Usual Manual of AI-m6ARS

About AI-m6ARS

A machine learning model for Accurate Identification of M6A RNA modification Sites (AI-m6ARS) is a recent interpretable machine learning model that accurately identifies m6A RNA modification sites using high-resolution experimental data. AI-m6ARS effectively encodes RNA sequences through four types of features - one-hot encoding, iFeatures, conservation scores, and geographic features - to enhance the site characterisation.

For further information, please refer to:

AI-m6ARS: Machine learning-driven m6A RNA methylation site discovery with integrated sequence, conservation, and geographical descriptors.

Korawich Uthayopas, Alex G. C. de Sá, David B. Ascher. bioRxiv 2024.06.17.599439; doi: https://doi.org/10.1101/2024.06.17.599439

Marked sections in the above toolbox, include:

(A) main page of AI-m6ARS. Users are directed to the job submission page by clicking on **Prediction (1)** page at the top menu. **Help page (2)**

is available for instructions on how to use AI-m6ARS. **Data page (3)** provides the user with all data used in training, cross-validation, and evaluation of the model.



accuracy. This study presents AI-m6ARS, a novel interpretable machine learning method for accurately identifying m6A RNA modification sites using high-resolution experimental data. AI-m6ARS effectively encodes RNA sequences through four types of variables: one-hot encoding, iFeatures, conservation scores and geographic features, overall improving the site characterisation. With the use of two robust feature selection methods and an ensemble-based machine learning algorithm, the resultant model for AI-m6ARS demonstrated generalisable predictive performance across different validation sets.

Job Submission Page



a machine learning model for Accurate Identification of m6A RNA modification Sites

Choose an option		
	\sim	
Step 2: Please provide an input file		
Input file		
Choose file No file chosen		
Please refer to the help section or download example files for format guidance[Example files].	r input	
Step 3: Fill your email addres (optional)		
E-mail address (for sending a notice with the res	ılt link):	
thanks for using simCare@unimelh.edu.au	•	

In the Job submission Page (B), users need to provide the input file for a prediction. AI-m6ARS runs in two modes (1) - transcript scanning and transcript coordinate. In transcript scanning mode, users need to provide one column of transcript identifiers (in ensembl id (ENST)) (2). Our model will return a prediction for all DRACH motifs site within the transcripts. Another mode, transcript coordinate, allows the predictions on provided positions. Users are required to provide the inputs with two columns, transcript identifiers and modified position (2). The examples file can be downloaded under the [Example files] buttons. Please ensure that the header for columns of the provided input files matches the example files. An email address can be provided by the user (optional) **(5)**. The user will be notified with an access link when the analysis is finished.

Waiting Page



After submitting the analysis, users will be redirected to a **waiting page (C)**. Users can follow the progress of their submission with the progress bar **(1)**. If the email address is provided in **(B)**, users can receive the link to the result page in the email.

Result Page - Main



The **Result Page (D)** provides the information on predicted m6A sites for each provided transcript IDs. The **provided identifiers (1)**, **transcripts IDs (2)**, **gene IDs (3)** are provided with **a number of predicted potential m6A sites (4)**, as well as **their position** in terms of **transcript (5)** and **chromosomal (6)**.

For **further details** of predictions in each IDs, please click 'View Details' **(7)**, and the server will redirect you to the **Detailed result page (E)**.

Three buttons in D allow users to **run another prediction (8)**, **download results - summary table (9)**, **download results - detailed table (10)**, and **download an error log (11)**. For each file, please see the Downloadable **results section** for more detail.

Result Page - Detailed result

E AI-mOARS Analysis - ENST0000004531

G	General Information								
	Provided ID		Transcript ID		Gene ID		Chromosome	Length	
	ENST0000004531		ENST0000004531		ENSG0000003989		8	7560	
	Start		op	Strand		Transcript type	No. predicted m6A sites		S
	17538777	1757	0561	+		mRNA		3	2

Prediction result of m6A modification	sites 4	5	6	7
m6A position (transcript)	m6A position (genome)	Prediction	Probability	DRACH motif
60	17538836	No	0.344	AAACA
79	17538855	No	0.36	AGACA
108	17538884	No	0.18	GAACA
235	17543406	Yes	0.503	GGACA
247	17543418	Yes	0.522	AGACA

The distribution of m6A modification sites : ENST00000004531



Result page - detailed result (E) shows detailed information on all predictions for each transcript. It consists of three parts - **General Information**, **Prediction Results**, and the **Distribution of m6A modification sites**.

In **General Information (1)**, information on transcripts is provided, including their position, length, and **number of predicted potential m6A sites (2)** from AI-m6ARS. For further details, clicking the link will direct users to the Ensembl database (https://www.ensembl.org).

The prediction result table includes m6A position in terms of transcript (3) and chromosomal/genomic (4), prediction result (Yes/No) (5), probabilities (0 - 1) (6), and DRACH motif sequence (7).

The figure of the **distribution of m6A modification sites (8)** shows the distribution of all predicted m6A sites for all DRACH motifs with respective probabilities.

Error Page

F

Error					
Error was found during the analyis. Sorry for an inconvenience. If possible, please submit the inquiry to our group website at https://biosig.lab.uq.edu.au/contact					
← Back to submission					

The **error page (F)** is shown when your input is incorrect. Please check the example files for an input format. Common problems are 1) Wrong format in input files 2) Wrong header of columns 3) Setting input mode incorrectly.

If you have any inquiries, please do not hesitate to contact us via our **Group website** (<u>https://biosig.lab.uq.edu.au/contact</u>).

Downloadable results

From the result page, three files can be downloaded. They contain the similar information as the result pages.

 Summary table: An overview table summarising provided ID, transcript ID, gene ID, chromosome, start position, stop position, strand, transcript type, number of predicted m6A sites, site coordinates (transcript), and site coordinates (chromosomal).

- 2) Detailed table: A detailed table for all predicted m6A sites. It includes provided ID, transcript ID, gene ID, chromosome, site coordinate (transcript), site coordinate (chromosomal), the state of being m6A site (Yes/No), calculated probability, DRACH motif sequence, and length of a transcript.
- 3) Error logs: The error entries and related causes will be reported in this file. Please note that the analysis submission will still be completed even though some error entries are found. The error entries will be ignored, and the analysis will be performed on the remaining correct entries.