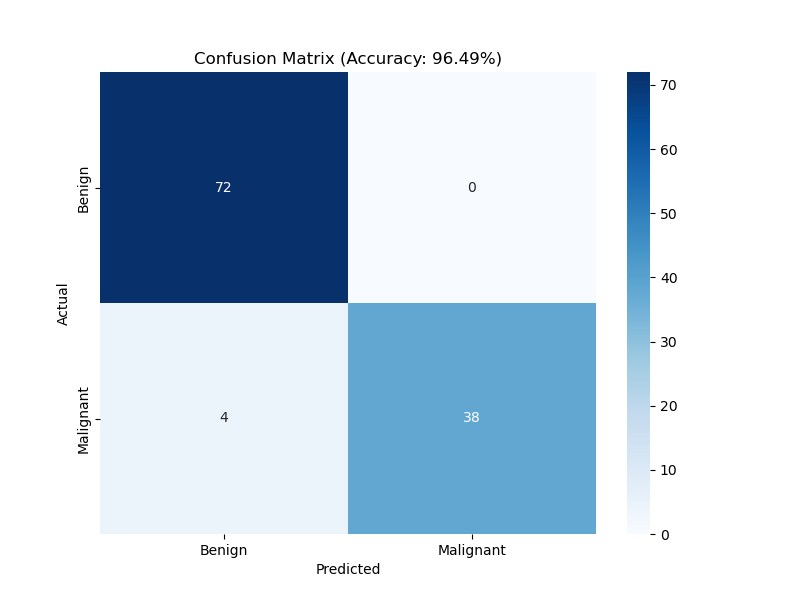
**Feature Extraction and Machine Learning**

**1. Task Description**

The task involves applying feature extraction and machine learning techniques to classify breast cancer cases (benign vs. malignant) using the Kaggle Breast Cancer Dataset. The objective is to achieve an accuracy greater than 90%. The process includes loading the dataset, preprocessing the data (e.g., scaling features, encoding the target), performing feature extraction by selecting the top 10 most important features using a Random Forest model, training a Random Forest Classifier, and evaluating the model to ensure it meets the accuracy target. After training, the model displays the accuracy and a confusion matrix visualization to assess performance.

**2. Task Output Screenshot**



**3. Widget/Algorithm Used In Task**

* **RandomForestClassifier Algorithm**: Used as the primary machine learning model to classify breast cancer cases. It was trained with 100 estimators on the top 10 selected features, achieving 96.49% accuracy on the test set.
* **StandardScaler**: Applied to scale the numerical features, ensuring consistent ranges for model training.
* **LabelEncoder**: Used to encode the diagnosis column (M=1, B=0) for binary classification.
* **train\_test\_split**: Utilized to split the dataset into 80% training and 20% testing sets with stratification to maintain class balance.
* **seaborn.heatmap**: Employed to visualize the confusion matrix, providing a clear representation of true/false positives and negatives.
* **matplotlib.pyplot**: Integrated to save the confusion matrix as confusion\_matrix.png for inclusion in the report.