PhD thesis title: **Leveraging representation learning for drug discovery**
Name: **Yogesh Kalakoti**
Entry Number: **2018BEZ7512**

ABSTRACT

Deciphering the secrets of cellular machinery and disease progression is critical for developing solutions in prognosis and *in silico* drug discovery. The ability to extract quantifiable information about cellular processes allows the profiling that is necessary to devise strategies for mitigating aberrations induced in a diseased state. Advances in high-throughput technologies have allowed characterisation of this information in the form of large-scale biological datasets at multiple levels of subcellular machinery. Traditional methods used to examine this data do not have sufficient sensitivity to identify correlations between molecular patterns and distinguishing characteristics of a disease. Moreover, such profiling methods do not allow for the stratification of high-risk malignancies at a molecular level. This is one of the primary drivers of research in personalized medicine. Over the years, machine learning algorithms have found their utility in analysing biological datasets due to their ability to model complex distributions and identify data patterns. However, there is a gap in methods to effectively represent biological datasets in a way that maximises the amount of information presented to the ML agent. This thesis aims at developing frameworks capable of extracting relevant information from omics, sequence and structural datasets efficiently for downstream ML-based solutions in prognosis modelling and drug discovery. Workflows, analysis and results presented in this thesis demonstrate the utility of representation learning in developing solutions for survival estimation, motif discovery, drug-target interaction and protein fingerprinting in an effective manner.