frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 13 August 2021
doi: 10.3389/fcell.2021.730475

An Ovarian Cancer Susceptible Gene Prediction Method Based on Deep Learning Methods

Lu Ye[1], Yi Zhang[1], Xinying Yang[1], Fei Shen[2] and Bo Xu[2]*

[1] Department of Gynecology, Guangdong Second Provincial General Hospital, Guangzhou, China, [2] Department of Thyroid Surgery, Guangzhou First People's Hospital, School of Medicine, South China University of Technology, Guangzhou, China

Ovarian cancer (OC) is one of the most fatal diseases among women all around the world. It is highly lethal because it is usually diagnosed at an advanced stage which may reduce the survival rate greatly. Even though most of the patients are treated timely and effectively, the survival rate is still low due to the high recurrence rate of OC. With a large number of genome-wide association analysis (GWAS)-discovered risk regions of OC, expression quantitative trait locus (eQTL) analyses can explore candidate susceptible genes based on these risk loci. However, a large number of OC-related genes remain unknown. In this study, we proposed a novel gene prediction method based on different omics data and deep learning methods to identify OC causal genes. We first employed graph attention network (GAT) to obtain a compact gene feature representation, then a deep neural network (DNN) is utilized to predict OC-related genes. As a result, our model achieved a high AUC of 0.761 and AUPR of 0.788, which proved the accuracy and effectiveness of our proposed method. At last, we conducted a gene-set enrichment analysis to further explore the mechanism of OC. Finally, we predicted 245 novel OC causal genes and 10 top related KEGG pathways.

Keywords: ovarian cancer, gene prediction, omics data, deep learning method, pathway analysis

OPEN ACCESS

Edited by:
Lei Deng,
Central South University, China

Reviewed by:
Shihua Zhang,
Wuhan University of Science and Technology, China
Hong Ju,
Heilongjiang Vocational College of Biology Science and Technology, China

****Correspondence:***
Lu Ye
yu2223@163.com

Specialty section:
This article was submitted to Molecular and Cellular Pathology, a section of the journal Frontiers in Cell and Developmental Biology

Received: *25 June 2021*
Accepted: *22 July 2021*
Published: *13 August 2021*

Citation:
Ye L, Zhang Y, Yang X, Shen F and Xu B (2021) An Ovarian Cancer Susceptible Gene Prediction Method Based on Deep Learning Methods. Front. Cell Dev. Biol. 9:730475. doi: 10.3389/fcell.2021.730475

INTRODUCTION

Ovarian cancer (OC) is one of the major lethal diseases for women, despite ranking tenth in morbidity rate, it is the fifth leading cause of death among cancers (Siegel et al., 2011). Usually, OC is diagnosed at an advanced stage which induced a high death rate. However, even patients got primary treatment such as surgical resection and adjuvant drug therapy, the high rate of recurrence, and high incidence of advanced stage disease eventually caused a high mortality rate (Armstrong, 2002). In terms of treating OC and reducing the high fatality rate to improve survival in OC patients, studies have been exploring the development of new treatment, and effective chemotherapies (Badgwell and Bast, 2007; Kobayashi et al., 2012). While efforts and contributions have been made to improve the treatment, cure rates have not been raised significantly. Thus, it is more important to explore the mechanism and underlying biological pathogenic factors of OC to better understand the disease, and find a better treatment.

Genome-wide association analysis (GWAS) have identified hundreds of risk genetic variants (SNPs) associated with OC (Song et al., 2009; Bolton et al., 2010; Goode et al., 2010; Permuth-Wey et al., 2013; Pharoah et al., 2013; Shen et al., 2013). However, they can only explain a small fraction of disease risk regions in a functional way (Pomerantz et al., 2010; Grisanzio et al., 2012; Bojesen et al., 2013). It is widely known that most of the risk alleles are located in the nonprotein coding

regions of the genome, indicating that most of them are functional regulators of the expression of target genes (Hazelett et al., 2014). Thus, it is not comprehensive to identify disease-related genes by merely being dependent on GWAS datasets. To provide additional evidence for exploring risk genes, expression quantitative trait locus (eQTL) analysis is a direct method to explore candidate genes at risk loci. Since most transcripts are regulated by genes, eQTL can identify genetic variants related to the expression level of genes. eQTL analyses have identified multiple causal genes for different cancer types such as prostate, breast, colorectal, and kidney cancers (Loo et al., 2017; Yang et al., 2017; Beesley et al., 2020; Bicak et al., 2020). Therefore, it is more creditable to identify OC-related genes based on the combination of GWAS and eQTL data.

Besides, over the past decades, numerous noncoding RNAs (ncRNAs), such as lncRNA, siRNA, piRNA, and miRNA have been detected to execute the regulation function by interacting with target genes (Esteller, 2011; Lee and Young, 2013). In humans, it is estimated that the number of ncRNA genes are more than twice as many as that of protein-coding genes (Bunch et al., 2016). Thus, ncRNAs have been considered key regulators of multiple biological processes and development. Along with the rapid advancement of high-throughput sequence analyses of ncRNAs, more and more transcriptional mechanisms have been illustrated. ncRNAs should also be regarded as a major factor to explore the pathologies of OC due to its regulation function of gene expression. Hence, it is important to take into consideration the role of regulatory ncRNAs to identify OC-related causal genes.

