DeepciRGO: functional prediction of circular RNAs through hierarchical deep neural networks using heterogeneous network features

CURRENT STATUS: UNDER REVIEW

BMC Bioinformatics BMC Series

Lei Deng
Central South University

Wei Lin
Central South University

Jiacheng Wang
Central South University

Jingpu Zhang
Henan University of Urban Construction

zhangjp@csu.edu.cn*Corresponding Author*
ORCiD: https://orcid.org/0000-0002-8006-747X

DOI:
10.21203/rs.2.20476/v1

SUBJECT AREAS
Bioinformatics

KEYWORDS
Gene ontology, Representation learning, Hin2vec, Multi-label hierarchical classication

Abstract

Circular RNAs (circRNAs) are special noncoding RNA molecules with closed loop structures. Compared with the traditional linear RNA, circRNA is more stable and not easily degraded. Many studies have shown that circRNAs are involved in the regulation of various diseases and cancers. Determining the functions of circRNAs in mammalian cells is of great significance for revealing their mechanism of action in physiological and pathological processes, diagnosis and treatment of diseases. However, determining the functions of circRNAs on a large scale is a challenging task because of the high experimental costs.

In this paper, we present a hierarchical deep learning model, DeepciRGO, which can effectively predict gene ontology functions of circRNAs. We build a heterogeneous network containing circRNA co-expressions, protein-protein interactions (PPIs) and protein-circRNA interactions. The topology features of proteins and circRNAs are calculated using a novel representation learning approach Hin2vec across the heterogeneous network. Then, a deep multi-label hierarchical classification model is trained with the topology features to predict the biological process (BP) function in the Gene Ontology (GO) for each circRNA. In particular, we manually curated a benchmark dataset containing 185 GO annotations for 62 circRNAs, namely, circRNA2GO-62. The DeepciRGO achieves promising performance on the circRNA2GO-62 dataset with a maximum F-measure of 0.412, a recall score of 0.4, and an accuracy of 0.4, which are significantly better than other state-of-the-art RNA function prediction methods. In addition, we demonstrate the considerable potential of integrating multiple interactions and association networks.

Full Text

Due to technical limitations, full-text HTML conversion of this manuscript could not be completed. However, the manuscript can be downloaded and accessed as a PDF.

