WideDTA:
deep drug-target binding affinity prediction

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

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

1 INTRODUCTION

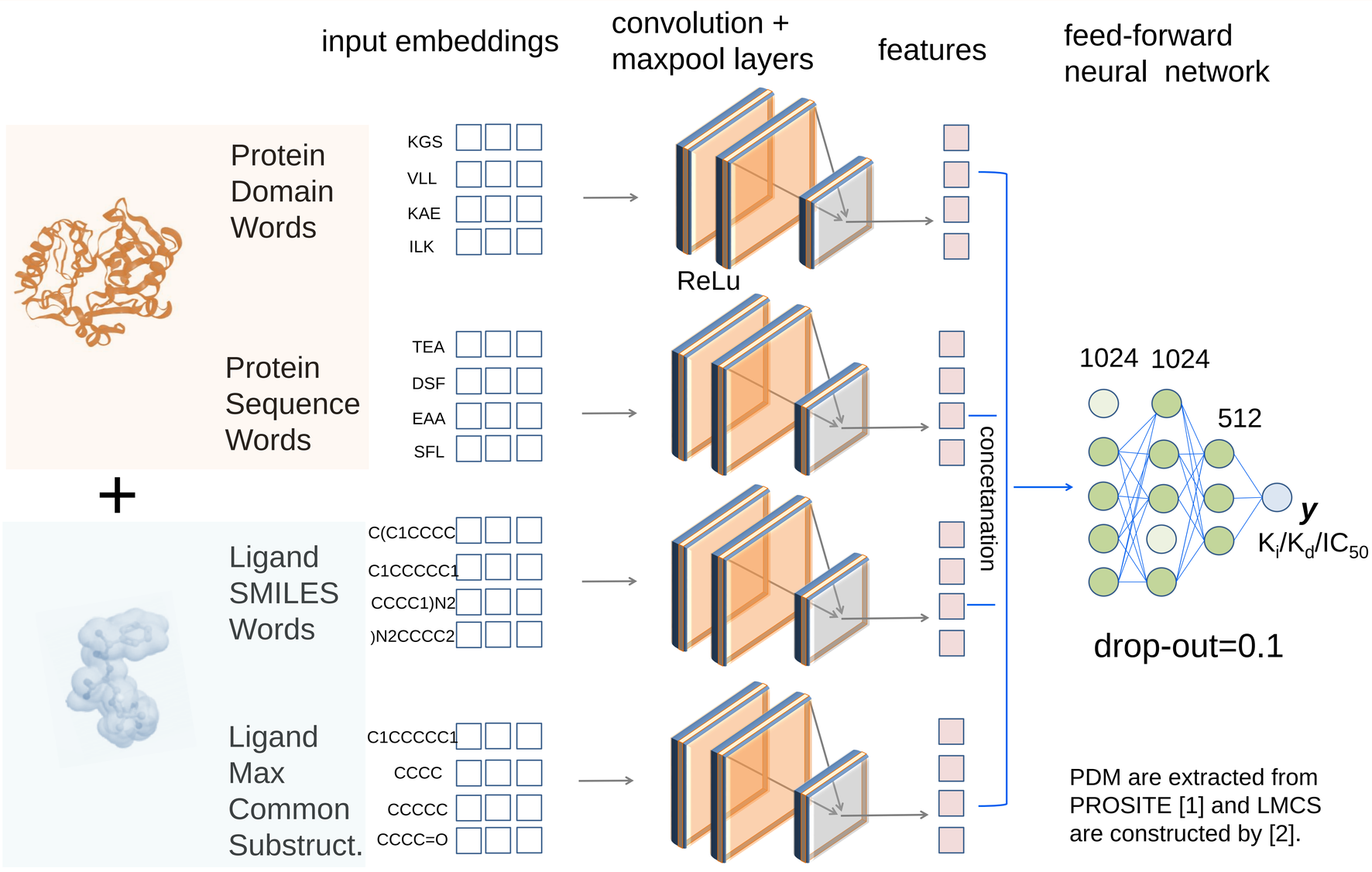
Prediction of the interaction affinity between proteins and compounds is a major challenge in the drug discovery process.

WideDTA is a deep-learning based prediction model that employs four text-based information sources:

- protein sequence (**PS**),
- ligand SMILES (**LS**),
- protein domains and motifs (**PDM**),
- maximum common substructure (**LMCS**) words

to predict binding affinity.