However, along with the rapid development of understanding the mechanisms of complex disease, there are a few computational methods to predict disease genes based on various gene features. In this study, we aimed to identify susceptibility genes associated with OC based on integrated gene features. We first employed graph attention network (GAT) to learn the compact gene feature from a gene interaction network with gene features, then employed a deep neural network (DNN) to predict OC-related susceptibility genes. To further explore the mechanism of OC, we also performed a gene-set enrichment analysis to predict more related pathways in OC process.

MATERIALS AND METHODS

Work Frame

In this study, our method contains four main parts, feature extraction, compact gene feature learning based on GAT, and OC-related susceptibility gene prediction based on DNN and OC-related pathway analysis. In the first section, we extracted gene features based on integrated GWAS data, eQTL data, and published data of gene-related lncRNAs and miRNAs. In total, we extracted a 2,664-dimensional feature representation from four gene features. We then utilized a graph neural network with attention mechanism (GAT) to learn the compact gene feature for a low-dimension feature representation in order to obtain a better classification performance in the prediction process. The low-dimension feature matrix is considered the input of DNN to train the model and conduct the prediction process. After obtaining the predicted causal genes related to OC, we further performed a pathway analysis based on enrichR (Chen et al., 2013; Kuleshov et al., 2016), and a gene-set enrichment tool to find more related kyoto encyclopedia of genes and genomes (KEGG) pathways for a better understanding of the mechanism of OC. The workflow is presented in **Figure 1**.

Data Collection

We first downloaded and verified 3,181 OC-related genes from DisGeNET database (Piñero et al., 2017, 2020) as a positive dataset. To build a gene-gene interaction network, we downloaded gene interaction information from HumanNet database (Hwang et al., 2019). For constructing a balanced training set, we randomly selected 3,171 genes which have interactions with positive genes from HumanNet as a negative set. To extract ncRNA-gene interaction feature, we downloaded gene-lncRNA association and gene-miRNA association information from LncRNA2Target, and miRTarBase, respectively (Hsu et al., 2011; Jiang et al., 2015; Cheng et al., 2019; Huang et al., 2020). MiRTarBase is a database providing comprehensive information based on experimentally verified miRNA-target interactions curated from published articles; it accumulated over 13,404 validated associations. LncRNA2Target is a database storing comprehensive lncRNA-target interactions inferred from published articles and experiments.

The GWAS data providing OC susceptibility loci was downloaded from GWAS catalog database, accession ID GCST90011821 (MacArthur et al., 2017). They sampled from 1,259 European ancestry cases and 410,350 controls providing genetic variant loci related to OC. eQTL data in ovary tissue was downloaded from GTEx database (Lonsdale et al., 2013). Finally, our training set is constructed based on 3,181 positive genes and 3,171 negative genes for further feature extraction.

Feature Extraction

We extracted gene features from four aspects, susceptibility loci derived from GWAS, eQTL data from ovary tissue, interactions between genes, and miRNAs/lncRNAs. We first obtained the detail location information of training genes containing chromosome name, start position, and end position. Then, we mapped the genes to the susceptibility loci and sorted by p-value provided by GWAS data. We extracted the top five significant p-values as GWAS feature of the gene. Thus, the GWAS feature of a gene can be represented as a 5-D vector:

$$F_{\mathrm{GWAS}} = [D_1, D_2, D_3, D_4, D_5] \tag{1}$$

For those genes that cannot be mapped to five SNPs, the feature value is set to one. To extract eQTL-based gene feature, we mapped the genes to eQTL data based on gene location information, and extracted the top five significant p-values as eQTL feature, set feature value to one if a gene cannot map to five SNPs. Thus, the eQTL feature can also be represented as a 5-D vector:

$$F_{\mathrm{eQTL}} = [D_1, D_2, D_3, D_4, D_5] \tag{2}$$

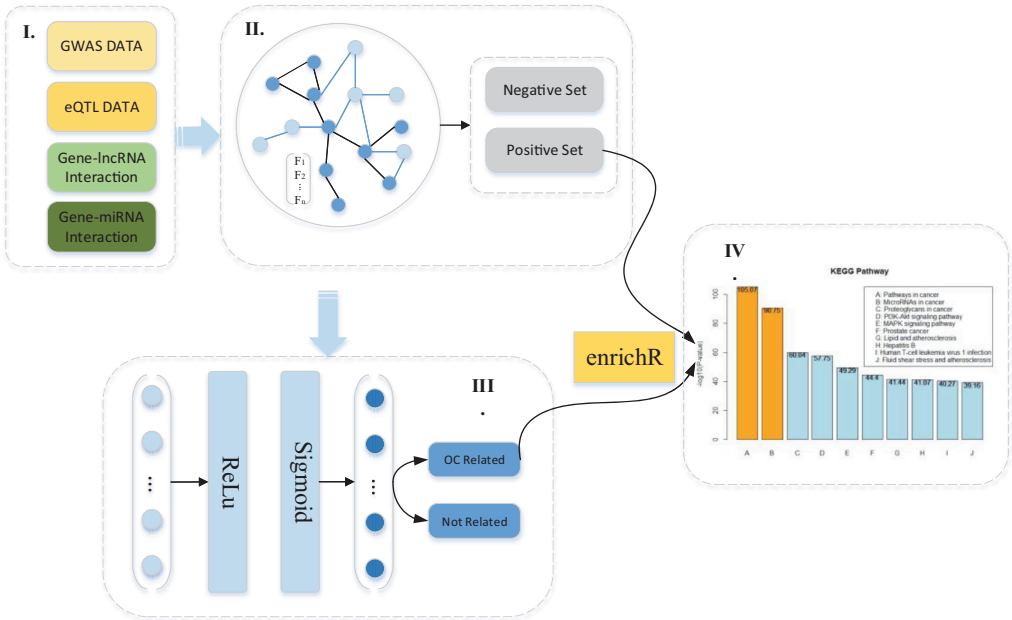

FIGURE 1 | The pipeline of ovarian cancer (OC) causal gene prediction method.

From the gene-lncRNA interaction obtained from lncRNA2Target, we filtered the interactions to make sure each of the training genes is related to at least one lncRNA. As a result, 59 lncRNAs are preserved. Thus, the lncRNA feature of a gene can be denoted as a 59-D vector, where the value is 1 if the gene is related to lncRNA[i], the value is set to 0 vice versa.

$$F_{\text{lncRNA}} = [D_1, D_2, D_3, , D_{59}] \tag{3}$$

$$D_i = \begin{cases} 0, & \textit{if gene is not related to lnRNA}_i \\ 1. & \textit{if gene is related to lnRNA}_i \end{cases} \tag{4}$$

We then filtered the gene-miRNA interactions to make sure each of the training genes is related to at least one miRNA. As a result, 2,595 miRNAs are preserved. Thus, the miRNA feature of a gene can be denoted as a 2,595-D vector:

$$F_{\text{miRNA}} = [D_1, D_2, D_3, , D_{2595}] \tag{5}$$

$$D_i = \begin{cases} 0, & \textit{if gene is not related to miRNA}_i \\ 1. & \textit{if gene is related to miRNA}_i \end{cases} \tag{6}$$

Therefore, the feature representation of each gene in training set can be denoted as a 2,664 dimensional vector. Since the feature vector could be very sparse, we need to learn the compact feature representation to obtain a better classification performance.

Compact Feature Learning Based on GAT

Sparse matrix is a matrix composed of mostly zero values, which often induces a poor classification performance in machine learning methods. Thus, we need to reconstruct the gene feature to get a low-dimensional feature representation. Since we can build a gene-gene interaction network based on gene association information obtained from HumanNet. A 6358*6358 dimensional adjacent matrix can be constructed based on the network. Besides, each gene of the Internet also has a feature representation of itself. Thus, the gene network with gene features can be regarded as a graph-structured data; to make the gene prediction method more general, and we utilized a GAT as a feature learning model. GAT addressed the shortcomings of requiring costly matrix operation and dependency on preknowledge of graph structure by stacking layers in which nodes are participant in features of neighborhoods, and arranging attention weights to each nodes. Consider the node features being denoted as: $h = \left\{\vec{h_1}, \vec{h_2}, , \vec{h_N}\right\}, \vec{h_i} \in \mathbb{R}^F$, where N is the size of training set and F is the dimension of gene features. The output of graph attentional layer is a new set of node features (of a low-dimension F'), denoted as: $h' = \left\{\vec{h_1}', \vec{h_2}', , \vec{h_N}'\right\}, \vec{h_i}' \in \mathbb{R}^{F'}$.

We then performed a self-attention on each node with a shared attentional mechanism a:$\mathbb{R}^{F'} \times \mathbb{R}^{F'}$ to compute attention coefficients:

$$e_{ij} = a(W\vec{h_i}, W\vec{h_j}) \tag{7}$$

where, $W \in \mathbb{R}^{F' \ F}$ is a weight matrix applied to each node. e_{ij} denotes the importance of node j to node i. Based on this formulation, the model allows each node to participate with every other node and dropping structural information. For each node j

