Check for updates

XGBG: A Novel Method for Identifying Ovarian Carcinoma Susceptible Genes Based on Deep Learning

Ke Feng Sun[1†], Li Min Sun[2†], Dong Zhou[3†], Ying Ying Chen[4], Xi Wen Hao[5], Hong Ruo Liu[2*], Xin Liu[2*] and Jing Jing Chen[6*]

[1] Department of Obstetrics and Gynecology, First Affiliated Hospital, Heilongjiang University of Chinese Medicine, Harbin, China, [2] Department of Oncology, The Second Affiliated Hospital of Dalian Medical University, Dalian, China, [3] Department of Oncology, Affiliated Zhongshan Hospital of Dalian University, Dalian, China, [4] Department of Nephrology, The First Affiliated Hospital of Heilongjiang University of Chinese Medical, Harbin, China, [5] Heilongjiang University of Chinese Medicine, Harbin, China, [6] Department of Rheumatology and Immunology, The First Hospital Affiliated to Army Medical University, Chongqing, China

Ovarian carcinomas (OCs) represent a heterogeneous group of neoplasms consisting of several entities with pathogenesis, molecular profiles, multiple risk factors, and outcomes. OC has been regarded as the most lethal cancer among women all around the world. There are at least five main types of OCs classified by the fifth edition of the World Health Organization of tumors: high-/low-grade serous carcinoma, mucinous carcinoma, clear cell carcinoma, and endometrioid carcinoma. With the improved knowledge of genome-wide association study (GWAS) and expression quantitative trait locus (eQTL) analyses, the knowledge of genomic landscape of complex diseases has been uncovered in large measure. Moreover, pathway analyses also play an important role in exploring the underlying mechanism of complex diseases by providing curated pathway models and information about molecular dynamics and cellular processes. To investigate OCs deeper, we introduced a novel disease susceptible gene prediction method, XGBG, which could be used in identifying OC-related genes based on different omics data and deep learning methods. We first employed the graph convolutional network (GCN) to reconstruct the gene features based on both gene feature and network topological structure. Then, a boosting method is utilized to predict OC susceptible genes. As a result, our model achieved a high AUC of 0.7541 and an AUPR of 0.8051, which indicates the effectiveness of the XGPG. Based on the newly predicted OC susceptible genes, we gathered and researched related literatures to provide strong support to the results, which may help in understanding the pathogenesis and mechanisms of the disease.

Keywords: ovarian cancer, susceptible genes, XGBG, deep learning, pathway analyses

INTRODUCTION

Ovarian carcinomas (OCs) are one of the most fatal cancers in women; a scientific study of the disease is of vital priority due to its high death rate (1). A better understanding of the entities and molecules that contribute to the pathogenesis and progression of OC is essential to improve the diagnostics and treatment of the disease. Although the etiologic causes of OCs have not been recognized well, genetic factors that caused mutations in the disease have been examined profoundly with the help of many genetic approaches. However, there are still many disease susceptible genes not identified, and it is of vital importance to explore the mechanism and underlying pathogenic factors to better understand the disease and make a contribution in treating the disease.

A genome-wide association study (GWAS) is an approach utilized in genetics research to associate specific genetic variants [single-nucleotide polymorphisms (SNPs)] with a specific disease. It has identified hundreds of risk genetic variants (SNPs) that may result in ovarian cancers (2–6). However, these studies can only explain a small fraction of disease-related regions in a functional point of view (7–9). Since many risk alleles may locate in the non-protein-coding regions to regulate the expression of target genes (10), though GWAS provides strong support in revealing the associations between variants and traits, it is not comprehensive to discover the disease-related genes or gene regulators merely based on GWAS datasets.

Expression quantitative trait loci (eQTLs) are genomic loci that explain variation in expression levels of genes, which can be regarded as an additional evidence for identifying disease-related genes. eQTLs indicate the chromosomal loci that can explain variance in expression traits. These distinguishing characteristics from most expression quantitative trains are not the product of the expression of a single gene. With the help of eQTL analyses, a lot of causal genes for multiple types of cancers have been identified, such as kidney cancers, prostate cancers, breast cancers (9, 11, 12), and other complex diseases such as Alzheimer's disease and schizophrenia (13, 14). Therefore, it is more worthy to discover disease causal genes based on the integration of both GWAS and eQTL datasets.

