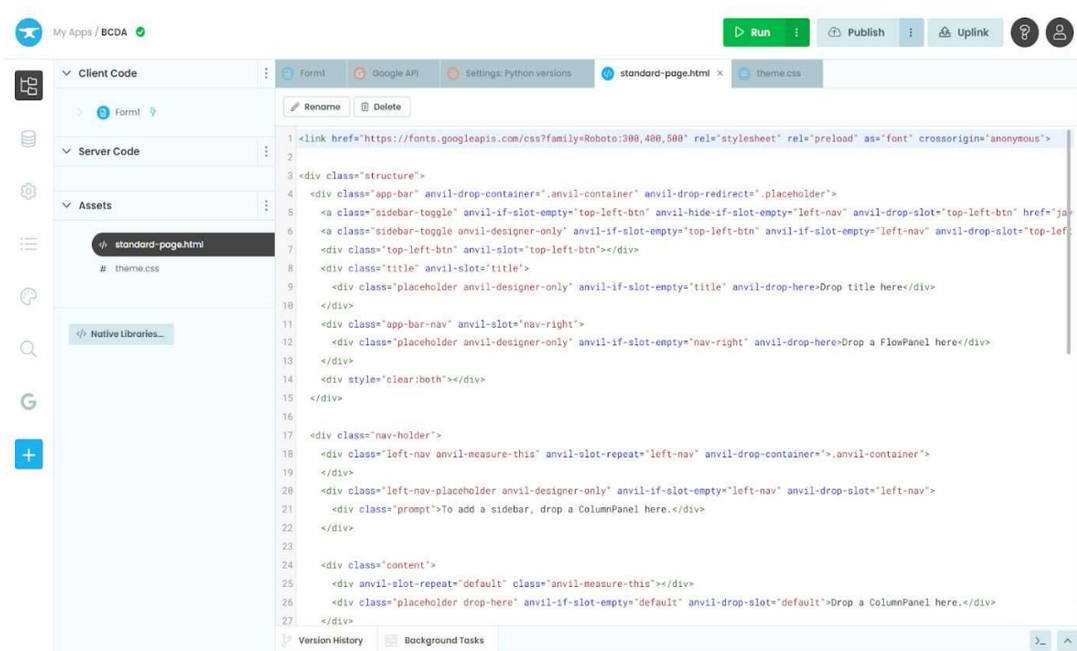
4. IMPLEMENTATION AND RESULTS

Front End – Graphical User Interface

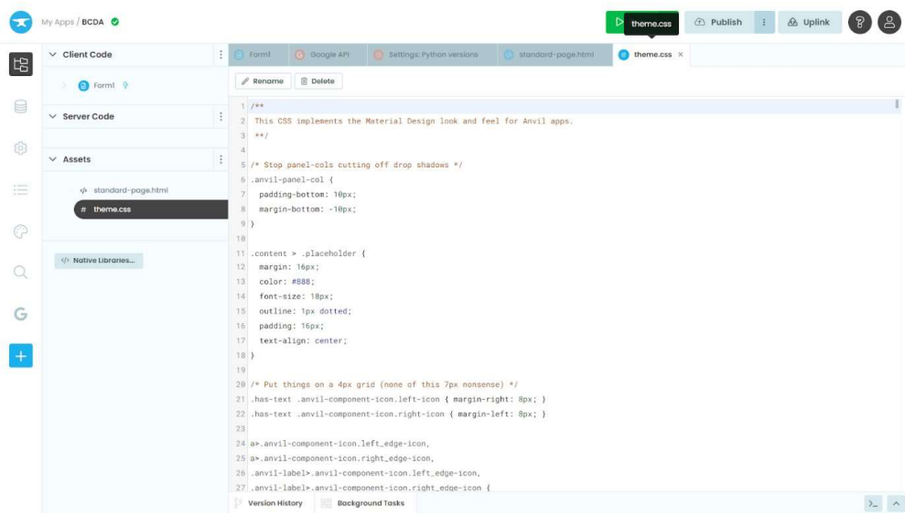
As specified earlier, the front-end has been developed using the Anvil. The Graphical interface allows the user to access the system, upload a CSV file that contains a patient's data and obtain the diagnostic. The main objective of the system is to simplify the diagnosis process as much as possible. The Interface has been designed with that in mind; anyone who knows how to use a computer can use the system. Below is a figure displaying a snapshot of the Graphical Interface:



Next is a figure showing the HTML development interface used to build the User Interface.

