



*Research article*

## **UPFPSR: a ubiquitylation predictor for plant through combining sequence information and random forest**

**Shuwan Yin<sup>1</sup>, Jia Zheng<sup>1</sup>, Cangzhi Jia<sup>1,\*</sup>, Quan Zou<sup>2,3</sup>, Zhengkui Lin<sup>4,\*</sup> and Hua Shi<sup>5,\*</sup>**

<sup>1</sup> School of Science, Dalian Maritime University, Dalian 116026, China

<sup>2</sup> Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, Quzhou, China

<sup>3</sup> Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, China

<sup>4</sup> School of Maritime Economics and Management, Dalian Maritime University Dalian 116026, China

<sup>5</sup> School of Opto-electronic and Communication Engineering, Xiamen University of Technology, Xiamen, China

\* **Correspondence:** Email: [cangzhijia@dlmu.edu.cn](mailto:cangzhijia@dlmu.edu.cn), [dalianjx@163.com](mailto:dalianjx@163.com), [shihua@xmut.edu.cn](mailto:shihua@xmut.edu.cn).

## **Supplementary**

---

**Table S1.** Selection of optimal RF parameter according to results of the ten-fold cross-validation.

<b>NumTrees</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>
50	0.787	0.786	0.788	0.786
100	0.795	0.797	0.794	0.795
150	0.803	0.804	0.801	0.802
200	0.811	0.810	0.812	0.811
250	0.804	0.805	0.803	0.803
300	0.804	0.806	0.803	0.804

**Table S2.** Selection of optimal SVM parameter according to results of the ten-fold cross-validation.

<b>C</b>	<b><math>\gamma</math></b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>
2 <sup>2</sup>	0.01	0.721	0.720	0.723	0.721
2 <sup>2</sup>	0.1	0.747	0.748	0.745	0.746
2 <sup>2</sup>	1	0.702	0.818	0.523	0.636
2 <sup>3</sup>	0.01	0.721	0.720	0.723	0.721
2 <sup>3</sup>	0.1	0.747	0.748	0.745	0.746
2 <sup>3</sup>	1	0.702	0.818	0.523	0.636
2 <sup>4</sup>	0.01	0.720	0.721	0.721	0.720
2 <sup>4</sup>	0.1	0.743	0.744	0.740	0.742
2 <sup>4</sup>	1	0.700	0.817	0.520	0.634
2 <sup>5</sup>	0.01	0.722	0.721	0.725	0.722
2 <sup>5</sup>	0.1	0.745	0.746	0.744	0.744
2 <sup>5</sup>	1	0.701	0.815	0.526	0.637

**Table S3.** Selection of optimal KNN parameter according to results of the ten-fold cross-validation.

<b>k</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>
3	0.663	0.638	0.751	0.690
5	0.663	0.638	0.751	0.690
7	0.666	0.641	0.757	0.694
9	0.673	0.646	0.769	0.702
11	0.676	0.645	0.781	0.707
13	0.675	0.646	0.776	0.705
15	0.674	0.643	0.781	0.705

**Table S4.** Performance comparison between CNN+word2vec and UPFPSR for identifying plant ubiquitylation sites on the independent test set.

<b>Method</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>	<b>AUC</b>
CNN+word2vec	0.756	0.733	0.767	0.749	0.81
UPFPSR	0.773	0.750	0.817	0.782	0.84