Breast Cancer Diagnosis Analysis

February 05, 2026

--- title: "Breast Cancer Diagnosis Analysis" subtitle: "Predict breast cancer diagnosis (malignant vs benign) from digitized cell nuclei images using machin" author: "Jotty SwarmMLComprehensive" date: "February 05, 2026" geometry: "margin=0.9in" fontsize: 11pt documentclass: article classoption: twoside colorlinks: true linkcolor: NavyBlue urlcolor: NavyBlue toccolor: NavyBlue toc-depth: 3 numbersections: true header-includes: Typography - \usepackage{fontspec} - \setmainfont{DejaVu Seri

Tables - \usepackage{booktabs} - \usepackage{longtable} - \usepackage{array} - \usepackage{multirow} - \usepackage{float} - \usepackage{tabularx} - \usepackage{colortbl} - \renewcommand{\arraystretch}{1.3}

Graphics - \usepackage{graphicx} - \usepackage{adjustbox}

Colors - \usepackage{xcolor} - \definecolor{NavyBlue}{RGB}{26,54,93} - \definecolor{TableHeader}{RGB}{44,82,130} - \definecolor{TableAlt}{RGB}{240,245,250} - \definecolor{AccentBlue}{RGB}{49,130,206} - \definecolor{SuccessGreen}{RGB}{56,161,105} - \definecolor{WarningGold}{RGB}{214,158,46} - \definecolor{DangerRed}{RGB}{229,62,62}

Header/Footer - \usepackage{fancyhdr} - \pagestyle{fancy} - \fancyhf{} - \fancyheadLE,RO{\small\thepage} - \fancyheadLO{\small\textit{Breast Cancer Diagnosis Analysis}} - \fancyheadRE{\small\textit{Jotty SwarmMLComprehensive}} - \fancyfootC{\small\textcolor{gray}{Jotty ML Comprehensive Report}} - \renewcommand{\headrulewidth}{0.4pt} - \renewcommand{\footrulewidth}{0.2pt}

Title page styling - \usepackage{titling} - \pretitle{\begin{center}\LARGE\bfseries\color{NavyBlue}} - \posttitle{\par\end{center}\vskip 0.5em} - \preauthor{\begin{center}\large} - \postauthor{\par\end{center}} - \predate{\begin{center}\large} - \postdate{\par\end{center}}

Section styling - \usepackage{titlesec} - \titleformat{\section}{\Large\bfseries\color{NavyBlue}}{\thesection}{1em}{} - \titleformat{\subsection}{\large\bfseries\color{TableHeader}}{\thesubsection}{1em}{} - \titleformat{\subsubsection}{\normalsize\bfseries}{\thesubsubsection}{1em}{}

Captions - \usepackagefont=small,labelfont=bf,textfont=it{caption}

Hyperref (load last) - \usepackage{hyperref} - \hypersetup{pdfauthor={Jotty SwarmMLComprehensive}, pdftitle={Breast Cancer Diagnosis Analysis}, pdfsubject={Machine Learning Analysis}}

\thispagestyle{empty} \begin{center} \vspace{2cm} {\Huge\bfseries\color{NavyBlue} Breast Cancer Diagnosis Analysis} \vspace{0.5cm}

{\Large\textit{Predict breast cancer diagnosis (malignant vs benign) from digitized cell nuclei images using machin}} \vspace{2cm}

{\large Jotty SwarmMLComprehensive} \vspace{0.3cm}

{\large February 05, 2026} \vspace{3cm}

\includegraphicswidth=0.3\textwidth{professionalreports/figures/featureimportance.png} \vfill

{\small\textit{Generated by Jotty SwarmMLComprehensive}} \vspace{0.5cm}

{\small\textcolor{gray}{Comprehensive ML Analysis Report}} \end{center} \newpage

\tableofcontents \newpage

Executive Summary

Predict breast cancer diagnosis (malignant vs benign) from digitized cell nuclei images using machine learning.

