Application of vision transformers to protein-ligand affinity prediction.

Jakub Poziemski* and Pawel Siedlecki

Laboratory of Cheminformatics and Molecular Modeling, IBB PAS, Pawinskiego 5a, 02-106 Warsaw, Poland

jakub.poziemski@ibb.waw.pl, pawel@ibb.waw.pl

70 YEARS
INSTITUTE
OF BIOCHEMISTRY
AND BIOPHYSICS
POLISH ACADEMY
OF SCIENCES

1. Introduction

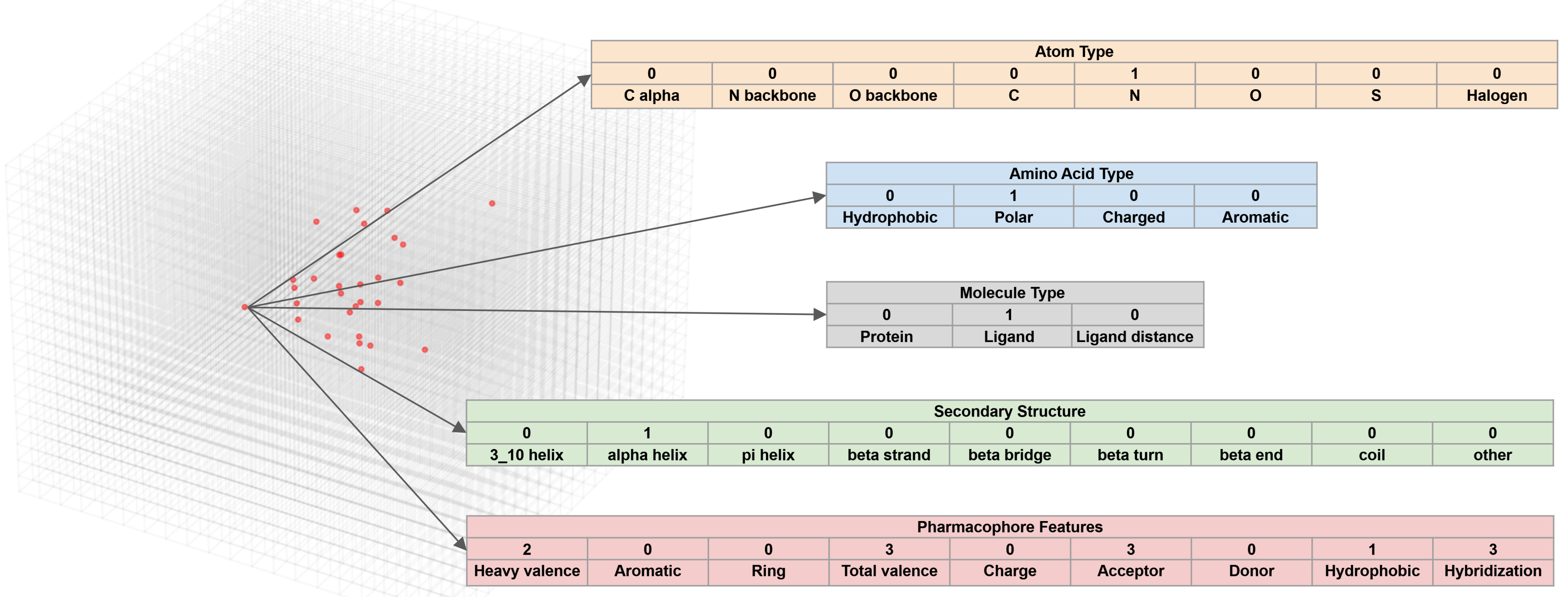
Vision Transformers (ViTs) have found wide application in various areas of computer vision, including image classification, object detection and image generation; typically achieving results that outperform traditional convolutional architectures. Due to their strong ability to model global dependencies, ViTs have also successfully applied to advanced tasks such as medical image analysis or video analysis. In this work, we show that Vision Transformers can also be successfully applied to problems in the area of computer-aided drug design (CADD). We present the use of ViTs in the problem of predicting protein-ligand affinity, a very important problem drug development. Protein-ligand affinity is the strength with which a chemical (ligand) binds to a molecular target (protein), expressed as a real number. This phenomenon is the basis for predicting the activity of new potential drugs. In our approach, we use a set of 3D structures of protein-ligand complexes.

2. Data

The training, test, and validation sets were acquired from the PDBBind 2019 collection (excluding peptides and complexes raising errors).
Data split (13,599 complexes):

- **Training set** - 90 % (12 212 complexes)
- **Validation set** - 10 % (1 387 complexes)
- **Test set** - CASF'13 (195 complexes) and CASF'16 (285 complexes)

We represent the protein-ligand 3D complex with a 21x21x21Å grid of 1Å (9261 grid points). Each grid point is a vector of 33 features (input tensor shape: **21x21x21x33**).