Figures

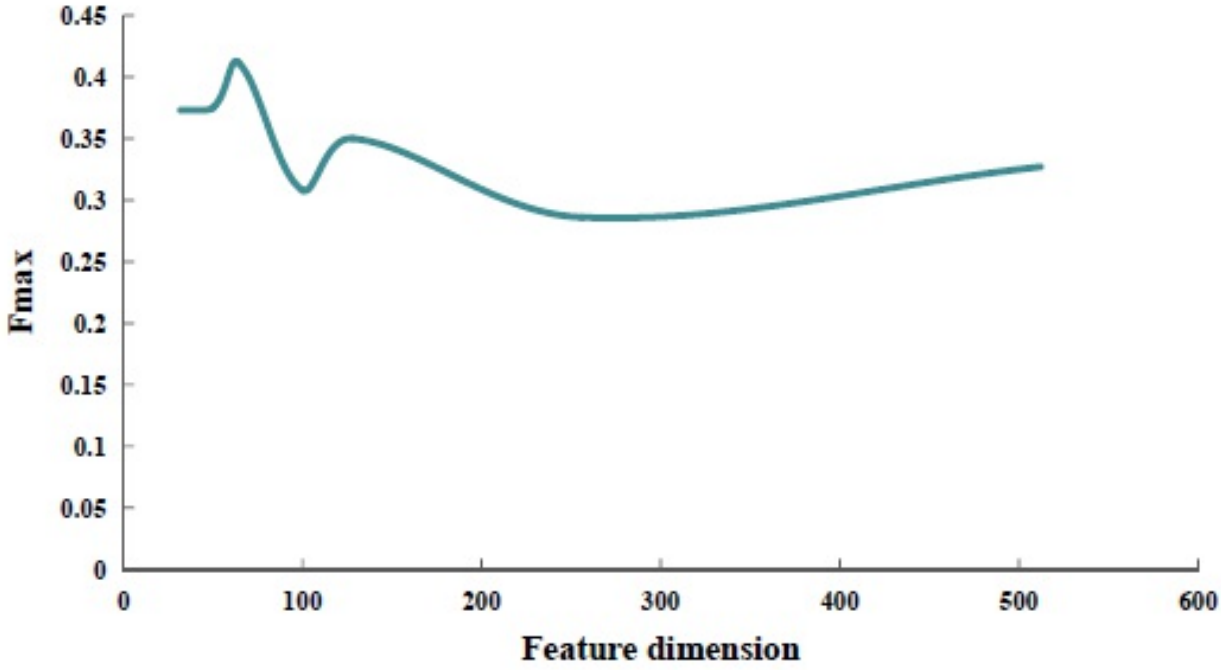


Figure 1

The Fmax values when using different dimensions.

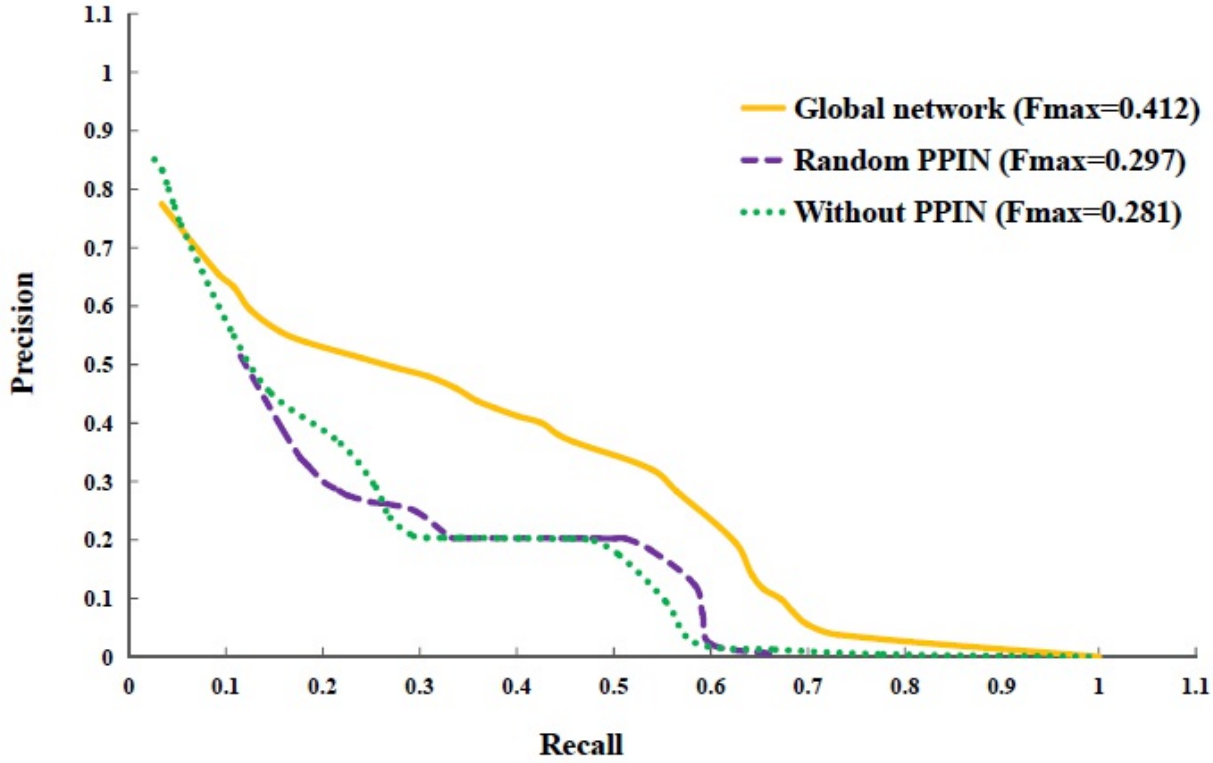


Figure 2

The precision-recall curves of circRNA2GO-62 biological process prediction on different networks (global network, without PPI network and random PPI network).

