NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE

JOSÉ DIOGO TEOTÓNIO PINTO ROMANO

BSc in Computer Science and Engineering

LEARNING TO MAP MICRORNA BIOMARKERS SIGNATURES TO BREAST CANCER SUBTYPES

MASTER IN COMPUTER SCIENCE AND ENGINEERING

NOVA University Lisbon

Draft: *June 13, 2025*

NOVA SCHOOL OF
SCIENCE & TECHNOLOGY

DEPARTMENT OF
COMPUTER SCIENCE

LEARNING TO MAP MICRORNA BIOMARKERS SIGNATURES TO BREAST CANCER SUBTYPES

JOSÉ DIOGO TEOTÓNIO PINTO ROMANO

BSc in Computer Science and Engineering

Adviser: David Semedo
Assistant Professor, NOVA School of Science and Technology

Co-adviser: Bárbara Mendes
Post-Doctoral Researcher, NOVA Medical School

MASTER IN COMPUTER SCIENCE AND ENGINEERING

NOVA University Lisbon

Draft: *June 13, 2025*

Abstract

Regardless of the language in which the dissertation is written, usually there are at least two abstracts: one abstract in the same language as the main text, and another abstract in some other language.

Keywords: One keyword, Another keyword, Yet another keyword, One keyword more, The last keyword

Resumo

Independentemente da língua em que a dissertação está escrita, geralmente esta contém pelo menos dois resumos: um resumo na mesma língua do texto principal e outro resumo numa outra língua.

Palavras-chave: Primeira palavra-chave, Outra palavra-chave, Mais uma palavra-chave, A última palavra-chave

Contents

List of Figures

Acronyms

AI Artificial Intelligence *(p. 2)*

BC Breast Cancer *(pp. 1, 2)*

DL Deep Learning *(p. 2)*

miRNAs microRNAs *(pp. 1, 2)*

ML Machine Learning *(p. 2)*

1

Introduction

1.1 Motivation & Problem Statement

1.1.1 Breast Cancer - A Global Health Challenge

Breast Cancer (BC) is currently one of the biggest public health challenges worldwide. In 2022, more than 2.3 million new cases of BC were diagnosed, resulting in around 665,000 global deaths [4] (Bray *et al.*, 2024). Other studies estimate that BC will continue to not only be the most commonly diagnosed cancer but also to increase in incidence, with projections indicating that by 2040, the number of deaths will almost double and the number of new cases will be around 3.2 million [1] (Arnold *et al.*, 2022). These figures underline the high incidence and mortality associated with the disease, highlighting the ongoing need to develop more effective strategies for its diagnosis and treatment.

BC is characterized by marked biological heterogeneity, manifested in multiple molecular subtypes that exhibit distinct clinical behaviors [8] (Perou *et al.*, 2000). Each subtype exhibits substantial differences in terms of tumor aggressiveness, metastatic potential, and behavior to specific therapies [9] (Prat *et al.*, 2015). Thus, accurate classification of these subtypes is essential to enable personalized therapeutic approaches, with a direct impact on treatment efficacy and patient prognosis. [10] (Testa *et al.*, 2020).

1.1.2 Can we improve the classification of Breast Cancer subtypes?

Among the emerging candidates for robust biomarkers for the classification of BC subtypes are microRNAs (miRNAs), small non-coding RNA molecules that play a crucial regulatory role in gene expression. They are estimated to modulate the expression of about one-third of the genes in the human genome [5] (Hammond *et al.*, 2015) and are implicated in the regulation of multiple physiological and pathological processes, including various human diseases [6] (Ho *et al.*, 2022).

Given their regulatory nature, several studies have demonstrated a significant association between miRNAs expression profiles and relevant clinical characteristics in the

context of BC, including processes such as tumor progression and metastasis development [6] [7] (Muñoz *et al.*, 2023) [3] (Blenkiron *et al.*, 2007). In addition to these aspects, a seminal study by Blenkiron *et al.* (2007) demonstrated that miRNAs expression profiles can effectively distinguish between different molecular subtypes of BC, highlighting their potential as a precise subtyping tool. This ability to discriminate between subtypes reinforces the value of miRNAs as promising clinical biomarkers.

