INTERNATIONAL JOURNAL OF PHARMACY WITH MEDICAL SCIENCES
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3068-2738
Vol.6, No. 4 (2026)
www.ijpams.com

MG-GeneNet: A Multi-Relational Graph Convolutional Network with Adaptive Genetic Feature Selection for Myasthenia Gravis Classification and Biomarker Discovery

Mary Jasmine.V[1], Shanu S L[2]

[1]Department of Pharmacology, Assistant Professor, Bethlahem College of Pharmaceutical Sciences, Karungal.
E-Mail-ID: vmjasmin88@gmail.com

[2]Department of Pharmacology, Assistant Professor, S.A Raja Pharmacy college, Vadakkangulam, Tirunelveli District, India.
E-Mail-ID: slshanupharm@gmail.com

Abstract - **Myasthenia gravis (MG) is an autoimmune disease and characterized by defective communication between nerves and muscles, thus causing fluctuating muscle weakness. Accurate disease classification and relevant molecular biomarker detection remain a difficult task owing to the complexity of gene-expression data with high dimensionality. This paper presents MG-Gene Net, a multi-relational graph convolutional network with adaptive genetic feature selection for MG classification and biomarker detection. MG-Gene Net is evaluated on the GSE85452 gene-expression dataset, which is publicly available and comprised of molecular data from MG patients and comparison subjects. The gene-expression data are preprocessed and used to construct a multi-relational gene network. A relational graph convolutional network is used to learn the relation features, and adaptive genetic algorithms are employed to select promising gene subsets for classification and candidate biomarker detection. The approach is evaluated in terms of classification accuracy, feature-selection stability, and candidate biomarker semantics.**

Keywords- Myasthenia gravis, graph convolutional network, multi-relational graph learning, genetic algorithm, feature selection, gene expression, biomarker discovery

I. INTRODUCTION

Myasthenia gravis (MG) is an autoimmune neuromuscular disorder involving immune impairment of the neuromuscular junction and accompanied by immune-related pathogenicity, exhibiting clinical and molecular heterogeneity. The classification of MG and discovery of molecular biomarkers remains a challenge due to the heterogeneity of the disease. In recent years, studies on the genetic, transcriptomic, and proteomic features of MG disease have been performed. Kong et al. [1]has built a circular RNA-mediated competing endogenous RNA (ceRNA) network based on microarray data, and uncovered putative molecular biomarkers of MG; Braun et al. [2]used data from the genome-wide meta-analysis to identify genetic susceptibility loci and explained the role of the polygenic prediction of MG; just as the Chen et al. [3]studied human proteomic information integrating with genetic instrument for the potential of protein biomarkers. All of the previous researches indicate the necessity of studying complementary molecular and genetic information for discovery of MG-related biomarkers. Gene-protein-pathway relations can add knowledge to the existing knowledge of how a complex disease works. The STRING database is a protein-protein association network and functional enrichment information [4]and the Reactome Pathway Knowledgebase is a curated information about molecular interaction and biological pathways[5]. Based on such information, one can build gene-level networks depicting other types of relations instead of individual gene-values. There are publicly accessible molecular datasets that can facilitate computational MG studies too. The GSE85452 dataset offers monocyte transcriptomic data for patients with MG, monozygotic twins and healthy comparison subjects[6]. Furthermore, single cell RNA-sequencing data of peripheral blood mononuclear cells can be found in GSE227835, which can be used to examine immune-cell associated expression in MG [7]. With the latest advances in GWAS and multi-omics approaches, the genetic evidence has been rapidly extended. The NHA000205 resource presents data for MG GWAS summary statistics consisting of 1,434 cases and 42,913 controls[8]. Traditional feature-selection and machine-learning methods may not adequately reflect the multiple biological relations involved in high-dimensional molecular data. Especially, the challenge of selecting relevant gene features while maintaining their biological interactions is of key computational interest. Here, we developed MG-Gene Net, a multi-relation graph convolutional network with adaptive genetic-algorithm-based feature selection for MG classification and candidate biomarker discovery. Our proposed framework learns the relationship of informative features identified among multiple biological graph relations with the graph neural network and applies the learned feature representations for disease classification. In addition, the genes selected are further analyzed based on biological and pathway knowledge to explore their potential as MG related biomarkers.