In addition to the genetic information derived from GWAS and eQTL datasets to understand the mechanisms of complex diseases, investigation and identification of molecular pathways are also important in exploring the underlying mechanism of diseases. Pathway analysis is a typical efficient analysis to explore the biology of genes and proteins that are differentially expressed in biological processes. There are many widely accepted pathway databases such as KEGG and BioCarta that can provide illustrative information to study diseases from the view of pathway system (15, 16). According to the information of molecular dynamics and cellular processes, genes and gene products are annotated based on different functions and characteristics (17). Since complex diseases are not only caused by a single gene or a single biological process, it is important to understand the diseases and identify disease causal genes from the point of view of a pathway system.

In this article, we proposed a novel OC causal gene identification method, XGPG, integrating gene features from both genomic point and pathway annotation point. We first employed the graph convolutional network (GCN) to reconstruct the gene feature based on both gene feature and network topological structure, then utilized a boosting method, extreme gradient boosting (XGBoost), to predict OC-related susceptible genes as a binary classification problem. By applying this method, we built an efficient gene prediction model and prioritized more putative genes associated with OCs.
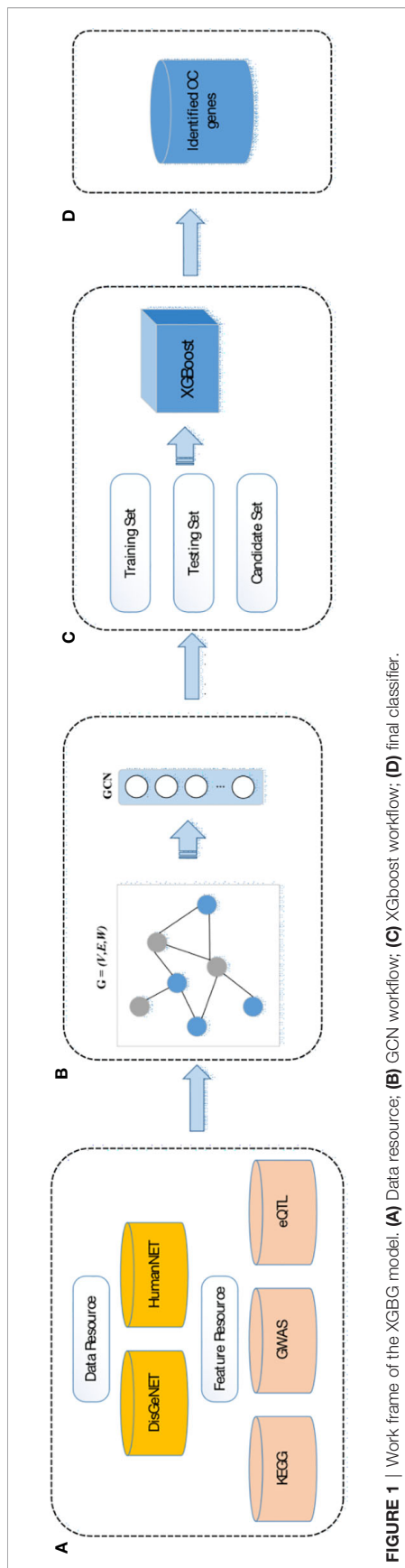
METHODS

Framework

Our method, XGPG, contains 4 main parts, data collection, feature extraction, gene feature reconstruction based on both gene feature and network topology structure, and OC causal gene prediction based on the constructed XGBoost model. In the first section (A), we manually collected different types of ovarian diseases including OC-related genes from the DisGeNET database (18) and then we obtained gene features from the GWAS Catalog, GTEx Portal, and KEGG database for different features (19, 20). Furthermore, we collected gene interaction information from the HumanNet database (21). (B) Thus, we extracted gene features from GWAS data, eQTL data, and pathway annotations, and then extracted gene network structure topological features based on the gene–gene interaction network. (C) After the feature extraction process, we utilized the GCN model to reconstruct the integrated gene features based on both gene feature and topological structure for a more precise representation of collected genes. (D) In the disease gene prediction part, a boosting model, XGBoost, is employed for constructing the prediction model and to prioritize OC-related genes. The work frame is shown in **Figure 1**.

