



A Joint Model of Random Forest and Artificial Neural Network for the Diagnosis of Endometriosis

Jiajie She^{1,2}, Danna Su¹, Ruiying Diao^{1*} and Liping Wang^{1*}

¹Reproductive Medicine Centre, Shenzhen Second People's Hospital, The First Affiliated Hospital of Shenzhen University, Shenzhen, China, ²Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

OPEN ACCESS

Edited by:

Tapas Si,
Bankura Unnayani Institute of
Engineering, Bankura, India

Reviewed by:

Soumadip Ghosh,
Institute of Engineering and
Management (IEM), India
Soumita Seth,
Aliah University, India

*Correspondence:

Liping Wang
wlp18665070696@163.com
Ruiying Diao
15889753127@163.com

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: 04 January 2022

Accepted: 28 January 2022

Published: 08 March 2022

Citation:

She J, Su D, Diao R and Wang L (2022)
A Joint Model of Random Forest and
Artificial Neural Network for the
Diagnosis of Endometriosis.
Front. Genet. 13:848116.
doi: 10.3389/fgene.2022.848116

Endometriosis (EM), an estrogen-dependent inflammatory disease with unknown etiology, affects thousands of childbearing-age couples, and its early diagnosis is still very difficult. With the rapid development of sequencing technology in recent years, the accumulation of many sequencing data makes it possible to screen important diagnostic biomarkers from some EM-related genes. In this study, we utilized public datasets in the Gene Expression Omnibus (GEO) and Array-Express database and identified seven important differentially expressed genes (DEGs) (*COMT*, *NAA16*, *CCDC22*, *EIF3E*, *AHI1*, *DMXL2*, and *CISD3*) through the random forest classifier. Among these DEGs, *AHI1*, *DMXL2*, and *CISD3* have never been reported to be associated with the pathogenesis of EMs. Our study indicated that these three genes might participate in the pathogenesis of EMs through oxidative stress, epithelial–mesenchymal transition (EMT) with the activation of the Notch signaling pathway, and mitochondrial homeostasis, respectively. Then, we put these seven DEGs into an artificial neural network to construct a novel diagnostic model for EMs and verified its diagnostic efficacy in two public datasets. Furthermore, these seven DEGs were included in 15 hub genes identified from the constructed protein–protein interaction (PPI) network, which confirmed the reliability of the diagnostic model. We hope the diagnostic model can provide novel sights into the understanding of the pathogenesis of EMs and contribute to the clinical diagnosis and treatment of EMs.

Keywords: endometriosis, random forest, artificial neural network, diagnostic model, diagnostic efficacy

INTRODUCTION

Endometriosis (EM) is an estrogen-dependent inflammatory disorder, which afflicts about 10%–15% of women of childbearing age (Parasar et al., 2017). It is defined as the presence of endometrial-like tissue outside of the uterine cavity, which can lead to chronic pelvic pain, and infertility (Drabble et al., 2021). However, the true prevalence of EMs is uncertain as visual laparoscopy is the gold standard for the diagnosis of EMs (Taylor et al., 2018). At the moment, Sampson's theory of menstrual blood reflux observed in most patients is commonly accepted in the pathophysiology of EMs, while only a small portion will develop into this disease (Burney and Giudice, 2012). However, it could only explain a portion of EMs. Therefore, it's necessary to further investigate a comprehensive understanding of the pathogenesis of EMs and find effective molecular biomarkers to improve the early diagnosis and treatment of EMs.

DNA microarray technology is a high-throughput detection method that can be used to provide gene expression profiles and thus can help to screen disease-related genes and biomarkers (Yoo et al.,

2009). With the rapid development of DNA microarray technology, a large amount of high-throughput data has accumulated available on public platforms. However, how to make effective use of these data to screen critical disease-related genes for the diagnosis of EMs is a great challenge. At present, random forest and neural network are widely applied for disease prediction (Yigit and Isik, 2018; Khan et al., 2019; Shaia et al., 2019; Kugunavar and Prabhakar, 2021). Among them, random forest algorithm can perform random sampling to screen the target variables (Schonlau and Zou, 2020) and has high predicted accuracy (Byeon, 2019; Chen et al., 2020). Furthermore, the artificial neural network can be used to evaluate the accuracy of predicted model with divided training and validation datasets (Curchoe et al., 2020). Currently, there are some useful visualization and analysis tools for neural networks, such as NeuralNetTools (Beck, 2018), spiking neuronal networks (Galindo et al., 2020), and Net2Vis (Bauerle et al., 2021).

Therefore, the combination of random forest and artificial neural network would have better classification performance and more meaningful selected features (Kong and Yu, 2018; Tian et al., 2020). In this study, we firstly identified some differentially expressed genes (DEGs) between EMs and normal samples from public datasets in the Gene Expression Omnibus (GEO) database. Through the random forest classifier, we screened these DEGs and obtained seven important DEGs (*COMT*, *NAA16*, *CCDC22*, *EIF3E*, *AH11*, *DMXL2*, and *CISD3*). Then, we put these seven DEGs into an artificial neural network to construct a novel diagnostic model and verified its diagnostic efficacy in two public datasets (See the detailed process in **Figure 1**). We hope this diagnostic model can provide novel sights into the pathogenesis of EMs and improve the early diagnosis and treatment of EMs.

MATERIALS AND METHODS

Data Download and Processing

The GSE51981, GSE6364, and GSE7307 datasets were downloaded by the R package “GEOquery” (2.60.0) (Davis and Meltzer, 2007) to obtain the expression profile data. Then, the E-MTAB-694 dataset was downloaded through the Array-Express database. The related annotation information including the platforms, the probes, and ID conversion was obtained from the GEO database. When multiple probes corresponded to one gene symbol, the average expression level of multiple probes was used as the expression level of the corresponding gene. ID conversion was conducted with the R package “org.Hs.eg.db” (v3.13.0). Furthermore, the “removeBatchEffect” function in the R package “LIMMA” (v3.48.3) (Ritchie et al., 2015) was used to adjust batch effects, which were evaluated by principal component analysis (PCA).

Differential Expression and Functional Enrichment Analysis

Differential expression analysis was conducted on 77 EM disease and 71 normal samples of the GSE51981 dataset

through the Bayesian analysis of the R package “LIMMA”. The $\log_2\text{FoldChange} > 1.5$ and $p\text{-value} < 0.05$ were set as the threshold of DEGs. The R package “pheatmap” (v1.0.12) was used to perform clustering analysis of DEGs for the heatmap. To explore the biological significance of these DEGs in the pathogenesis of EMs, GO and KEGG pathway enrichment analyses were performed through the R package “clusterProfiler” (v4.1.3) (Wu et al., 2021) to identify significantly enriched GO terms and significantly enriched KEGG pathways with the threshold of $p\text{-value} < 0.05$.

