Pep2TCR: accurate prediction of CD4 T cell receptor binding specificity through transfer learning and ensemble approach

Kaixuan Diao[1,2,3#]， Tao Wu[1]， Xiangyu Zhao[1]， Nan Wang[1]， Die Qiu[1]，
Wei-Liang Wang[4]， Xinxiang Li[5]， Xue-Song Liu[1,6*]

[1]School of Life Science and Technology, ShanghaiTech University；

[2]Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences；

[3]University of Chinese Academy of Sciences； [4]Department of Dermatology,

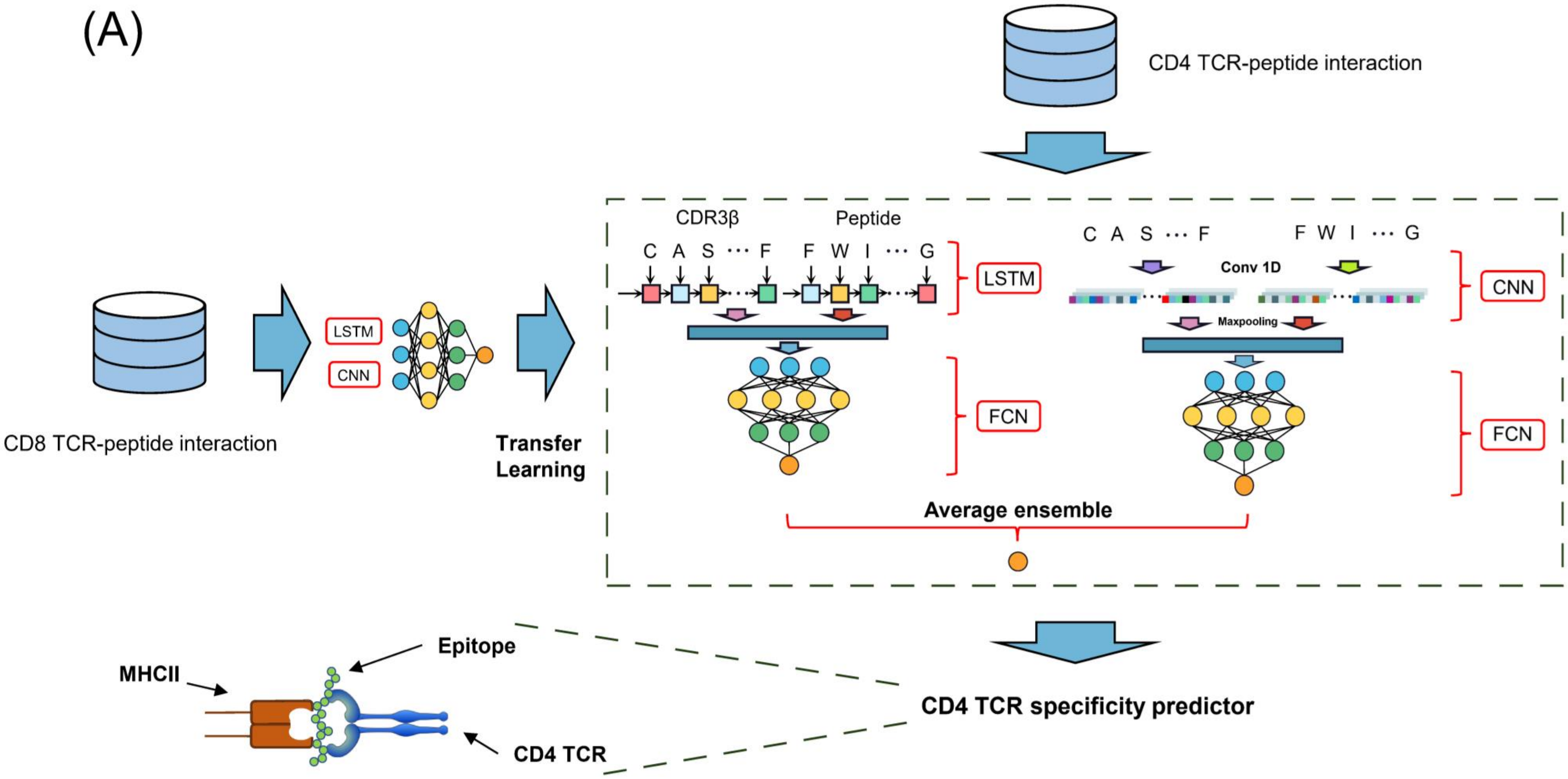
Yangjiang People's Hospital affiliated to Guangdong Medical University；

