Ovarian Cancer Analysis

Introduction:-

The ovaries are the part of the female reproductive system which produce women eggs and female hormones. The eggs which are produced are known as *ova* and hormones such as *estrogen*, *progesterone* and *testosterone*. Ovarian cancer refers to cancerous growth that beings in this ovaries. Ovarian cancer occurs when abnormal cells in the ovary multiply out of control and form a tumor. If this tumor is not treated in the early stages or left as it is this tumor can spread to other parts of the body. We don't yet know exactly what causes most ovarian cancers but there are some risk factors which can sometimes leads to formation of tumor.

Risk Factors:-

- Being overweight or obese.
- Having family cancer syndrome.
- Having children late (*after age 28*) or never having full term pregnancy.
- Taking hormone therapy after menopause

Types of ovarian Cancer:-

There are over 30 types of ovarian cancer depending upon type of cell in which it starts.

In most of the cases around 90% of time its **ovarian epithelial cancer**. It starts on the outside of the ovaries and it's named for the cells that's makes up the surface. These are called epithelial cells.

Diagnosing ovarian cancer:

Test that may be used to diagnose ovarian cancer may include

- Complete blood count
- CA 125 test
- CA 19-9 test
- HE4 test
- Creatinine test
- CEA test
- ALT test

Treatment:

The most important aspect of treating the ovarian cancer is it stage .Stage refers to how far the cancer has spread. So detecting the ovarian cancer in the early stages can be beneficial in its treatment and have high probability of getting it cured. Also the treatment of ovarian cancer depends on type of ovarian cancer and whether the female wants to have children in future. The initial treatment for ovarian cancer is surgery to remove the tumor.

Methodology:-

There is high probability of getting cured from ovarian cancer if it is diagnose in the earlier stages. One have to go through following test to diagnose ovarian cancer

- **Complete blood test (CBC):** The CBC test usually includes white blood cell (WBC) count, Red blood cell count (RBC), Platelet count and Hemoglobin.

- **CA 125:** Carbohydrate antigen 125 (CA 125) measures the amount of cancer antigen 125 in blood.
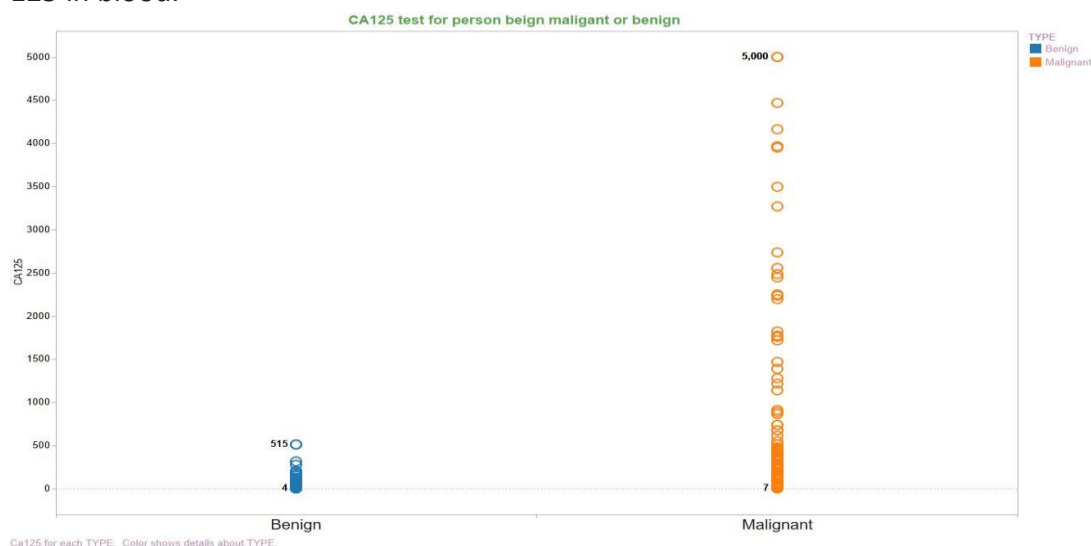

Fig.1

The Fig.1 shows that when CA125 test values is higher there is high probability that tumor is malignant but CA125 test isn't accurate enough to use for ovarian cancer screening in general because many noncancerous condition can increase the CA125 level. As one can see in Fig.1 that tumor is benign but CA125 level is high (515). Many different condition can cause increase in CA125, including normal condition such as menstruation.

