Summary

Biomarkers play a crucial role in modern medicine, serving as measurable indicators of biological processes, disease states, or responses to treatment. Their importance spans from early disease detection and diagnosis to personalized treatment strategies and identification of predisposition characteristics, significantly improving patient outcomes. Advancements in molecular biology and technology have led to the discovery of a wide range of biomarkers, transforming fields such as oncology, neurology, and infectious diseases. Furthermore, beyond molecular biomarkers, other types such as clinical (e.g., blood pressure, LDL cholesterol, BMI, assessment scores), signal-based (e.g., EEG and MRI characteristics), and sensor biomarkers (e.g., glucose measuring sensors) further enhance diagnostic and monitoring capabilities in medicine.

Bioinformatics has become essential for analyzing and interpreting complex conditions with the emergence and increasing availability of high-throughput data. Computational tools enable researchers to identify new biomarkers, assess their reliability, and integrate them into clinical practice. Moreover, new machine learning algorithms have enhanced biomarker research by avoiding statistical assumptions, thus allowing for pattern recognition and the development of predictive models. Leveraging large datasets, like omics, machine learning algorithms can identify subtle correlations that may be overlooked by traditional statistical methods, ultimately leading to more accurate clinical diagnostics and improved treatment planning.

Despite these advancements, several challenges persist that this thesis seeks to address. In bioinformatics, new tools and methods are necessary in biomarker discovery which is often hindered by biological variability, limited sample sizes, and issues with reproducibility. In clinical practice, it is important to foresee the early evolution of diseases to prevent or cure them in time and create better treatments.

This thesis focuses on the identification of biomarkers in different types of clinical and biological data through the use of bioinformatics and machine learning Furthermore, it addresses the role and methods of identification for coding and non-coding molecular markers, and it also discusses the challenges around them at the beginning to introduce the importance of molecular biomarkers. A major part of this work is dedicated to the harmonization of microarray datasets to increase the sample sizes of conditions and improve the statistical power of biomarkers' findings. A method that can help in cases of rare conditions and valuable past studies combination. Emphasis was given on neurodevelopmental disorders and blood samples to face both the problems of minimal invasiveness and data scarcity. Additionally, an application of machine learning modelling on cross-sectional clinical data is presented, proving that it is beneficial in clinical decision making by identifying important markers that play a role in the manifestation of diseases. This is extremely useful in prognostics and precision medicine. To this end, an application for early risk prediction of cardiovascular disease in people with thalassemia as a clinical decision support system is proposed with a high predictive efficiency. Finally, this thesis presents a functional and interactive unified platform that integrates the above-mentioned tools into an easy-to-use online application. This open-access platform allows for the practical use of the methods developed on the user's datasets.

In summary, the present work is presented as an innovation in the field of finding potential biomarkers of different types through data-driven approaches while also creating clinical decision support systems, following the dictates of personalized medicine and risk prediction methods.