1.1.3 How to identify the most relevant miRNAs?

The identification of the most relevant miRNAs for BC subtyping represents a major analytical challenge due to the complexity of these high dimensional regulatory molecules, non-linearity interactions between and clinical phenotypes require advanced computational approaches to be effectively modelled. Recent advances in Artificial Intelligence (AI), particularly in Machine Learning (ML) and Deep Learning (DL), have demonstrated remarkable potential in extracting meaningful patterns from high-dimensional and heterogeneous (data from distinct nature) biomedical data. These approaches enable not only the accurate classification of BC subtypes but also the identification of discriminative miRNAs signatures, supporting their integration as actionable biomarkers in clinical workflows.

In this context, ML and DL models are particularly well suited for the task of robustly characterizing and explaining the profiles of miRNAs-based biomarkers — should such biomarkers exist — with the potential to effectively discriminate between different BC subtypes, as already seen in a study done by [2] Azari *et al.* 2023 where ML algorithms identified potential diagnostic and prognostic n gastric cancer, showing high accuracy in the identification of reliable biomarkers for this disease.

This reality reinforces the urgency of developing advanced computational tools that can enable more precise molecular characterization and guide personalized therapeutic decisions, ultimately improving clinical outcomes for patients with aggressive and hard-to-treat BC subtypes.

2

Background

3

State of the Art

4

Premilinary Results

Bibliography

[1] M. Arnold et al. "Current and future burden of breast cancer: Global statistics for 2020 and 2040". In: *The Breast : Official Journal of the European Society of Mastology* 66 (2022), pp. 15–23. DOI: 10.1016/j.breast.2022.08.010 (cit. on p. 1).

[2] H. Azari et al. "Machine learning algorithms reveal potential miRNAs biomarkers in gastric cancer". In: *Scientific Reports* 13 (2023). DOI: 10.1038/s41598-023-32332-x (cit. on p. 2).

[3] C. Blenkiron et al. "MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype". In: *Genome Biology* 8 (2007), R214–R214. DOI: 10.1186/gb-2007-8-10-r214 (cit. on p. 2).

[4] F. Bray et al. "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries". In: *CA: A Cancer Journal for Clinicians* 74.3 (2024), pp. 229–263. DOI: https://doi.org/10.3322/caac.21834. eprint: https://acsjournals.onlinelibrary.wiley.com/doi/pdf/10.3322/caac.21834. URL: https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.3322/caac.21834 (cit. on p. 1).

[5] S. Hammond. "An overview of microRNAs." In: *Advanced drug delivery reviews* 87 (2015), pp. 3–14. DOI: 10.1016/j.addr.2015.05.001 (cit. on p. 1).

[6] P. T. Ho, I. Clark, and L. Le. "MicroRNA-Based Diagnosis and Therapy". In: *International Journal of Molecular Sciences* 23 (2022). DOI: 10.3390/ijms23137167 (cit. on pp. 1, 2).

[7] J. P. Muñoz et al. "The Role of MicroRNAs in Breast Cancer and the Challenges of Their Clinical Application". In: *Diagnostics* 13 (2023). DOI: 10.3390/diagnostics13193072 (cit. on p. 2).

[8] C. M. Perou et al. "Molecular portraits of human breast tumours". In: *Nature* 406.6797 (2000), pp. 747–752. DOI: 10.1038/35021093. URL: https://doi.org/10.1038/35021093 (cit. on p. 1).

[9] A. Prat et al. “Clinical implications of the intrinsic molecular subtypes of breast cancer.” In: *Breast* 24 Suppl 2 (2015), S26–35. DOI: 10.1016/j.breast.2015.07.008 (cit. on p. 1).

[10] U. Testa, G. Castelli, and E. Pelosi. “Breast Cancer: A Molecularly Heterogenous Disease Needing Subtype-Specific Treatments”. In: *Medical Sciences* 8 (2020). DOI: 10.3390/medsci8010018 (cit. on p. 1).

A

Appendix 1 Covers Showcase

B

Appendix 2 Lorem Ipsum

I

Annex 1 Lorem Ipsum