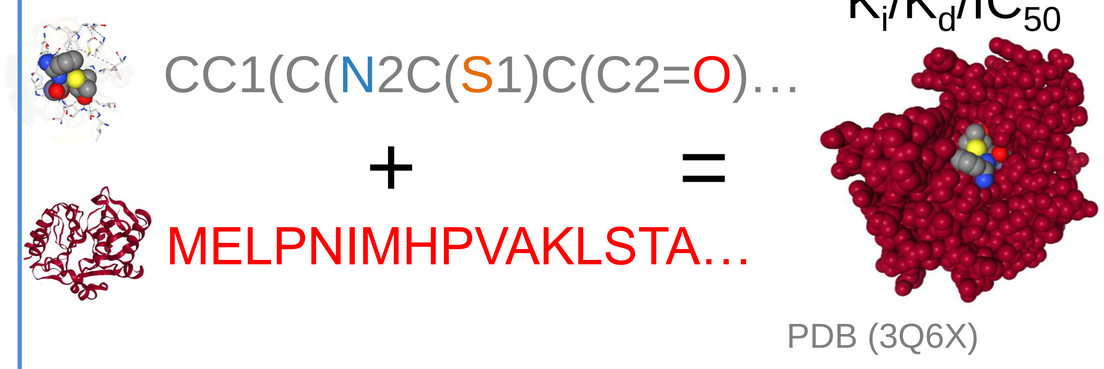
DeepDTA: deep drug-target binding affinity prediction

Hakime Öztürk[1], Arzucan Özgür[1] and Elif Ozkirimli[2]

Departments of { **1** Computer Engineering, **2** Chemical Engineering }, Boğaziçi University, İstanbul, Turkey
{ hakime.ozturk, elif.ozkirimli, arzucan.ozgur } @boun.edu.tr

Introduction

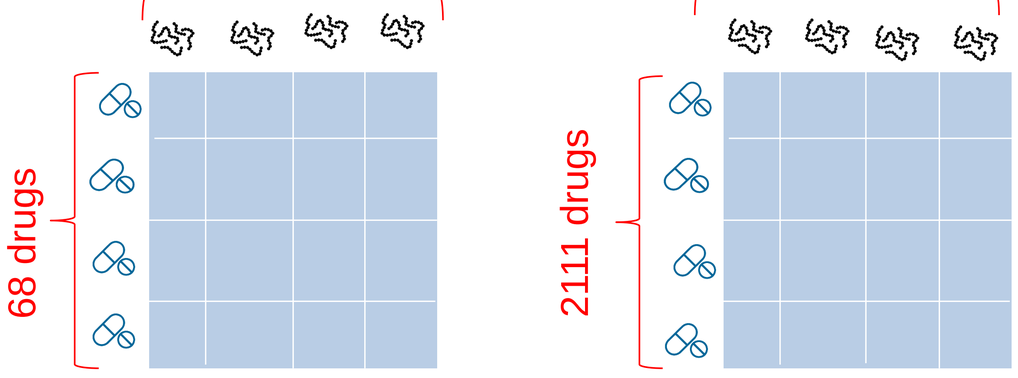
The identification of novel drug-target (DT) interactions is a substantial part of the drug discovery process. Formulating the DT prediction as a binding affinity prediction problem brings in the advantage of predicting an approximate value for the strength of the interaction between the drug and target