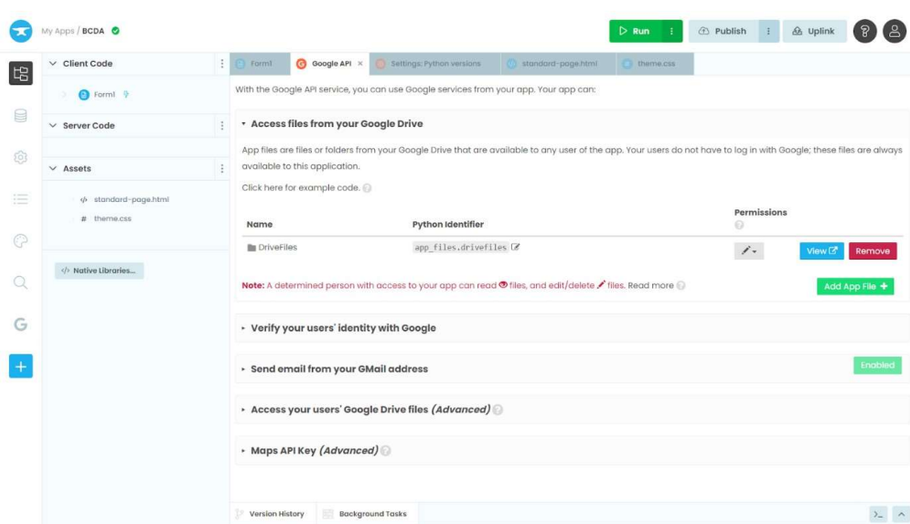


Finally, here is a figure showing the CSS development interface used to build the User Interface.