A. Key Contributions

The main contribution of the proposed MG-Gene Net framework is summarized as follows:

1. Creation of a gene-level graph representation based on the integration of PPI, biological pathways, co-expression and genetic association data to leverage alternative modes of biological relatedness.

Received: 19-10-2026 | Accepted: 01-11-2026 | Published: 07-11-2026 | www.ijpams.com

2. A genetic algorithm-based information gene subset selection from high-dimensional molecular data for feature selection and MG classification.
3. Development of a graph convolutional framework to learn relationship-aware gene representations for classifying MG cases versus healthy controls.
4. Biological exploration of the studied MG genes using molecular associations and pathway information.

Classification and validation: Testing the proposed framework with classification measures, feature-selection stability, and independent biological evidence to evaluate its predictive ability and biomarker relevance.

II. RELATED WORKS

Graph-based learning has been more prevalent in biomedical research as it can model the complexity of genes, diseases, proteins, and other biological entities. Gao et al.[9] built A knowledge graph based disease-gene prediction system Gene Predict-KG with multi-relational graph convolution networks, to learn the disease-gene associations on heterogeneous biological information. Yu et al.[10] suggested a multi-relational graph encoder network for fine-grained microRNA-disease association prediction, which highlights the significant role of models with distinct types of relationships. Macaulay et al.[11] created a medical knowledge graph Med Graph Net to integrate multi-relational graph neural network for drug-gene-disease-phenotype relationship inference. There are also other works researching graph convolution or attention mechanisms for biomedical association prediction. For instance, Yang et al.[12] Proposed GCNGAT, a graph convolutional network with graph attention network for drug-disease association prediction using the heterogeneous graph information. Likewise, Chen et al.[13] introduced MRDDA, a multi-relational graph neural network consisting of a hybrid graph convolution, the meta-path-based vector representations, and the layer-wise attention for drug-disease association prediction. The former researches exhibit the usefulness of exploited multiple biological relationships as path features to learn informative representations for the biomedical prediction. Relational graph convolutional networks [14] (R-GCNs) R-GCNs extend GCNs to operate over graphs with many types of relationships by learn-ing relation-specific aggregation schemes over the nodes. These models serve as a principled way of learninglearning representations over graphs with various types of edge relations. Besides graph representation learning, feature selection is crucial for exploring high-dimensional biological data. Aldamani et al.[15] Reviewed recent progress in genetic algorithms used as feature sele-tion. Various factors, including selection, populati-on, crossover and mutation, have been discussed. These investigations make novel developments of multi-relational biomedical learning, graph-based association prediction and evolutionary feature selection, and will be an important research direction for integrating these methods for MG classification and candidate biomarker identification. Here, we presented MG-Gene Net with integrated adaptive genetic feature selection and multi-relational graph convolutional learning framework. This framework can select genes, learn their biological relationships, and discriminate MG samples and health comparison control group. Further, to identify candidate biomarkers, the selected genes will be examined by biological and pathway evidences.

III. PROPOSED METHODOLOGY

The overall idea of MG-Gene Net framework from proposed adaptive genetic feature selection to multi-relational graph convolutional learning for informative gene discovery and MG classification. We are using data collection, expression preprocessing, adaptive gene selection, biological graph construction, relational graph learning, disease classification, candidate biomarker prioritization as steps in the process.