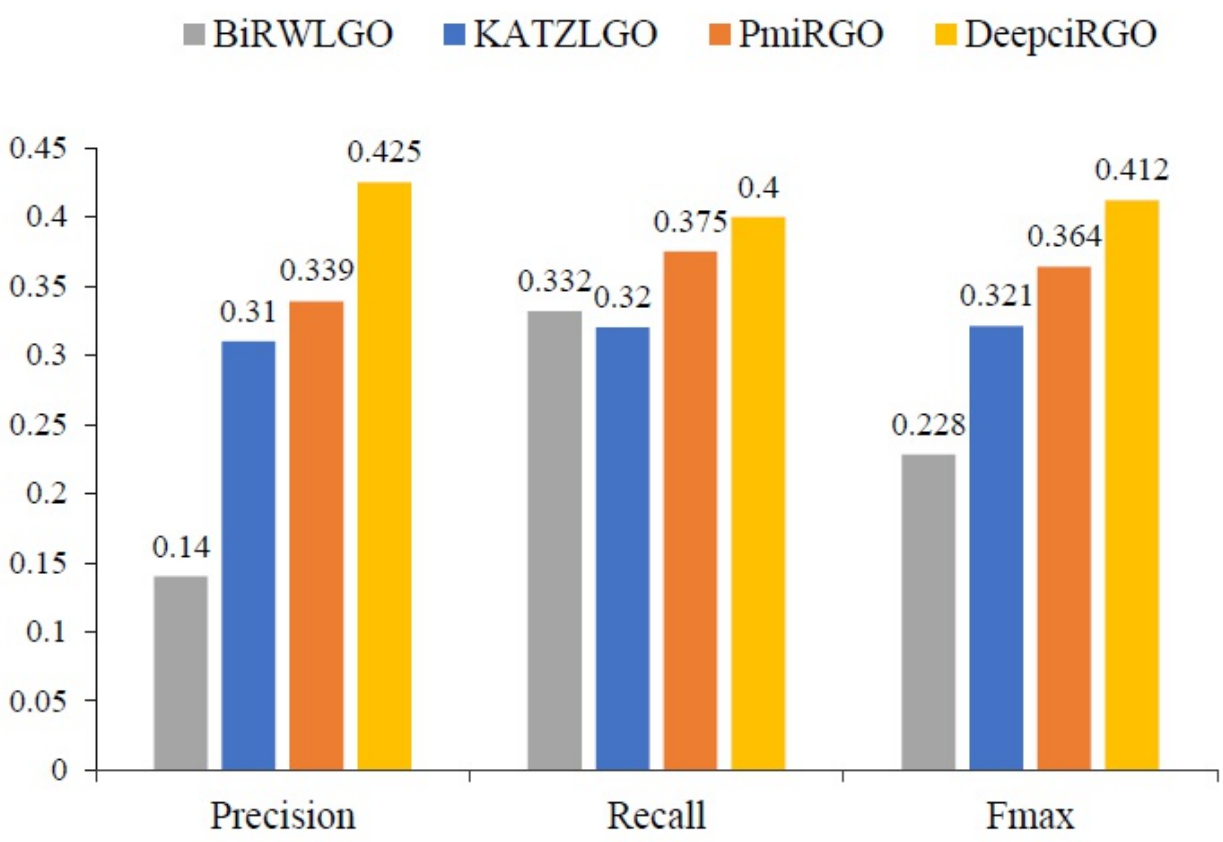


Figure 3

Comparison of the performance of Deep-ciRGO and other existing methods in BP terms of recall, precision and Fmax.