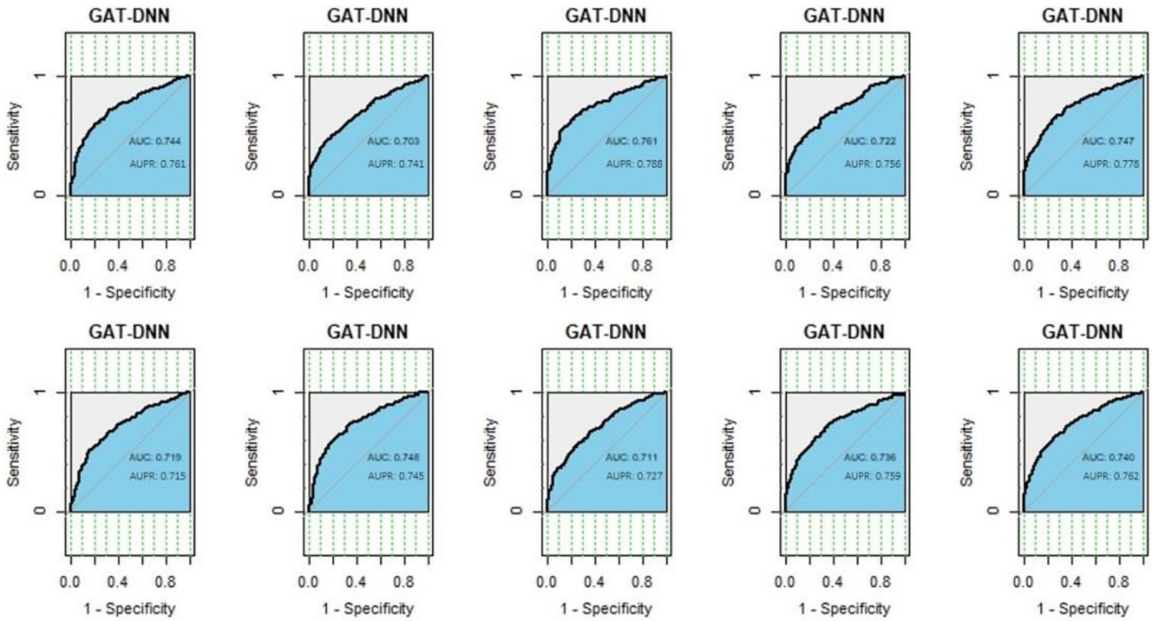

FIGURE 2 | The performance of graph attention network-deep neural network (GAT-DNN) across a 10-fold cross-validations.

in the neighborhood of node i (denoted as $\mathcal{N}_i$), we performed a softmax function to normalize the coefficients e_{ij}:

$$\alpha_{ij} = \text{softmax}\left(e_{ij}\right) = \frac{\exp\left(\text{e}_{\text{ij}}\right)}{\sum_{k\in\mathcal{N}_i}\exp\left(\text{e}_{\text{ik}}\right)} \tag{8}$$

After being activated by LeakyReLU function, e_{ij} can be denoted as:

$$e_{ij} = \text{LeakyReLU}(\vec{a}^T\left[W\vec{h_i}, ||W\vec{h_j}\right]) \tag{9}$$

where, $\vec{a} \in \mathbb{R}^{2F'}$ is a weight vector; $||$ denotes the concatenation operation. Once obtained, the output feature of each node can be computed as a linear combination of the neighborhood node features with e_{ij}:

$$\vec{h_i}' = \delta(\sum j \in \mathcal{N}_i \alpha_{ij} W \vec{h_j}) \tag{10}$$

where, δ denotes a nonlinear transition. Thus, we obtained a low-dimension feature representation of the genes based on GAT.

DNN Model Construction

Deep neural network has been regarded as a powerful tool in many domains of machine-learning applications. In this part, a binary-classification DNN model is used to predict OC-related genes based on the gene features derived from the output of GAT layer. The gene features were input to the DNN. The DNN model contains one hidden layer with a ReLU activation function and an output layer with a sigmoid activation function and a dropout technique. The sigmoid activation formulation is:

$$\sigma(x) = 1/(1+e^{-x}) \tag{11}$$

We used the Adam optimizer and binary cross-entropy function as the loss function. The loss function is:

$$\text{loss} = -\sum_{i=1}^{n} y'_i \log\left(y_i\right) + (1-y'_i)\log(1-y'_i) \tag{12}$$

Training and Testing

To verify the performance of our GAT-DNN gene prediction model, we conducted a 10-fold cross-validation method on the training dataset containing 3,181 positive samples and 3,171 negative samples. The training set was randomly divided into 10 groups, nine of the 10 are regarded as training samples, and one left group is regarded as the test samples. The training samples were used to train the model and the last one was used to test the classification performance. This process was repeated 10 times to make sure the credibility of the verification.

RESULTS

Measurement of Model Performance

The performance of 10-fold cross-validation was assessed by area under curve (AUC) and the area under precision-recall curve (AUPR), as shown in **Figure 2**. As demonstrated in **Figure 2**, the AUC and AUPR are both over 0.7 across a 10-fold cross-validation, and which is a good performance for a classification model. The best performance is shown in the third validation with the AUC of 0.761 and the AUPR of 0.788, and which is chosen as the final prediction model to identify OC-related genes.