[5]Department of Colorectal Surgery, Fudan University Shanghai Cancer Center；

[6]Shanghai Clinical Research and Trial Center；

Kaixuan Diao, Tao Wu, Xiangyu Zhao, Nan Wang, Die Qiu, Wei-Liang Wang, Xinxiang Li, Xue-Song Liu. 2024. Pep2TCR: accurate prediction of CD4 T cell receptor binding specificity through transfer learning and ensemble approach. ***iMetaOmics*** 1: e43. https://doi.org/10.1002/imo2.43

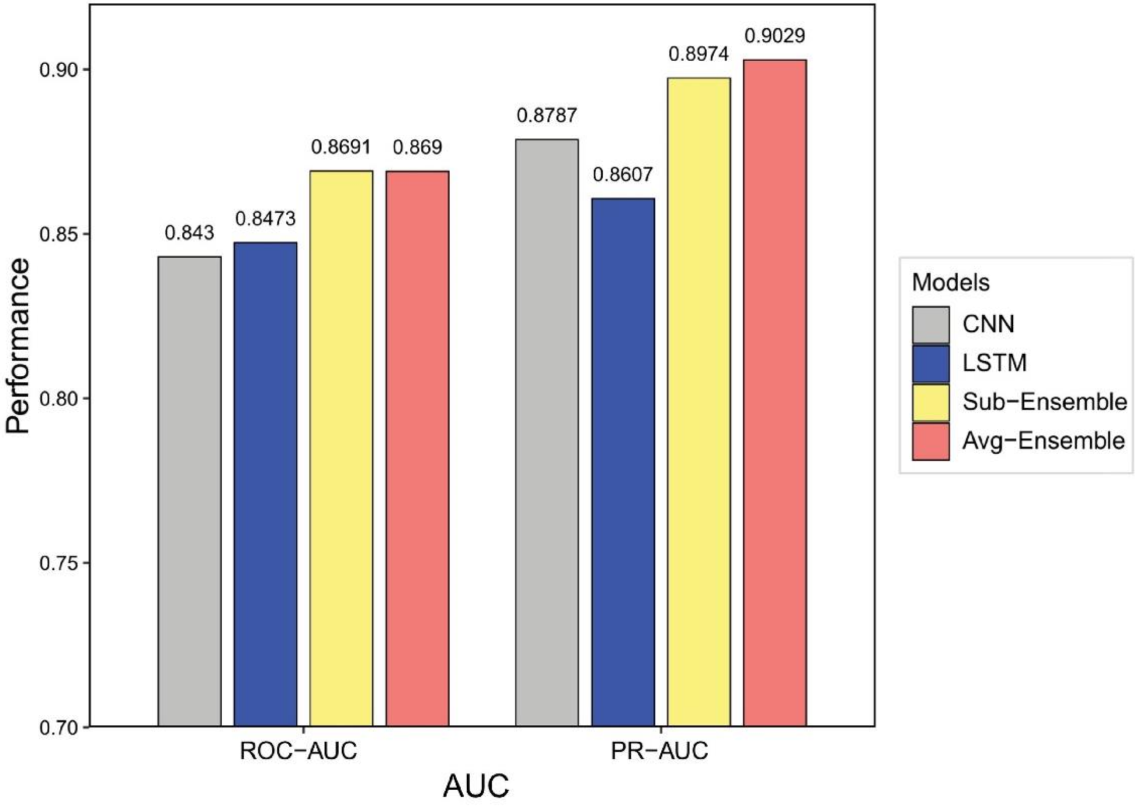
