

Bachelor thesis

Comparative Analysis of Classifiers for Breast Cancer Detection with
Visualizations



Iván Sotillo del Horno

**UNIVERSIDAD AUTÓNOMA DE MADRID
ESCUELA POLITÉCNICA SUPERIOR**



Bachelor as Ingeniería Informática (modalidad bilingüe)

BACHELOR THESIS

**Comparative Analysis of Classifiers for Breast
Cancer Detection with Visualizations**

**Author: Iván Sotillo del Horno
Advisor: Alejandro Bellogín Kouki**

febrero 2024

All rights reserved.

No reproduction in any form of this book, in whole or in part
(except for brief quotation in critical articles or reviews),
may be made without written authorization from the publisher.

© February 2024 by UNIVERSIDAD AUTÓNOMA DE MADRID
Francisco Tomás y Valiente, nº 1
Madrid, 28049
Spain

Iván Sotillo del Horno

Comparative Analysis of Classifiers for Breast Cancer Detection with Visualizations

Iván Sotillo del Horno

PRINTED IN SPAIN

A mi madre y a mi abuela, cuya lucha contra el cáncer de mama me ha inspirado a realizar este trabajo.

DRAFT

RESUMEN

Esta tesis presenta un análisis comparativo de clasificadores para la detección del cáncer de mama y el uso de Inteligencia Artificial Explicable (XAI) para interpretar los resultados. En la fase inicial se realizará la construcción y optimización de los modelos de clasificación, estos clasificadores analizarán los resultados de las biopsias de aguja fina y clasificarán las muestras como benignas o malignas.

Posteriormente, se realiza una comparación de rendimiento comparando métricas como la puntuación F1 o la *recall*. El objetivo es identificar el mejor clasificador de acuerdo a nuestras métricas. Una vez encontrado el mejor modelo, nos adentramos más en él para entender cómo funciona. Para esto, utilizaremos SHAP (SHapley Additive exPlanations), un método de XAI que nos permite ver la importancia de cada característica y cómo contribuyen a la decisión final del modelo. Esto nos permitirá no solo clasificar las muestras, sino también entender por qué el modelo ha tomado esa decisión, lo que puede ser un avance en la comprensión de los modelos de IA para fines médicos.

PALABRAS CLAVE

Detección de Cáncer de Mama, Clasificadores, Análisis Comparativo, Interpretabilidad, SHAP, IA Explicable, Visualización

ABSTRACT

This thesis presents a comparative analysis of base and ensemble classifiers for breast cancer detection and the use of eXplainable AI (XAI) to interpret the results. The initial phase involves constructing and optimizing the classifier models, these classifiers will analyze the results from fine needle biopsy aspirations and classify the samples as benign or malignant.

Following this, a performance comparison is conducted comparing metrics such as the F1 score or the recall. The aim is to identify the best classifier regarding our metrics. Once the best classifier model is found, we dive deeper into it to understand how it works. For this, we will use SHAP (SHapley Additive exPlanations), a method of XAI (eXplainable AI) that allows us to see the importance of each feature, and how they contribute to the final decision of the model. This will allow us to not only classify the samples but also to understand why the model has made that decision which can be a step forward in understanding AI models for medical purposes.

KEYWORDS

Breast Cancer Detection, Classifiers, Comparative Analysis, Interpretability, SHAP, eXplainable AI, Visualization

TABLE OF CONTENTS

DRAFT

LISTS

List of algorithms

List of codes

List of equations

List of figures

List of tables

DRAFT

INTRODUCTION

1.1. Motivation

1.2. Objectives

1.3. Structure of the document

STATE OF THE ART

2.1. Base and Ensemble Classifiers

2.1.1. Base Classifiers

2.1.2. Ensemble Classifiers

2.2. Classifier Optimization

2.3. Evaluation of Classifiers

2.4. Explainable AI

2.4.1. Explainable AI

2.4.2. SHAP

2.5. Web Application Development

DESIGN AND IMPLEMENTATION

- 3.1. Project Structure**
- 3.2. Exploratory Data Analysis**
 - 3.2.1. Descriptive Statistics**
 - 3.2.2. Data Visualization**
- 3.3. Data Preprocessing**
 - 3.3.1. Scaling and Normalization**
 - 3.3.2. Principal Component Analysis**
- 3.4. Building and Optimizing Classifiers**
- 3.5. SHAP Implementation**
- 3.6. Web Application Development**

EXPERIMENTS AND RESULTS

4.1. Classifier Comparison

4.1.1. Base Classifiers Comparison

4.1.2. Ensemble Classifiers Comparison

4.1.3. Choosing the Best Classifier

4.2. SHAP Analysis

4.2.1. Global Interpretability

4.2.2. Local Interpretability

CONCLUSIONS AND FUTURE WORK

5.1. Conclusions

5.2. Future Work



Universidad Autónoma
de Madrid