- **CA 19-9:** Carbohydrate antigen 19-9 (CA 19-9) is a type of antigen released by pancreatic cancer cells. It can also be referred as tumor marker. High level of CA-19-9 are often a sign of pancreatic cancer.
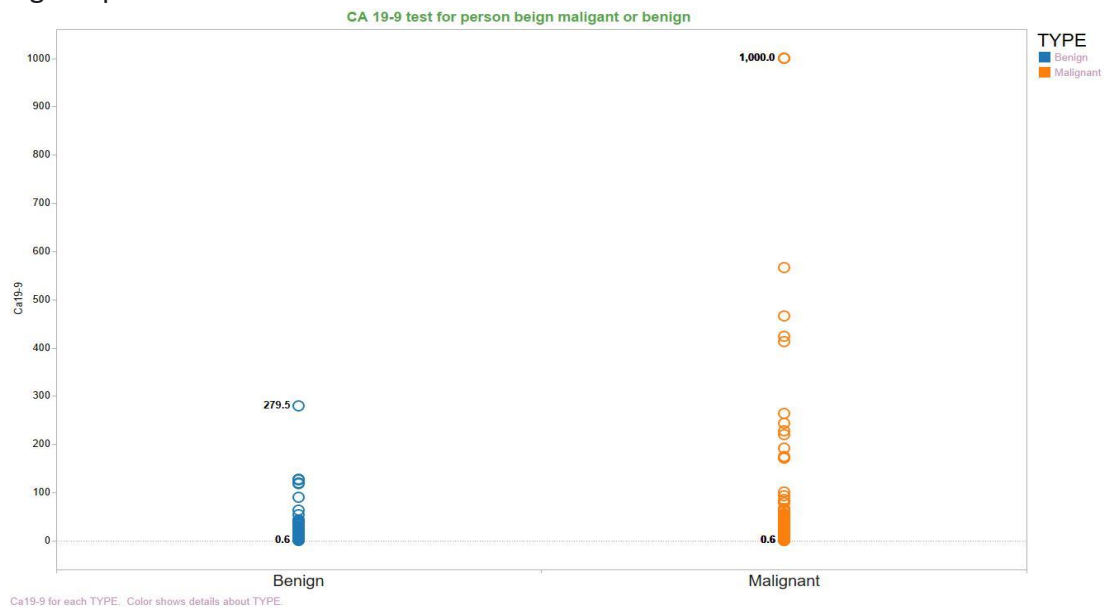
Fig.2

As we saw in CA 125 that it was not accurate enough same we can conclude in this case to .By analyzing the Fig.2 that this test is not accurate enough to diagnose ovarian cancer. High level of CA 19-9 can also be found in a person with liver disease.