Introduction

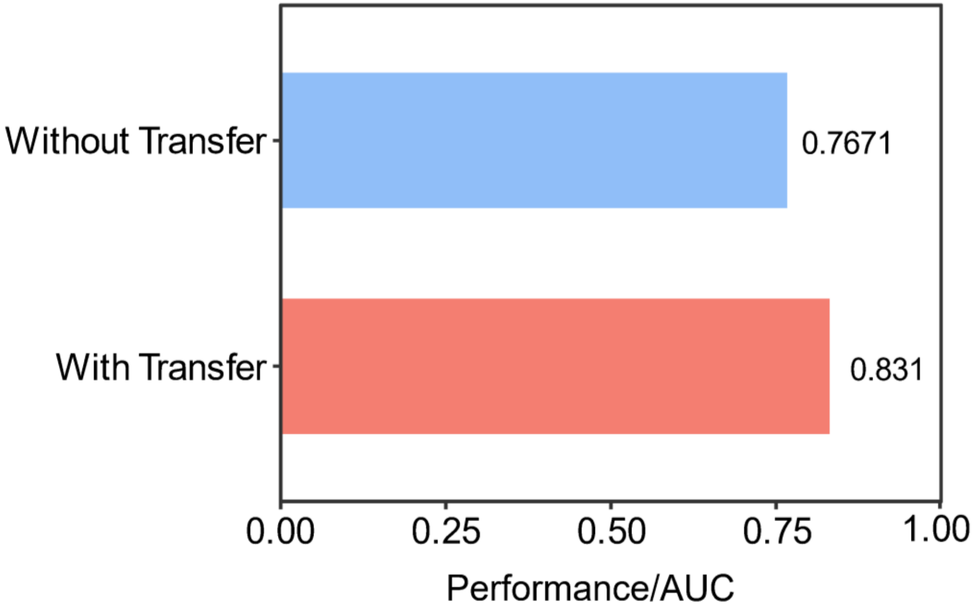


Highlights

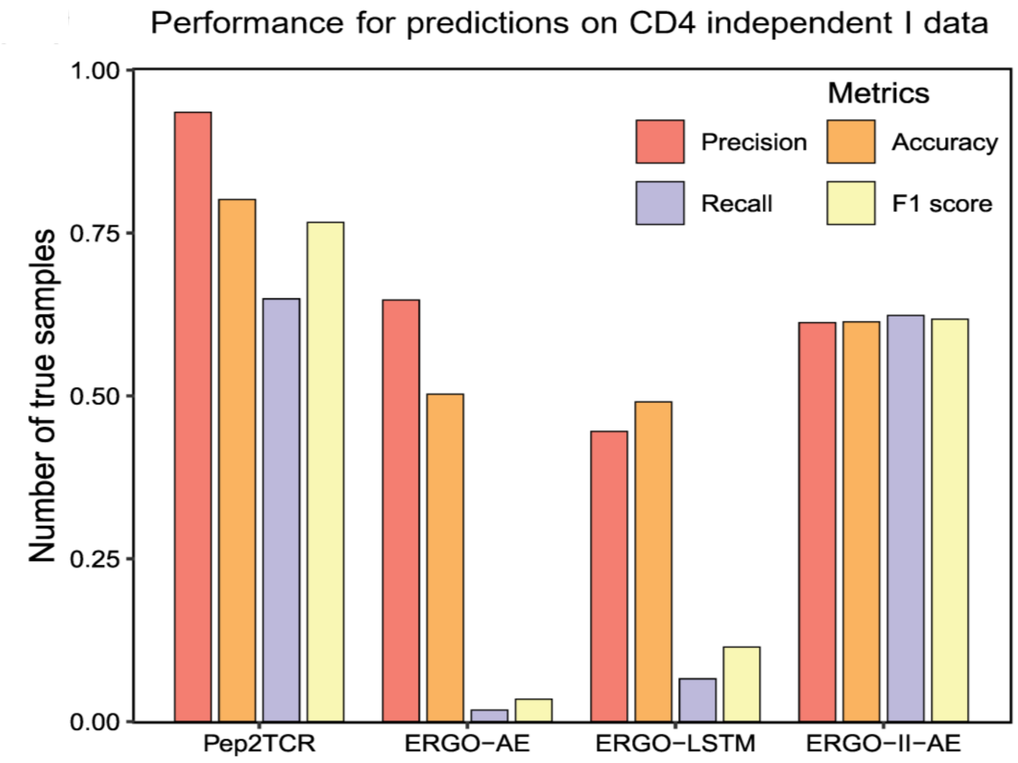
1. We developed a novel deep learning model named Pep2TCR, integrating transfer learning and ensemble methods, to predict the specificity of CD4 TCRs;
2. Pep2TCR outperforms existing models;
3. Pep2TCR can identify characteristics of CD4+ T cells reactive to neoantigens;
4. We created a user-friendly website and an accessible Docker container.