Data Collection

We first downloaded published verified ovarian cancer-related genes from the DisGeNET database; after filtering, the dataset contains 3,181 genes to be regarded as a positive gene set. To construct a balanced training set, we randomly selected 3,171 genes that have interactions with positive genes but have no associations with ovarian diseases. These genes are used to construct the negative gene set. Then, we downloaded gene interaction information from the HumanNet database to build the gene–gene interaction network. For the prediction of OC causal genes, we also downloaded 721 ovarian disease-related genes as candidate genes to construct the prediction gene set. To extract gene features, we downloaded GWAS data from the GWAS Catalog and obtained 9,793,553 susceptible loci associated with OC, and we downloaded eQTL data from the GTEx v8 database including 25,325 susceptible loci detected in ovary tissue based on gene expression level. Moreover, we downloaded gene-pathway information from the KEGG database, including 343 annotated pathways.

FIGURE 1 | Work frame of the XGBG model. **(A)** Data resource; **(B)** GCN workflow; **(C)** XGboost workflow; **(D)** final classifier.

Feature Extraction

We extracted gene features from three aspects, namely, GWAS data, eQTL, data and KEGG pathway information. We first obtained the detailed gene location information of the training and predictive gene data, including chromosome name, start position, and end position. Then, the genes are mapped to the SNPs provided by GWAS data. To construct the SNP feature, we sorted the gene-mapped SNPs by p-value and extracted the top 5 significant SNPs as the SNP feature of the gene. Thus, the SNP feature can be denoted as a 5-D vector:

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

For those genes that have less than 5 mapped SNPs, we set the value to 9×10^{-6} to avoid calculation error. For the expression feature, we mapped the genes to eQTL data based on gene location information and then extracted the top 5 significant eQTL p-values as expression feature. We also set the value to 9×10^{-6} for those genes mapped to less than 5 loci to avoid the calculation error. Thus, the expression feature can be denoted as a 5-D vector:

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

We then downloaded the KGML files from the KEGG database, representing the details for computational analysis and pathway relations in KEGG pathways. According to the KGML files, we can obtain the genes that participate in each KEGG annotated pathway. In total, the KEGG database has annotated 343 pathways; thus, the pathway feature of each gene can be denoted as a 343-D vector; the value is set to 1 if the gene is in the pathway process or set to 0 *vice versa*:

$$F_{path} = [D_1, D_2, \dots, D_{342}, D_{343}] \tag{3}$$

$$D_i = \begin{cases} 0, & \text{if gene is in pathway } i \\ 1, & \text{if gene is not in pathway } i \end{cases} \tag{4}$$

Thus, the primary feature representation of each gene can be denoted as a 353-D vector including the SNP feature, the expression feature, and the pathway feature. Since the feature matrix could be very sparse and is not comprehensive, we further utilized the GCN model to reconstruct the feature representation with the information of the gene interaction network topological structure.

Feature Reconstruction by GCN

We first downloaded the gene–gene interaction information from the HumanNet database and constructed a gene–interaction network of the training set with dimensions of 6,352 × 6,352. Then, the adjacent matrix can be constructed based on the topological structure of the net. Next, the gene interaction network with gene features is input to the GCN model to reconstruct the gene features to obtain a more comprehensive feature representation. Consider the graph G = (V, E, W), where V is the nodes, E is the edge, and W is the weight matrix encoding the associations between nodes. In the

GCN model, Rectified Linear Units (ReLU) is used as the activation function. We input the gene feature matrix X to the GCN model and then the gene feature can be extracted by the propagation rule of each layer:

$$H^{l+1} = \sigma\left(LH^{l}XW^{l}\right) \quad (5)$$

$$\text{ReLU}(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (6)$$

where σ is the non-linearity activation function; here, we used ReLU.

Lastly, gene feature representation is reconstructed by GCN.

Gene Prediction Based on XGBoost Model

XGBoost is a state-of-the-art boosting method that has been widely employed in many kinds of data mining problems. It can also be used in classification and regression problems. Boosting is an ensemble learning algorithm that firstly train a weak model and then train an enhanced model to improve the errors by iteration. By iteration, the new model can fit the residuals of the previous model. Here, we utilized the "xgboost" package in R to perform the training and prediction process. In order to evaluate our prediction model, we performed a 10-fold cross-validation on the 6,352 training set. Since the training set is composed of 3,181 positive samples and 3,171 negative samples, we randomly divided them into 10 groups, and 9 of them is used to train the model and the last one is used to test the model based on the labels at each time. Grid searches were performed to evaluate the best performance of the parameters of the model.