The Construction of Hub Gene Network

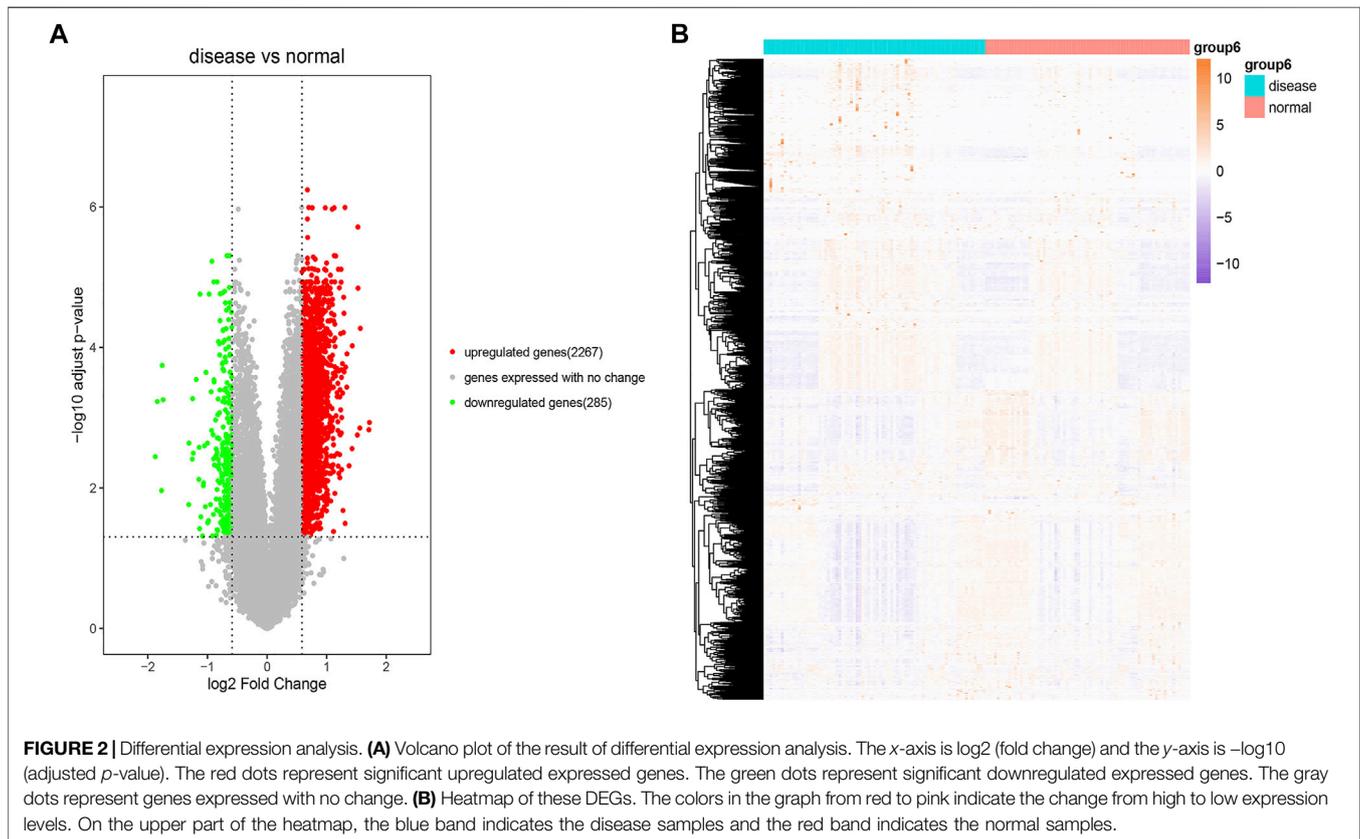
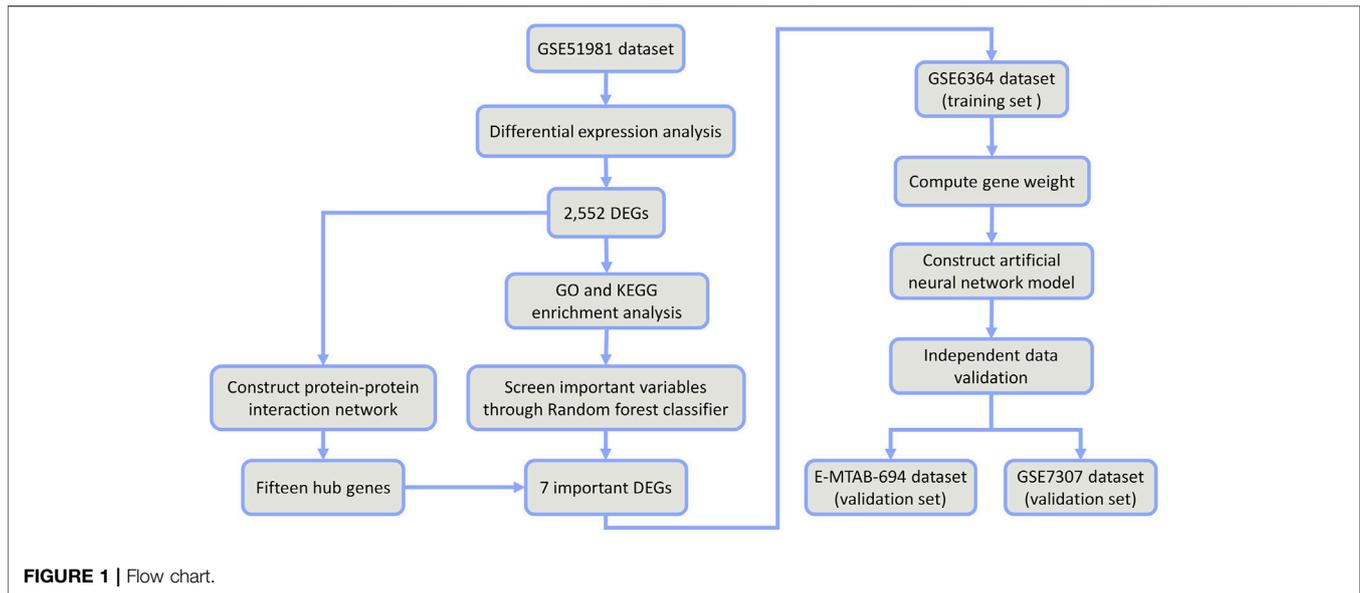
The STRING (v11.5) (<https://string-db.org/cgi/input.pl>) (Szklarczyk et al., 2021) has been widely applied to construct a protein–protein interaction (PPI) network. Based on those DEGs, the “Multiple proteins” option was selected. In the PPI network, the minimum required interaction score was set as “high confidence (0.700)”. Then, the cytoHubba (Chin et al., 2014) was employed to identify hub genes. The eccentricity algorithm was selected and 15 top-ranked genes were chosen as hub genes. Finally, the hub gene network was visualized with Cytoscape (v3.9.0) (Demchak et al., 2014).