Transfer learning and ensemble method are useful

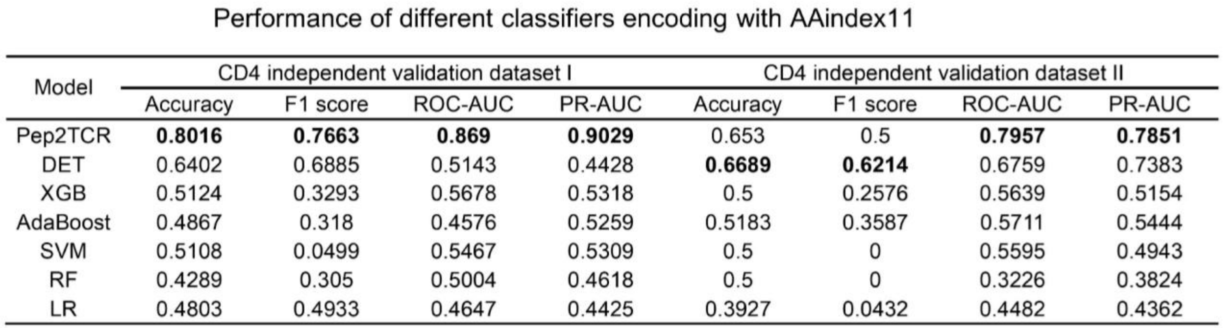




Pep2TCR outperforms existing methods

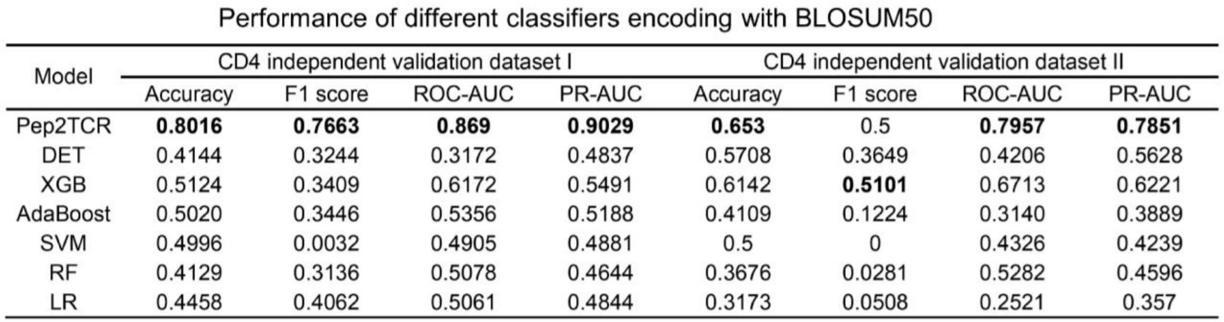


Performance for predictions on CD4 independent I data



Performance of different classifiers encoding with AAindex11

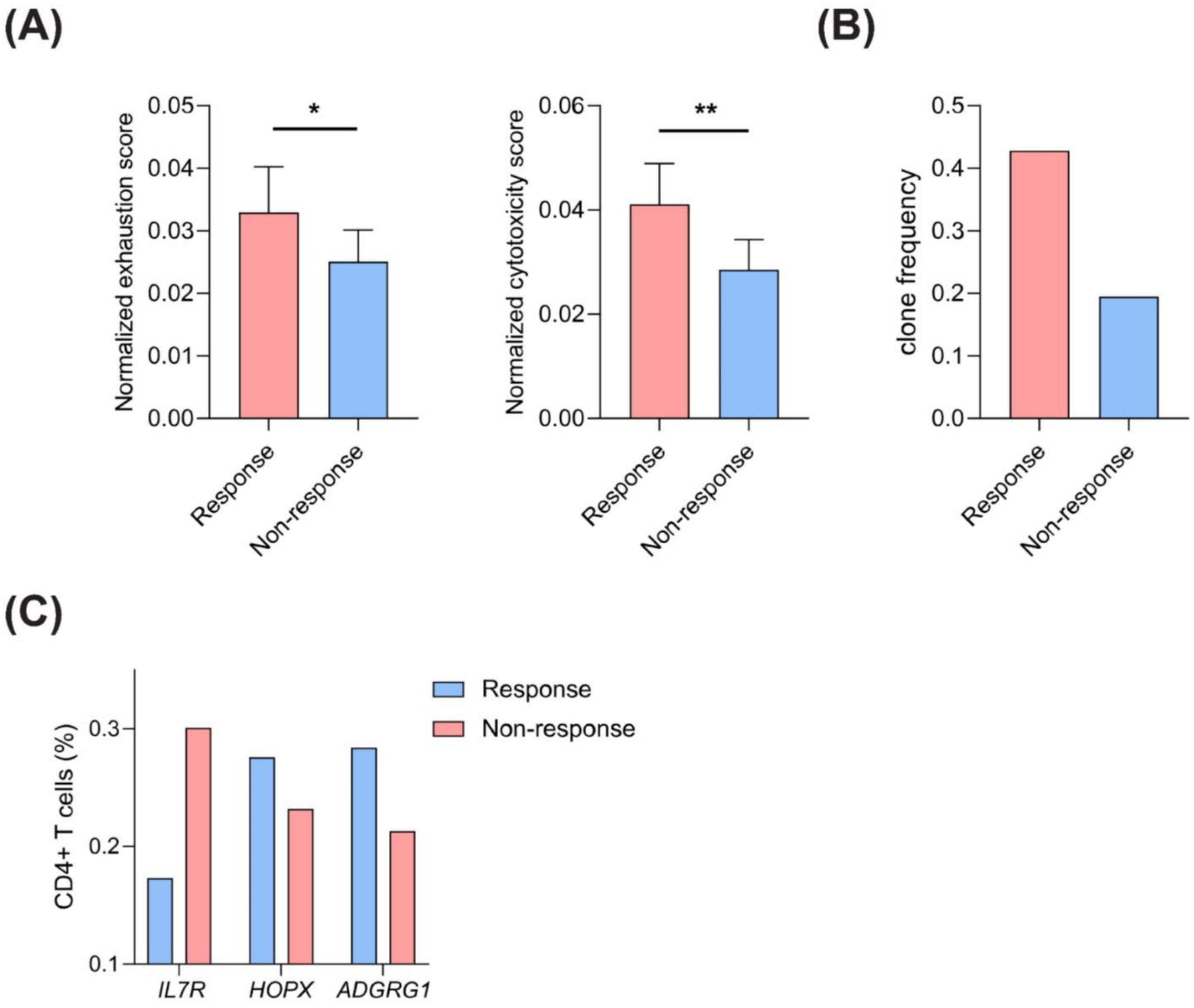
Model	CD4 independent validation dataset I				CD4 independent validation dataset II			
	Accuracy	F1 score	ROC-AUC	PR-AUC	Accuracy	F1 score	ROC-AUC	PR-AUC