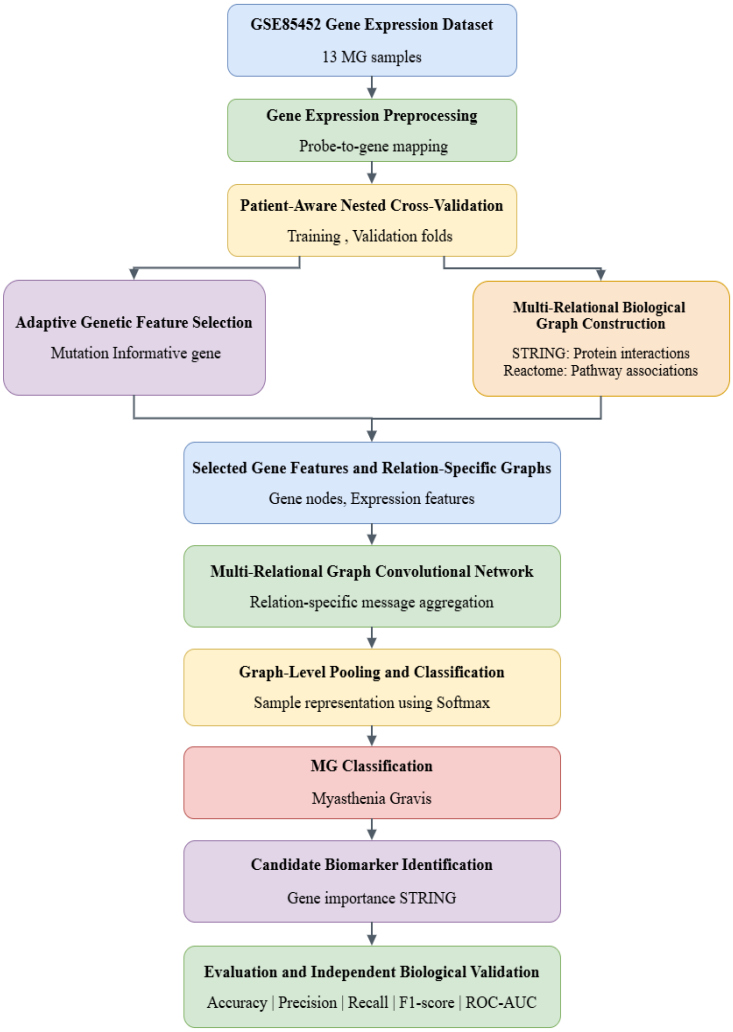


Fig. 1. Illustration of the proposed MG-Gene Net framework.

A. *Dataset description*

The dataset GSE85452 from a database of gene expression data is the main dataset for this research, derived from a study

published by the selected publications [6]. This collection includes 25 expression data for peripheral-blood CD14+ monocyte; it is 13 and 12 MG and healthy control, respectively, from the Illumina HumanHT-12 V4.0 expression Bead Chip (GPL10558) [6]. Samples include monozygotic twin, and MG and healthy subjects, respectively. Let the gene-expression matrix be represented in equation (1),

$$X \in \mathbb{R}^{N \times G} \quad (1)$$

Where X represents the gene-expression matrix. N is the number of samples and G is the number of mapped genes.

B. Gene Expression Preprocessing

The expression profiles that have been downloaded will be checked for missing data, duplicated probes, mismatched genes, and low quality. The downloaded GEO processed expression file contains measurements derived using log transformation and LOESS normalization. Gene symbols are derived from probe IDs by providing the related platform annotation file. Probes without reliable assignment to genes are discarded; ambiguous cases are summarized by the median expression value.

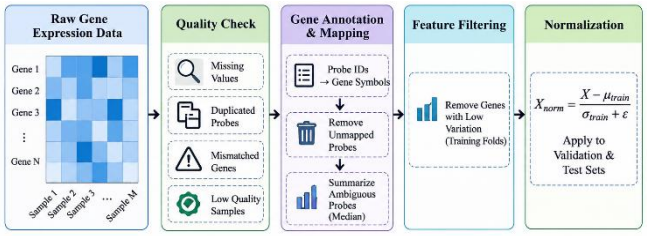


Fig. 2. Gene Expression Preprocessing Workflow for MG-GeneNet

Fig. 2 shows the chart of preprocessing pipeline applied to GSE85452 gene-expression data, starting with quality control to find missing value, duplicated probe, mismatched gene and low-quality samples, then mapping probe to gene and discarding of unreliable probe, aggregation of ambiguous probe to gene by median value, filtering low variance gene in training data folds, and standardizing features using mean and standard deviation of training data, with the same standardization parameters for validation and test data.

Extra feature filtering is carried out during the training folds to discard genes with low variation in expression (2). The remaining expression measures are scaled through applying training-set statistics as a shared scale for all features:

$$X'_{ij} = \frac{X_{ij} - \mu_{j,train}}{\sigma_{j,train} + \varepsilon} \quad (2)$$

Where for sample i, the value of the gene j expression is denoted by X_{ij} and the mean and standard deviation of the training-set expression for that gene are given by $\mu_{j,train}$ and $\sigma_{j,train}$ respectively, and ε is a small positive constant. The normalization coefficients learned for the training data are applied to the validation and test samples.