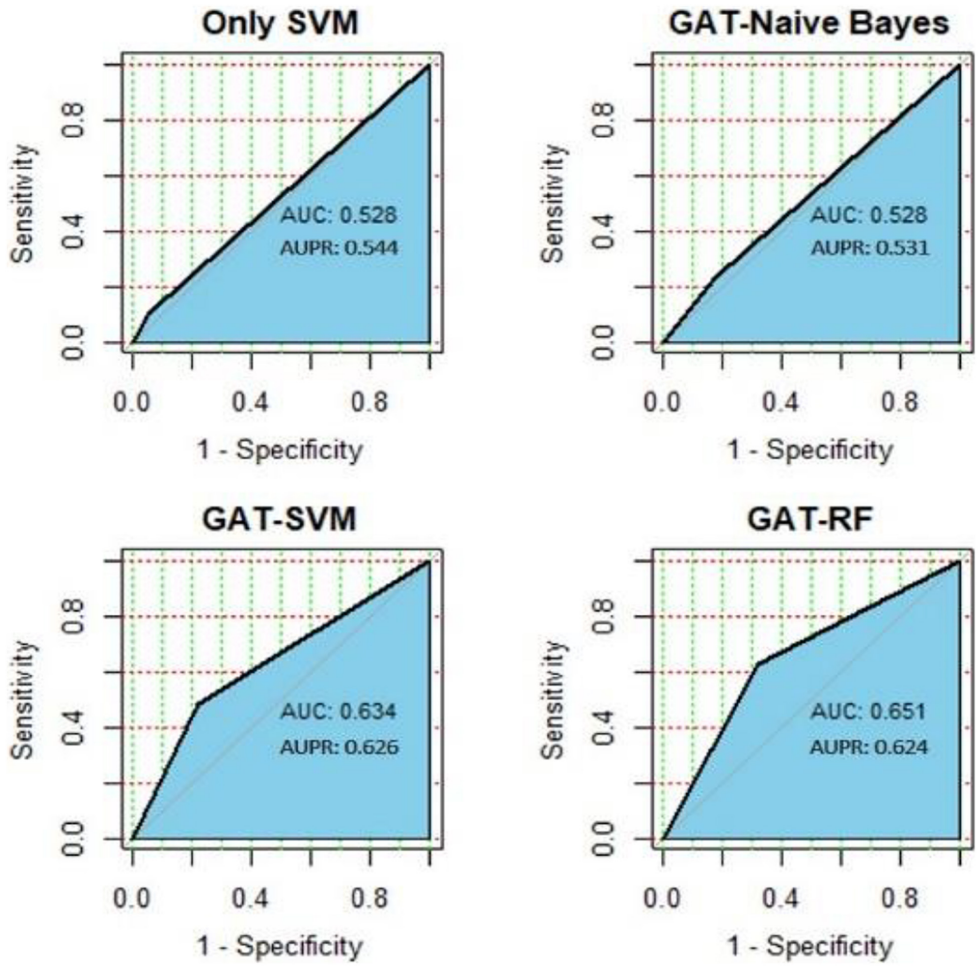

FIGURE 3 | Results of method comparison.

TABLE 1 | Top 20 predicted genes associated with ovarian cancer (OC).

Gene	Score	Gene	Score
HS6ST1	0.98354	TEAD2	0.88879
C2orf83	0.97384	PCSK1	0.88736
TM4SF4	0.95513	MTRR	0.88176
ARTN	0.95246	H6PD	0.87361
WDPCP	0.92387	EIF2B2	0.87341
KCNJ11	0.92073	SOX5	0.85771
TBL2	0.91197	NNMT	0.85176
AGTR2	0.90641	MIR324	0.84883
ATF1	0.90194	MIR33B	0.83557
PKD1	0.89846	TBPL2	0.83491

Performance Comparison

To better illustrate the effectiveness and credibility of our method, we compared our model with other four combinations of machine learning methods with the same training set we used in model training part. We compared our model with GAT-SVM, only SVM (which means the gene features are not operated with GAT), GAT-Random Forest, and GAT-Naïve Bayes. The results are shown in **Figure 3**. As shown in the results, the performance is significantly poorer than our model. The best model is GAT-RF, with an AUC of 0.651 and AUPR of 0.624, which is approximately 0.1 lower than our GAT-DNN model. However, indicated from the performance of only SVM and GAT-SVM, it is obvious that the classification performance has been improved significantly after compact feature learning by GAT layer. Therefore, our model is the best to predict OC-related genes.