Atom Type							
0	0	0	0	1	0	0	0
C alpha	N backbone	O backbone	C	N	O	S	Halogen

Amino Acid Type			
0	1	0	0
Hydrophobic	Polar	Charged	Aromatic

Molecule Type		
0	1	0
Protein	Ligand	Ligand distance

Secondary Structure								
0	1	0	0	0	0	0	0	0
3_10 helix	alpha helix	pi helix	beta strand	beta bridge	beta turn	beta end	coil	other

Pharmacophore Features								
2	0	0	3	0	3	0	1	3
Heavy valence	Aromatic	Ring	Total valence	Charge	Acceptor	Donor	Hydrophobic	Hybridization

Figure 1: Grid point feature representation

3. Model Architecture

We adopted the architecture proposed in [1], modifying it to accommodate 3D grid data (See Figure 2).

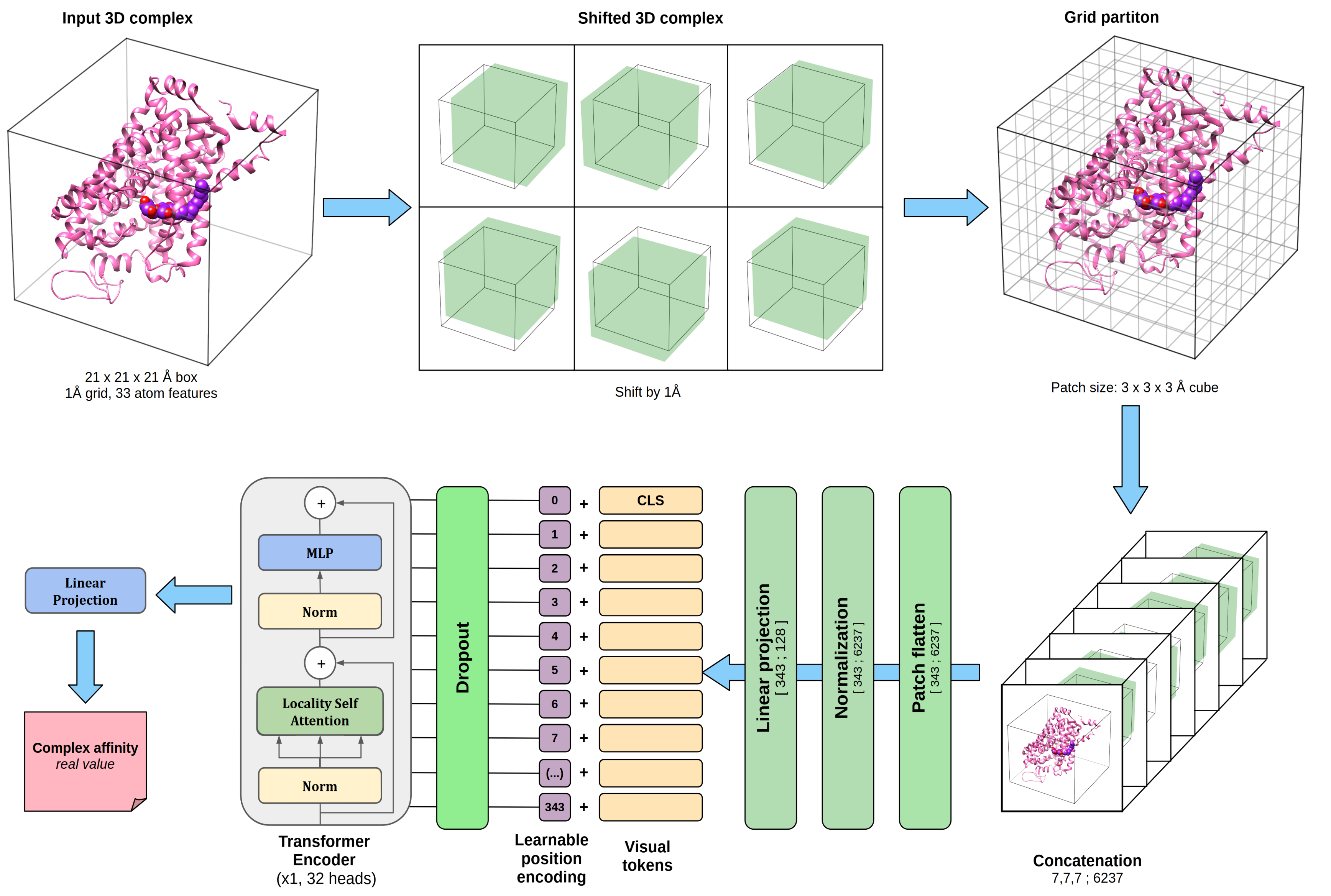


Figure 2: Model Architecture

[1] "Vision Transformer for Small-Size Datasets". Seung HL, et. al. 2021

4. Results

We compared the performance of the ViT model with other state-of-the-art methods and architectures available in the public domain.