Screening Differentially Expressed Genes With the Random Forest Model

The R package “randomForest” (v4.6.14) (Liaw et al., 2014) was used to construct a random forest model to screen DEGs. The number of random seeds and decision trees was set as 1–5,000 and 3,000 in the random forest classifier originally, respectively. Finally, the number of random seeds and decision trees was set as 4,543 and 219, respectively, which represented higher accuracy of the constructed model and stable model error. The Gini coefficient method was used to obtain the dimensional importance value of all variables from the constructed random forest model. Those DEGs with an importance value greater than 4 were screened as important genes of EMs for subsequent model construction and verification. The R package “pheatmap” was used to perform clustering analysis of the screened important genes for the heatmap in this dataset.

The Construction and Verification of the Artificial Neural Network Model

The GSE6364 dataset downloaded through the R package “GEOquery” was selected as the training set for the construction of the artificial neural network model. After the data normalization, the R package “neuralnet” (v1.44.2) (Fritsch and Guenther, 2016) was used to construct an artificial neural network model of those important variables. The number of hidden neuron layers should be two-thirds of the number of the input layer plus two-thirds of the number of the output layer. Therefore, six hidden layers were set as the model parameter to construct a classification model of EMs through the predicted gene weight information. The R packages “pROC” (v1.18.0) (Robin et al., 2011) and



“ggplot2” (v3.3.5) (Gómez-Rubio, 2017) were used to calculate the verification results of AUC classification performance and draw the ROC curve. Another two datasets E-MTAB-694 and GSE7307 were used to verify the accuracy of the constructed neural network model for the

diagnosis of EMs. The R package “pROC” was used to draw ROC curves for each dataset, and the AUC value was calculated to verify the classification efficiency. Meanwhile, the sensitivity and specificity in distinguishing the disease samples from normal samples were calculated.

RESULTS

Data Processing and Differential Expression Analysis

The R package “GEOquery” was used to download the GEO dataset GSE51981 (77 EM disease samples and 71 normal samples) and obtain detailed information. We used the “removeBatchEffect” function in the R package “LIMMA” to adjust batch effects and then conducted principal component analysis (PCA) analysis to evaluate the performance of batch effect adjustment. PCA results (**Supplementary Figure S1**) indicated that the disease samples were mixed with the normal samples, which suggested the challenge of diagnosing. We also used the R package “LIMMA” to perform differential expression analysis for the dataset GSE51981 through the Bayesian test. We finally identified 2,267 significantly upregulated and 285 significantly downregulated expressed genes between the disease samples and the normal samples with the threshold of fold change values of >1.5 and $p < 0.05$. The detailed information of all DEGs is listed in **Supplementary Table S1**. The results of these DEGs and the heatmap of these DEGs are visualized in **Figures 2A** and **2B**, respectively.

Functional Enrichment Analysis for DEGs and the Construction of PPI Network

To explore the biological significance of these DEGs in the pathogenesis of EMs, we performed GO and KEGG pathway enrichment analyses through the R package ‘clusterProfiler’. GO terms were classified into three categories: biological process (BP), cellular component (CC), and molecular function (MF). The top five GO terms of genes with significantly upregulated and downregulated expression levels were visualized in **Figures 3A,B**. The GO enrichment analysis results indicated that these significantly upregulated expressed genes were mainly involved in the transmembrane transporter activity, ATPase activity, metalloproteinase activity, aldehyde dehydrogenase NADP⁺ activity, and lipid transporter activity (**Supplementary Table S2**), while these significantly downregulated expressed genes were mainly involved in the flavin adenine dinucleotide binding, acyl-CoA dehydrogenase activity, phosphatidylcholine transporter activity, extracellular matrix structural constituent, and ATPase-coupled intramembrane lipid transporter activity (**Supplementary Table S3**). For KEGG pathway enrichment analysis, the results indicated that these upregulated expressed genes were significantly associated with the cAMP signaling pathway, adrenergic signaling in cardiomyocytes, aldosterone synthesis and secretion, ABC transporters, and salivary secretion (**Supplementary Table S4**), while these downregulated expressed genes were significantly associated with fatty acid degradation and metabolism; valine, leucine, and isoleucine degradation; lysosome; the PPAR signaling pathway; and the Hippo signaling pathway (**Supplementary Table S5**). Furthermore, we constructed a PPI network through the STRING database. The hub genes selected from the PPI network are shown in **Supplementary Figure S2**.

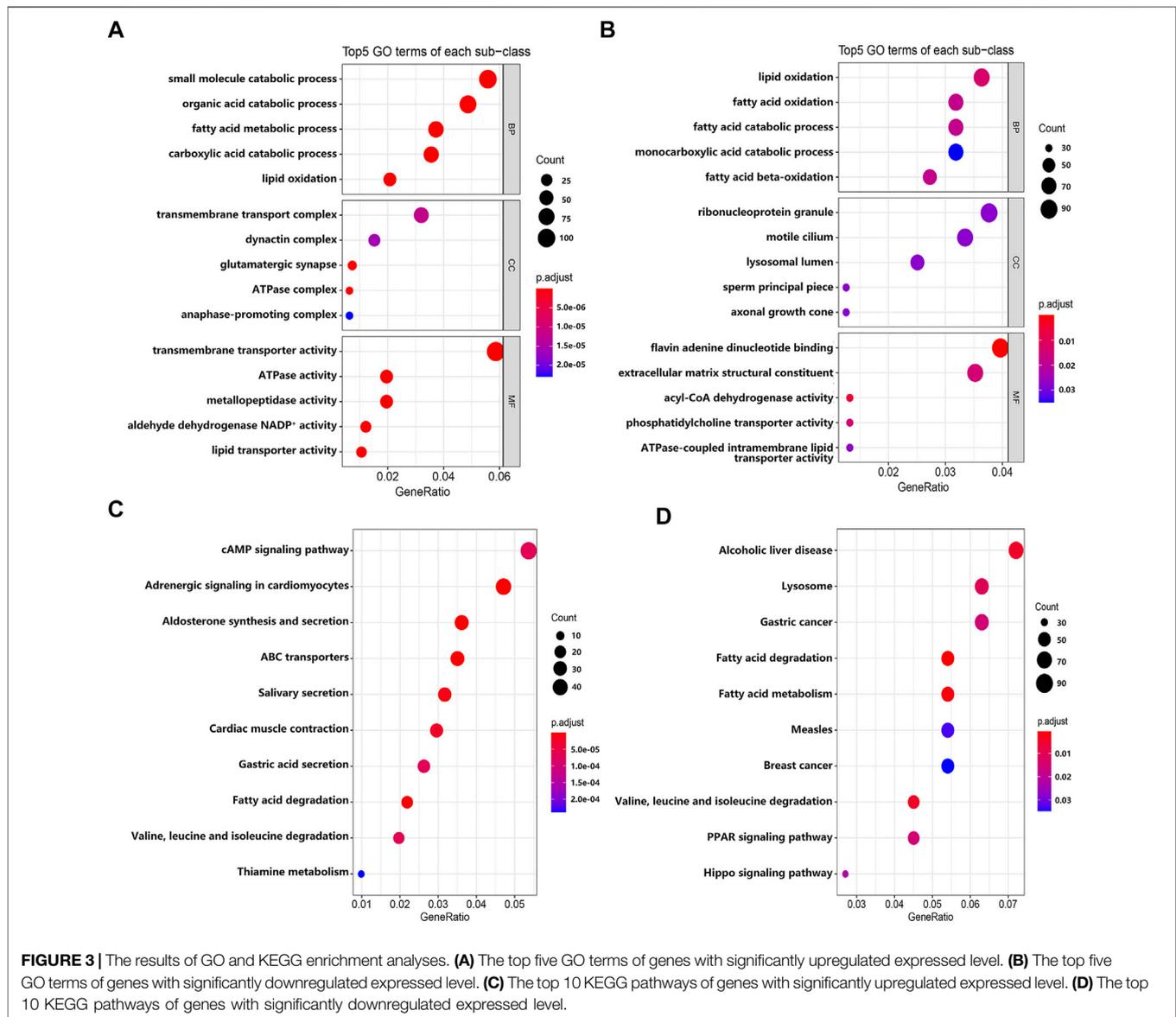
According to the eccentricity scores, we identified 15 hub genes from the network, which had highest confidence scores.