- **HE4:** HE4 test measures the amount of human epididymis protein 4 in blood. HE4 are frequently present in the blood of the women having epithelial ovarian cancer.
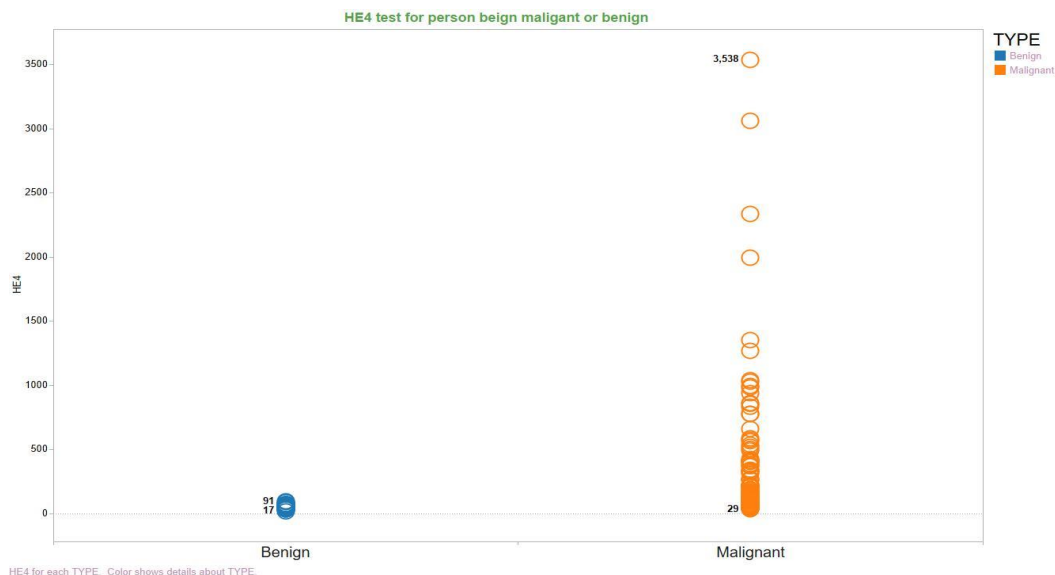
Fig.3

In Fig.3 it is clearly visible that Female with low level of HE4 are benign and one with high level of human epididymis protein 4 (HE4) in blood are having malignant tumor. But in ovarian cancer screening it is possible that cancerous condition can have low level of HE4.

Note: HE4 test should not be used as screening test for monitoring person with mucinous or germ cell ovarian cancer. These types of cancer rarely express HE4.

- **Creatinine:** Creatinine can be measured to check how well the kidney are working. Higher level of creatinine could mean that kidney are not working well and that may be due to cancer.

Creatinine test for tumor beign maligant or benign
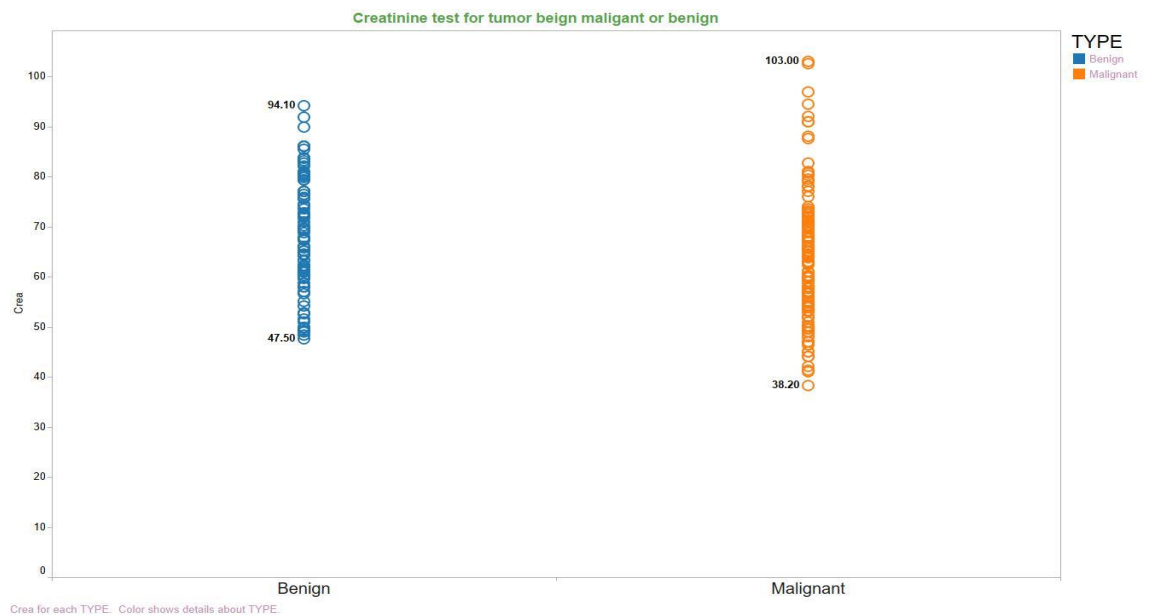
Fig.4

Creatinine is also not significant to diagnose ovarian cancer because there is nearly same range of creatinine for tumor benign or malignant.