Pep2TCR	**0.8016**	**0.7663**	**0.869**	**0.9029**	0.653	0.5	**0.7957**	**0.7851**
DET	0.6402	0.6885	0.5143	0.4428	**0.6689**	**0.6214**	0.6759	0.7383
XGB	0.5124	0.3293	0.5678	0.5318	0.5	0.2576	0.5639	0.5154
AdaBoost	0.4867	0.318	0.4576	0.5259	0.5183	0.3587	0.5711	0.5444
SVM	0.5108	0.0499	0.5467	0.5309	0.5	0	0.5595	0.4943
RF	0.4289	0.305	0.5004	0.4618	0.5	0	0.3226	0.3824
LR	0.4803	0.4933	0.4647	0.4425	0.3927	0.0432	0.4482	0.4362



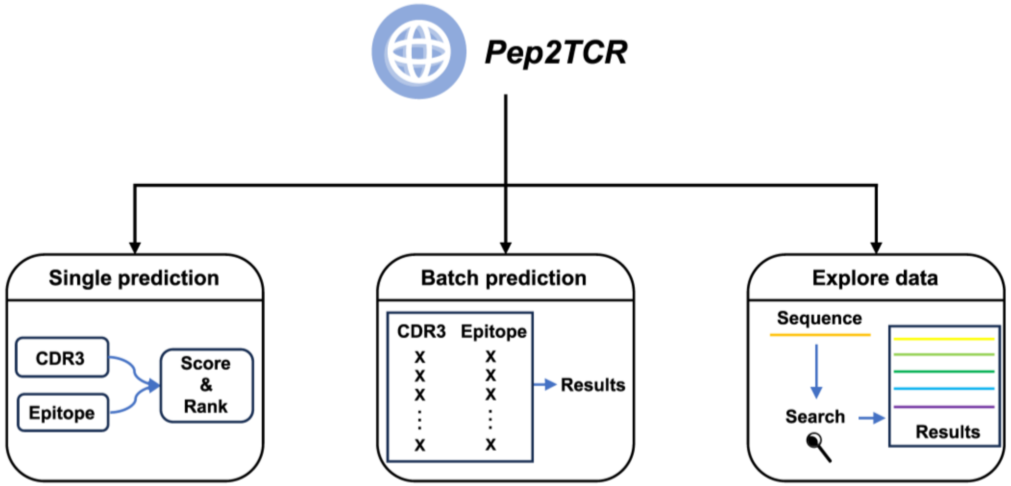
Performance of different classifiers encoding with BLOSUM50

