# An in silico deep learning approach to multi-epitope vaccine design: a SARS-CoV-2 case study

## OVERVIEW

In silico vaccine design

Virus protein sequence

(Traditional method)

Predict individually through different tools –

- B cell epitope
- CTL epitope
- HTL epitope
- Antigenicity

Evaluate and select the best subunits

Cloning the best vaccine subunits

Multiepitope vaccine

**DeepVacPred**

Not a replacement to in silico vaccine design, but aims to make it more efficient. It is able to analyze best of all 4 features at once from one peptide sequence as input in less than 1 sec. (So, skip 95% of unnecessary predictions)

The result gives us less than 30 potential vaccine subunits. It predicts their secondary and 3D structure.

In addition, it can predict emerging threats by RNA mutations too, proving whether the given subunit

Codon optimization and in silico cloning is performed for checking effective expression.

## IMPLEMENTATION PROCESS SUMMARY

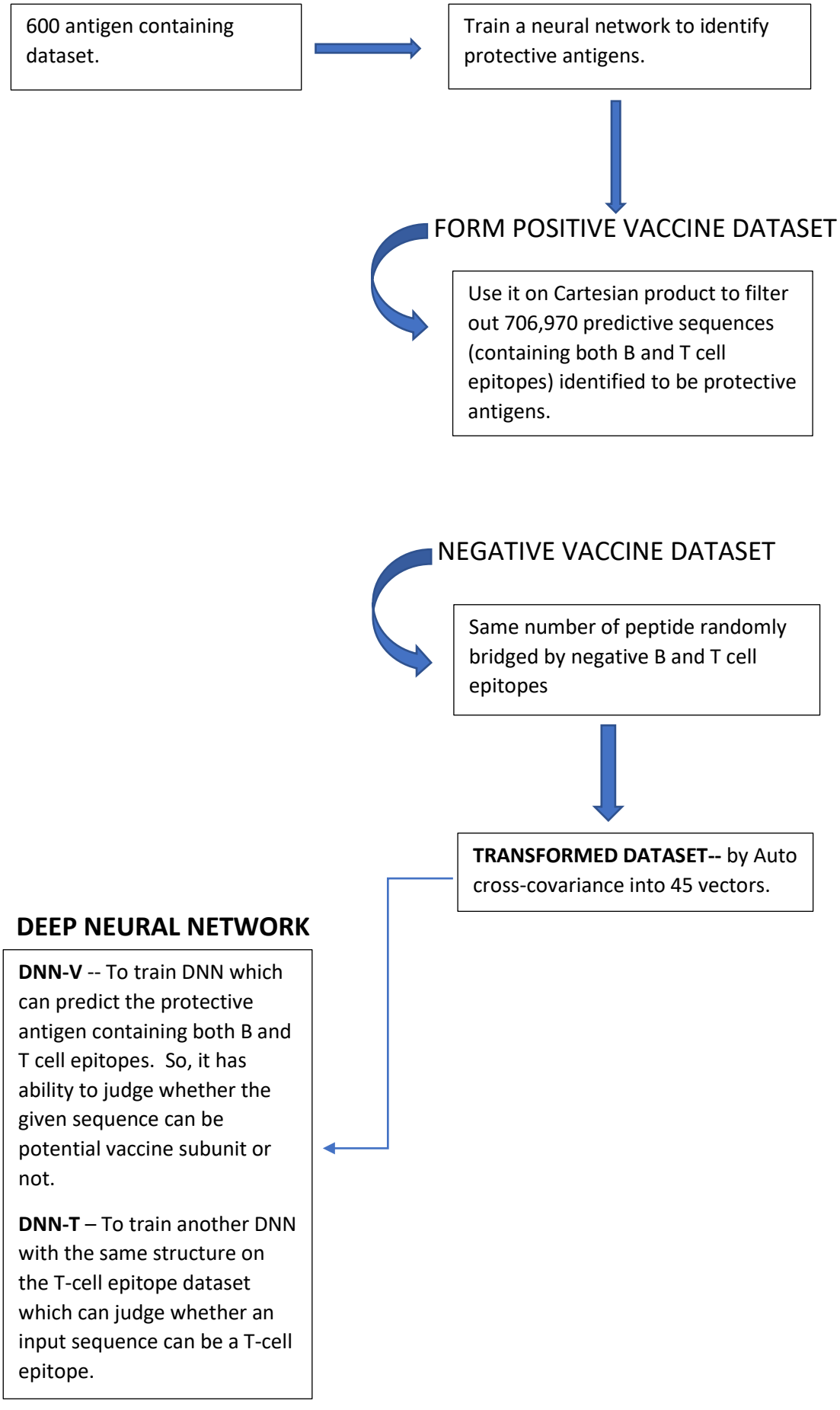
DATA COLLECTED (randomly selected)–

1. 5000 known B cell epitopes + 7000 non-epitope proteins
2. 2000 known T cell epitopes
3. 100 latest protective viral antigens
4. 100 latest non-protective viral antigens
5. 400 previous(old) viral antigens

Dataset containing epitopes and non-epitopes.

Dataset containing 600 antigens.

CARTESIAN PRODUCT – contains all the possible combinations between B and T cell epitopes.



## DNN-V dataset should be validated.

So, to test the trained DNN-V dataset we divide that into train and test set. Both contain 200 protein sequences having 100 positive ones and 100 negative ones.

Test set contain same type of protein sequence of known epitopes but not from our original dataset.

We keep ROC threshold to low value to decrease our false positive number which in turn decrease our non-candidates.

### 132 total potential vaccine subunits were selected by this DNN.

When cross-checked with Vaxijen, only 14 were known to be nonprotective.

From these 132 potential subunits, the ones containing less than 8 T cell epitopes were discarded.

**It left us with 14 potential subunits**.

Now 3 more subunit vaccines were discarded with poor performance in CTL and HTL sequence predictions.

**It left us with 11 potential subunits.**

By population coverage analysis tool, they figured out that their 25 HLA to predict T cell epitope can cover almost entire human population.

- **Multi-epitope vaccine construction**
- **Secondary structure formation**
- **3D structure formation**