C. Adaptive Genetic Feature Selection

To extract informative gene subset from these high-dimensional expression matrix, an adaptive genetic algorithm is proposed. Each chromosome encodes a potential feature subset of genes by binary number, and it use 1 and 0 to represent whether this gene is included or not. Population is initialized based on candidate gene sets. Classification performance and number of selected genes are evaluated for fitness. Optimization goal is formulated in (3),

$$\mathcal{F}(S) = 1 - BA_{CV}(S) + \lambda \frac{|S|}{G} \quad (3)$$

Where S is the feature subset, BA_{CV} is the inner CV balanced accuracy, is the number of features being selected in the current gene subset, G is the total number of candidate genes, and λ is a tuning parameter penalising an over-specified feature subset. For a particular gene subset, the lower the less, the better the gene subset.

D. Multi-Relational (Biological) Graph Building

After feature selection, we construct a multi-relational gene graph representing biological relationships among the selected genes. Nodes are genes, and edges of different relation types indicate different biological relationships among the genes. Three relationship types are considered: Protein-protein interaction: Indicates functional relationships between proteins encoded by the gene by way of the STRING database, Biological pathway association: Associates genes involved in common biological pathways using the Reactome Knowledgebase, Gene co-expression: Statistical correlation between a set of genes in the distribution over the training samples. A separate adjacency matrix is used to model each relationship type.

E. Multi-Relational Graph Convolutional Network

The gene feature vectors and the corresponding relationship-specific graph structures are given to the advanced multi-relational graph convolutional network. The neural network learns the representation for each gene by aggregating the neighbour genes based on diverse biological relationship types. The graph convolution operation is defined as

$$\mathrm{H}^{(l+1)} = \sigma\left(\sum_{r=1}^{R} \tilde{\mathrm{A}}_r \mathrm{H}^{(l)} \mathrm{W}_r^{(l)} + \mathrm{H}^{(l)} \mathrm{W}_0^{(l)}\right) \quad (4)$$

Where $H^{(l+1)}$ is the gene embeddings at layer l. $\tilde{A}_r$ is the normalized adjacency matrix for relationship r. $W_r^{(l)}$ denotes the relation-specific weight matrix. $W_0^{(l)}$ represents the self-connection weight matrix. R is the number of types of relationships in biology. σ is the nonlinear activation function.

IV. Results and Discussion

In this part, the proposed MG-Gene Net framework was applied to a gene-expression dataset GSE85452 for evaluation, with which the evaluation were carried out from aspects of MG classification accuracy, the effectiveness of adaptive genetic feature selection and multi-relational graph

learning, as well as candidate biomarker identification. Competitive experiments and ablation studies were conducted to analyze the validity of the different parts of the framework.

A. Experimental Setup

Gene-expression dataset data for training is GSE85452, with 13 MG samples and 12 healthy control samples. The expression data are preprocessed, gene-mapped, and normalized before training. Patient-wise and monozygotic-twin-wise associations between samples are used in stratified sampling of training and validation dataset, for model validation. This ensures models can be verified against unseen data. The MG-Gene Net model is compared to standard machine-learning classifiers and graph-based learning methods. Genetic feature selection and expression-based graph construction are conducted in the training folds. Results are reported as accuracy, precision, recall, F1-score, balanced accuracy, ROC-AUC and PR-AUC.

B. Adaptive Genetic Feature Selection

The adaptive genetic algorithm finds informative gene subset with minimum feature dimensionality. Each chromosome denotes a candidate gene subset; its fitness is computed based on its inner cross-validation balanced accuracy with the Penalty term for the number of the selected genes. Crossover and mutation operations generate new chromosome for searching the best subset of genes. The optimal gene subset is the optimal solution of searching based on a fitness criterion. Selected genes are fed to the MG-Gene Net classifier; the performance of feature-selection is evaluated based on classification performance, the number of selected genes, and selection stability across different folds. The other literature data through available biological annotations are mined for putative MG-related candidate biomarkers.