Model	CD4 independent validation dataset I				CD4 independent validation dataset II			
	Accuracy	F1 score	ROC-AUC	PR-AUC	Accuracy	F1 score	ROC-AUC	PR-AUC
Pep2TCR	**0.8016**	**0.7663**	**0.869**	**0.9029**	**0.653**	0.5	**0.7957**	**0.7851**
DET	0.4144	0.3244	0.3172	0.4837	0.5708	0.3649	0.4206	0.5628
XGB	0.5124	0.3409	0.6172	0.5491	0.6142	**0.5101**	0.6713	0.6221
AdaBoost	0.5020	0.3446	0.5356	0.5188	0.4109	0.1224	0.3140	0.3889
SVM	0.4996	0.0032	0.4905	0.4881	0.5	0	0.4326	0.4239
RF	0.4129	0.3136	0.5078	0.4644	0.3676	0.0281	0.5282	0.4596
LR	0.4458	0.4062	0.5061	0.4844	0.3173	0.0508	0.2521	0.357

Pep2TCR identifies characteristics of reactive CD4+ T cells against neoantigens



Pep2TCR website



Pep2TCR

Single prediction

CDR3

Epitope

Score & Rank

Batch prediction

CDR3	Epitope
X	X
X	X
X	X
⋮	⋮
X	X

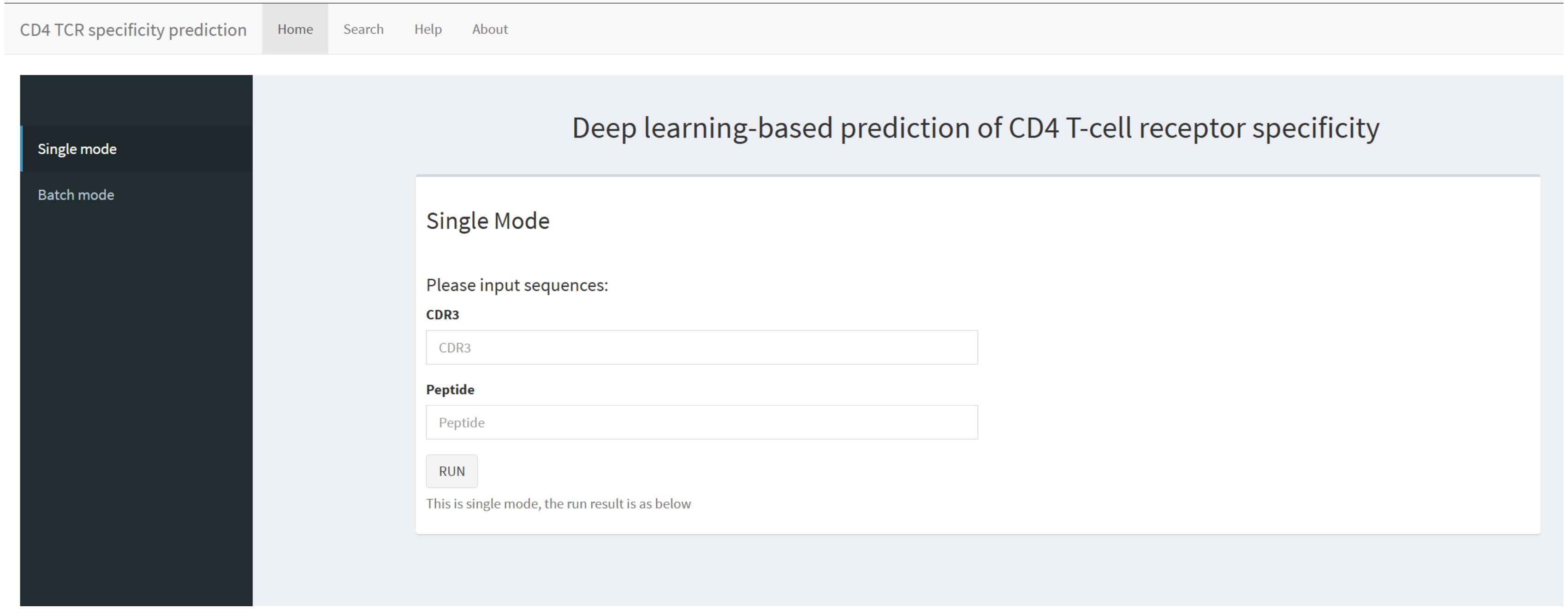
Results

Explore data

Sequence

Search

Results



Deep learning-based prediction of CD4 T-cell receptor specificity

Single Mode

Please input sequences:

CDR3: CDR3

Peptide: Peptide

This is single mode, the run result is as below

http://pep2tcr.liuxslab.com/

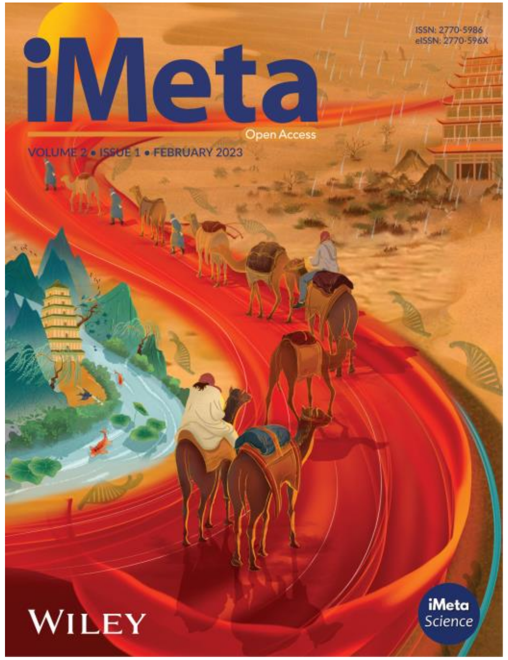
Summary

- ❑ Pep2TCR is a novel deep learning tool specifically designed to predict CD4 T cell receptor binding specificity, enhancing predictive accuracy;
- ❑ This study employs transfer learning and ensemble methods to address the issue of limited CD4 TCR data;
- ❑ Pep2TCR demonstrates excellent performance in identifying characteristics of reactive CD4+ T cells against neoantigens;
- ❑ The tool offers a user-friendly website and Docker container, facilitating CD4 TCR specificity predictions for researchers.