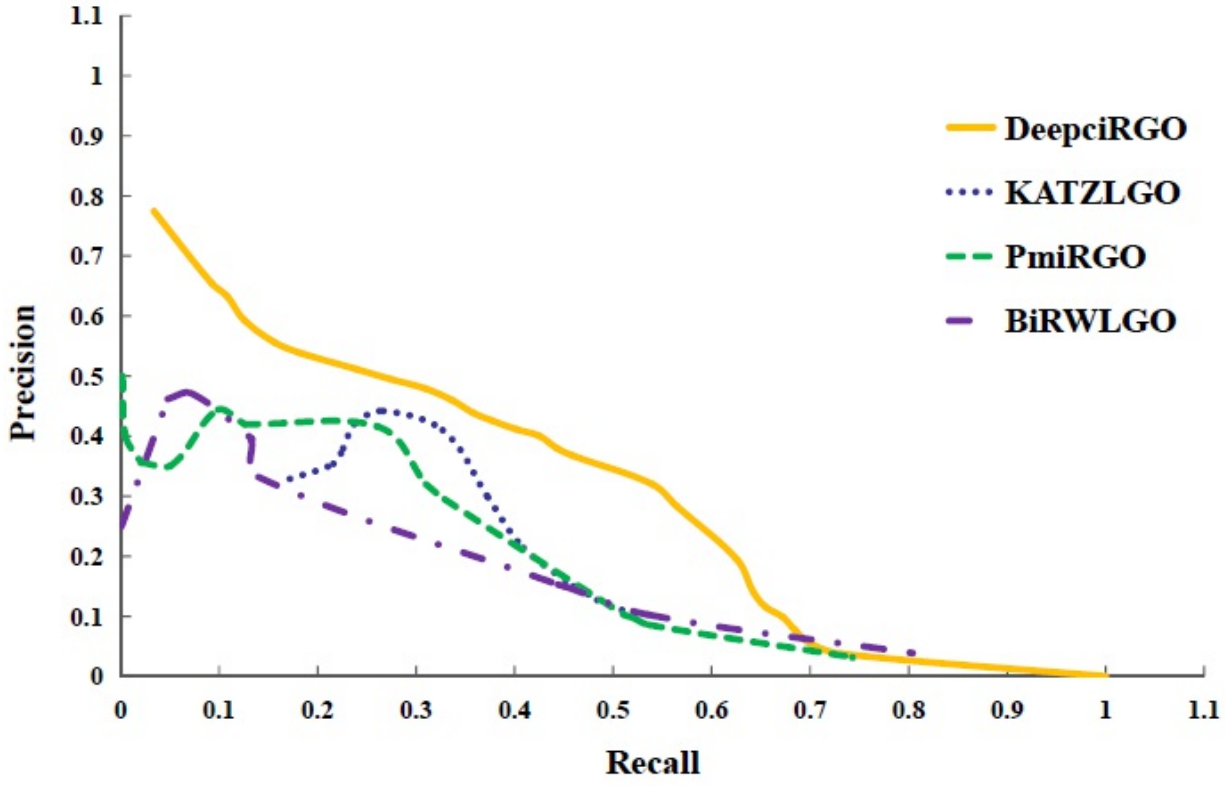


Figure 4

Precision-recall curves of DeepciRGO and other methods (BiRWLGO, PmiRGO and KAT-ZLGO) on the circRNA2GO-62 dataset for BP terms.