Model	**CASF2016**	**CASF2013**
Onionnet2	1.164	1.29
SS-GNN	1.181	1.347
ViT	**1.248**	**1.396**
CAPLA	1.200	1.446
DCML	1.255	1.432
AGL	1.271	1.45
OnionNet	1.278	1.503
HPC/HWPC	1.307	1.483
RF-Score v3	1.39	1.51
Pafnucy	1.418	1.620
PerSpect ML	1.724	1.956
DeepDTAF	1.355	2.103

Table 1. RMSE of different models obtained on benchmark datasets.

5. Explainable AI (XAI)

The obtained results indicate ViT is able to learn from valid interaction patterns and its attention focuses on relevant molecular information.

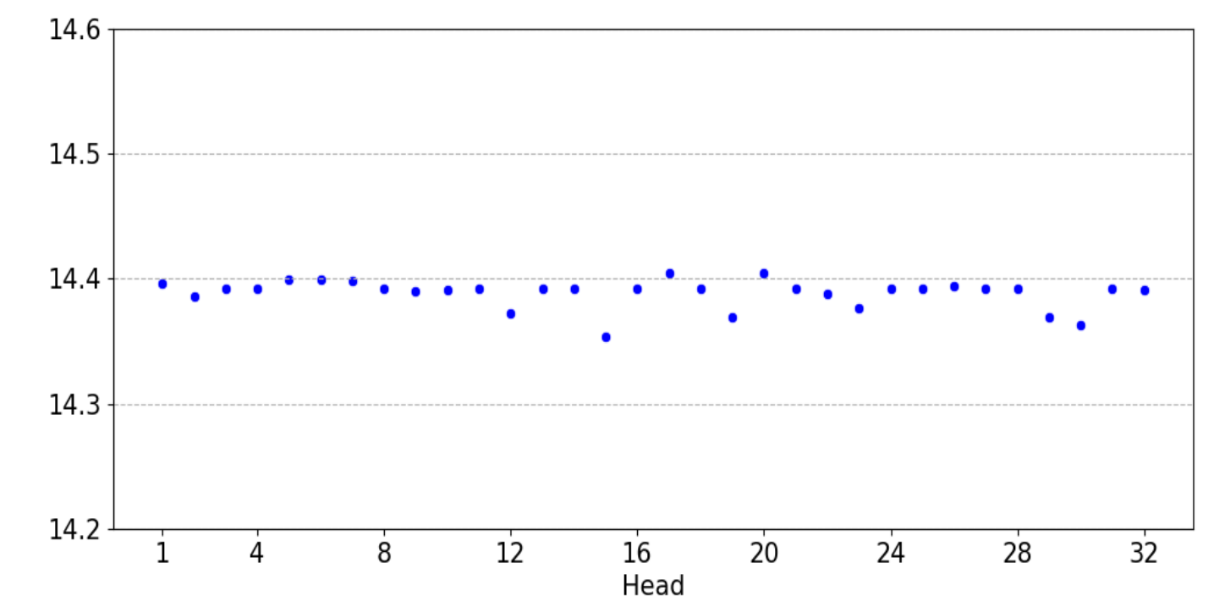