Kaixuan Diao, Tao Wu, Xiangyu Zhao, Nan Wang, Die Qiu, Wei-Liang Wang, Xinxiang Li, Xue-Song Liu. 2024. Pep2TCR: accurate prediction of CD4 T cell receptor binding specificity through transfer learning and ensemble approach. ***iMetaOmics*** 1: e43. https://doi.org/10.1002/imo2.43

***iMeta*:** Integrated meta-omics to change the understanding of the biology and environment

WILEY



"***iMeta***" is a Wiley partner journal launched by iMeta Science Society in 2022, first impact factor (IF) **23.8** in 2024, ranking 2/165 in the microbiology field. It aims to publish innovative and high-quality papers with broad and diverse audiences. **Its scope is similar to *Nature Biotechnology, Nature Methods, Nature Microbiology , Nature Food,* etc**. Its unique features include video abstract, bilingual publication, and social media dissemination, with more than 600,000 followers. It has published 220+ papers and been cited for 5000+ times, and has been indexed by **ESCI/WOS**, **PubMed**, **Google Scholar**, and **Scopus**.

"***iMetaOmics***" is a sister journal of "***iMeta***" launched in 2024, with a **target IF>10, and its scope is similar to *Nature Communications***, *Microbiome, ISME J, Nucleic Acids Research, Briefings in Bioinformatics,* etc. All contributes are welcome!

Society: http://www.imeta.science
Publisher: https://wileyonlinelibrary.com/journal/imeta
Submission: https://wiley.atyponrex.com/journal/IMT2
https://wiley.atyponrex.com/journal/IMO2

office@imeta.science
imetaomics@imeta.science

Promotion Video

iMetaScience

iMetaScience