1. Introduction

Problem Statement and Research Motivation (100 words)

Breast cancer can progress rapidly, with malignant tumors posing the greatest risk because they can spread to other parts of the body, while benign tumors typically remain confined. Early detection and accurate classification are essential for effective treatment. One important metric is the average tumor radius (radius_mean), as malignant tumors tend to have larger radii compared to benign ones. This study aims to examine whether the difference in radius_mean between malignant and benign tumors is significant enough to serve as a practical marker for diagnosis, potentially simplifying early detection and supporting medical decision-making.

The Dataset (75 words)

This study uses the Kaggle "Cancer Data" dataset, which contains 569 patient records with 29 variables, including both categorical and numerical features. It includes patient IDs, diagnostic labels (M for Malignant, B for Benign), and various measurements of cell nuclei properties. Any missing values were removed during pre-processing to maintain data integrity. This dataset provides a solid basis for analyzing the relationship between tumor size and cancer classification.

Research Question (50 words)

RQ: "Does the mean tumor radius (radius_mean) significantly differ between Malignant and Benign breast tumors?"

This research explores whether radius_mean alone can serve as an effective indicator for distinguishing between tumor types, which could provide a simple, practical method for early detection and support clinical decision-making.

---

2. Null Hypothesis and Alternative Hypothesis (H_o/H_1) (100 words)

- **Null Hypothesis (H_o):** There is no significant difference in the average tumor radius between Malignant and Benign tumors.
  This suggests that tumor size alone may not be sufficient to distinguish cancer types.
- **Alternative Hypothesis (H_1):** There is a significant difference in the average tumor radius between Malignant and Benign tumors.
  If confirmed, this would indicate that radius_mean can act as a simple and reliable diagnostic feature, allowing doctors to quickly identify high-risk tumors and prioritize timely treatment.

---

3. Background Research

Relevant Studies (200 words)

1. **Abo El Nasr et al. (2024):** This study analyzed tumor features with a focus on radius_mean. Using the Kaggle dataset, the authors explored correlations and applied Variance Inflation Factor (VIF) analysis. They found that perimeter, area, smoothness, compactness, and concave points were strongly linked to radius_mean, emphasizing its central role among tumor metrics.
2. **Girdhar & Raju (2022):** The researchers investigated whether nuclear measurements could differentiate malignant from benign tumors. Using ImageJ software, they measured cell nuclei and concluded that nuclei in malignant tumors were consistently larger than those in benign samples, highlighting tumor size as a key distinguishing factor.
3. **Gity et al. (2018):** This work examined MRI measurements to classify breast tumors. Statistical tests comparing lesion sizes showed that malignant tumors were larger than benign ones, supporting the relevance of tumor size for diagnostic purposes.

Why This Research Question Matters (100 words)

Previous studies show that malignant tumors are generally larger than benign ones, and radius_mean correlates strongly with other tumor features. However, many studies use multiple features or complex methods for classification. This research focuses solely on radius_mean to investigate whether a single, easily measurable metric can effectively differentiate tumor types. By simplifying the classification process, it could reduce computational complexity, minimize reliance on expensive diagnostic tools, and provide a practical, quick method for early detection, supporting clinicians in making informed decisions.

Description

I have worked on this three topics.

Nihanth Reddy Maram

24065329