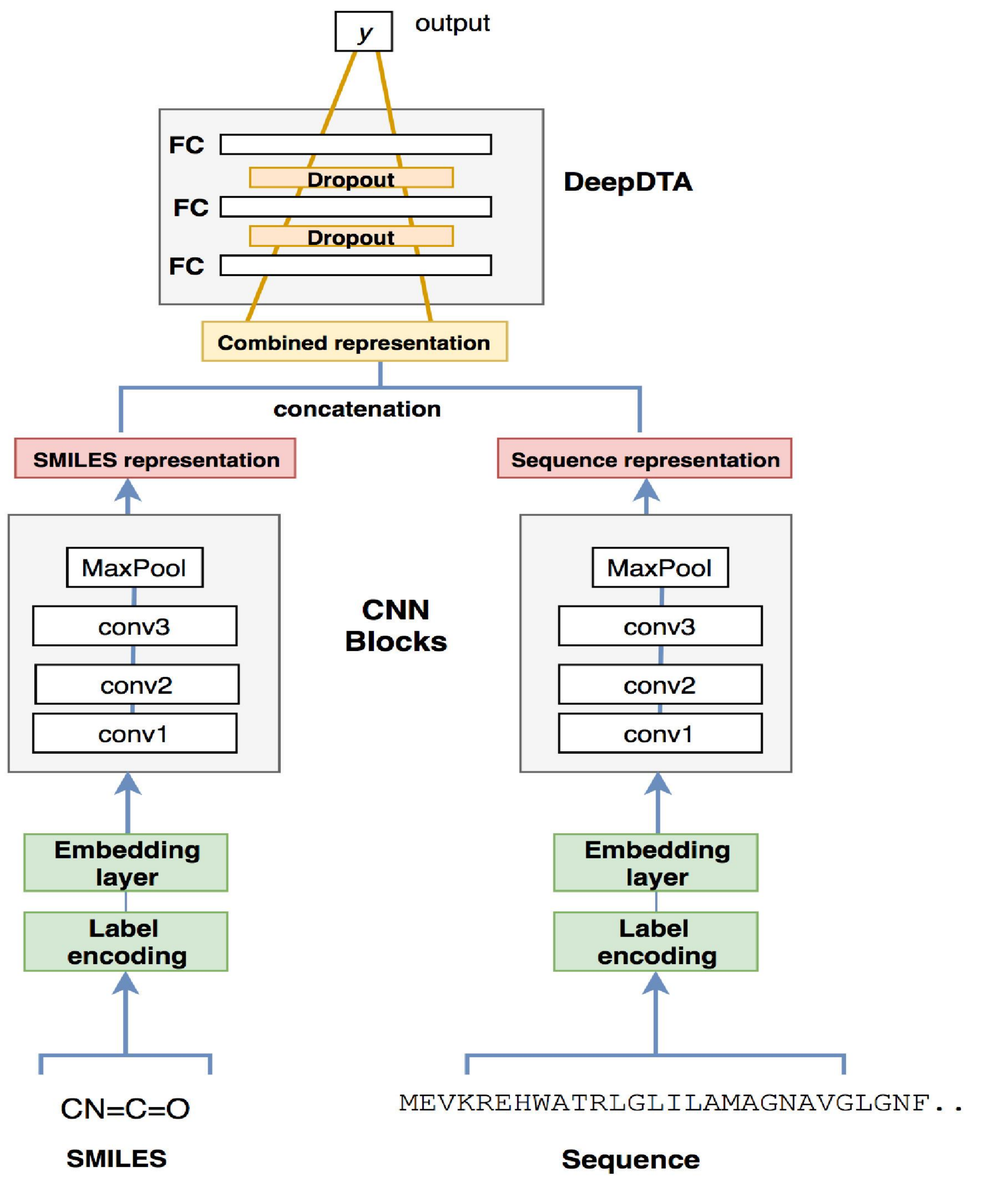
We propose to predict drug-target binding affinity by using raw compound SMILES and protein SEQUENCE with a deep-learning based approach.

METHOD

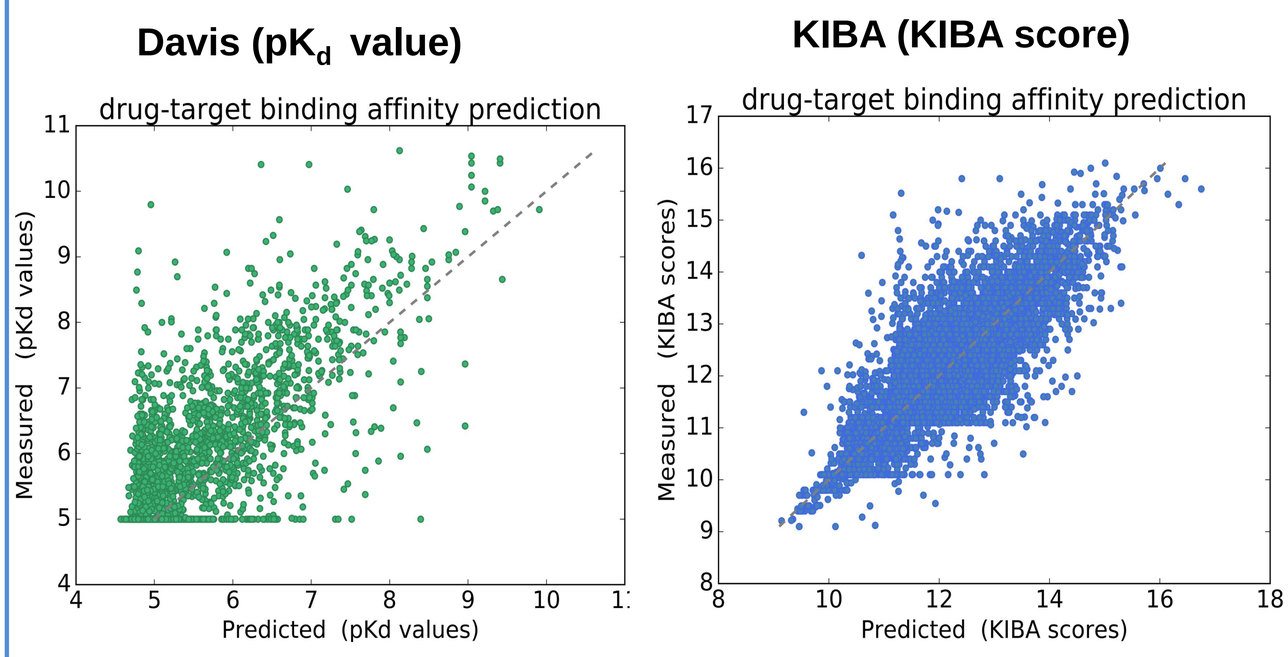
Davis Kinase (pK_d) [1] **KIBA Kinase (KIBA score)** [2]