Figure 3: Mean Attention Distance

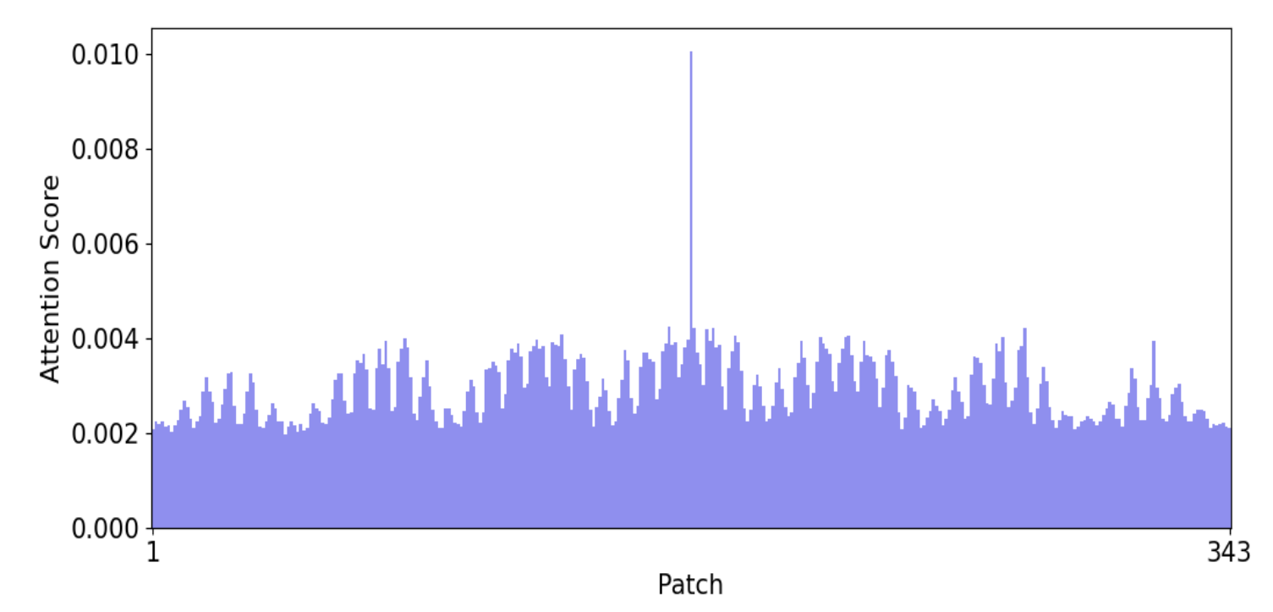


Figure 4: Attention scores for CLS

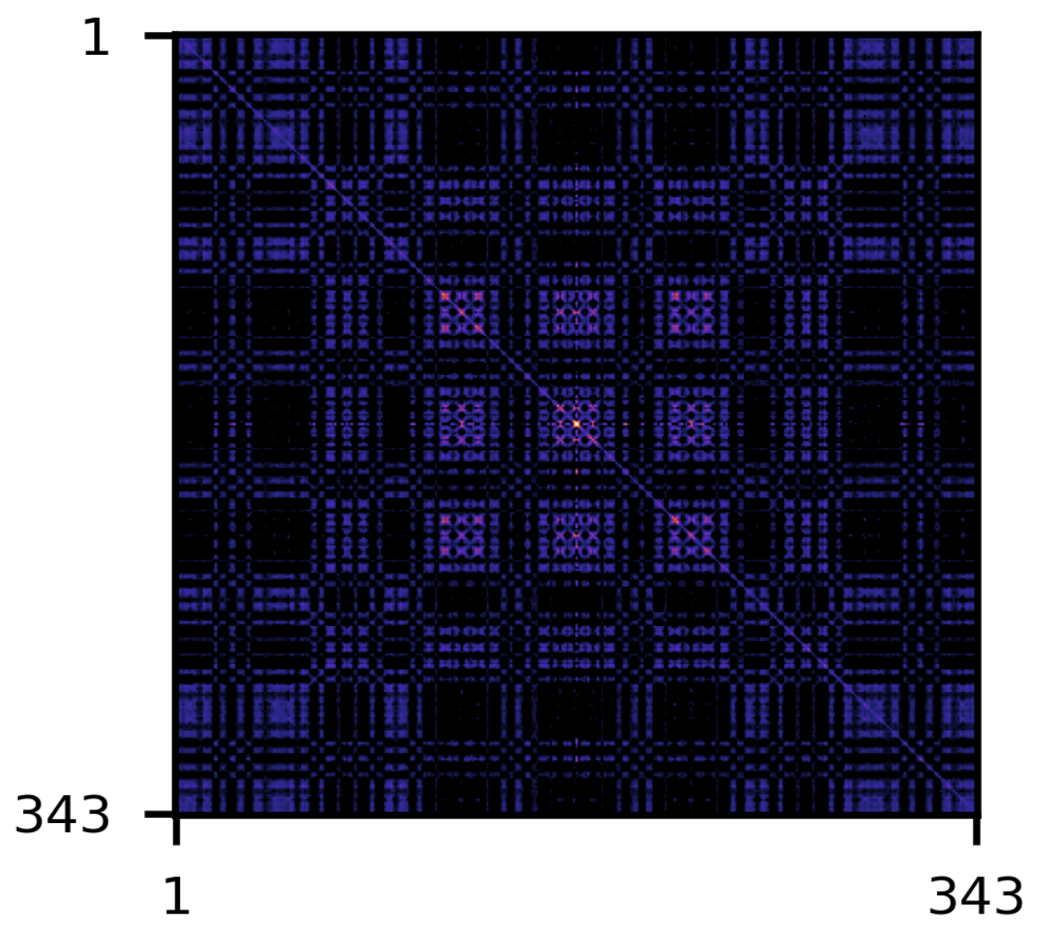


Figure 5: Patches similarity

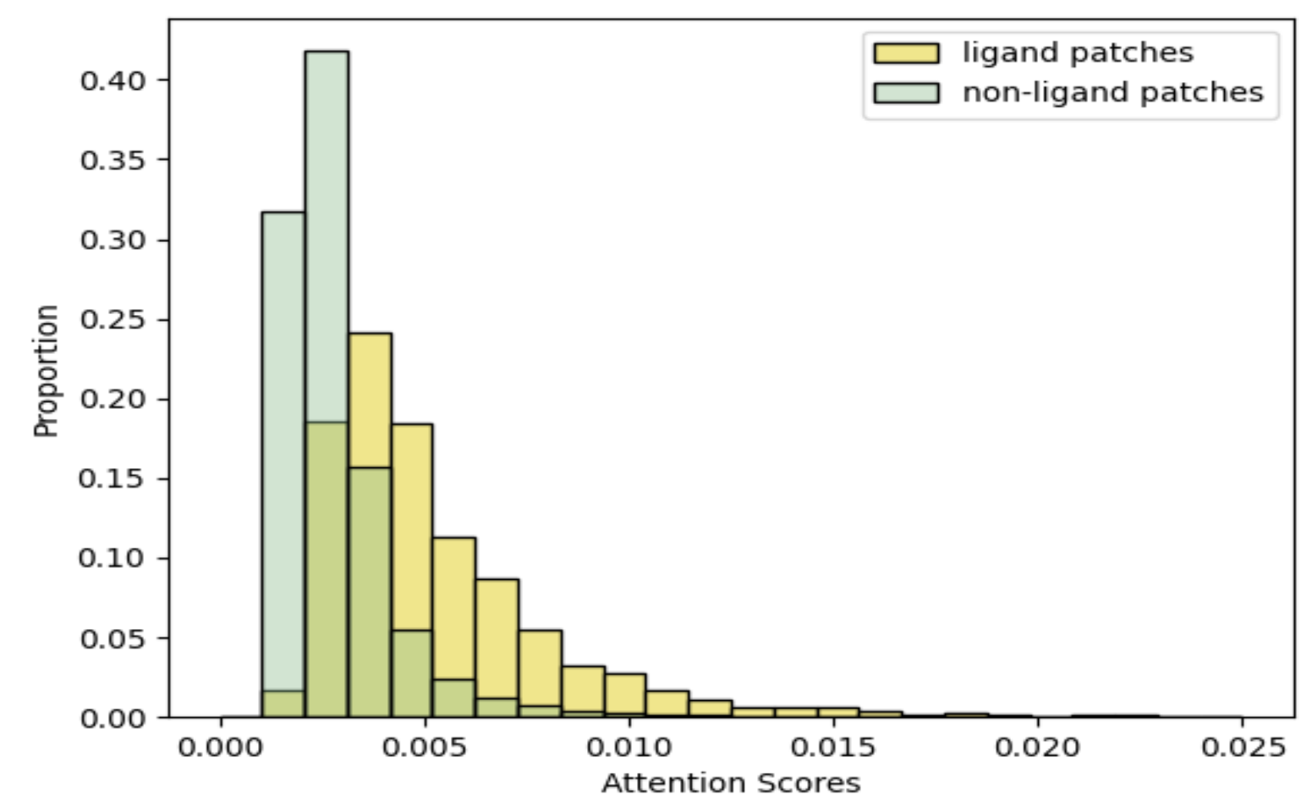


Figure 6: CLS token attention score

6. Conclusions

We have demonstrated the successful application of ViT in the protein-ligand prediction problem, an important task for early drug discovery. With the use of a relatively simple architecture, we obtained results comparable to the best models reported in the literature.

1. Despite the complex nature of the problem (small data set, high data sparsity, activity cliffs), ViT can be effectively applied to the problem of protein-ligand affinity prediction.
2. The obtained XAI results indicate ViT is able to learn from valid interaction patterns and its attention focuses on relevant molecular information.
3. The ViT architecture presented here offers potential for further optimization, which may lead to enhanced performance.

Our results can serve as foundation for the use of ViT in other medically relevant problems hindered by data scarceness such as ligand-RNA interactions.