# Soft computing for constructive peptide design and peptide activity prediction


**Goran Mauša[1,2,3] Erik Otović[1,3] Marko Njirjak[1,3] Ivan Erjavac[1] Daniela Kalafatovic[2,3,4]**

[1] University of Rijeka, Faculty of Engineering, Department of Computing, Vukovarska 58, 51000 Rijeka, Croatia
[2] University of Rijeka, Center for Advanced Computing and Modeling, Radmile Matejčić 2, 51000 Rijeka, Croatia
[3] University of Rijeka, Center for Artificial Intelligence and Cybersecurity, , Radmile Matejčić 2, 51000 Rijeka, Croatia
[4] University of Rijeka, Department of Biotechnology, Radmile Matejčić 2, 51000 Rijeka, Croatia


## Soft Computing

- **Soft computing is a set of probabilistic algorithms**
  - robust to imprecision and tolerant to uncertainty,
  - enable us to grapple with analytically intractable problems,
  - make up for the lack of theoretical knowledge.
- **We apply a wide range of soft computing models to:**
  - predict peptide activity,
  - construct novel peptides,
  - cover the chemical search space.

## Project goals

- **In our project we tackle the problems of:**
  1. Sensivity of highly accurate predictive models [1],
  2. Informative representation schemes for peptides [2],
  3. Motif identification in sequence composition patterns [3],
  4. Building predictive models with low amount of available data,
  5. Interpretability of neural network-based classifiers,
  6. Ability to generate new peptide sequences,
  7. Coverage-based parallel exploration of chemical space.

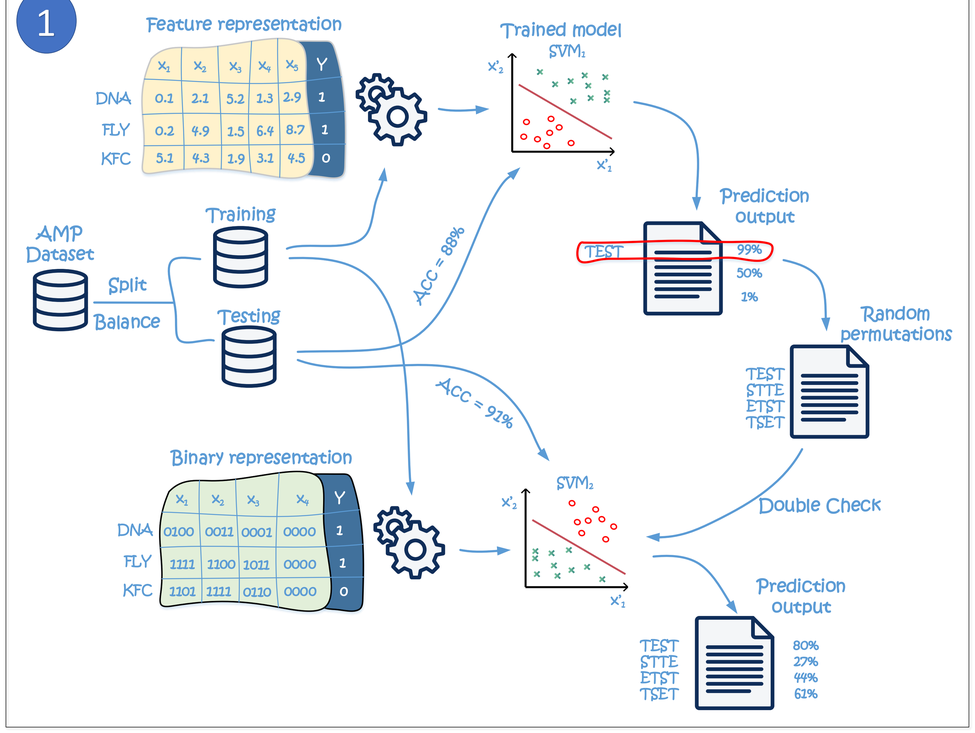
## Publications (2022)