OC Gene Prediction Process

Since we have demonstrated the effectiveness of our classification model based on a 10-fold cross-validation and comparison with other classification models, and we then performed the gene prediction process based on our built model on 721 ovary disease-related genes. These 721 candidate genes were downloaded from DisGeNET which is associated with ovary diseases but not OC. We extracted the gene features as mentioned in the section "Feature Extraction". After compact feature learning by GAT layer, we reduced the dimension of gene features from 2,664 to 100. We then input the compact gene features to the DNN

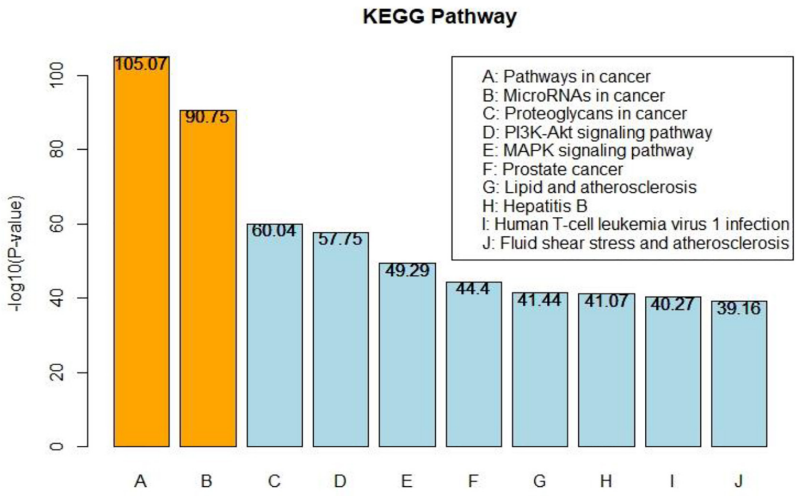

FIGURE 4 | Pathway analysis based on OC-related genes.

model; we finally predicted 245 of 721 candidate genes to be positively related to OC.

Case Study

According to the results we obtained from gene prediction process, 245 of 721 candidate genes were predicted to be associated with OC. We listed the top 20 genes in **Table 1**. Backen et al. (2007) indicated that HS6ST1 are aberrant overexpression in carcinoma of ovary compared with normal ovaries. Natanzon et al. (2018) observed a significant association between methylation WDPCP expression in OC. KCNJ11 could be considered a favorable prognostic factor since they are observed to be expressed in OC according to the investigation of Fukushiro-Lopes et al. (2020). TBL2 was identified by Kim et al. (2012) as a DNA methylation regulated cancer antigen in OC. Park et al. (2014) investigated the expression of AGTR1 and AGTR2 in OC and revealed that the dual regulation of AGTR1/2 may be a therapeutic strategy since AGTR2 could antagonize the cancer cell proliferation induced by AGTR1.

Pathway Analysis

After predicting the causal genes by our proposed model, we combined the published OC-related genes and our predicted genes with a total number of 3,426. We performed a pathway analysis on KEGG pathways using enrichr in order to further understand the mechanism of OC. Enrichr is a gene-set enrichment method to identify pathway enrichment among genes related to OC. The top 10 enriched pathways resulting from enrichr are shown in **Figure 4**. Enriched pathways are ordered by –log(p-value), obtained from a Fisher's exact test.

Consistent with KEGG DISEASE database, top 2 OC-related pathways named Pathways in cancer (hsa05200) and MicroRNAs in cancer (hsa05206) are enriched among the predicted OC-related genes. Pathway proteoglycans (PGs) in cancer are known as a key pathway in understanding cancers since PGs in the tumor microenvironment are indicated to play important roles in contributing to biology of multiple types of cancer. The MAPK and PI3K-AKT pathway have been frequently observed to be important in OC, and both of the pathways are involved in OC cell migration (Jeong et al., 2013; Li et al., 2014). Understanding the pathways related to OC are important in revealing the underlying mechanism of OC.

DISCUSSION

In this article, we proposed an OC causal gene prediction method based on deep learning methods. We first extracted gene features considering ncRNA regulation function of gene expression and integrated GWAS and eQTL data. To learn a compact feature representation, we utilized a GAT, which can learn node features from a graph-structured data format without a preknowledge of the graph structure. After GAT layer, the feature dimension is reduced from 2,664 to 100. The new feature representations were then input to a DNN model which can learn gene features and perform a binary classification task. To demonstrate the performance of our proposed method, we conducted a 10-fold cross-validation on training set and made a comparison with other four integrated machine learning models. As a result, our model is significantly better than other models and achieved a high AUC of 0.761 and the AUPR of 0.788. We then employed the constructed model to predict causal genes and obtained 245 related genes. From the result of KEGG pathway analysis, we identified more OC-related pathways which are potential favorable evidence in understanding the mechanism of OC and provide new ideas for diagnosis and treatment.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and

accession number(s) can be found in the article/supplementary material.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

LY, YZ, and BX participated in its design. LY, YZ, XY, and FS analyzed the data. LY, YZ, and FS wrote the manuscript. All authors read and approved the final manuscript.

