



---

*Research article*

## **SoftVoting6mA: An improved ensemble-based method for predicting DNA N6-methyladenine sites in cross-species genomes**

**Zhaoting Yin<sup>1</sup>, Jianyi Lyu<sup>1</sup>, Guiyang Zhang<sup>1</sup>, Xiaohong Huang<sup>1</sup>, Qinghua Ma<sup>2,3</sup> and Jinyun Jiang<sup>1,\*</sup>**

<sup>1</sup> College of Information Science and Engineering, Shaoyang University, Shaoyang 422000, China

<sup>2</sup> College of Information Science and Engineering, Hohai University, Nanjing 210000, China

<sup>3</sup> Faculty of Information Technology, University of Jyvaskyla, Jyvaskyla, Finland

\* **Correspondence:** Email: [tjjy86@ecjtu.edu.cn](mailto:tjjy86@ecjtu.edu.cn).

---

**Table S1.** Performance comparison of deep learning model with varying Filter and Kernel Sizes.

Filters	Kernel_size	ACC
32	5	0.723
32	9	0.730
32	13	0.722
64	5	0.729
64	9	0.727
64	13	0.708
128	5	0.744
128	9	0.745
128	13	0.733

**Table S2.** Performance comparison of deep learning model with different Dropout.

Dropout	ACC
0.5	0.829
0.7	0.801
0.8	0.786

**Table S3.** Performance comparison of deep learning model with different Dense.

Dense	ACC
4	0.826
8	0.836
16	0.827

**Table S4.** The performance of single representations using XGBoost.

Feature Type	SN	SP	ACC	MCC	AUC
EIIP	0.833	0.756	0.795	0.591	0.867
One-hot	0.824	0.759	0.792	0.585	0.866
NCP	0.825	0.751	0.788	0.577	0.864
Kmer	0.644	0.694	0.669	0.339	0.736
RCKmer	0.614	0.647	0.630	0.261	0.688
PseDNC	0.604	0.652	0.628	0.256	0.686
DPCP	0.594	0.653	0.623	0.247	0.678
ANF	0.603	0.630	0.617	0.234	0.658
DACC	0.612	0.609	0.610	0.220	0.657
NAC	0.573	0.638	0.606	0.212	0.647
DCC	0.606	0.603	0.605	0.210	0.640
DAC	0.573	0.586	0.579	0.159	0.618
TCC	0.539	0.572	0.555	0.110	0.564
TACC	0.520	0.584	0.552	0.104	0.574
TAC	0.520	0.559	0.539	0.079	0.555

**Table S5.** The performance of single representations using AdaBoost.

Feature Type	SN	SP	ACC	MCC	AUC
EIIP	0.757	0.776	0.767	0.534	0.842
One-hot	0.758	0.774	0.766	0.532	0.842
NCP	0.756	0.774	0.765	0.530	0.842
Kmer	0.656	0.721	0.689	0.378	0.752
RCKmer	0.615	0.674	0.645	0.290	0.705
DPCP	0.601	0.688	0.644	0.290	0.706
PseDNC	0.596	0.673	0.635	0.270	0.697
NAC	0.589	0.659	0.624	0.249	0.677
DACC	0.606	0.630	0.618	0.237	0.666
DCC	0.599	0.629	0.614	0.229	0.651
DAC	0.584	0.620	0.603	0.206	0.640
TAC	0.507	0.640	0.574	0.149	0.597
TACC	0.494	0.632	0.563	0.127	0.586
TCC	0.462	0.650	0.556	0.114	0.571
ANF	0.494	0.573	0.534	0.068	0.548

**Table S6.** The performance of single representations using SVM.

Feature Type	SN	SP	ACC	MCC	AUC
One-hot	0.789	0.759	0.774	0.546	0.843
NCP	0.788	0.759	0.773	0.5473	0.844
EIIP	0.890	0.624	0.757	0.534	0.762
PseDNC	0.664	0.617	0.640	0.281	0.704
NAC	0.656	0.613	0.634	0.269	0.689
RCKmer	0.607	0.647	0.627	0.255	0.683
DACC	0.637	0.615	0.626	0.252	0.673
DPCP	0.620	0.628	0.624	0.247	0.681
DCC	0.623	0.610	0.616	0.233	0.662
Kmer	0.292	0.916	0.604	0.270	0.719
DAC	0.597	0.594	0.596	0.191	0.640
TAC	0.516	0.647	0.582	0.165	0.614
TACC	0.526	0.637	0.581	0.163	0.616
TCC	0.500	0.652	0.576	0.153	0.601
ANF	0.499	0.609	0.554	0.109	0.570

**Table S7.** Comparative performance metrics in rice, mouse, and combined species datasets.

Data	SN	SP	ACC	MCC	AUC
Rice	0.883	0.866	0.874	0.749	0.956
Mouse	1.000	0.933	0.966	0.935	0.941
Fusion	0.777	0.808	0.792	0.585	0.789



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)