RESULTS

Measurement of Model Performance

Since we have assessed the performance of our model based on 10 CVs with training sets, the ROC curve and PR curve are used to measure the performance of the model; the curves of 10 CVs are shown in **Figure 2**. The AUC and AUPR of 10 CVs are shown in **Table 1**. As a result, we obtained the average AUPR of 0.8051 and the average AUC of 0.7541. We chose the best performance model with an AUPR of 0.8301 and an AUC of 0.7770 to predict the OC causal genes.

Performance Comparison Between Models

Although we have proved the performance of XGPG by 10 CVs on the training set, there have been many other machine learning and deep learning methods used in classification problems, such as random forest (RF), Naïve Bayesian (NB), support vector machine (SVM), and deep neural network (DNN). To better illustrate the effectiveness and credibility of XGPG, we also compared it with SVM, RF, Naïve Bayes, and DNN. In order to ensure the consensus of the input to each model, all the gene features are reconstructed by GCN. The results are shown in **Figure 3**. As shown in the figure, SVM and RF perform better than NB and DNN, but they are far behind the XGBoost model.
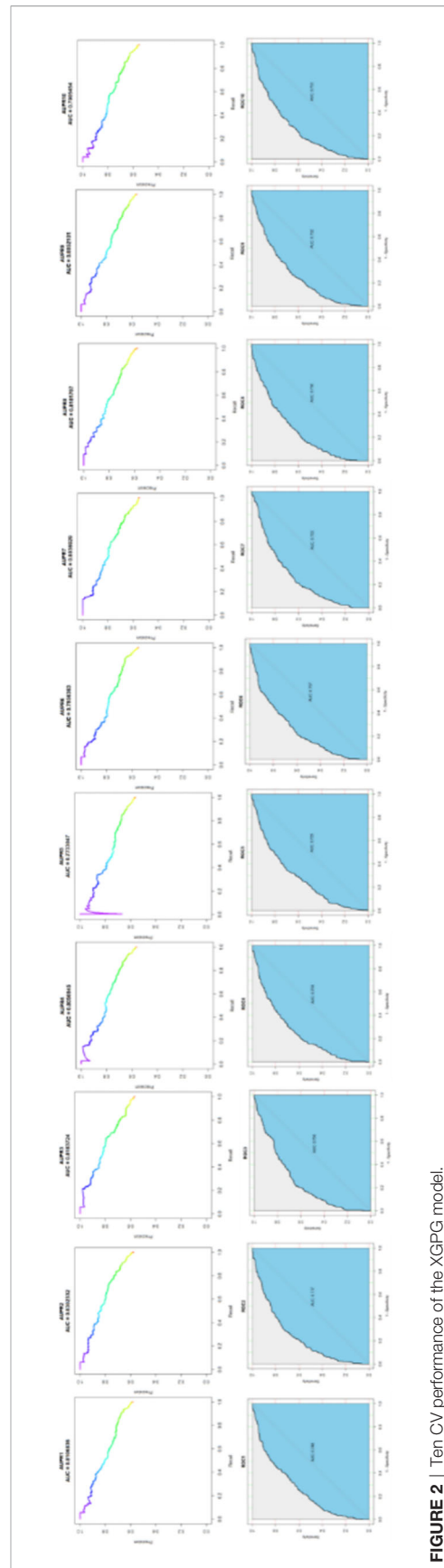

FIGURE 2 | Ten CV performance of the XGPG model.