Constructing the Random Forest Model to Screen Differentially Expressed Genes

To screen DEGs, we put these DEGs into the random forest classifier and set the number of random seeds to 4,543. By referring to the relationship between the model error and the number of decision trees (**Figure 4A**), we selected 219 trees as the parameter of the random forest model, which represented a stable error in the model. In the modeling process, we used the Gini coefficient method to measure the importance of all variables according to decreased mean square error and model accuracy (**Figure 4B**). Finally, we selected seven DEGs (*AH11*, *DMXL2*, *NAA16*, *CCDC22*, *CISD3*, *COMT*, and *EIF3E*) with a mean decrease of Gini index greater than 4 as important variables for subsequent analysis. Interestingly, all these DEGs were included in the 15 hub genes identified from the constructed PPI network. Among these variables, *AH11* was the most important, with the mean decrease of the Gini index being much higher than other variables (**Supplementary Table S6**). A small number of variables meant a small out-of-band error, which represented a high accuracy of the constructed random forest model. Based on these seven variables, we performed the *k*-means clustering of the dataset. The results suggested that these seven genes could be used to distinguish the disease sample from the normal samples (**Figure 4C**). Furthermore, *AH11*, *DMXL2*, and *NAA16* genes were clustered as a group with low expression in the normal sample and high expression in the disease sample. On the contrary, *CCDC22*, *CISD3*, *COMT*, and *EIF3E* were clustered as another group with high expression in the normal sample and low expression in the disease sample.

The Construction of the Artificial Neural Network Model and the Evaluation of the ROC Curve

Based on the R package ‘neuralnet’, we use the GSE6364 dataset (21 disease samples and 21 normal samples) as the training set to construct the artificial neural network model. Firstly, we performed the preprocessing and normalization of this dataset. According to the output results of the neural network model (**Figure 5A**), it is illuminated that the entire training was performed in 11,684 steps. Among the output results, the predicted weights of each hidden neuron layer were -3.97906 , 1.04457 , 2.76611 , -2.00181 , -11.84206 , and -0.90829 (**Supplementary Table S7**). Next, we drew the ROC curve to evaluate the predicted performance; the AUC values of *AH11*, *COMT*, *DMXL2*, *CISD3*, *NAA16*, *EIF3E*, and *CCDC22* were 0.7150 , 0.7809 , 0.6927 , 0.7266 , 0.7217 , 0.7093 , and 0.7050 , respectively (**Figure 5B**). The larger the AUC value of each DEG is, the higher the credibility of the constructed diagnostic model will be. We also used another two datasets E-MTAB-694 (18 disease samples and 17 normal samples) and GSE7307 (18 disease samples and 23 normal samples) to verify the accuracy of the constructed neural network model. In the E-MTAB-694

