1. Introduction

As one of the deadliest diseases and a major health concern, the estimated number of cancer-related cases and deaths worldwide has reached approximately 19.3 million and 10 million in 2020 [Ferlay et al. (2021)]. Together with cardiovascular diseases, cancer has become the leading cause of death in 127 countries [Bray et al. (2021)]. Cancer is the result of an accumulation of inherited and somatic mutations in oncogenes and tumor suppressor genes, as stated by Jonsson and Bates (2006). In simple terms, cancer occurs when there are mutations that make cells multiply uncontrollably. Many factors play a role in cancer, but they can be traced down to molecules, such as proteins for example. One example is the well-known tumor suppressor protein p53, where mutations of this protein and the gene that produces it, the TP53 gene, have been linked to more than 50% of cancer cases [Duffy et al. (2022)]. There have also been other findings of cancer-related proteins in different studies. Some other studies have also found proteins that have links to cancer. A study on the CCN protein family found that the expression of CCN proteins may have a role in the regulation of cancer cell growth, and another study on the ErbB/HER protein kinase family showed that mutations in this family can lead to malignancy in some cancers [Perbal (2003); Roskoski (2014)].

Proteins, referred to as the building blocks of life, are complex organic molecules made up of chains of amino acids usually consisting of 50 or more [LaPelusa and Kaushik (2023)]. One of the many properties of proteins is that they like to form bonds with other molecules, either through physical or chemical means. This is often the case in the human body, where in human cells alone, there are more than 39,000 identified protein interactions (PIs) [Gonzalez and Kann (2012)]. These interactions allow proteins to achieve their functions in bioprocesses such as cell communication, signal transduction, metabolism, and immune system regulation [Jonsson and Bates (2006)]. Therefore, disruptions caused by internal or external factors to these essential functions can initiate or further develop a disease, including cancer [Kuzmanov and Emili (2013)]. Among protein interactomes, peptides have emerged as promising candidates for therapeutic agents due to their safety, good efficacy, high selectivity, ease of synthesis, and good biocompatibility [Matijass and Neundorf (2021); Fosgerau and Hoffmann (2015)]. The use of peptides today has become widespread, with the history of their use as drugs beginning in 1922 and growing in popularity. It is estimated that in 2019, they accounted for 5% of the global pharmaceutical market or approximately $50 billion, and within this market share, 17% was dedicated to oncology [Muttenthaler et al. (2021)]. Considering the various side effects and limited effectiveness of conventional treatments in this domain, attempts to use peptides as cancer treatments may be a viable option [Cavalcanti and Soares (2021)].

Uncovering possible PIs is key in revealing suitable drug targets [Rao et al. (2014)]. Several in vivo and in vitro approaches are commonly used for this purpose, although there are technical difficulties along with poor scalability, inefficiency, and time constraints that have plagued such approaches [Lee et al. (2019)]. Over the years, computational approaches, otherwise referred to as in silico approaches, have always been the first choice in assisting this important task. Recently,

the rapid advancement of artificial intelligence (AI) has made a great impact in revolutionizing industries and various fields of study. In the field of bioinformatics, the trend of integrating deep learning methods has seen a significant increase over the past decade [Min et al. (2017)]. Compared with traditional machine learning, deep learning methods can extract a higher level of data representation from inputs with stacked processing layers [Ahmed et al. (2023)]. Following the trend, considerations regarding our data, and the lack of implementation for this method in bioinformatics, we propose the use of TabNet to provide peptide-protein interaction (PepPI) prediction in cancer. We also performed a comparison with common deep learning and machine learning methods that have been adapted previously for this task.