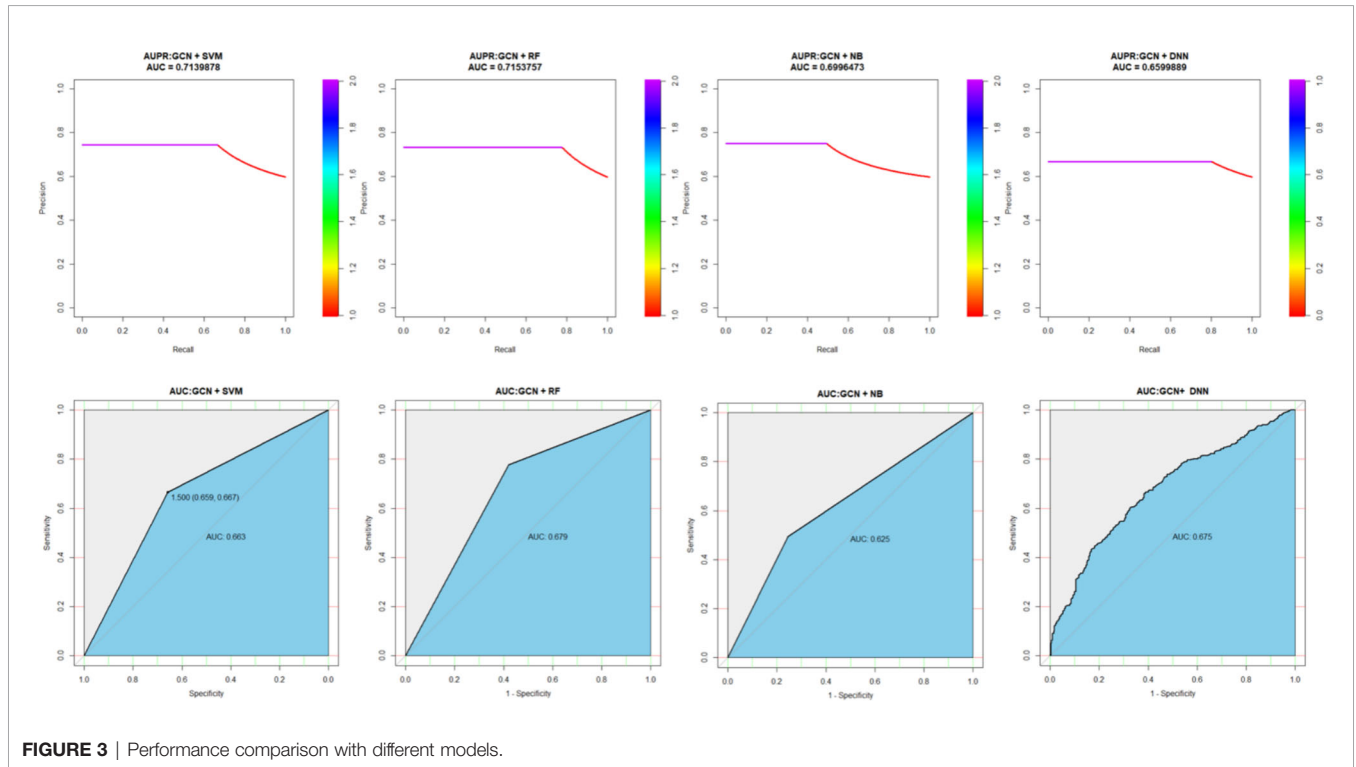|	1	2	3	4	5	6	7	8	9	10	Ave
AUPR	0.8102	0.8301	0.8187	0.8050	0.7731	0.7959	0.8034	0.8191	0.8050	0.7904	0.8051
AUC	0.7484	0.7770	0.7558	0.7592	0.7294	0.7568	0.7533	0.7564	0.7516	0.7532	0.7541

FIGURE 3 | Performance comparison with different models.

OC Gene Prediction Process

Since we have demonstrated the performance of our method and chose the best model to predict the OC genes, we then performed the gene prediction process with 721 verified ovary disease-related genes obtained from DisGeNET to further identify genes that are significantly associated with OCs. We also extracted the gene features as mentioned in the *Feature Extraction* section and built the gene interaction network to obtain the topological structure. After the gene prediction process by XGPG, we finally prioritize the candidate genes by the score resulting from the XGBoost model.

CASE STUDY

According to the results, our method predicted 148 (score threshold is 0.8) and 45 (score threshold is 0.9) OC causal genes from 721 candidate susceptible genes. We listed the top 20 genes in **Table 2**. As shown in **Table 2**, some of predicted genes have been reported to have direct or indirect associations with OC. Studies have indicated that KNG1 is highly related to the gonadotropin-releasing hormone (GnRH) (22), which is a hypothalamic neuropeptide that plays an important role in the reproductive system. Investigators have made a great effort to

TABLE 2 | Top 20 predicted OC causal genes.