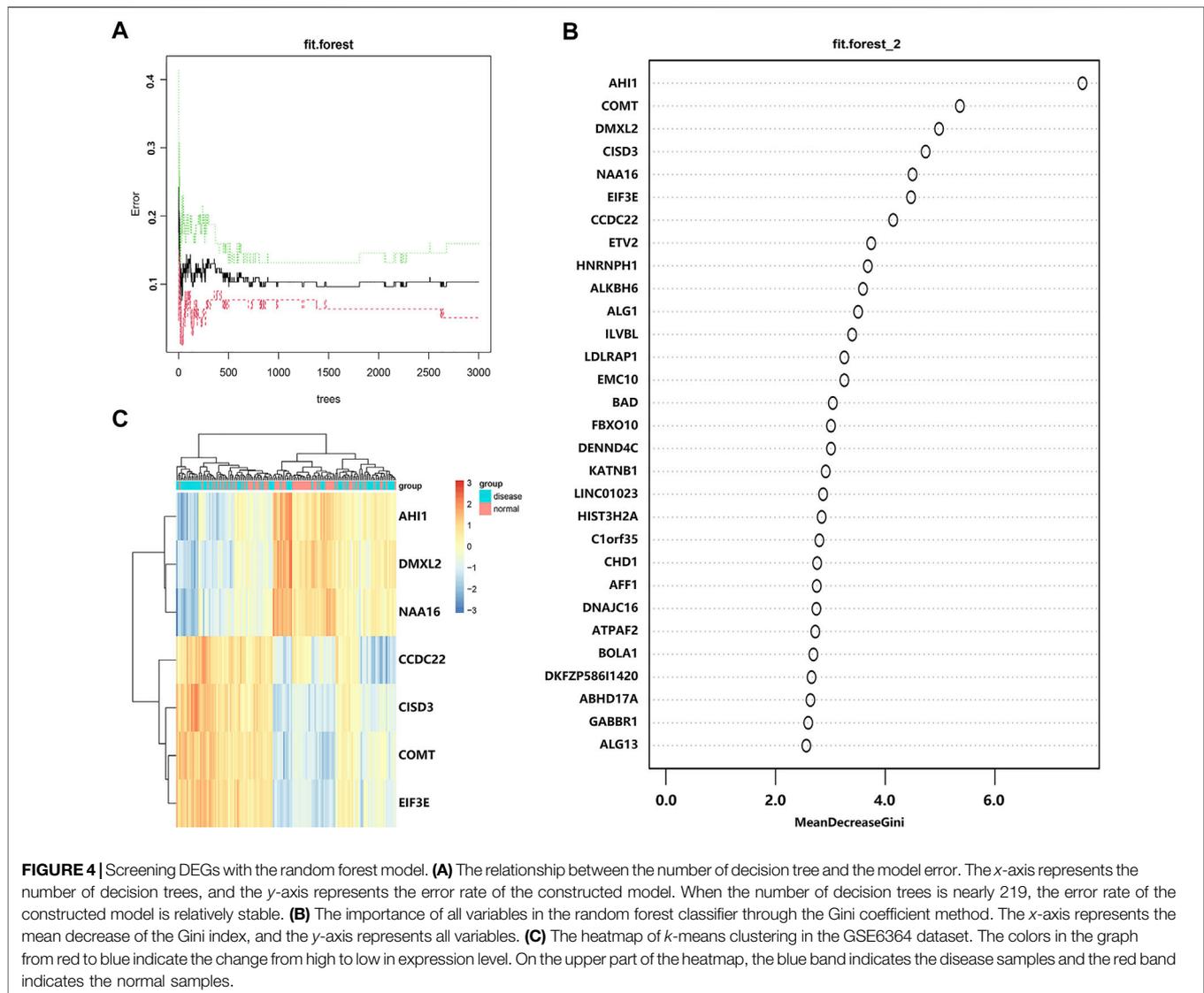


dataset (**Figure 5C**), the AUC values of the seven DEGs were 0.8226, 0.6623, 0.6836, 0.6625, 0.8367, 0.8471, and 0.8617. In the verification results of the GSE7307 dataset (**Figure 5D**), the AUC values of the seven DEGs were 0.7464, 0.6484, 0.7020, 0.6300, 0.9075, 0.8295, and 0.8327. In general, we constructed a novel diagnostic model of EMs and verified its diagnostic efficacy through the constructed artificial neural network in two public datasets.

DISCUSSION

The combination of random forest and artificial neural network can be used to construct a reliable predictive model for the diagnosis of some diseases, such as polycystic ovary syndrome (PCOS) (Xie et al., 2020) and ulcerative colitis

(Li et al., 2020). In this study, we identified 2,552 DEGs associated with EMs in the GSE51981 dataset. Based on the random forest classifier, seven important candidate DEGs (*COMT*, *NAA16*, *CCDC22*, *EIF3E*, *AH11*, *DMXL2*, and *CISD3*) were screened. Then, we used the GSE6364 dataset as the training set to construct the artificial neural network model and evaluated the classification efficacy of the model in E-MTAB-694 and GSE7307 datasets. The AUC values of the ROC curve were about 0.7, which had great efficiency and verified the diagnostic efficacy of the model. Furthermore, we constructed a 15-hub-gene-based PPI network and confirmed the reliability of the prediction model. Compared with the Nnet package, we found that the neuralnet package had higher accuracy of the predicted model (86.5% vs 81.1%). In total, the constructed diagnostic model could provide new insight into our understanding of the pathogenesis of EMs and identify



crucial biomarkers as diagnostic and therapeutic targets of EMs.

Among these seven genes, *COMT*, *NAA16*, *CCDC22*, and *EIF3E* have been reported to be associated with the pathogenesis of EMs. Catechol-*O*-methyltransferase (*COMT*) is highly expressed in the placental, adrenal gland, ovary, and other tissues. The degradative pathways of the catecholamine transmitters can relieve painful uterine contractions (D'Astous-Gauthier et al., 2021). *COMT* polymorphism may contribute to the risk of EMs and adenomyosis (Li et al., 2018) and has a relationship with EM susceptibility (Ji et al., 2017; Zhai et al., 2019). *N*-alpha-acetyltransferase 16 (*NAA16*) is highly enriched in bone marrow, testis, endometrium, and other tissues. It can alter NAT 2 enzyme activity and thus contribute to the susceptibility of EMs (Nakago et al., 2001). Coiled-coil domain containing 22 (*CCDC22*), a membrane-binding protein, is highly enriched in the spleen, lymph node, and

other tissues. Studies have demonstrated that there is also a relationship between *CCDC22* polymorphisms and EM susceptibility (de Oliveira Francisco et al., 2017). Eukaryotic translation initiation factor 3 subunit E (*EIF3E*) is highly expressed in the ovary, lymph node, endometrium, and other tissues. Its downregulation may be involved in epithelial-mesenchymal transition (EMT) in EMs, possibly through the preferential translation of snail (an inhibitor of E-cadherin) (Cai et al., 2018) and involved in the development of adenomyosis through activating the TGF- β 1 signaling pathway (Cai et al., 2019).

Interestingly, we identified another three important genes (*AHI1*, *DMXL2*, and *CISD3*), which have never been reported to be involved in the pathogenesis of EMs. Abelson helper integration site 1 (*AHI1*) is highly enriched in testis, adrenal gland, brain, prostate, endometrium, and other tissues, which has upregulated expression level in EMs. The *AHI1* protein participates in reactive oxygen species (ROS) production in the