Key Results

Best Model: Logistic Regression

Performance Metrics:

| Metric | Value | |--------|-------| | Accuracy | 0.9825 | | Precision | 0.9861 | | Recall | 0.9861 | | F1 | 0.9861 | | Auc Roc | 0.9954 |

Dataset: 30 features analyzed

Data Quality Analysis

A comprehensive analysis of data quality, identifying potential issues before modeling.

Dataset Overview

| Metric | Value | |--------|-------| | Total Samples | 114 | | Total Features | 30 | | Numeric Features | 30 | | Categorical Features | 0 | | Features with Missing | 0 | | Total Missing Values | 0 (0.00%) |

Distribution Analysis

| Feature | Skewness | Kurtosis | Assessment | |---------|----------|----------|------------| | mean radius | 1.03 | 1.35 | Right-skewed | | mean texture | 0.25 | -0.40 | Symmetric | | mean perimeter | 1.07 | 1.56 | Right-skewed | | mean area | 1.93 | 5.46 | Right-skewed, Heavy-tailed | | mean smoothness | 0.53 | 1.50 | Right-skewed | | mean compactness | 0.91 | 0.54 | Right-skewed | | mean concavity | 1.44 | 2.46 | Right-skewed | | mean concave points | 1.15 | 1.18 | Right-skewed | | mean symme

Feature Distributions

!Feature Distributions(professionalreports/figures/distributions.png)

Outlier Analysis

Method: Interquartile Range (IQR) with 1.5x multiplier

Total Outliers Detected: 96 across 27 features

| Feature | Outliers | % of Data | Min | Max | |---------|----------|-----------|-----|-----| | area error | 10 | 8.8% | 8.61 | 542.20 | | fractal dimension error | 7 | 6.1% | 0.00 | 0.01 | | concavity error | 6 | 5.3% | 0.00 | 0.09 | | mean concavity | 5 | 4.4% | 0.00 | 0.36 | | worst symmetry | 5 | 4.4% | 0.20 | 0.48 | | mean area | 4 | 3.5% | 181.00 | 2501.00 | | radius error | 4 | 3.5% | 0.12 | 2.55 | | perimeter error | 4 | 3.5% | 0.77 | 18.65 | | smoothness error | 4 | 3.5% | 0.00 | 0.02 |

Outlier Distribution

!Outlier Boxplot(professionalreports/figures/outlierboxplot.png)

Correlation & Multicollinearity Analysis

Understanding feature relationships is critical for model interpretation and feature selection.

Correlation Matrix

!Correlation Matrix(professionalreports/figures/correlationmatrix.png)

Highly Correlated Feature Pairs (|r| >= 0.7)

| Feature 1 | Feature 2 | Correlation | |-----------|-----------|-------------| | mean radius | mean perimeter | 0.998 | | worst radius | worst perimeter | 0.994 | | mean perimeter | mean area | 0.983 | | mean radius | mean area | 0.983 | | worst radius | worst area | 0.980 | | radius error | perimeter error | 0.979 | | worst perimeter | worst area | 0.976 | | mean perimeter | worst radius | 0.969 | | mean perimeter | worst perimeter | 0.969 | | mean radius | worst radius | 0.968 | | mean area |

Variance Inflation Factor (VIF)

VIF measures multicollinearity. VIF > 5 indicates moderate, VIF > 10 indicates severe multicollinearity.

| Feature | VIF | Assessment | |---------|-----|------------| | mean radius | 79262.63 | Critical | | mean perimeter | 72990.12 | Critical | | worst radius | 22342.62 | Critical | | worst perimeter | 12079.97 | Critical | | worst area | 2216.27 | Critical | | mean area | 1946.14 | Critical | | worst fractal dimension | 1185.56 | Critical | | mean fractal dimension | 1108.49 | Critical | | worst smoothness | 849.75 | Critical | | mean smoothness | 811.67 | Critical | | worst texture | 667.77 | Cr