[1] I. Erjavac, D. Kalafatovic, G. Mauša. **Coupled encoding methods for antimicrobial peptide prediction: How sensitive is a highly accurate model?**, Artificial Intelligence in the Life Sciences, Vol. 2, 100034

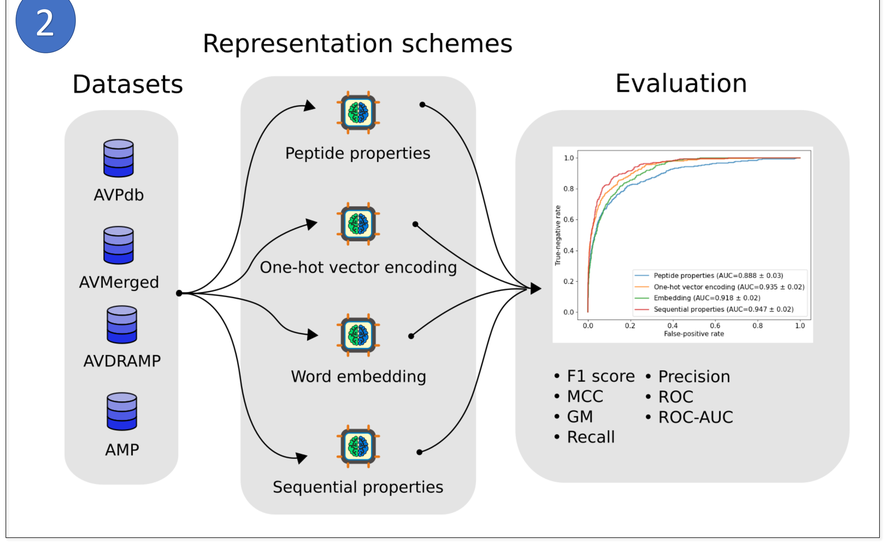
[2] E. Otović, M. Njirjak, D. Kalafatovic, G. Mauša. **Sequential Properties Representation Scheme for Recurrent Neural Network-Based Prediction of Therapeutic Peptides**, Journal of Chemical Information and Modeling, Vol. 62, 12, pp. 2961–2972

[3] M. Babić, P. Janković, S. Marchesan, G. Mauša, D. Kalafatovic. **Esterase Sequence Composition Patterns for the Identification of Catalytic Triad Microenvironment Motifs**, Journal of Chemical Information and Modeling