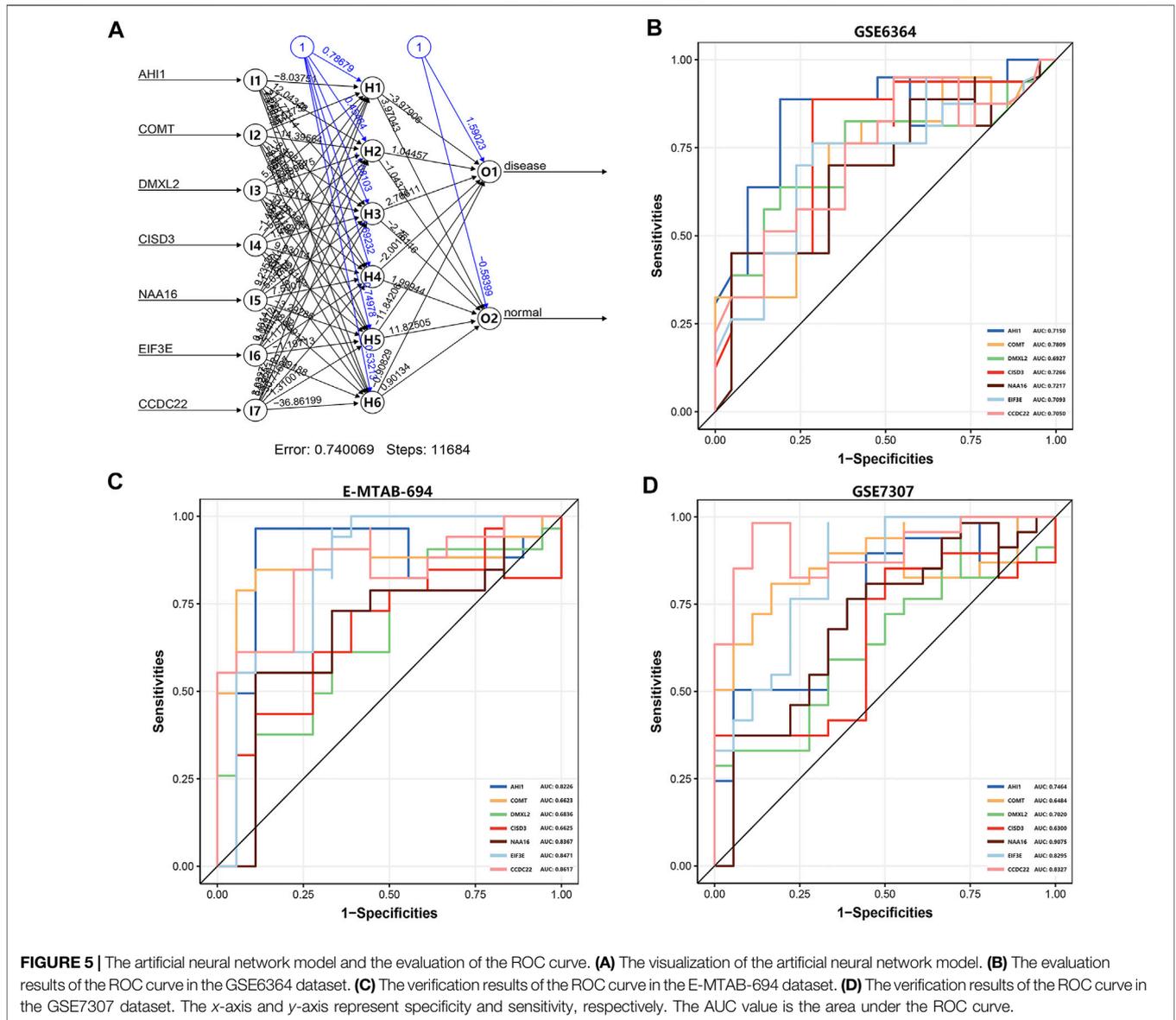


FIGURE 5 | The artificial neural network model and the evaluation of the ROC curve. **(A)** The visualization of the artificial neural network model. **(B)** The evaluation results of the ROC curve in the GSE6364 dataset. **(C)** The verification results of the ROC curve in the E-MTAB-694 dataset. **(D)** The verification results of the ROC curve in the GSE7307 dataset. The x-axis and y-axis represent specificity and sensitivity, respectively. The AUC value is the area under the ROC curve.

form of protein complexes (Liu et al., 2017). Excessive production of ROS can result in oxidative stress (OS) and overall immune activation and inflammation (Newsholme et al., 2016). OS represents an imbalance between ROS and antioxidants, which may have an essential role in the endometriosis pathogenesis in the peritoneal cavity (Samimi et al., 2019). Hence, the *AH11* protein may participate in the EMs pathogenesis through multiple processes such as OS and immune and inflammatory response.

Dmx like 2 (*DMXL2*) encodes a protein with 12 WD domains, which has relatively low expression in endometrium tissue and downregulated expression in EMs. The *DMXL2* protein is demonstrated to participate in the regulation of the Notch signaling pathway (Sethi et al., 2010) and acts as a transmembrane protein, which can

promote EMT through hyperactivation of the Notch signaling pathway (Faronato et al., 2015). Interestingly, decreased Notch signaling can contribute to impaired decidualization through the downregulation of FOXO1 (a downstream target of Notch signaling) and thus lead to the pathogenesis of EMs (Su et al., 2015). Furthermore, studies indicate that a circRNA with downregulated expression can regulate EMT in EMs via the Notch signaling pathway (Zhang et al., 2019). Therefore, the downregulated expression of *DMXL2* may activate the Notch signaling pathway, contribute to EMT through the interaction with circRNA, and thus lead to the pathogenesis of EMs.

CDGSH iron sulfur domain 3 (*CISD3*) is a member of the CDGSH domain-containing family, whose expression is upregulated in EMs. The *CISD3* protein is redox active and

is thought to play an important role in mitochondrial homeostasis (Geldenhuys et al., 2019). Studies indicate that mitochondrial homeostasis can be considered as the therapeutic target for the treatment of EMs via limiting ESC migration and promoting apoptosis (Suliman and Piantadosi, 2016; Zhao et al., 2018). Furthermore, excessive mitochondrial fission can initiate caspase 9-related mitochondrial apoptosis and thus lead to cell death (Fuhrmann and Brüne, 2017; Zhou et al., 2017). Therefore, upregulated expression of *CISD3* may affect mitochondrial homeostasis and thus play an important role in the pathogenesis of EMs.

In this study, based on random forest and artificial neural network algorithm, we established a novel reliable diagnostic model and screened out three important DEGs that have never been reported to be involved in the pathogenesis of EMs. We aimed at the supplement of existing methods and provided an alternative marker panel for further research in the early screening of EMs. However, there are some limitations for this study. Firstly, all samples are only classified as EM (disease) and non-EM (normal) groups, which may affect the final screening results of DEGs. Secondly, the diagnostic model is only verified in two public datasets, which need more samples for verification. Thirdly, we conduct data analysis only at the mRNA level in the tissue samples of EMs, which require further validation at the mRNA and protein levels. In general, our approach has a certain clinical value, which can be beneficial for the early screening of EMs.

REFERENCES

- Bauerle, A., Van Onzenoort, C., and Ropinski, T. (2021). Net2Vis - A Visual Grammar for Automatically Generating Publication-Tailored CNN Architecture Visualizations. *IEEE Trans. Vis. Comput. Graph.* 27, 2980–2991. doi:10.1109/TVCG.2021.3057483
- Beck, M. W. (2018). NeuralNetTools: Visualization and Analysis Tools for Neural Networks. *J. Stat. Soft.* 85 (11), 1–20. doi:10.18637/jss.v085.i11
- Burney, R. O., and Giudice, L. C. (2012). Pathogenesis and Pathophysiology of Endometriosis. *Fertil. Sterility* 98, 511–519. doi:10.1016/j.fertnstert.2012.06.029
- Byeon, H. (2019). Developing a Random forest Classifier for Predicting the Depression and Managing the Health of Caregivers Supporting Patients with Alzheimer's Disease. *Technol. Health Care* 27, 531–544. doi:10.3233/THC-191738
- Cai, X., Shen, M., Liu, X., and Guo, S.-W. (2018). Reduced Expression of Eukaryotic Translation Initiation Factor 3 Subunit e and Its Possible Involvement in the Epithelial-Mesenchymal Transition in Endometriosis. *Reprod. Sci.* 25, 102–109. doi:10.1177/1933719117702248
- Cai, X., Shen, M., Liu, X., and Nie, J. (2019). The Possible Role of Eukaryotic Translation Initiation Factor 3 Subunit e (eIF3e) in the Epithelial-Mesenchymal Transition in Adenomyosis. *Reprod. Sci.* 26, 377–385. doi:10.1177/1933719118773490
- Chen, J., Li, Q., Wang, H., and Deng, M. (2020). A Machine Learning Ensemble Approach Based on Random forest and Radial Basis Function Neural Network for Risk Evaluation of Regional Flood Disaster: A Case Study of the Yangtze River delta, China. *Int. J. Environ. Res. Publ. Health* 17, 49. doi:10.3390/ijerph17010049
- Chin, C.-H., Chen, S.-H., Wu, H.-H., Ho, C.-W., Ko, M.-T., and Lin, C.-Y. (2014). cytoHubba: Identifying Hub Objects and Sub-networks from Complex Interactome. *BMC Syst. Biol.* 8 Suppl 4(Suppl 4), S11. doi:10.1186/1752-0509-8-S4-S11