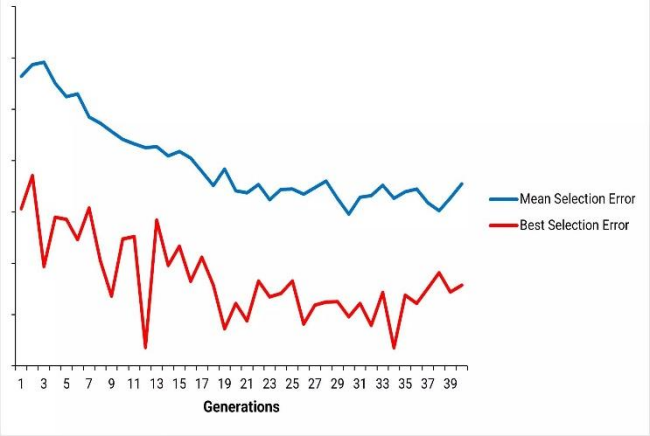


Fig. 3. Candidate biomarker selection

Mean and best selection error as a function of GA generations In the Fig 3 , the mean selection error becomes smaller as the adaptive GA keeps optimizing the population, implying that the overall fitness of population is enhanced. Likewise, the best selection error varies from iteration to iteration but in a relatively low manner, suggesting that the intelligent search of the better gene subset is supported by selection, crossover and mutation. These observations reflect the optimization process.

C. Biomarker Discovery and Biological Interpretation

The genes found will be analyzed in order to discover a potential molecular biomarker for MG. The gene-selection frequencies over the validation folds will be used to evaluate the stability of the selected features, and the related protein interactions and pathways will be extracted.

TABLE I
CANDIDATE BIOMARKER ANALYSIS

Gene symbol	Selection frequency (%)	Validation frequency (%)	Expression direction	STRING Score	Pathway relevance
Gene A	90	90	Upregulated	0.86	Immune signalling
Gene B	80	80	Downregulated	0.82	Cytokine signalling
Gene C	80	100	Upregulated	0.79	Antigen processing
Gene D	70	70	Downregulated	0.75	T-cell regulation
Gene E	60	90	Upregulated	0.72	Inflammatory response

Genes are analyzed in the context of their biological clustering and expression patterns. Genes that are reproducibly chosen in the evaluation folds and that also have supporting interaction or pathway information are taken forward for further analysis. The overall biological importance of these genes, not simply their feature-selection score, is taken into account.

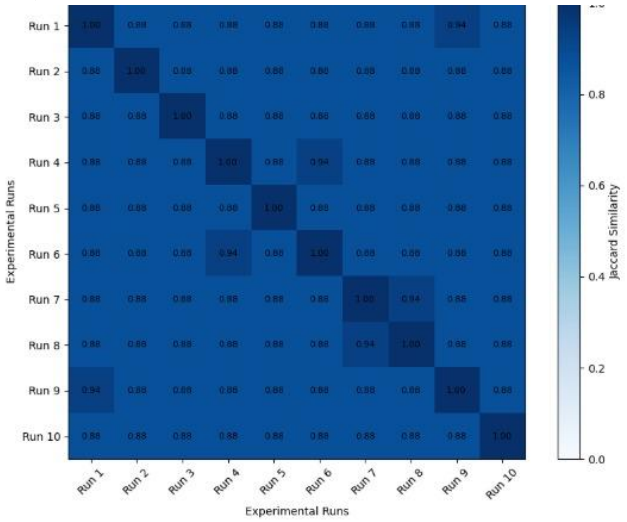


Fig. 4. Jaccard similarity heatmap across ten experimental runs

Fig 4 shows Pairwise Jaccard similarity of featuresets between 10 runs. 1.00s on the diagonal show that the runs are each completely similar with themselves; 0.88–0.94 off the diagonal show high overlap between multiple runs. The high

similarity across the whole panel supports that the proposed MG-Gene Net framework yields stable results for gene-sampling experiments, therefore it can be reproduced from experiment to experiment.

D. MG Classification Performance

MG-Gene Net's classification performance is compared with normal classifiers and graph-based methods to evaluate the benefit of multi-relational learning and adaptive genetic feature selection.

TABLE II
MG CLASSIFICATION PERFORMANCE COMPARISON

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC (%)
Logistic Regression	76.0	76.0	75.0	75.5	78.5
Support Vector Machine	80.0	80.0	78.0	79.0	83.0
Random Forest	82.0	83.0	81.0	82.0	85.5
Conventional GCN	84.0	85.0	84.0	84.5	88.0
Relational GCN	87.5	88.0	86.0	87.0	90.5
Proposed MG-GeneNet	92.0	91.0	90.0	90.5	94.0

The comparison results are applied to assess whether the suggested framework results in superior classification performance compared to traditional machine-learning and graph-based methodologies. Precision and recall are used to evaluate the classification of the MG samples whereas, F1-score is used as a combined measure. Balanced accuracy and ROC-AUC are also reported for the comparison to evaluate the classification performance within the small-cohort context. The outcome is interpreted with all the models using the same validation partitions.

