**BREAST CANCER PREDICTION USING MACHINE LEARNING ALGORITHMS**

OVERVIEW

Breast cancer is the most common cancer among women in the world. It account for 25% of all cancer cases, and affected over 2.1 Million people in 2015. It starts when cells in the breast begin to grow out of control. These cells usually form a tumor that can often be seen on an X-ray or felt as a lump.

Early diagnosis significantly increases the chances of surviver. The key challenges against it's detection is how to classify tumors into malignant(Cancer) or benign(not cancer). A tumor is considered malignant (Cancer) if the cells can grow into surrounding tissues or spread to distant areas of the body. A benign tumor does not invade nearby tissue or spread to other parts of the body the way cancer can. But benign tumors can be serious if they press on vital structures such as blood vessel or nerves.

Machine Learning technique can dramatically improve the level of diagnosis in breast cancer. Research shows that experience physicians can detect cancer by 79% accuracy, while 91%(up to 97%) accuracy can be achieved using Machine Learning techniques.

SUMMARY :

The breast cancer detection model using different techniques achieved the following results:

1. Logistic Regression:
   * The model achieved an accuracy of 96%.
   * The precision for class 0 (non-cancerous) was 96%, indicating that 96% of the predicted non-cancerous cases were correct.
   * The recall for class 0 was 97%, suggesting that the model correctly identified 97% of the actual non-cancerous cases.
   * The precision for class 1 (cancerous) was 96%, indicating that 96% of the predicted cancerous cases were correct.
   * The recall for class 1 was 94%, suggesting that the model correctly identified 94% of the actual cancerous cases.
2. Decision Tree:
   * The model achieved an accuracy of 93%.
   * The precision for class 0 was 94%, indicating that 94% of the predicted non-cancerous cases were correct.
   * The recall for class 0 was 96%, suggesting that the model correctly identified 96% of the actual non-cancerous cases.
   * The precision for class 1 was 93%, indicating that 93% of the predicted cancerous cases were correct.
   * The recall for class 1 was 91%, suggesting that the model correctly identified 91% of the actual cancerous cases.
3. Random Forest:
   * The model achieved an accuracy of 97%.
   * The precision for class 0 was 97%, indicating that 97% of the predicted non-cancerous cases were correct.
   * The recall for class 0 was 99%, suggesting that the model correctly identified all the actual non-cancerous cases.
   * The precision for class 1 was 98%, indicating that all the predicted cancerous cases were correct.
   * The recall for class 1 was 96%, suggesting that the model correctly identified 96% of the actual cancerous cases.
4. SVM (Support Vector Machines):
   * The model achieved an accuracy of 97%.
   * The precision for class 0 was 100%, indicating that all the predicted non-cancerous cases were correct.
   * The recall for class 0 was 95%, suggesting that the model correctly identified 95% of the actual non-cancerous cases.
   * The precision for class 1 was 94%, indicating that 94% of the predicted cancerous cases were correct.
   * The recall for class 1 was 100%, suggesting that the model correctly identified all the actual cancerous cases.

Conclusion:

In this study, multiple Machine Learning techniques were employed to detect breast cancer. Overall, the models demonstrated good performance, with accuracy ranging from 91% to 97%.

Among the models evaluated, Random Forest achieved the highest accuracy of 97%. It exhibited excellent precision and recall for both non-cancerous and cancerous cases, with no false negatives for non-cancerous cases.

SVM also performed well, achieving an accuracy of 96%. It exhibited high precision and recall for both classes, indicating its ability to accurately classify breast cancer cases.

Logistic Regression and Decision Tree achieved similar accuracies of 96% and 93%, respectively. While their precision and recall scores were generally good, they were slightly lower compared to Random Forest and SVM.

In conclusion, both Random Forest and SVM show promise for breast cancer detection. These models can aid in early detection and contribute to more effective treatment strategies. However, further evaluation on larger datasets and validation is necessary to assess their generalizability and robustness. Thus as we mentioned earlier in overview that using Machine learning we can attain an accuracy of around 97%.