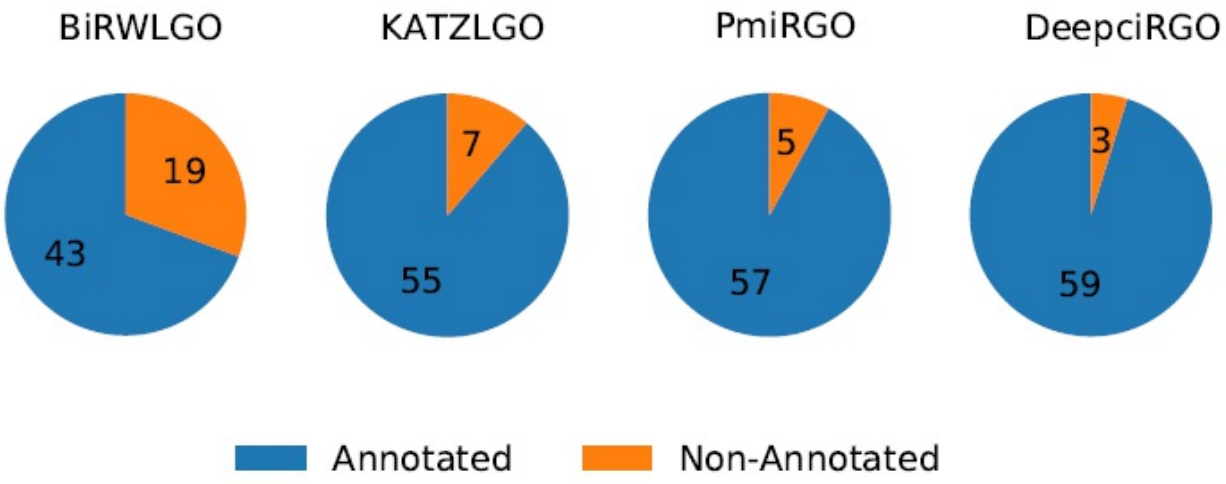
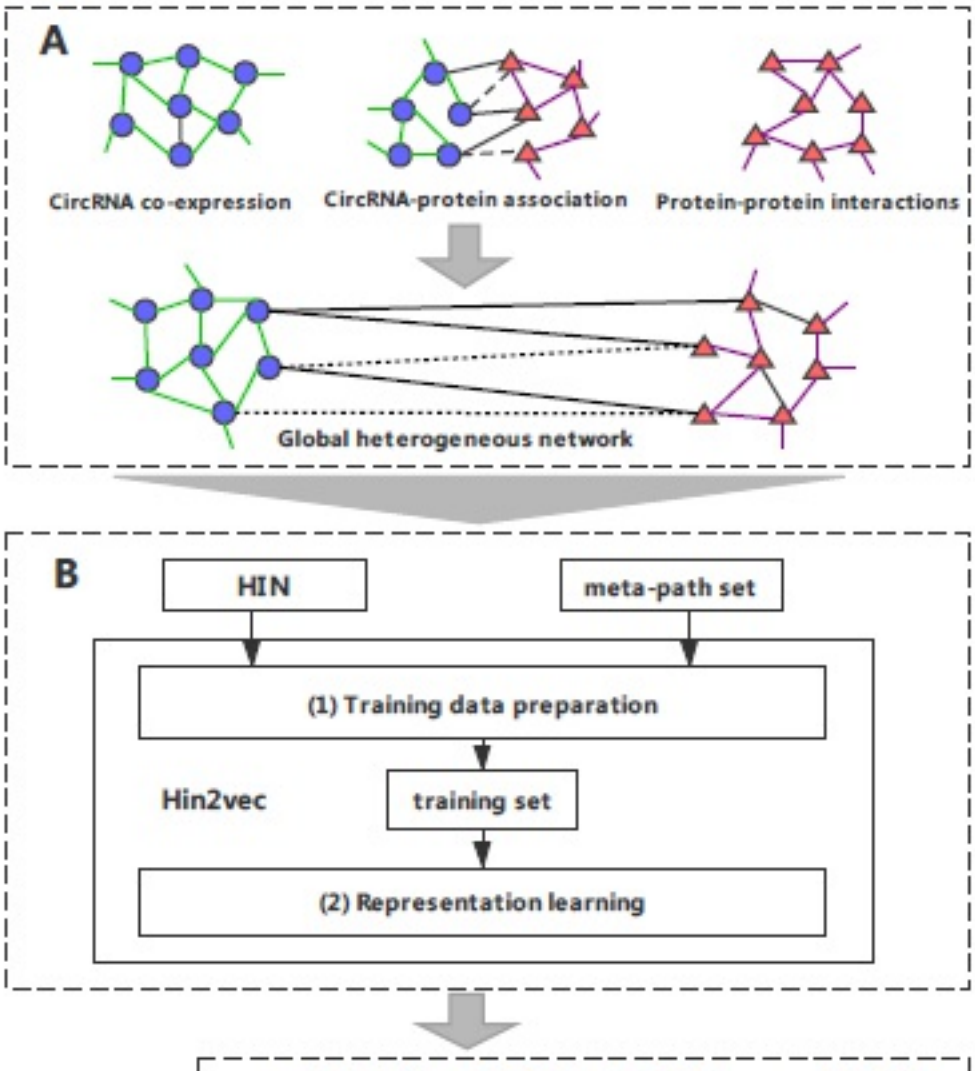


Figure 5

Performance comparison of coverage on the independent dataset circRNA2GO-62 by DeepciRGO and the other three methods (PmiRGO, KATZLGO and BiRWLGO).



