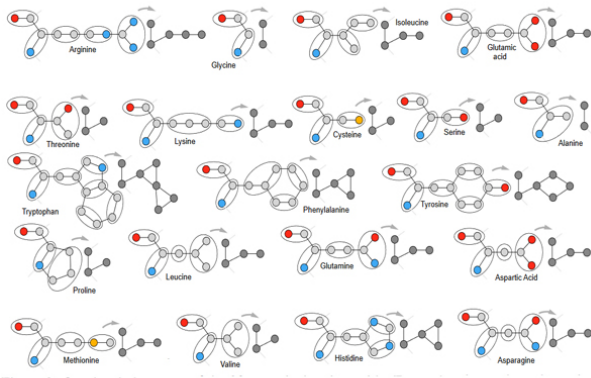


ProteiNN: Intrinsic-Extrinsic Convolution and Pooling for Scalable Deep Protein Analysis

Hermosilla, Pedro; Schafer, M.; Lang, M.; Fackelmann, G.; Vazquez, Pere-Pau; Kozlikova, B.; Krone, M.; Ritschel, T.; Ropinski, T.



Proteins perform a large variety of functions in living organisms, thus playing a key role in biology. As of now, available learning algorithms to process protein data do not consider several particularities of such data and/or do not scale well for large protein conformations. To fill this gap, we propose two new learning operations enabling deep 3D analysis of large-scale protein data. First, we introduce a novel convolution operator which considers both, the intrinsic (invariant under protein folding) as well as extrinsic (invariant under bonding) structure, by using n-D convolutions defined on both the Euclidean distance,

as well as multiple geodesic distances between atoms in a multi-graph. Second, we enable a multi-scale protein analysis by introducing hierarchical pooling operators, exploiting the fact that proteins are a recombination of a finite set of amino acids, which can be pooled using shared pooling matrices. Lastly, we evaluate the accuracy of our algorithms on several large-scale data sets for common protein analysis tasks, where we outperform state-of-the-art methods.