REFERENCES

Armstrong, D. K. (2002). Relapsed ovarian cancer: challenges and management strategies for a chronic disease. *Oncologist* 7, 20–28. doi: 10.1634/theoncologist.7-suppl_5-20

Backen, A. C., Cole, C. L., Lau, S. C., Clamp, A. R., McVey, R., Gallagher, J. T., et al. (2007). Heparan sulphate synthetic and editing enzymes in ovarian cancer. *Br. J. Cancer* 96, 1544–1548. doi: 10.1038/sj.bjc.6603747

Badgwell, D., and Bast, R. C. JR. (2007). Early detection of ovarian cancer. *Dis. Mark.* 23, 397–410. doi: 10.1155/2007/309382

Beesley, J., Sivakumaran, H., Marjaneh, M. M., Shi, W., Hillman, K. M., Kaufmann, S., et al. (2020). eQTL colocalization analyses identify NTN4 as a candidate breast cancer risk gene. *Am. J. Hum. Genet.* 107, 778–787. doi: 10.1016/j.ajhg.2020.08.006

Bicak, M., Wang, X., Gao, X., Xu, X., Väänänen, R.-M., Taimen, P., et al. (2020). Prostate cancer risk SNP rs10993994 is a trans-eQTL for SNHG11 mediated through MSMB. *Hum. Mol. Genet.* 29, 1581–1591. doi: 10.1093/hmg/ddaa026

Bojesen, S. E., Pooley, K. A., Johnatty, S. E., Beesley, J., Michailidou, K., Tyrer, J. P., et al. (2013). Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat. Genet.* 45, 371–384. doi: 10.1038/ng.2566

Bolton, K. L., Tyrer, J., Song, H., Ramus, S. J., Notaridou, M., Jones, C., et al. (2010). Common variants at 19p13 are associated with susceptibility to ovarian cancer. *Nat. Genet.* 42, 880–884. doi: 10.1038/ng.666

Bunch, H., Lawney, B. P., Burkholder, A., Ma, D., Zheng, X., Motola, S., et al. (2016). RNA polymerase II promoter-proximal pausing in mammalian long non-coding genes. *Genomics* 108, 64–77. doi: 10.1016/j.ygeno.2016.07.003

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., et al. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.* 14:128. doi: 10.1186/1471-2105-14-128

Cheng, L., Wang, P., Tian, R., Wang, S., Guo, Q., Luo, M., et al. (2019). LncRNA2Target v2. 0: a comprehensive database for target genes of lncRNAs in human and mouse. *Nucleic Acids Res.* 47, D140–D144. doi: 10.1093/nar/gky1051

Esteller, M. (2011). Non-coding RNAs in human disease. *Nat. Rev. Genet.* 12, 861–874. doi: 10.1038/nrg3074

Fukushiro-Lopes, D., Hegel, A. D., Russo, A., Senyuk, V., Liotta, M., Beeson, G. C., et al. (2020). Repurposing Kir6/SUR2 channel activator minoxidil to arrests growth of gynecologic cancers. *Front. Pharmacol.* 11:577. doi: 10.3389/fphar.2020.00577

Goode, E. L., Chenevix-Trench, G., Song, H., Ramus, S. J., Notaridou, M., Lawrenson, K., et al. (2010). A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. *Nat. Genet.* 42, 874–879. doi: 10.1038/ng.668

Grisanzio, C., Werner, L., Takeda, D., Awoyemi, B. C., Pomerantz, M. M., Yamada, H., et al. (2012). Genetic and functional analyses implicate the NUDT11, HNF1B, and SLC22A3 genes in prostate cancer pathogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11252–11257. doi: 10.1073/pnas.1200853109

Hazelett, D. J., Rhie, S. K., Gaddis, M., Yan, C., Lakeland, D. L., Coetzee, S. G., et al. (2014). Comprehensive functional annotation of 77 prostate cancer risk loci. *PLoS Genet.* 10:e1004102. doi: 10.1371/journal.pgen.1004102

Hsu, S.-D., Lin, F.-M., Wu, W.-Y., Liang, C., Huang, W.-C., Chan, W.-L., et al. (2011). miRTarBase: a database curates experimentally validated microRNA–target interactions. *Nucleic Acids Res.* 39, D163–D169. doi: 10.1093/nar/gkq1107

Huang, H.-Y., Lin, Y.-C.-D., Li, J., Huang, K.-Y., Shrestha, S., Hong, H.-C., et al. (2020). miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database. *Nucleic Acids Res.* 48, D148–D154. doi: 10.1093/nar/gkz896

Hwang, S., Kim, C. Y., Yang, S., Kim, E., Hart, T., Marcotte, E. M., et al. (2019). HumanNet v2: human gene networks for disease research. *Nucleic Acids Res.* 47, D573–D580. doi: 10.1093/nar/gky1126

Jeong, G. O., Shin, S. H., Seo, E. J., Kwon, Y. W., Heo, S. C., Kim, K.-H., et al. (2013). TAZ mediates lysophosphatidic acid-induced migration and proliferation of epithelial ovarian cancer cells. *Cell. Physiol. Biochem.* 32, 253–263. doi: 10.1159/000354434

