United International University
Department of Computer Science and Engineering

Mid-Term Report

ACP-DL: Predicting Anticancer Peptides using Traditional Classifiers

---

Course Title: **Machine Learning**
Course Code: CSE 489
Section: **A**

Submitted To:
Dr. Khondaker Abdullah Al Mamun
Professor, Dept of CSE, UIU

Submitted By:
Nasif Ishtiaque Islam (011 171 223)
Saad Ibn Rouf (011 133 004)

Date of Submission: June 5, 2020

1. ABSTRACT

Cancer imposes a global health burden as it represents one of the leading causes of morbidity and mortality while also giving rise to significant economic burden owing to the associated expenditures for its monitoring and treatment. Anticancer peptides open a promising perspective for cancer treatment, and they have various attractive advantages. As the traditional experiments are expensive and inefficient for predicting ACP, the researchers have proposed A deep learning long short-term memory(LSTM) neural network model, ACP-DL, to effectively predict novel anticancer peptides. In this project, we have used some traditional classifiers to predict anticancer peptides.

2. PROBLEM DEFINITION AND BACKGROUND

Cancer is one of the most devastating killers of human beings, accounting for millions of deaths around the world each year. The discovery of anticancer peptides(ACPs), a kind of short peptide around the world each year. The discovery of anticancer peptides(ACPs), a kind of short peptide generally with a length of less than 50 amino acids and most of which are derived from antimicrobial peptides(AMPs). ACPs open a promising perspective for cancer treatment, and they have various attractive advantages over traditional methods like chemotherapy, radiation therapy, and others. It reduces cost and time, besides increasing specificity.

With the huge therapeutic importance on ACPs, there is an urgent need to develop highly efficient prediction techniques.

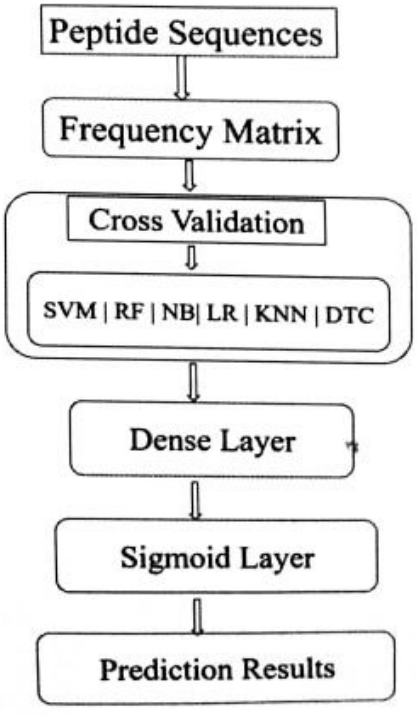

Figure 1. Flowchart of our used methods.

3. RESULTS AND DISCUSSION

We have used ACP740 and ACP240, two benchmarks dataset of anticancer peptide. Here, anticancer peptides and non-anticancer peptides features are available. The results for ACP740 are shown in Table 1, the composition of all 20 amino acids in these peptides were counted and compared.

Dataset	Model	ACC(Paper)	ACC(Implemented)
ACP740	SVM	64.59	79
	RF	76.36	76
	NB	69.73	67
	LR	--------	75
	KNN	--------	77
	DTC	--------	73

Table 1. Actual Performance of Comparison Models

4. PERFORMANCE AND EVALUATION CRITERIA

We followed the widely used evaluation criteria (Table 2), including accuracy (ACC), Sens or recall, Spec, Prec, and MCC, defined as follows:

$$\text{Acc} = \frac{TN + TP}{TN + TP + FN + FP}$$

$$\text{Sens} = \frac{TP}{TP + FN}$$

$$\text{Spec} = \frac{TN}{TN + FP}$$

$$\text{Prec} = \frac{TP}{TP + FP}$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Where TN indicates the true negative numbers, TP denotes true positive numbers, FN denotes false negative numbers, FP denotes false positive numbers, and MCC denotes Matthew's Correlations Coefficient.

Dataset	Model	Sens(%)		Spec(%)		Prec(%)		MCC(%)	
ACP240	SVM	85.89	70	70.68	68	85.65	71	57.31	38
	RF	73.53	57	76.09	69	67.63	67	44.38	27
	NB	72.26	77	79.94	44	63.95	60	44.38	20
	LR	----	76	----	65	----	71	----	41
	KNN	----	71	----	53	----	62	----	24
	DTC	----	59	----	58	----	61	----	17

Table 2. The 5-Fold Cross-Validation Details in the ACP240 Dataset

5. CONCLUSION

There are some differences between our implementation and the selected paper. We extracted features using frequency mapping and the paper used K-mer Spares Matrix and Binary profile. The other difference is on SVM model kernel choosing, using the rbf kernel we got better results.

6. REFERENCES

1. Hai-Cheng Yi, Zhu-Hong You, Xi Zhou, Li Cheng, Xiao Li, Tong-Hai Jiang, and Zhan-Heng Chen, ACP-DL: A Deep Learning Long Short-Term Memory Model to Predict Anticancer Peptides Using High-Efficiency Feature Representation

2. Wei, L., Ding, Y., Su, R., Tang, J., and Zou, Q. (2018). Prediction of human protein subcellular localization using deep learning. J. Parallel Distrib. Comput. 117, 212–217.