- **CEA:** A carcinoembryonic antigen (CEA) test is a blood test used to help diagnose and manage cancers.
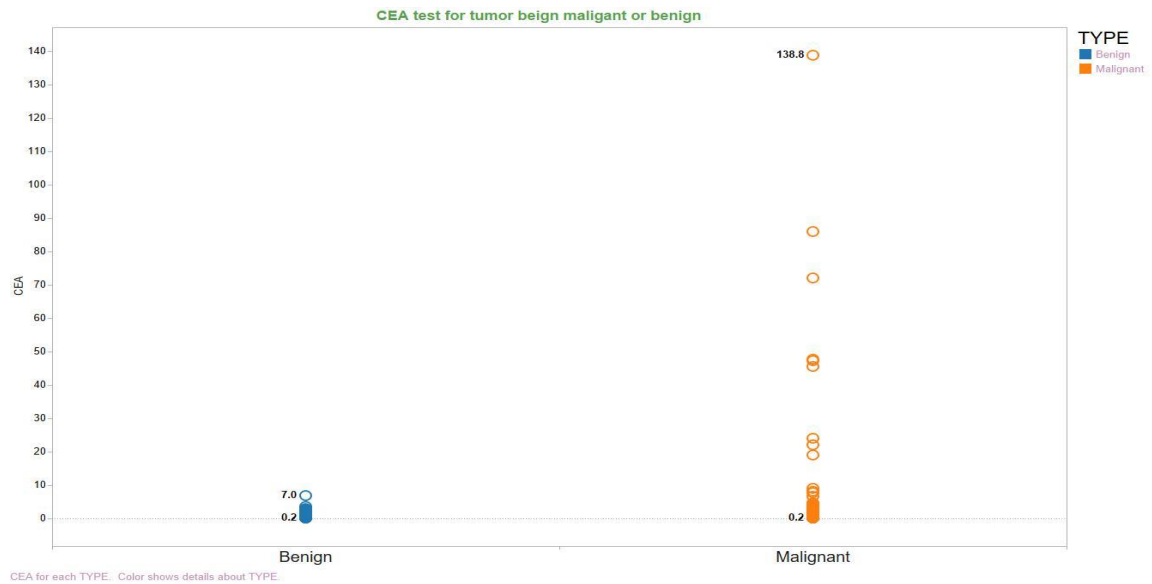
Fig.5

High level of CEA can be sign of certain types of cancer. By analyzing the figure it can be conclude that there is also probability of tumor being malignant even if CEA level is low.