Jiang, Q., Wang, J., Wu, X., Ma, R., Zhang, T., Jin, S., et al. (2015). LncRNA2Target: a database for differentially expressed genes after lncRNA knockdown or overexpression. *Nucleic Acids Res.* 43, D193–D196. doi: 10.1093/nar/gku1173

Kim, K.-M., Song, M.-H., Kim, M.-J., Daudi, S., Miliotto, A., Old, L., et al. (2012). A novel cancer/testis antigen KP-OVA-52 identified by SEREX in human ovarian cancer is regulated by DNA methylation. *Int. J. Oncol.* 41, 1139–1147. doi: 10.3892/ijo.2012.1508

Kobayashi, E., Ueda, Y., Matsuzaki, S., Yokoyama, T., Kimura, T., Yoshino, K., et al. (2012). Biomarkers for screening, diagnosis, and monitoring of ovarian cancer. *Cancer Epidemiol. Prevent. Biomark.* 21, 1902–1912. doi: 10.1158/1055-9965.epi-12-0646

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, W90–W97. doi: 10.1093/nar/gkw377

Lee, T. I., and Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. *Cell* 152, 1237–1251. doi: 10.1016/j.cell.2013.02.014

Li, H., Zeng, J., and Shen, K. (2014). PI3K/AKT/mTOR signaling pathway as a therapeutic target for ovarian cancer. *Arch. Gynecol. Obstetr.* 290, 1067–1078. doi: 10.1007/s00404-014-3377-3

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., et al. (2013). The genotype-tissue expression (GTEx) project. *Nat. Genet.* 45, 580–585. doi: 10.1038/ng.2653

Loo, L. W., Lemire, M., and Le Marchand, L. (2017). In silico pathway analysis and tissue specific cis-eQTL for colorectal cancer GWAS risk variants. *BMC Genom.* 18:381. doi: 10.1186/s12864-017-3750-2

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., et al. (2017). The new NHGRI-EBI catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* 45, D896–D901. doi: 10.1093/nar/gkw1133

Natanzon, Y., Earp, M., Cunningham, J. M., Kalli, K. R., Wang, C., Armasu, S. M., et al. (2018). Genomic analysis using regularized regression in high-grade serous ovarian cancer. *Cancer Inform.* 17:1176935118755341. doi: 10.1177/1176935118755341

Park, Y.-A., Choi, C. H., Do, I.-G., Song, S. Y., Lee, J. K., Cho, Y. J., et al. (2014). Dual targeting of angiotensin receptors (AGTR1 and AGTR2) in epithelial ovarian carcinoma. *Gynecol. Oncol.* 135, 108–117. doi: 10.1016/j.ygyno.2014.06.031

Permuth-Wey, J., Lawrenson, K., Shen, H. C., Velkova, A., Tyrer, J. P., Chen, Z., et al. (2013). Identification and molecular characterization of a new ovarian cancer susceptibility locus at 17q21. 31. *Nat. Commun.* 4:1627. doi: 10.1038/ncomms2613

Pharoah, P. D., Tsai, Y.-Y., Ramus, S. J., Phelan, C. M., Goode, E. L., Lawrenson, K., et al. (2013). GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat. Genet.* 45, 362–370. doi: 10.1038/ng.2564

Piñero, J., Bravo, À, Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., et al. (2017). DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* 45, D833–D839. doi: 10.1093/nar/gkw943

Piñero, J., Ramírez-Anguita, J. M., Saüch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., et al. (2020). The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* 48, D845–D855. doi: 10.1093/nar/gkz1021

Pomerantz, M. M., Shrestha, Y., Flavin, R. J., Regan, M. M., Penney, K. L., Mucci, L. A., et al. (2010). Analysis of the 10q11 cancer risk locus implicates MSMB and NCOA4 in human prostate tumorigenesis. *PLoS Genet.* 6:e1001204. doi: 10.1371/journal.pgen.1001204

Shen, H., Fridley, B. L., Song, H., Lawrenson, K., Cunningham, J. M., Ramus, S. J., et al. (2013). Epigenetic analysis leads to identification of HNF1B as a subtype-specific susceptibility gene for ovarian cancer. *Nat. Commun.* 4:1628. doi: 10.1038/ncomms2629

Siegel, R., Ward, E., Brawley, O., and Jemal, A. (2011). Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *Cancer J. Clin.* 61, 212–236. doi: 10.3322/caac.20121

Song, H., Ramus, S. J., Tyrer, J., Bolton, K. L., Gentry-Maharaj, A., Wozniak, E., et al. (2009). A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22. 2. *Nat. Genet.* 41, 996–1000. doi: 10.1038/ng.424

Yang, M. Q., Li, D., Yang, W., Zhang, Y., Liu, J., and Tong, W. (2017). A gene module-based eQTL analysis prioritizing disease genes and pathways in kidney cancer. *Comput. Struct. Biotechnol. J.* 15, 463–470. doi: 10.1016/j.csbj.2017.09.003

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 © 2021 Ye, Zhang, Yang, Shen and Xu. 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.