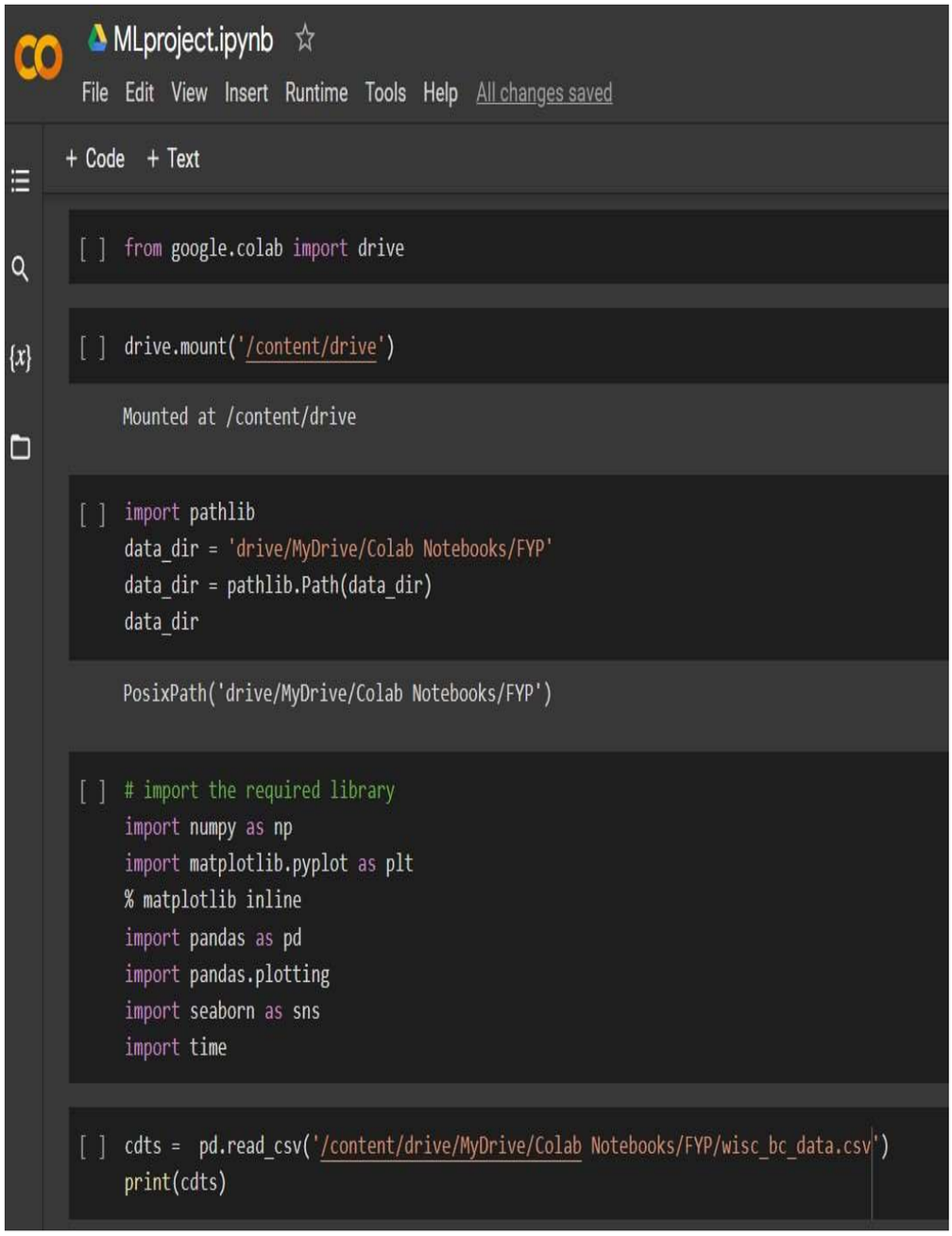


Front-End And Back-End Integration

For our system to work seamlessly, the Anvil front-end platform, the google drive storage platform and the google collab back-end platform have to beconnected to each other and exchange information. First, the Google Drive Application Programming Interface (API) is used to communicate connect to the storage platform; here is a snapshot that shows how it is done:



Next, the back-end platform has to be connected to the Storage platform to access the training data. The figure below shows it is achieved:



The back-end platform also as to fetch the user data for the diagnosis phase. For that, a special function has been created in the anvil platform to trigger that feature. That function will be called on the back-end side and will run every time there is a new file to process. The figures below show it is achieved:

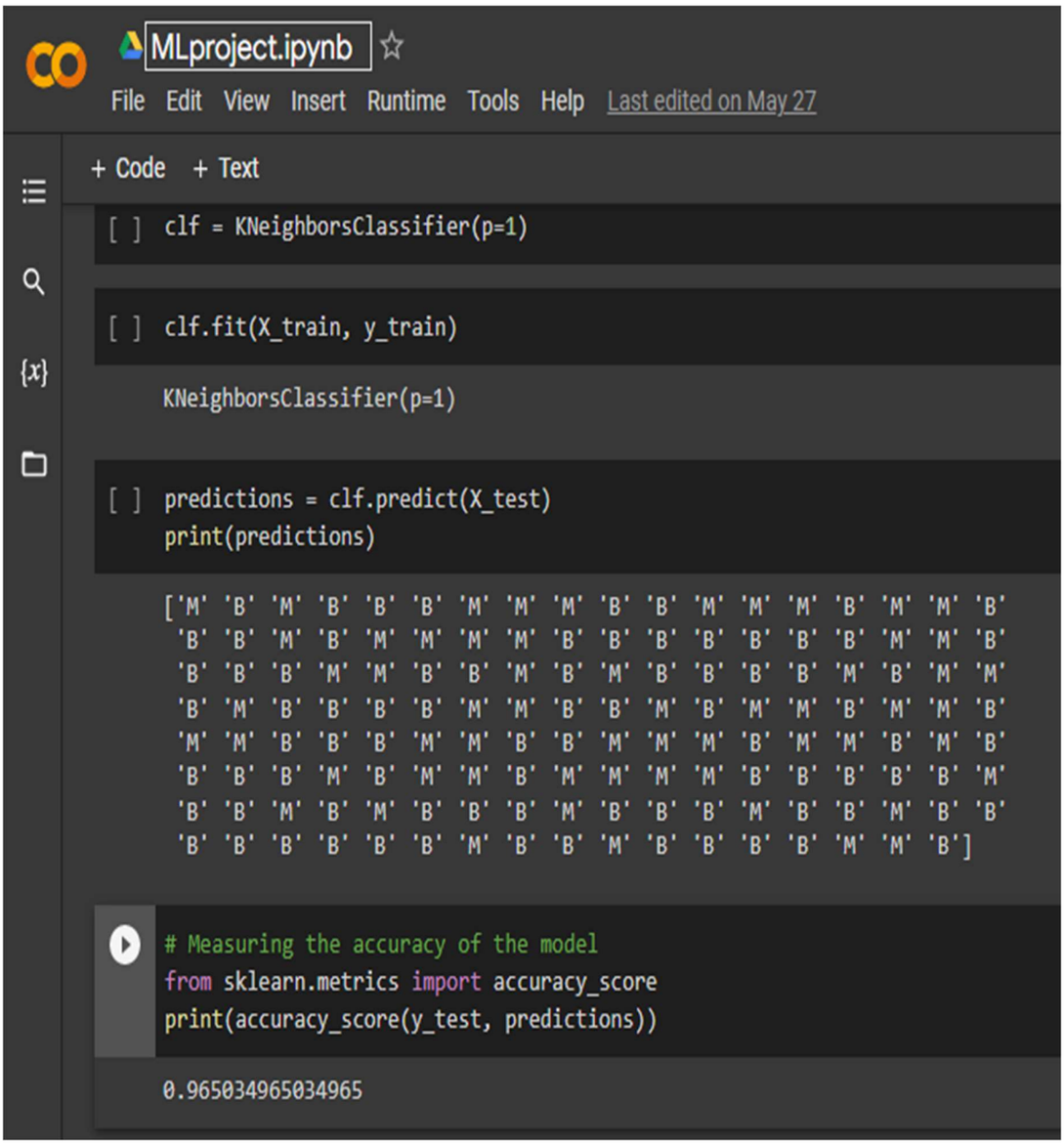
```
[ ] @anvil.server.callable
    def predict_cancer():
      time.sleep(15)
      cdtsp =  pd.read_csv('/content/drive/MyDrive/Colab Notebooks/DriveFiles/patientTest1.csv')
      print(cdtsp)
      cdtsp.pop('id')
      z = cdtsp.loc[:, cdtsp.columns]
      predictions2 = clf.predict(z)
      print(predictions2)
```

Finally, the back end had to be conned to the anvil server using the uplink key. This allows the back-end process running on Google colab to listen to a function call made fron the anvil platform and act on it. This completes the front-end and back-end integration phase. Next is a snapshot showing that process:

```
[ ] !pip install anvil-uplink

    Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
    Collecting anvil-uplink
      Downloading anvil_uplink-0.3.42-py2.py3-none-any.whl (64 kB)
         |████████████████████████████████| 64 kB 2.2 MB/s
    Requirement already satisfied: future in /usr/local/lib/python3.7/dist-packages (from anvil-uplink) (0.16.0)
    Collecting ws4py
      Downloading ws4py-0.5.1.tar.gz (51 kB)
         |████████████████████████████████| 51 kB 222 kB/s
    Collecting argparse
      Downloading argparse-1.4.0-py2.py3-none-any.whl (23 kB)
    Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from anvil-uplink) (1.15.0)
    Building wheels for collected packages: ws4py
      Building wheel for ws4py (setup.py) ... done
      Created wheel for ws4py: filename=ws4py-0.5.1-py3-none-any.whl size=45229 sha256=fd3321b26a8184fddc3b8a045e1643cdccd916e1f016dea9c59201d05c9705b6
      Stored in directory: /root/.cache/pip/wheels/29/ea/7d/3410aa0aa0e4402ead9a7a97ab2214804887e0f5c2b76f0c96
    Successfully built ws4py
    Installing collected packages: ws4py, argparse, anvil-uplink
    Successfully installed anvil-uplink-0.3.42 argparse-1.4.0 ws4py-0.5.1

[ ] import anvil.google.auth, anvil.google.drive
    from anvil.google.drive import app_files

    import anvil.server

    anvil.server.connect("server_AN4T5G7BH7FMBYLBRBM5S6GX-BMDC6XCVRCVTJ7FA")  # Server Uplink Key

    Connecting to wss://anvil.works/uplink
    Anvil websocket open
    Connected to "Published" as SERVER
```

