**An Ensemble approach for Circular RNA-Disease Association prediction using Variational Autoencoder and Genetic Algorithm**

Salooja C.M.a,\*, Arjun Sankera, Deepthi Kb, Jereesh A.S.a,\*

a *Bioinformatics Lab, Department of Computer Science, Cochin University of Science and Technology, Kerala-682022, India*

b *Department of Computer Science,*[*Central University of Kerala*](https://scholar.google.com/citations?view_op=view_org&hl=en&org=803784772741901364)*(Central Govt. of India), Kerala-671316, India*

\* Corresponding author*.*

*E-mail addresses*: saloojacm@pg.cusat.ac.in (Salooja C.M.), jereesh@cusat.ac.in (Jereesh A.S.)

**Supplementary Data**

**Table 1** Study of different models on CircR2Disease dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sl no.** | **Paper** | **Methodology** | **AUC** | **General Dataset** | **Observations** |
| 1. | Integrating random walk with restart and k-Nearest Neighbour to identify novel circRNA-disease association [16] | Random walk with restart  K-Nearest Neighbour | 0.9297 | CircR2Disease | * Disease semantic similarity, circRNA function similarity and GIP kernel similarity * Simple to implement. * It is not applicable to new circRNAs or diseases without any known association |
| 2. | Predicting novel CircRNA-disease associations based on random walk and logistic regression model [17] | Random walk and logistic regression | 0.9600 | CircR2Disease | * CircRNA function similarity * Simple to implement and fast running * It does not include enough information of diseases. It is not applicable to new circRNAs or diseases without any known association. |
| 3. | GCNCDA: A new method for predicting circRNA-disease associations based on Graph Convolutional Network Algorithm [18] | Graph Convolutional Network | 0.9090 | CircR2Disease | * GIP kernel similarity and disease semantic similarity * It effectively integrates network characteristics * GIP kernel similarity relies heavily on known circRNA-disease associations, thus causing inevitable bias. |
| 4. | Inferring Potential CircRNA–Disease Associations via Deep Autoencoder‑Based Classification [20] | Deep Auto encoder,  Random forest | 0.9486 | CircR2Disease | * Integrated circRNA, disease similarities, and experimentally verifed circRNA–disease associations and extracted the hidden characteristics from them with the deep autoencoder. * Powerful computational tool for identifying novel circRNA–disease associations * The real negative samples were hard or even impossible to acquire. * Experimentally confrmed circRNA–disease relationships used in this study are far from enough. |
| 5. | SGANRDA: semi-supervised generative adversarial networks for predicting circRNA–disease associations [21] | semi-supervised generative adversarial networks | 0.9411 | CircR2Disease | * Integrates a variety of biological information. * Can identify new disease-associated circRNAs effectively. * Features extracted by natural language processing may not be obvious * Not easy to interpret the initialization and optimization process of hyperparameters in GAN network. * the randomly selected samples will bring some errors to the prediction results. |
| 6. | GATNNCDA: A Method Based on Graph Attention Network and Multi-Layer Neural Network for Predicting circRNA-Disease Associations [22] | Graph Attention Network.  Multi-Layer Neural Network | 0.9613 | CircR2Disease | * Integrates circRNA functional similarity, disease semantic similarity and the GIP similarities. * linear transformation to project the integrated similarity matrices into the same space. * Applies a graph attention network to extract dense representations of nodes in the heterogeneous circRNA–disease graph. * The initial node features may not be perfect. * Known interactions between circRNA–disease associations are insufficient. |
| 7. | CircWalk: a novel approach to predict CircRNA‑disease association based on heterogeneous network representation learning [24] | Deep Walk algorithm | 0.9777 | CircR2Disease  Mir2Disease | * Integrate multiple resources to construct a heterogeneous network from circRNAs, mRNAs, miRNAs, and diseases. * Each dataset uses unique identifiers and converting these identifiers sometimes can be impossible. * The lack of validated negative classes |