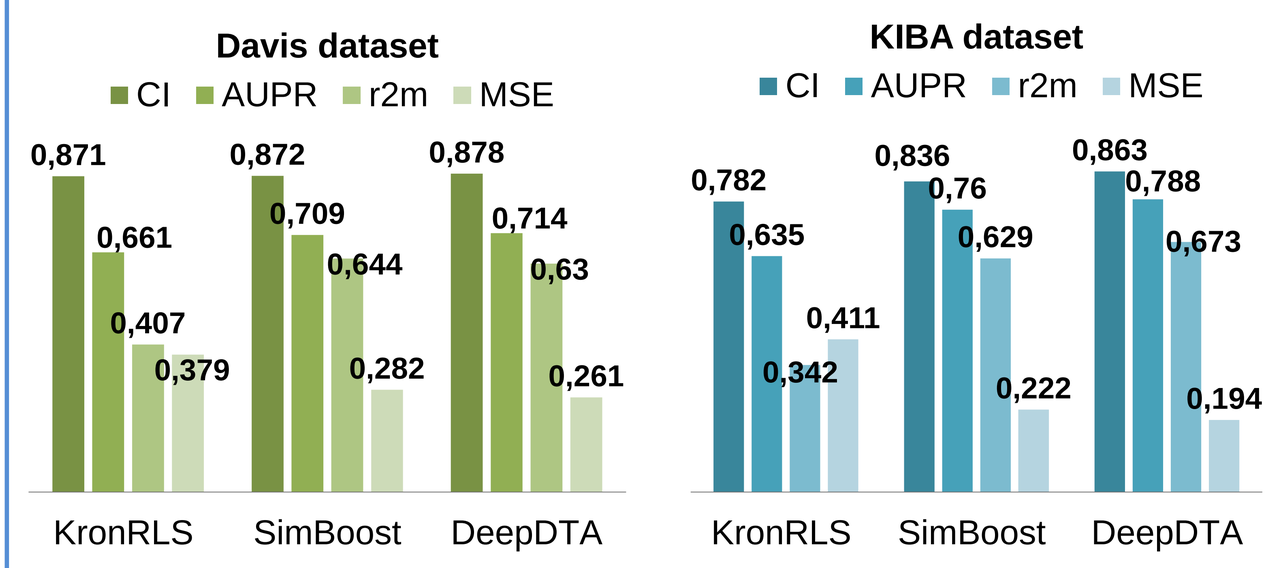
- We used a model that combines Convolutional Neural Networks (CNN) to learn high-level representations from protein sequences and drug SMILES strings and fully connected layers to predict binding affinity.



RESULTS



- We compared the performance of CNN-based DeepDTA with tw state-of-art approach in drug-target binding affinity prediction that we refer to as KronRLS [3] and SimBoost [4].



- For Davis dataset we used pK_d value of 7 as threshold and for KIBA dataset we used the threshold KIBA value of 12.1 to compute AUPR.

- DeepDTA outperformed both baselines on the KIBA data set with statistical significance (p-value of 0.0001 for both). The results suggest that both SimBoost and DeepDTA are acceptable models for affinity prediction in terms of r^2_m value.

Conclusion

- The major contribution of this study is the presentation of a novel deep learning-based model for drug - target affinity prediction that uses only character representations of proteins and drugs.

- By simply using raw sequence information for both drugs and targets, we were able to achieve similar or better performance than the baseline methods that depend on multiple different tools and algorithms to extract features.

References

[1] Davis, M. I. *et al.* (2011). *Nature biotechnology*.
[2] Tang, J. *et al.* (2014). *Journal of Chemical Information and Modeling*,
[3] Pahikkala, T. *et al.* (2014). *Briefings in bioinformatics*,
[4] He, T. *et al.* (2017). *Journal of cheminformatics*.

We gratefully acknowledge TUBITAK-BIDEB 2211 & BAGEP Academy of Science Award.