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

JS performed data preprocessing and data analysis and wrote the first draft. DS gave advice on the data analysis. RD and LW gave constructive advice for the whole project.

FUNDING

This work was supported by grants from the Science and Technology Innovation Committee of Shenzhen (JCYJ20190806165007495), the Shenzhen Foundation of Science and Technology (JCYJ20210324103606017), and the Guangdong Basic and Applied Basic Research Foundation (2019A1515011693).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.848116/full#supplementary-material>

- Curchoe, C. L., Flores-Saiffe Farias, A., Mendizabal-Ruiz, G., and Chavez-Badiola, A. (2020). Evaluating Predictive Models in Reproductive Medicine. *Fertil. Sterility* 114, 921–926. doi:10.1016/j.fertnstert.2020.09.159
- D'Astous-Gauthier, K., Graham, F., Paradis, L., Des Roches, A., and Bégin, P. (2021). Beta-2 Agonists May Be Superior to Epinephrine to Relieve Severe Anaphylactic Uterine Contractions. *J. Allergy Clin. Immunol. Pract.* 9, 1232–1241. doi:10.1016/j.jaip.2020.10.047
- Davis, S., and Meltzer, P. S. (2007). GEOquery: A Bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 23, 1846–1847. doi:10.1093/bioinformatics/btm254
- de Oliveira Francisco, D., de Paula Andres, M., Gueuvoghlian-Silva, B. Y., Podgaec, S., and Fridman, C. (2017). CCDC22 Gene Polymorphism Is Associated with Advanced Stages of Endometriosis in a Sample of Brazilian Women. *J. Assist. Reprod. Genet.* 34, 939–944. doi:10.1007/s10815-017-0936-0
- Demchak, B., Hull, T., Reich, M., Liefeld, T., Smoot, M., Ideker, T., et al. (2014). Cytoscape: the Network Visualization Tool for GenomeSpace Workflows. *FI000Res* 3, 151. doi:10.12688/fi000research.4492.2
- Drabble, S. J., Long, J., Alele, B., and O' Cathain, A. (2021). Constellations of Pain: a Qualitative Study of the Complexity of Women's Endometriosis-Related Pain. *Br. J. Pain* 15, 345. doi:10.1177/2049463720961413
- Faronato, M., Nguyen, V. T. M., Patten, D. K., Lombardo, Y., Steel, J. H., Patel, N., et al. (2015). DMXL2 Drives Epithelial to Mesenchymal Transition in Hormonal Therapy Resistant Breast Cancer through Notch Hyper-Activation. *Oncotarget* 6, 22467–22479. doi:10.18632/oncotarget.4164
- Fritsch, S., and Guenther, F. (2016). Neuralnet: Training of Neural Networks. R Package Version 1.33. Available at: <https://CRAN.R-project.org/package=neuralnet>. 2010-006.
- Fuhrmann, D. C., and Brüne, B. (2017). Mitochondrial Composition and Function under the Control of Hypoxia. *Redox Biol.* 12, 208–215. doi:10.1016/j.redox.2017.02.012
- Galindo, S. E., Toharia, P., Robles, Ó. D., Ros, E., Pastor, L., and Garrido, J. A. (2020). Simulation, Visualization and Analysis Tools for Pattern Recognition