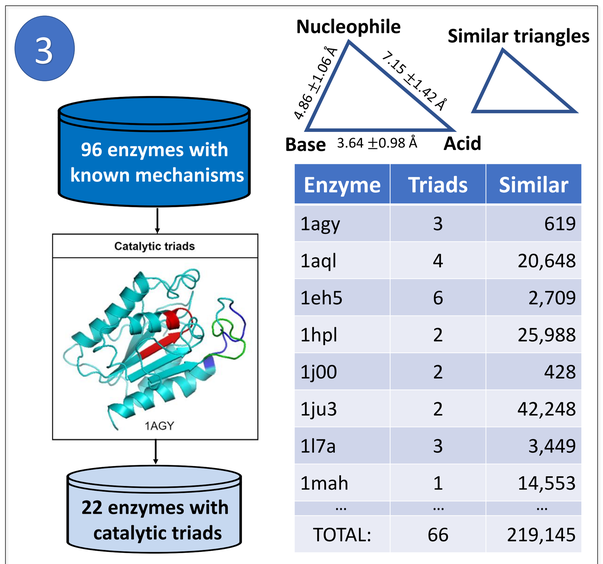

## 1




## 2

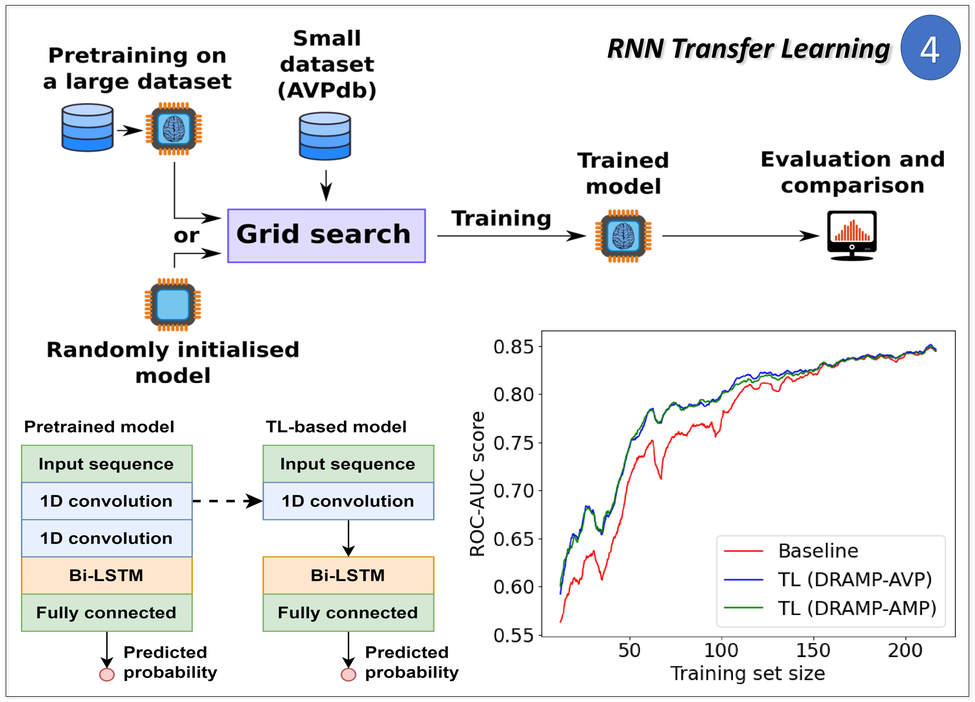



## 3

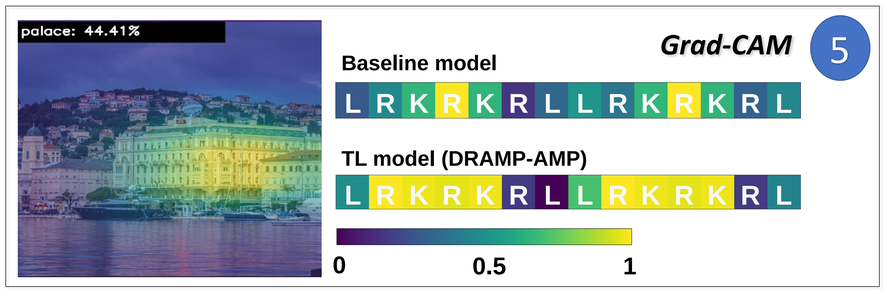



| Enzyme | Triads | Similar |
|---|---|---|
| 1agy | 3 | 619 |
| 1aql | 4 | 20,648 |
| 1eh5 | 6 | 2,709 |
| 1hpl | 2 | 25,988 |
| 1j00 | 2 | 428 |
| 1ju3 | 2 | 42,248 |
| 1l7a | 3 | 3,449 |
| 1mah | 1 | 14,553 |
| ... | ... | ... |
| TOTAL: | 66 | 219,145 |

