



Universidad La Salle

# Prediction of peptide MHC presentation using transformers and transfer learning in cancer immunology context

MSc. Vicente Machaca Arceda

2023

# Content



Problem

Related Works

Proposal

Preliminary Results

Models and databases  
Comparison

Conclusions

# Content



Problem

Related Works

Proposal

Preliminary Results

Models and databases  
Comparison

Conclusions

# Problem



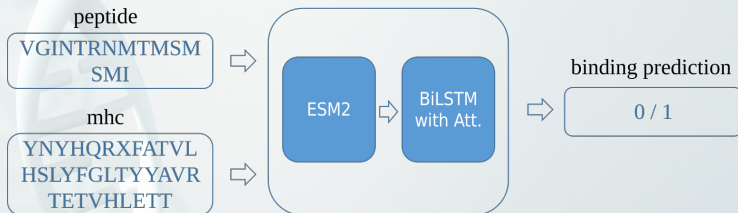
**Less than 5%** of detected neoantigens (peptides binded to MHC) succeed in activating the immune system [1]. Moreover, recent proposals only achieve 0.6 precision and 0.4 recall [2].

This is a **binary classification problem**. A peptide could be represented like:  $p = \{A, \dots, Q\}$  and a MHC like:  $q = \{A, N, \dots, Q, E\}$ . Finally, we need to know the probability of affinity between  $p$  and  $q$  (pMHC)

# Problem



Figure: pMHC binding prediction problem.



**Figure:** Proposal for pMHC binding and presentation prediction.

# Content



Problem

Related Works

Proposal

Preliminary Results

Models and databases

Comparison

Conclusions



We used the dataset from NetMHCIIpan3.2 [3] and HLAB [4].

**Table:** Number of samples used in training, evaluation and testing.

	<b>NetMHCIIpan3.2</b>	<b>HLAB</b>
<b>Train</b>	107424	539019
<b>Validation</b>	13428	179673
<b>Testing</b>	13429	172580





Instead of ESM2 [5] model, we used TAPE [6] because it is smaller and easier to train. Moreover, the Bi-LSTM with attention layer is based on HLAB [4].

**Table:** Models used in experiments.

	Description
<b>BERTMHC-LINEAR</b>	BERT architecture followed by a linear layer
<b>BERTMHC-RNN</b>	BERT architecture followed by a BiLSTM layer and then a Linear layer
<b>BERTMHC-RNN-ATT</b>	BERT architecture followed by a BiLSTM layer with attention and then a Linear layer

# Content



Problem

Related Works

Proposal

Preliminary Results

Models and databases  
Comparison

Conclusions

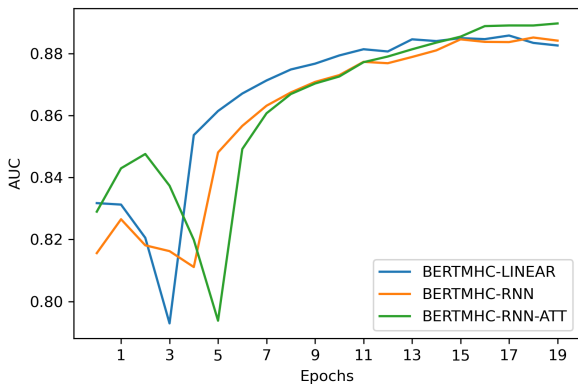


Figure: AUC per epoch of models.

# Comparison

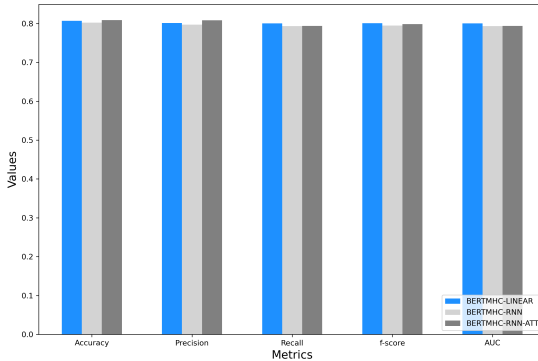


Figure: Metrics comparison.

**Table:** Metrics comparison of BERTMHC-LINEAR, BERTMHC-RNN and BERTMHC-RNN-ATT

Model	Acc	Precision	Recall	Fscore	AUC
LINEAR	0.8070	0.8012	<b>0.8005</b>	<b>0.8009</b>	<b>0.8005</b>
RNN	0.8023	0.7972	0.7932	0.7949	0.7932
RNN-ATT	<b>0.8086</b>	<b>0.8082</b>	0.7937	0.7985	0.7937

# Comparison

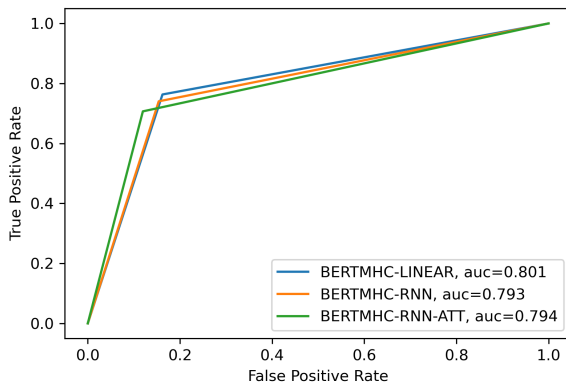


Figure: ROC curve.

# Conclusions



We evaluated a BERT architecture (transformer) with transfer learning from TAPE in these preliminary results. We choose TAPE because it is smaller and easier to train. In future experiments, we will evaluate ESM2.

According to experiments, BERTMHC-LINEAR and BERTMHC-RNN-ATT got better results in netMHCIIpan3.2 dataset. It happens because we evaluated these models in a small dataset. In future experiments, we will evaluate these models in a larger dataset.



- [1] L. Mattos, M. Vazquez, F. Finotello, R. Lepore, E. Porta, J. Hundal, P. Amengual-Rigo, C. Ng, A. Valencia, J. Carrillo *et al.*, “Neoantigen prediction and computational perspectives towards clinical benefit: recommendations from the esmo precision medicine working group,” *Annals of oncology*, vol. 31, no. 8, pp. 978–990, 2020.
- [2] N. A. Mill, C. Bogaert, W. van Criekinge, and B. Fant, “neoms: Attention-based prediction of mhc-i epitope presentation,” *bioRxiv*, 2022.
- [3] K. K. Jensen, M. Andreatta, P. Marcatili, S. Buus, J. A. Greenbaum, Z. Yan, A. Sette, B. Peters, and M. Nielsen, “Improved methods for predicting peptide binding affinity to mhc class ii molecules,” *Immunology*, vol. 154, no. 3, pp. 394–406, 2018.





- [4] Y. Zhang, G. Zhu, K. Li, F. Li, L. Huang, M. Duan, and F. Zhou, “Hlab: learning the bilstm features from the protbert-encoded proteins for the class i hla-peptide binding prediction,” *Briefings in Bioinformatics*, 2022.
- [5] Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli *et al.*, “Evolutionary-scale prediction of atomic-level protein structure with a language model,” *Science*, vol. 379, no. 6637, pp. 1123–1130, 2023.
- [6] R. Rao, N. Bhattacharya, N. Thomas, Y. Duan, P. Chen, J. Canny, P. Abbeel, and Y. Song, “Evaluating protein transfer learning with tape,” *Advances in neural information processing systems*, vol. 32, 2019.