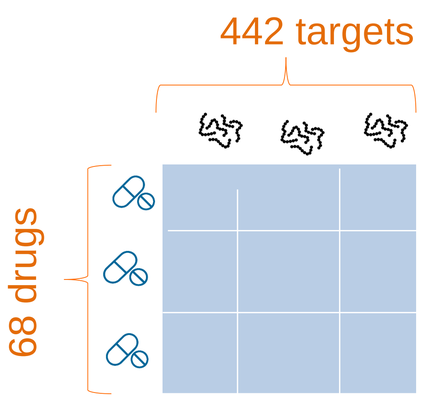
2 WideDTA Architecture



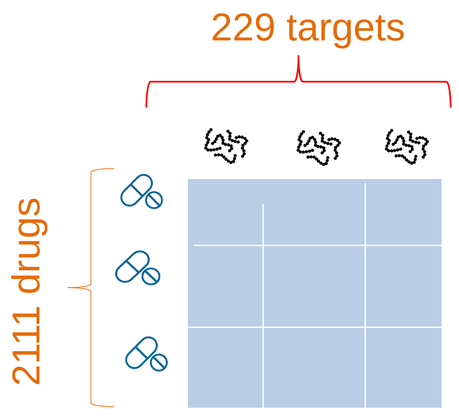
3 EXPERIMENTS & RESULTS

DATASETS

Davis Kinase (pK_d) [3]

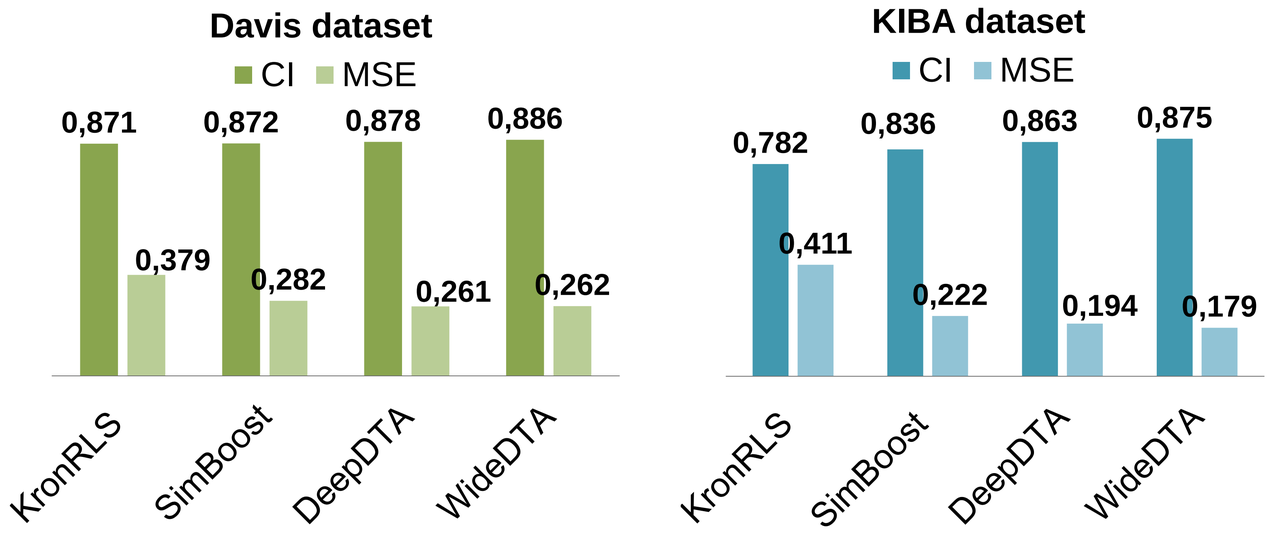


KIBA Kinase (KIBA score) [4]



We compared the performance of WideDTA with a state-of-art deep learning based approach, DeepDTA [5] and two traditional machine learning based models in drug-target that we refer to as KronRLS [6] and SimBoost [7].

SMILES sequences are represented in **DeepSMILES** [8] which yielded better performance in terms of evaluation metrics.



WideDTA model that combined all textual sources protein PS + PDM + LS + LMCS achieved better performance than other possible combinations.

4 CONCLUSION

- We proposed a deep-learning based approach to predict drug-target binding affinity, WideDTA, that combines four different textual pieces of information related to proteins and ligands.
- The protein sequence, domains and motifs were all represented as a set of 3-residue words, whereas for ligands, 8-character chemical words and maximum common substructure (MCS) words were extracted from the complete SMILES strings.
- DeepDTA is a character based approach, whereas WideDTA uses word representations. Our results suggest that the word based approach can be an alternatve to the character based approach.

References

[1] Sigrist, C. *et al.* (2009) Nucleic acids research
[2] Wozniak, M. *et al.* (2018) *Scientific reports*
[3] Davis, M. I. *et al.* (2011). *Nature biotechnology*.
[4] Tang, J. *et al.* (2014). *Journal of Chemical Information and Modeling*
[5] Öztürk, H. *et al.* (2018) *Bioinformatics*
[6] Pahikkala, T. *et al.* (2014). *Briefings in bioinformatics*
[7] He, T. *et al.* (2017). *Journal of cheminformatics*.
[8] O'Boyle & Dalke (2018) ChemRxiv

Travel funding to ISMB/ECCB 2019 was generously provided by **ISCB**.

We gratefully acknowledge TUBITAK-BIDEB 2211.