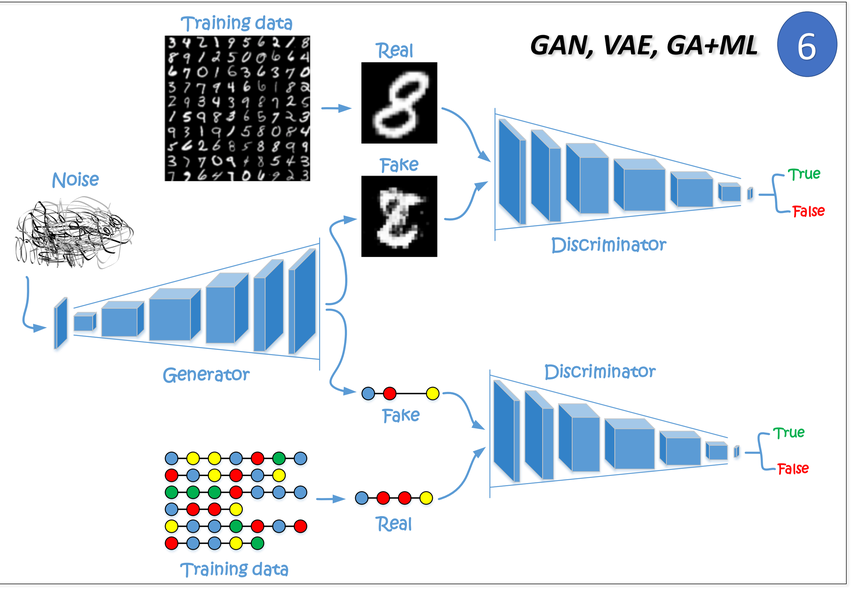
## 4 RNN Transfer Learning



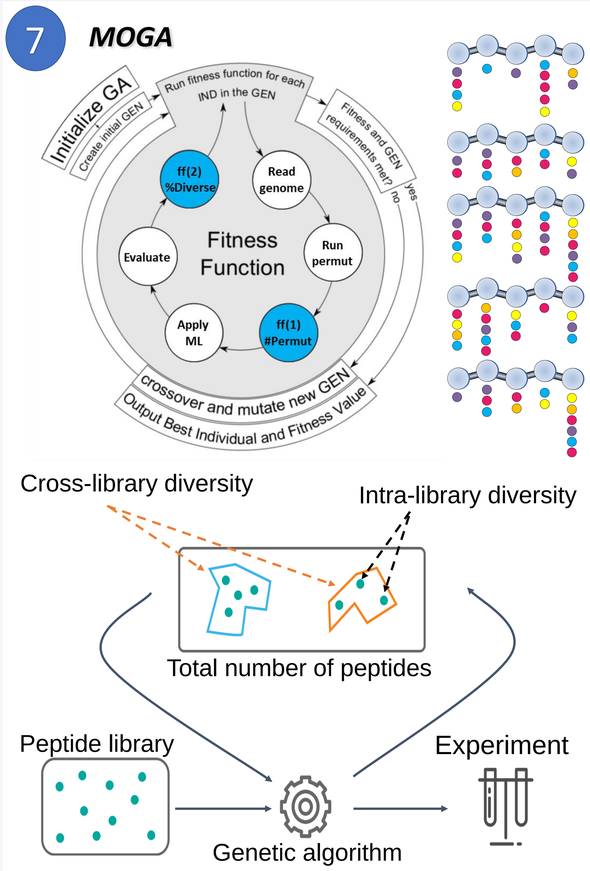

## 5 Grad-CAM




## 6 GAN, VAE, GA+ML




## 7 MOGA




## Design of Short Peptides (DeShPet) Team


web: www.deshpetlab.uniri.hr
web app: www.deshpet.riteh.hr
e-mail: deshpet@riteh.hr


## Conclusion

- **Artificial intelligence is changing the concepts of discovery**
- **Soft computing offers insight into sequence – function relationship**
  - We envision these strategies will expedite peptide research
- **Large and complex search space impossible to analyze manually**
  - Unbiased search may help discover unexplored regions
- **Prediction models are not to be taken for granted**

## This project is supported by:


Croatian Science Foundation
UIP-2019-04-7999
uniri-pr-tehnic-19-10