Results

The model performed very well on the dataset with minimal configuration adjustments. An accuracy score of 0.965034965034965 (96%) was obtained using the accuracy score function accuracy_score from the scikit-learn metrics library



5. SUMMARY, CONCLUSION AND RECOMMENDATIONS

Summary

A The use of Machine Learning offers exceptional benefits in most activities and sectors. In Healthcare, it allows for simpler faster diagnostic, medical pattern discovery, analysis and prediction; it is most useful for researchers and healthcare professional as it eases various medical processes that would otherwise long, difficult and complex task. Despite the effort various institutions and individuals, the adoption of ICT and Machine Learning in the healthcare sector in Ghana is still in its early stage. The deliverable of this project is an attempt to simplify the breast cancer diagnosis process and contribute to democratization of ICT and Machine Learning in the Ghanian health sector. The application effectively allows any user to seamlessly upload and process breast cancer data points and obtain predictions/diagnostic for it. This work shows that machine learning can be used to improve enhance medical analysis and disease diagnosis.

Conclusion

Firstly, farmers who not only harvest crops but also data from their farms can educate farmers based on purely informal means. This can help measure the success or failure of crops and increase their yield. Also, farmers can be more intentional about the quality of their soils by doing soil tests to have appreciable data on what their soils need and the kind of crops to plant with the aid of soil-to-crop recommendation systems. There are initiatives that have subsidized these soil tests and farmers can take advantage of such avenues.

Recommendations and Suggestions for Further Research

Firstly, farmers who not only harvest crops but also data from their farms can educate farmers based on purely informal means. This can help measure the success or failure of crops and increase their yield. Also, farmers can be more intentional about the quality of their soils by doing soil tests to have appreciable data on what their soils need and the kind of crops to plant with the aid of soil-to-crop recommendation systems. There are initiatives that have subsidized these soil tests and farmers can take advantage of such avenues.

BIBLIOGRAPHY

1. Nana Kwame Gyamfi, Kumasi Technical University Department of Computer Science
2. David Otoo-Arthur, Presbyterian Women's College of Edu, Ghana
3. Building Ghana's Universal Electronic Health Records (EHR) With Medtrack https://www.modernghana.com/news/1028050/building-ghanas-universal-electronic-healthrecor.html
4. Achampong, Emmanuel Kusi. "Electronic Health Record System: A Survey in
5. Ghanaian Hospitals." Journal of Health & Medical Informatics, vol. 03, no. 02, 2012, 10.4172/scientificreports.164. Accessed 4 Aug. 2018.
6. Thomas, Abigail S., et al. "Breast Cancer in Ghana: Demonstrating the Need for Population-Based Cancer Registries in Low- and Middle-Income Countries." Journal of Global Oncology, vol. 3, no. 6, Dec. 2017, pp.
7. 765–772, 10.1200/jgo.2016.006098. Accessed 15 Feb. 2022.

8. Health providers' readiness for electronic health records adoption: A cross-sectional study of two hospitals in northern Ghana. PLoS ONE 15(6): e0231569. Kwasi Torpey, University of Ghana College of Health Sciences, GHANA https://doi.org/10.1371/journal.pone.0231569
9. Yusif, S., Hafeez-Baig, A., & Soar, J. (2020). Change management and adoption of health information technology (HIT)/eHealth in public hospitals in Ghana: A qualitative study. Applied Computing and Informatics. https://doi.org/10.1016/j.aci.2019.11.007
10. Kesse-Tachi, A., Asmah, A. E., & Agbozo, E. (2019). Factors influencing adoption of eHealth technologies in Ghana. DIGITAL HEALTH, 5,
11. 205520761987142. https://doi.org/10.1177/2055207619871425
12. Kesse-Tachi, A., Asmah, A. E., & Agbozo, E. (2019). Factors influencing adoption of eHealth technologies in Ghana. DIGITAL HEALTH, 5,
13. 205520761987142. https://doi.org/10.1177/2055207619871425