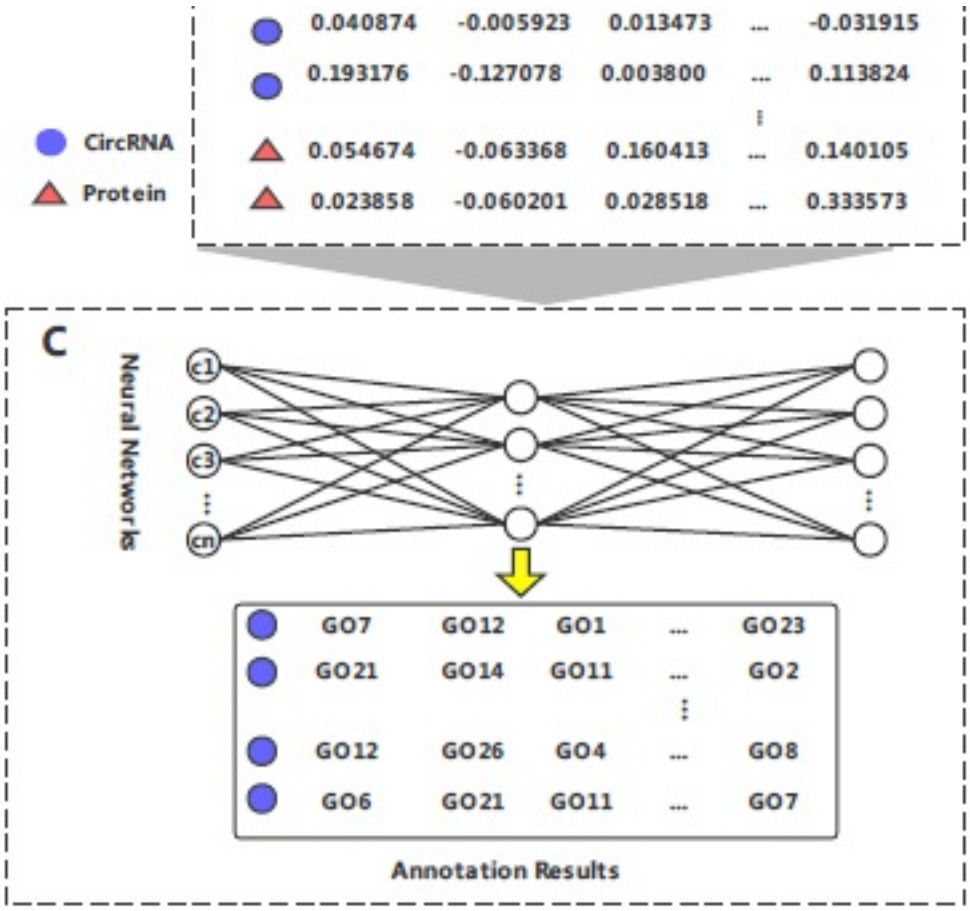


Figure 6

Flowchart of DeepciRGO, which includes three steps: A) build the global heterogeneous network by integrating three networks (circRNA co-expression network, circRNA-protein interaction network, and PPI network); B) employ HIN2vec to learn the latent representations of the nodes (circRNAs and proteins); C) train each GO class with the neural network model and annotate circRNAs