- Assessment with Spiking Neuronal Networks. *Neurocomputing* 400, 309–321. doi:10.1016/j.neucom.2020.02.114
- Goldenhuys, W. J., Skolik, R., Konkole, M. E., Menze, M. A., Long, T. E., and Robart, A. R. (2019). Binding of Thiazolidinediones to the Endoplasmic Reticulum Protein Nutrient-Deprivation Autophagy Factor-1. *Bioorg. Med. Chem. Lett.* 29, 901–904. doi:10.1016/j.bmcl.2019.01.041
- Gómez-Rubio, V. (2017). ggplot2 - Elegant Graphics for Data Analysis (2nd Edition). *J. Stat. Softw.* 77 (2), 678–679. doi:10.18637/jss.v077.b02
- Ji, F., Yang, X., He, Y., Wang, H., Aili, A., and Ding, Y. (2017). Aberrant Endometrial DNA Methylome of Homeobox A10 and Catechol-O-Methyltransferase in Endometriosis. *J. Assist. Reprod. Genet.* 34, 409–415. doi:10.1007/s10815-016-0862-6
- Khan, M. T., Kaushik, A. C., Ji, L., Malik, S. I., Ali, S., and Wei, D.-Q. (2019). Artificial Neural Networks for Prediction of Tuberculosis Disease. *Front. Microbiol.* 10, 395. doi:10.3389/fmicb.2019.00395
- Kong, Y., and Yu, T. (2018). A Deep Neural Network Model Using Random Forest to Extract Feature Representation for Gene Expression Data Classification. *Sci. Rep.* 8, 16477. doi:10.1038/s41598-018-34833-6
- Kugunavar, S., and Prabhakar, C. J. (2021). Convolutional Neural Networks for the Diagnosis and Prognosis of the Coronavirus Disease Pandemic. *Vis. Comput. Ind. Biomed. Art* 4, 12. doi:10.1186/s42492-021-00078-w
- Li, Y.-w., Wang, C.-x., Chen, J.-s., Chen, L., Zhang, X.-q., and Hu, Y. (2018). Catechol-O-methyltransferase 158G/A Polymorphism and Endometriosis/adenomyosis Susceptibility: A Meta-Analysis in the Chinese Population. *J. Can. Res. Ther.* 14, 980. doi:10.4103/0973-1482.188439
- Li, H., Lai, L., and Shen, J. (2020). Development of a Susceptibility Gene Based Novel Predictive Model for the Diagnosis of Ulcerative Colitis Using Random forest and Artificial Neural Network. *Aging* 12, 20471–20482. doi:10.18632/aging.103861
- Liaw, A., Yan, J., Li, W., Han, L., Schroff, F., Criminisi, A., et al. (2014). *Package “randomForest”*. R news.
- Liu, X., Rothe, K., Yen, R., Fruhstorfer, C., Maetzig, T., Chen, M., et al. (2017). A Novel AHI-1-BCR-ABL-DNM2 Complex Regulates Leukemic Properties of Primitive CML Cells through Enhanced Cellular Endocytosis and ROS-Mediated Autophagy. *Leukemia* 31, 2376–2387. doi:10.1038/leu.2017.108
- Nakago, S., Hadfield, R. M., Zondervan, K. T., Mardon, H., Manek, S., Weeks, D. E., et al. (2001). Association between Endometriosis and N-Acetyl Transferase 2 Polymorphisms in a UK Population. *Mol. Hum. Reprod.* 7, 1079–1083. doi:10.1093/molehr/7.11.1079
- Newsholme, P., Cruzat, V. F., Keane, K. N., Carlessi, R., and De Bittencourt, P. I. H. (2016). Molecular Mechanisms of ROS Production and Oxidative Stress in Diabetes. *Biochem. J.* 473, 4527–4550. doi:10.1042/BCJ20160503C
- Parasar, P., Ozcan, P., and Terry, K. L. (2017). Endometriosis: Epidemiology, Diagnosis and Clinical Management. *Curr. Obstet. Gynecol. Rep.* 6, 34–41. doi:10.1007/s13669-017-0187-1
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* 43, e47. doi:10.1093/nar/gkv007
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., et al. (2011). pROC: An Open-Source Package for R and S+ to Analyze and Compare ROC Curves. *BMC Bioinformatics* 12, 77. doi:10.1186/1471-2105-12-77
- Samimi, M., Pourhanifeh, M. H., Mehdizadehkashi, A., Eftekhar, T., and Asemi, Z. (2019). The Role of Inflammation, Oxidative Stress, Angiogenesis, and Apoptosis in the Pathophysiology of Endometriosis: Basic Science and New Insights Based on Gene Expression. *J. Cel. Physiol.* 234, 19384–19392. doi:10.1002/jcp.28666
- Schonlau, M., and Zou, R. Y. (2020). The Random forest Algorithm for Statistical Learning. *Stat. J.* 20, 3–29. doi:10.1177/1536867X20909688
- Sethi, N., Yan, Y., Quek, D., Schubach, T., and Kang, Y. (2010). Rabconnectin-3 Is a Functional Regulator of Mammalian Notch Signaling. *J. Biol. Chem.* 285, 34757–34764. doi:10.1074/jbc.M110.158634
- Shaia, K. L., Acharya, C. R., Smeltzer, S., Lyerly, H. K., and Acharya, K. S. (2019). Non-invasive Diagnosis of Endometriosis: Using Machine Learning Instead of the Operating Room. *Fertil. Sterility* 112, e80. doi:10.1016/j.fertnstert.2019.07.331
- Su, R.-W., Strug, M. R., Joshi, N. R., Jeong, J.-W., Miele, L., Lessey, B. A., et al. (2015). Decreased Notch Pathway Signaling in the Endometrium of Women with Endometriosis Impairs Decidualization. *J. Clin. Endocrinol. Metab.* 100, E433–E442. doi:10.1210/jc.2014-3720
- Suliman, H. B., and Piantadosi, C. A. (2016). Mitochondrial Quality Control as a Therapeutic Target. *Pharmacol. Rev.* 68, 20–48. doi:10.1124/pr.115.011502
- Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING Database in 2021: Customizable Protein-Protein Networks, and Functional Characterization of User-Uploaded Gene/measurement Sets. *Nucleic Acids Res.* 49, D605–D612. doi:10.1093/nar/gkaa1074
- Taylor, H. S., Adamson, G. D., Diamond, M. P., Goldstein, S. R., Horne, A. W., Missmer, S. A., et al. (2018). An Evidence-Based Approach to Assessing Surgical versus Clinical Diagnosis of Symptomatic Endometriosis. *Int. J. Gynecol. Obstet.* 142, 131–142. doi:10.1002/ijgo.12521
- Tian, Y., Yang, J., Lan, M., and Zou, T. (2020). Construction and Analysis of a Joint Diagnosis Model of Random forest and Artificial Neural Network for Heart Failure. *Aging* 12, 26221–26235. doi:10.18632/aging.202405
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: A Universal Enrichment Tool for Interpreting Omics Data. *Innovation* 2, 100141. doi:10.1016/j.xinn.2021.100141
- Xie, N.-N., Wang, F.-F., Zhou, J., Liu, C., and Qu, F. (2020). Establishment and Analysis of a Combined Diagnostic Model of Polycystic Ovary Syndrome with Random Forest and Artificial Neural Network. *Biomed. Res. Int.* 2020, 1–13. doi:10.1155/2020/2613091
- Yigit, A., and Isik, Z. (2018). “Application of Artificial Neural Networks in Dementia and Alzheimer’s Diagnosis,” in 26th IEEE Signal Processing and Communications Applications Conference, SIU 2018. Izmir, Turkey: May 2–5, 2018. doi:10.1109/SIU.2018.8404447
- Yoo, S. M., Choi, J. H., Lee, S. Y., and Yoo, N. C. (2009). Applications of DNA Microarray in Disease Diagnostics. *J. Microbiol. Biotechnol.* 19, 635. doi:10.4014/jmb.0803.226
- Zhai, J., Jiang, L., Wen, A., Jia, J., Zhu, L., and Fan, B. (2019). Analysis of the Relationship between COMT Polymorphisms and Endometriosis Susceptibility. *Med. (United States)* 98, e13933. doi:10.1097/MD.00000000000013933
- Zhang, M., Wang, S., Tang, L., Wang, X., Zhang, T., Xia, X., et al. (2019). Downregulated Circular RNA Hsa_circ_0067301 Regulates Epithelial-Mesenchymal Transition in Endometriosis via the miR-141/Notch Signaling Pathway. *Biochem. Biophys. Res. Commun.* 514, 71–77. doi:10.1016/j.bbrc.2019.04.109
- Zhao, Q., Ye, M., Yang, W., Wang, M., Li, M., Gu, C., et al. (2018). Effect of Mst1 on Endometriosis Apoptosis and Migration: Role of Drp1-Related Mitochondrial Fission and Parkin-Required Mitophagy. *Cell. Physiol. Biochem.* 45, 1172–1190. doi:10.1159/000487450
- Zhou, H., Zhu, P., Guo, J., Hu, N., Wang, S., Li, D., et al. (2017). Ripk3 Induces Mitochondrial Apoptosis via Inhibition of FUNDC1 Mitophagy in Cardiac IR Injury. *Redox Biol.* 13, 498–507. doi:10.1016/j.redox.2017.07.007

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 She, Su, Diao and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.