- **ALT:** The alanine aminotransferase (ALT) test is a blood test that checks for liver damage.
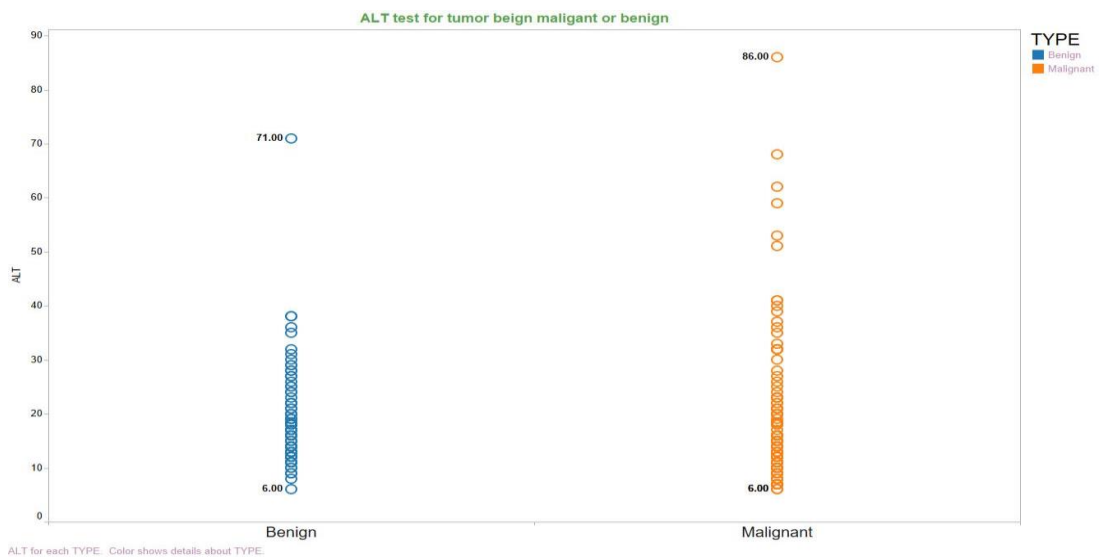
Fig.6

By analyzing the Fig.5 it is not possible to completely diagnose ovarian cancer using ALT test.

Now we have come to conclusion that we can't completely be reliable on any test for diagnosing ovarian cancer. So what if we combine all the test to diagnose ovarian cancer? But question arises how it is possible?

At this moment we can take the help of the technology known as Machine Learning in general terms Artificial Intelligence (AI).In Machine Learning based on the past data we train the model (Algorithm). We feed data (refer Fig.7) to the model on which it gets train and later on we can predict based on features (tests) if tumor is malignant or not.

TYPE	Age	CEA	CA125	CA19-9	HE4	CREA	ALT
0	47	1.40	15.36	36.48	208.552698	103.0	11.0
0	61	2.46	2444.00	19.98	934.100000	45.0	9.0
0	39	0.77	56.08	12.18	47.560000	48.0	9.0
0	45	0.82	2555.00	18.41	853.500000	65.7	16.0
0	45	0.42	1391.00	11.15	404.900000	70.3	21.0

Fig.7

 Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves.

Procedure:-

Data Collection: Collected ovarian cancer diagnose blood test data from kaggle. The dataset has 235 patients blood test record which can be adopted to build a model. Data initially had 49 columns (features) from which after doing exploratory data analysis we left of with just 8 columns which we made use to train a model which can predict if person has ovarian cancer or not.

Model Building: We build a model using Support Vector Machine (SVM)   algorithm which gave us around 94% accuracy which mean when we pass a particular observation to the data there are 94% chances that it is predicted correctly which is far better than using a single test to diagnose ovarian cancer. To evaluate the model we use confusion matrix.

		Actual	
		Malignant	Benign
Predicted	Malignant	20	5
	Benign	1	82

Fig.7

In Fig.7 confusion matrix 20 observation were malignant and model correctly predicted it as malignant same way 82 observation were benign and model predicted it as benign. But at some point our model made errors 5 observations were Benign but our model

predicted it as malignant (Type 1 error) and 1 observation was malignant and model predicted it as benign (Type 2 error). Our motive should be to reduce error especially type 2 error because if the person in malignant and model predicted it as benign than there are chances that tumor may spread to others parts of the body if proper treatment is not taken.

Conclusion:

It is difficult to diagnose ovarian cancer based on single test. So with the help of machine learning technology we build a model to predict if person has ovarian cancer or not. Also we have deployed model into API https://ovarian-cancer-diagnosis.herokuapp.com/ .

References

Brazier, Y. (2019). What is ovarian cancer?

Ka-ChunWong, & JunyiChen. (2019). Early Cancer Detection from Multianalyte Blood Test Results.

Lights, V. (2019). Ovarian Cancer.

Pietrangelo, A. (2020). What are the Early Signs of Ovarian Cancer and How Do You Detect Them?

team, T. A. (2018). What Is Ovarian Cancer?