Docket #: S13-394

Machine learning features for ligand-based drug discovery.

Stanford researchers have developed descriptors based on OpenEye Rapid Overlay of Chemical Structures (ROCS) that, when paired with machine learning methods improve virtual screening performance. Current practices of hit identification in drug discovery is often expensive and require high-throughput screens of large compound collections. However, computational shortcuts can help prioritize experiments and reduce screening costs by leveraging the structure of either the target or set of active compounds. This approach uses 3D shape and chemical similarity, which are natural descriptors for ligand-based approaches.

Applications

- Improving ligand-based virtual screening and lead optimization. Prioritization of screening libraries to reduce experimental costs, especially in an iterative "'active learning'' context.

Advantages

- Depending on the dataset used for validation, these descriptors can give improved receiver operating characteristic area under the curve (ROC-AUC), a measure of classification performance, compared to the default ROCS implementation. It should be noted that, depending on the dataset, these new descriptors do not necessarily outperform simple 2D similarity or a machine learning approach based on both standard ROCS similarity measures and 2D similarity.

Innovators

- Steven Kearnes
- Vijay Pande

Licensing Contact

Imelda Oropeza

Senior Licensing Manager, Physical Sciences

Email