Symbol	NCBI ID	Symbol	NCBI ID
3827	KNG1	2147	F2
2162	F13A1	5577	PRKAR2B
5921	RASA1	4086	SMAD1
657	BMPR1A	58	ACTA1
2688	GH1	2690	GHR
1489	CTF1	3489	IGFBP6
186	AGTR2	4879	NPPB
22806	IKZF3	10370	CITED2
8204	NRIP1	406954	MIR181A2
407021	MIR29A	407036	MIR32

develop GnRH agonists and antagonists for the treatment of tumors such as ovarian cancers (23). Coagulation factor II (F2) is found to be overexpressed in various epithelial neoplasms including ovarian cancer (24); F2 receptor, also known as PAR1, has been provided to be differentially expressed in ovarian cancer tissue (25). F13A, also known as coagulation factor XIII A, has been proven to have a significantly higher concentration in OC plasma, which may be a powerful tool for the clinical diagnosis and prognostic prediction of the disease (26). RASA1 is a member of the RAS-GAP family, which has been reported to play an important role in cell proliferation and

migration in several types of cancers, including OC, by inhibiting the malignant progression of OC cells in a high level (27). Furthermore, SMAD1 can regulate BMPs (such as BMPR1A), resulting in aberrant BMP signaling in ovarian cancer pathology (28, 29). The IGF system has been implicated in OC since it has a key role in normal growth and development. In the Yang study, they proved that IGFBP-6 may have profound effects on the migration of two ovarian cancer cell lines, which may help in developing an IGFBP-6-based therapeutic for ovarian cancers (30). Since AGTR1 has been demonstrated to be the main effector of RAS and AGTR1 protein was detected in 86% of OC tissues, AGTR2 is the antagonist of AGTR1, which means that it also plays an important role in the pathology of OC (31). NPPB is a secreted protein that has been proven to maintain a high level in the blood of women with ovarian cancer, which indicates that NPPB may be a novel biomarker for the detection of EOC (32).

DISCUSSION

OCs are one of the most dangerous cancers for women. It is important and essential to understand the mechanisms of the disease. In this study, we proposed an OC causal gene prediction method, XGPG, based on the deep learning method and the boosting method. Since GWASs have identified lots of susceptible loci associated with OC, due to the theory of linkage disequilibrium (LD), SNPs can regulate the pathologies of traits on the expression level of target genes. Thus, we integrated both GWAS and eQTL data to integrate the gene feature from both genetic and expression levels. Moreover, since complex diseases are not only caused by a single gene or SNP, it is important to also take gene–gene interaction into consideration. We built the gene interaction network to extract the gene network topological structure. Based on both gene feature and structure feature, we can reconstruct the gene feature representation by the GCN model and then perform the prediction process using the XGBoost model. We obtained a high AUPR 0.8051 of and an AUC of 0.7541 on the training set composed of 3,181 positive samples and 3,171 negative samples after 10-fold cross-validation. Compared with 4 other models, SVM, RF, NB and DNN, our model performed much better. Then, we performed the OC prediction process on the 721 candidate genes and derived a prioritized gene list. As a result, our method predicted 148 (score threshold is 0.8) and 45 (score threshold is 0.9) OC causal genes. From the results, prioritized genes such as F13A, RASA, SMAD, and AGTR2, and several other genes are published and proved to be associated with OC, which also proved the effectiveness of our method. In summary, our method is helpful in further understanding the etiology and pathology of OC, and may be used as a strong theoretical evidence for drug design.

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 authors.

AUTHOR CONTRIBUTIONS

KS, LS, and DZ designed the experiments, analyzed the data, and wrote the manuscript. YC and XH analyzed the bioinformatic data. HL and XL provided important ideas. This whole work is guided by JC. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Siegel R, Ward E, Brawley O, Jemal A. Cancer Statistics, 2011: The Impact of Eliminating Socioeconomic and Racial Disparities on Premature Cancer Deaths. *CA: Cancer J Clin* (2011) 61(4):212–36. doi: 10.3322/caac.20121

2. Song H, Ramus SJ, Tyrer J, Bolton KL, Gentry-Maharaj A, Wozniak E, et al. A Genome-Wide Association Study Identifies a New Ovarian Cancer Susceptibility Locus on 9p22. 2. *Nat Genet* (2009) 41(9):996–1000. doi: 10.1038/ng.424

3. Permuth-Wey J, Lawrenson K, Shen HC, Velkova A, Tyrer JP, Chen Z, et al. Identification and Molecular Characterization of a New Ovarian Cancer Susceptibility Locus at 17q21. 31. *Nat Commun* (2013) 4(1):1–12. doi: 10.1038/ncomms2613

