RESEARCH PROPOSAL

Breast Cancer Diagnosis

from mammograms using Deep Learning

María de los Ángeles Contreras Anaya - A01700284

TABLE OF CONTENTS

01

INTRODUCTION & BACKGROUND

¿What is breast cancer?

02

PROBLEM STATEMENT

¿What are the challenges of breast cancer diagnosis?

03

PROPOSAL & JUSTIFICATION

¿Why did I choose my proposed solution?

04

OBJECTIVES DEFINITION

¿What do I want to accomplish?

05

EXPECTED CONTRIBUTIONS

¿How will this research contribute to solving the problem?

06

METHODOLOGY & PLANNING

¿How are we going to accomplish the proposed objectives?

01. INTRODUCTION & BACKGROUND

Every

13 minutes

a woman **dies** of breast cancer in the U.S.

About

43,250

women, are expected to **die** of breast cancer in the U.S.

Worldwide

627,000

women, are expected to **die** of breast cancer.

Approximately,

287,850

of **invasive breast cancer** will be diagnosed in the U.S. by 2022.

02. PROBLEM STATEMENT

* **30% OF VISIBLE BREAST CANCERS ARE MISSED**

  Mammogram interpretation is a difficult task, that is why human readers can miss visible cancers on screening.

* **10 - 30% OF BREAST CANCERS ARE OCCULT IN MAMMOGRAPHY**

  Dense breast tissue, tissue superposition, false positioning at time of screening, among others, can lead to malignant lesions being ignored.

* **CAD TOOLS HAVE A FALSE-POSITIVE RATE OF 0.5 PER IMAGE**

  Every **four-image screening** mammogram processed with CAD will have two false-positive marks. This may be caused by the tools using programmed-in features, making them biased or error-prone.

02. PROBLEM STATEMENT

* **INTERVAL BREAST CANCERS ACCOUNT FOR ¼ OF POSITIVE DIAGNOSIS IN ROUTINELY SCREENED WOMEN**

  Interval cases are **6 times** more likely to be **grade III** and have **3.5 times** increased hazards of breast cancer death.

* **THE RATE OF MALIGNANCY OF LESIONS BI-RADS 4A IS 3 - 10%**

  Even tho lesions were benign patients received **unnecessary** biopsies or surgery, that is why we need techniques with higher specificity.

DESIGN AND IMPLEMENT A CNN MODEL USING THE CDD-CESM DATASET.

Collection of 2,006 low-energy images with their corresponding subtracted Contrast Enhanced Spectral Mammography images.

¿WHY CESM IMAGES?

- Better diagnostic sensitivity, specificity and accuracy than **conventional mammograms.**
- Better diagnostic accuracy than the combination of standard **mammograms and ultrasounds.**
- Similar sensitivity to **BMRI**, but at a lower cost and shorter examination time.

04. OBJECTIVES DEFINITION

Main objective

DIAGNOSE BREAST CANCER FROM MAMMOGRAMS USING DEEP LEARNING

This objective will be accomplished by completing the secondary objectives

OBJECTIVE 01

Select the best Deep Learning algorithm for breast cancer diagnosis.

OBJECTIVE 02

Design and implement a Deep Learning model for breast cancer diagnosis.

OBJECTIVE 03

Evaluate and document the results in a formal document.

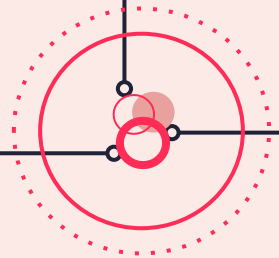
05. EXPECTED CONTRIBUTIONS

A BASELINE

For future research on the viability of using CSEM images and DL models for the diagnosis of breast cancer.

BREAST CANCER DIAGNOSIS MODEL

With a higher diagnostic sensitivity and specificity than the existing AI-software or CAD tools.



BREAST CANCER DIAGNOSIS MODEL

CNN model available to all public, enabling future researchers to try to improve its performance.

05. EXPECTED CONTRIBUTIONS

A. RESEARCH PROPOSAL

Document encompassing the problem statement, proposed solution, objectives definition, state of the art, methodology and evaluation metrics for the selected research topic.

B. DEEP LEARNING MODEL

Python code for the diagnosis of breast cancer using a Convolutional Neural Network algorithm.

D. UNDERGRADUATE DISSERTATION

Document encompassing the extensive research about breast cancer, the design and implementation of the proposed model for diagnosis and valuable results to the oncology and artificial intelligence field.

06. METHODOLOGY & PLANNING

OBJECTIVE 1

- Find 10 Q1 or Q2 articles.
- Select the metrics to measure my model's performance
- Select the DL algorithm to implement
- Search for pre-trained models.
- Document my decisions and findings in a research proposal.

OBJECTIVE 2

- Find a recent dataset
- Perform an EDA
- Preprocess the dataset if needed.
- Implement and tune the model
- Document the design and implementation of the model in a formal document

OBJECTIVE 3

- Perform the tests
- Evaluate the model according to the selected metrics
- Document the results and interpretations in a formal document.

03. METHODOLOGY & PLANNING

All the code generated for the proposed solution will be developed using in **Python**

PANDAS

For data analysis and manipulation

SEABORN

For data visualization

SCI-KIT LEARN

For the model's evaluation.

- → Recall
- → Precision
- → F1 score

SCI-KIT IMAGE

For image processing

KERAS

For the CNN design and implementation running on top of TensorFlow

THANKS

Do you have any questions?