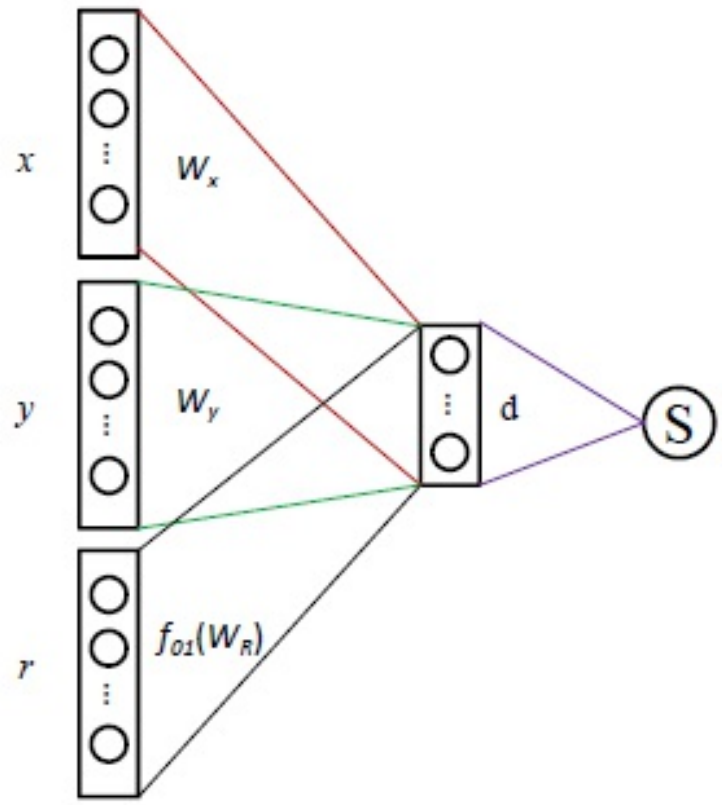


Figure 7

The HIN2vec NN model

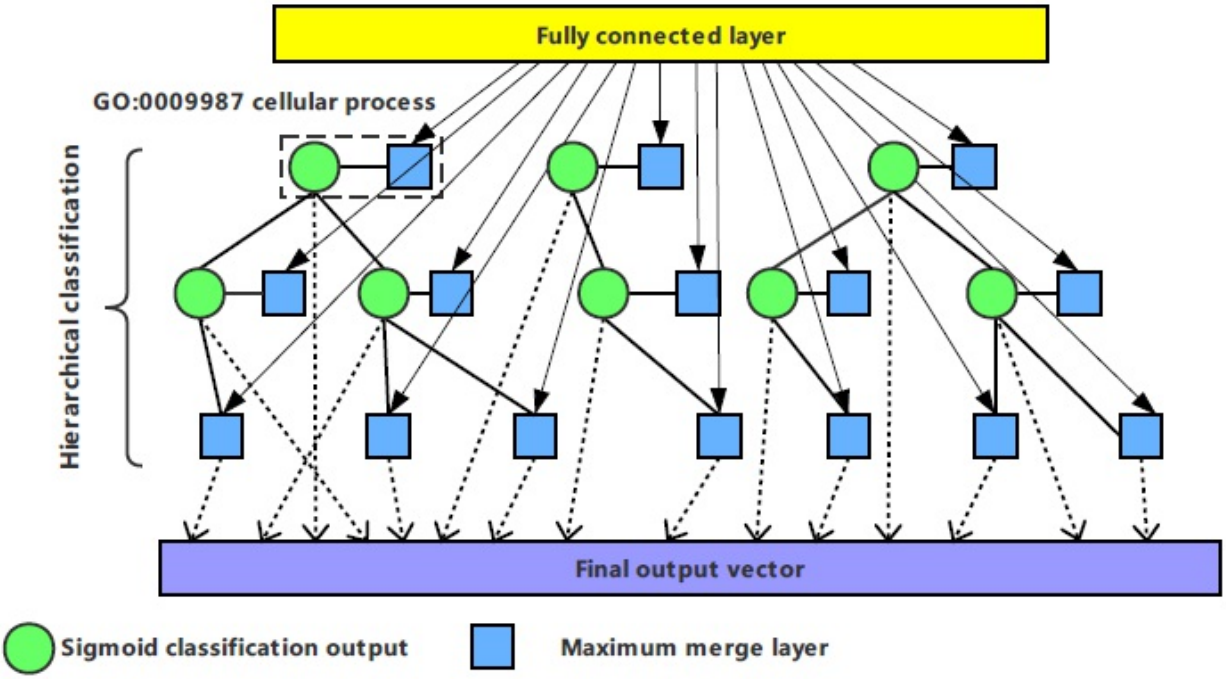


Figure 8

The hierarchical architecture of classica- tion in the neural network model.