Design of clinical decision support systems for cancer based upon clinical and molecular data

PhD Anneleen Daemen

The tsunami of data with clinical parameters and information from the genome, transcriptome, proteome and metabolome requires advanced mathematical approaches to transform the huge amount of data in meaningful support for clinicians. This dissertation is therefore directed towards analytical tools for the heterogeneous disease cancer, in order to move from bench to bedside with decision support systems in daily clinical practice.

Because none of the available data types provides a comprehensive understanding of the complex biology of cancer, complementary levels of biological information need to be integrated in order to efficiently translate cancer genome information into molecular markers and therapeutics. The way we approach data fusion differs from what recently has been described in the literature. We present a kernel-based integration framework in which multiple layers of experimental data are combined in one mathematical model for the development of more homogeneous classifiers.

An extra layer of processing is required for data types that have not been used for classification so far or that would benefit from an improved modeling to fully exploit its content. For array CGH data, we present a method for the conversion of copy number variations into features usable for classification. As histopathology will remain a cornerstone in cancer decision making, we improve modeling of clinicopathological parameters with a kernel function specifically adapted to the characteristics of clinical data.

Finally, our data integration approach can be improved with non-patient related data obtained from databases on different aspects of biological systems. We describe how prior biological knowledge can be incorporated in classification models, by exploiting relations between genes with similar functions but active in alternative pathways based on spectral graph theory. Our data integration approach, the extension towards more complex data types and the incorporation of prior interactome knowledge contribute to improved clinical decision support and deciphering biological deregulated processes in which key markers are involved.