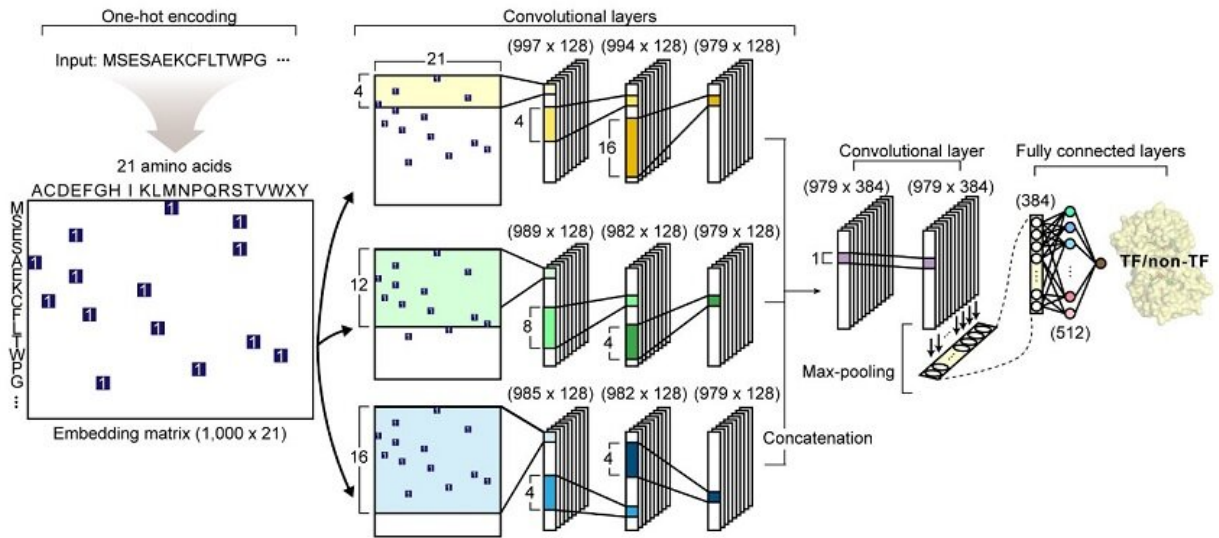


DeepTFactor predicts transcription factors

January 5 2021



The network architecture of DeepTFactor. An input protein sequence is processed using three parallel subnetworks. Credit: The Korea Advanced Institute of Science and Technology (KAIST).

A joint research team from KAIST and UCSD has developed a deep neural network named DeepTFactor that predicts transcription factors from protein sequences. DeepTFactor will serve as a useful tool for understanding the regulatory systems of organisms, accelerating the use of deep learning for solving biological problems.

A [transcription](#) factor is a protein that specifically binds to DNA sequences to control the transcription initiation. Analyzing

[transcriptional regulation](#) enables the understanding of how organisms control gene expression in response to genetic or environmental changes. In this regard, finding the transcription factor of an organism is the first step in the analysis of the transcriptional regulatory system of an organism.

Previously, [transcription factors](#) have been predicted by analyzing sequence homology with already characterized transcription factors or by data-driven approaches such as [machine learning](#). Conventional machine learning models require a rigorous feature selection process that relies on domain expertise such as calculating the physicochemical properties of molecules or analyzing the homology of biological sequences. Meanwhile, [deep learning](#) can inherently learn latent features for the specific task.

A joint research team comprised of Ph.D. candidate Gi Bae Kim and Distinguished Professor Sang Yup Lee of the Department of Chemical and Biomolecular Engineering at KAIST, and Ye Gao and Professor Bernhard O. Palsson of the Department of Biochemical Engineering at UCSD reported a deep learning-based tool for the prediction of transcription factors. Their research paper "DeepTFactor: A deep learning-based tool for the prediction of transcription factors," was published online in *PNAS*.

Their article reports the development of DeepTFactor, a deep learning-based tool that predicts whether a given protein sequence is a transcription factor using three parallel convolutional neural networks. The joint research team predicted 332 transcription factors of *Escherichia coli* K-12 MG1655 using DeepTFactor and the performance of DeepTFactor by experimentally confirming the genome-wide binding sites of three predicted transcription factors (YqhC, YiaU, and YahB).

The joint research team further used a saliency method to understand the

reasoning process of DeepTFactor. The researchers confirmed that even though information on the DNA binding domains of the transcription factor was not explicitly given in the training process, DeepTFactor implicitly learned and used them for prediction. Unlike previous transcription factor prediction tools that were developed only for protein sequences of specific organisms, DeepTFactor is expected to be used in the analysis of the transcription systems of all organisms at a high level of performance.

Distinguished Professor Sang Yup Lee said, "DeepTFactor can be used to discover unknown transcription factors from numerous [protein sequences](#) that have not yet been characterized. It is expected that DeepTFactor will serve as an important tool for analyzing the regulatory systems of organisms of interest."

More information: Gi Bae Kim et al, DeepTFactor: A deep learning-based tool for the prediction of transcription factors, *Proceedings of the National Academy of Sciences* (2020). [DOI: 10.1073/pnas.2021171118](https://doi.org/10.1073/pnas.2021171118)

Provided by The Korea Advanced Institute of Science and Technology (KAIST)

Citation: DeepTFactor predicts transcription factors (2021, January 5) retrieved 12 May 2024 from <https://phys.org/news/2021-01-deeptfactor-transcription-factors.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.