4. Pharoah PD, Tsai Y-Y, Ramus SJ, Phelan CM, Goode EL, Lawrenson K, et al. GWAS Meta-Analysis and Replication Identifies Three New Susceptibility Loci for Ovarian Cancer. *Nat Genet* (2013) 45(4):362–70. doi: 10.1038/ng.2564

5. Bolton KL, Tyrer J, Song H, Ramus SJ, Notaridou M, Jones C, et al. Common Variants at 19p13 are Associated With Susceptibility to Ovarian Cancer. *Nat Genet* (2010) 42(10):880–4. doi: 10.1038/ng.666

6. Shen H, Fridley BL, Song H, Lawrenson K, Cunningham JM, Ramus SJ, et al. Epigenetic Analysis Leads to Identification of HNF1B as a Subtype-Specific Susceptibility Gene for Ovarian Cancer. *Nat Commun* (2013) 4(1):1–10. doi: 10.1038/ncomms2629

7. Grisanzio C, Werner L, Takeda D, Awoyemi BC, Pomerantz MM, Yamada H, et al. Genetic and Functional Analyses Implicate the NUDT11, HNF1B, and SLC22A3 Genes in Prostate Cancer Pathogenesis. *Proc Natl Acad Sci* (2012) 109(28):11252–7. doi: 10.1073/pnas.1200853109

8. Pomerantz MM, Shrestha Y, Flavin RJ, Regan MM, Penney KL, Mucci LA, et al. Analysis of the 10q11 Cancer Risk Locus Implicates MSMB and NCOA4 in Human Prostate Tumorigenesis. *PloS Genet* (2010) 6(11):e1001204. doi: 10.1371/journal.pgen.1001204

9. Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, et al. Multiple Independent Variants at the TERT Locus Are Associated With Telomere Length and Risks of Breast and Ovarian Cancer. *Nat Genet* (2013) 45(4):371–84. doi: 10.1038/ng.2566

10. Hazelett DJ, Rhie SK, Gaddis M, Yan C, Lakeland DL, Coetzee SG, et al. Comprehensive Functional Annotation of 77 Prostate Cancer Risk Loci. *PloS Genet* (2014) 10(1):e1004102. doi: 10.1371/journal.pgen.1004102

11. Yang MQ, Li D, Yang W, Zhang Y, Liu J, Tong W. A Gene Module-Based eQTL Analysis Prioritizing Disease Genes and Pathways in Kidney Cancer. *Comput Struct Biotechnol J* (2017) 15:463–70. doi: 10.1016/j.csbj.2017.09.003

12. Loo LW, Lemire M, Le Marchand L. In Silico Pathway Analysis and Tissue Specific cis-eQTL for Colorectal Cancer GWAS Risk Variants. *BMC Genomics* (2017) 18(1):1–14. doi: 10.1186/s12864-017-3750-2

13. Patel D, Zhang X, Farrell JJ, Chung J, Stein TD, Lunetta KL, et al. Cell-Type-Specific Expression Quantitative Trait Loci Associated With Alzheimer

Disease in Blood and Brain Tissue. *Trans Psychiatry* (2021) 11(1):1–17. doi: 10.1038/s41398-021-01373-z

14. Cai L, Huang T, Su J, Zhang X, Chen W, Zhang F, et al. Implications of Newly Identified Brain eQTL Genes and Their Interactors in Schizophrenia. *Mol Therapy-Nucleic Acids* (2018) 12:433–42. doi: 10.1016/j.omtn.2018.05.026

15. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* (2000) 28(1):27–30. doi: 10.1093/nar/28.1.27

16. Nishimura D. BioCarta. *Biotech Softw Internet Rep: Comput Softw J Sci* (2001) 2(3):117–20. doi: 10.1089/152791601750294344

17. Ge H, Liu Z, Church GM, Vidal M. Correlation Between Transcriptome and Interactome Mapping Data From Saccharomyces Cerevisiae. *Nat Genet* (2001) 29(4):482–6. doi: 10.1038/ng776

18. Bauer-Mehren A, Rautschka M, Sanz F, Furlong LI. DisGeNET: A Cytoscape Plugin to Visualize, Integrate, Search and Analyze Gene–Disease Networks. *Bioinformatics* (2010) 26(22):2924–6. doi: 10.1093/bioinformatics/btq538

19. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of Published Genome-Wide Association Studies, Targeted Arrays and Summary Statistics 2019. *Nucleic Acids Res* (2019) 47(D1):D1005–D12. doi: 10.1093/nar/gky1120

20. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) Project. *Nat Genet* (2013) 45(6):580–5. doi: 10.1038/ng.2653

21. Hwang S, Kim CY, Yang S, Kim E, Hart T, Marcotte EM, et al. HumanNet V2: Human Gene Networks for Disease Research. *Nucleic Acids Res* (2019) 47 (D1):D573–D80. doi: 10.1093/nar/gky1126

22. Tripathi PH, Akhtar J, Arora J, Saran RK, Mishra N, Polisetty RV, et al. Quantitative Proteomic Analysis of GnRH Agonist Treated GBM Cell Line LN229 Revealed Regulatory Proteins Inhibiting Cancer Cell Proliferation. *BMC Cancer* (2022) 22(1):1–12. doi: 10.1186/s12885-022-09218-8

23. Ma S, Pradeep S, Villar-Prados A, Wen Y, Bayraktar E, Mangala LS, et al. GnRH-R–Targeted Lytic Peptide Sensitizes BRCA Wild-Type Ovarian Cancer to PARP Inhibition. *Mol Cancer Ther* (2019) 18(5):969–79. doi: 10.1158/1535-7163.MCT-18-0770

24. Grisaru-Granovsky S, Salah Z, Maoz M, Pruss D, Beller U, Bar-Shavit R. Differential Expression of Protease Activated Receptor 1 (Par1) and Py397fak in Benign and Malignant Human Ovarian Tissue Samples. *Int J Cancer* (2005) 113(3):372–8. doi: 10.1002/ijc.20607

25. Wang F-q, Fisher J, Fishman DA. MMP-1-PAR1 Axis Mediates LPA-Induced Epithelial Ovarian Cancer (EOC) Invasion. *Gynecolog Oncol* (2011) 120 (2):247–55. doi: 10.1016/j.ygyno.2010.10.032

26. Xu Y, Xu Y, Wang C, Xia B, Mu Q, Luan S, et al. Mining TCGA Database for Gene Expression in Ovarian Serous Cystadenocarcinoma Microenvironment. *PeerJ* (2021) 9:e11375. doi: 10.7717/peerj.11375

27. Hu J, Wang L, Chen J, Gao H, Zhao W, Huang Y, et al. The Circular RNA Circ-ITCH Suppresses Ovarian Carcinoma Progression Through Targeting miR-145/RASA1 Signaling. *Biochem Biophys Res Commun* (2018) 505 (1):222–8. doi: 10.1016/j.bbrc.2018.09.060

28. Herrera B, van Dinther M, Ten Dijke P, Inman GJ. Autocrine Bone Morphogenetic Protein-9 Signals Through Activin Receptor-Like Kinase-2/ Smad1/Smad4 to Promote Ovarian Cancer Cell Proliferation. *Cancer Res* (2009) 69(24):9254–62. doi: 10.1158/0008-5472.CAN-09-2912

29. Edson MA, Nalam RL, Clementi C, Franco HL, DeMayo FJ, Lyons KM, et al. Granulosa Cell-Expressed BMPR1A and BMPR1B Have Unique Functions in Regulating Fertility But Act Redundantly to Suppress Ovarian Tumor Development. *Mol Endocrinol* (2010) 24(6):1251–66. doi: 10.1210/me.2009-0461

30. Yang Z, Bach LA. Differential Effects of Insulin-Like Growth Factor Binding Protein-6 (IGFBP-6) on Migration of Two Ovarian Cancer Cell Lines. *Front Endocrinol* (2015) 5:231. doi: 10.3389/fendo.2014.00231

31. Park Y-A, Choi CH, Do I-G, Song SY, Lee JK, Cho YJ, et al. Dual Targeting of Angiotensin Receptors (AGTR1 and AGTR2) in Epithelial Ovarian Carcinoma. *Gynecolog Oncol* (2014) 135(1):108–17. doi: 10.1016/j.ygyno.2014.06.031

32. Lawrenson K, Grun B, Lee N, Mhawech-Fauceglia P, Kan J, Swenson S, et al. NPPB Is a Novel Candidate Biomarker Expressed by Cancer-Associated Fibroblasts in Epithelial Ovarian Cancer. *Int J Cancer* (2015) 136(6):1390–401. doi: 10.1002/ijc.29092