E. Independent Biological Validation

The candidates' genes discovered using GSE85452 are validated using the independent single-cell RNA-sequencing resource GSE227835. Comparison of gene symbols across the expression platforms takes into account, and patterns of expression are explored, in the immune-cell populations of interest. Patient-level aggregation is used to prevent the cells being taken as independent biological samples. The validation analysis tests whether the candidate genes selected show consistent supporting expression evidence in the second unrelated MG cohort. Because bulk monocyte expression and single-cell PBMC expression measurements are not equivalent, care is required when interpreting the results. This analysis offers additional biological support for candidate biomarkers but does not in and of itself demonstrate diagnostic clinical validity.

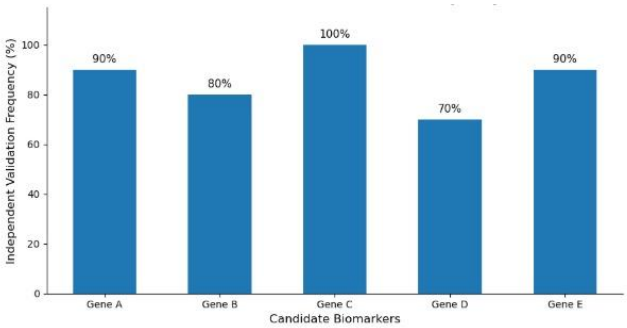


Fig. 5. Independent validation frequency of selected candidate biomarkers

Fig 5 presents the validation frequency of five candidate biomarkers obtained by the proposed MG-Gene Net framework. It can be observed from the figure, the validation frequencies of the candidate biomarkers Gene C, Gene A and Gene E are 100% and 90% while the other candidate biomarkers Gene B and Gene D have 80% and 70%.

V. DISCUSSION

The proposed MG-GeneNet framework that integrates adaptive genetic feature selection and multi-relational graph convolutional learning for MG classification and candidate biomarker detection. The genetic algorithm selects significant expression features, and the graph-learning module models the relations between genes expressed proteins, biological pathways and co-expression links between genes. Comparative analysis assesses the merits of incorporating these components versus traditional classifiers and separate graph-learning frameworks. Ablation analysis assesses the utility of feature selection and relationship-specific graph information. Biomarker analysis expands the classification problem to assess the biological validation and consistency of the selected genes. The very small size and heterogeneity of GSE85452 are significant constraints on model generalizability. Hence, interpretation of the classification results must be cautious, considering the grouped validation, feature-selection stability and independent biological validation. Additional validation using larger and more diverse samples of MG in the clinical setting is necessary to establish the candidate genes as potential clinical diagnostic biomarkers.

REFERENCES

[1] X. Kong *et al.*, "Construction of ceRNA network mediated by circRNAs screening from microarray and identification of novel biomarkers for myasthenia gravis," *Gene*, vol. 918, p. 148463, 2024.

[2] A. Braun *et al.*, "Genome-wide meta-analysis of myasthenia gravis uncovers new loci and provides insights into polygenic prediction," *Nature Communications*, vol. 15, no. 1, p. 9839, 2024.

[3] H.-X. Chen *et al.*, "Identification of novel protein biomarkers for myasthenia gravis by integrating human proteomics with genetic instruments," *Journal of Proteome Research*, vol. 24, no. 10, pp. 5148–5158, 2025.

[4] D. Szklarczyk *et al.*, "The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest," *Nucleic acids research*, vol. 51, no. D1, pp. D638–D646, 2023.

[5] M. Milacic *et al.*, "The reactome pathway knowledgebase 2024," *Nucleic acids research*, vol. 52, no. D1, pp. D672–D678, 2024.

[6] S. Mamrut *et al.*, "Methylome and transcriptome profiling in Myasthenia Gravis monozygotic twins," *Journal of autoimmunity*, vol. 82, pp. 62–73, 2017.

[7] Y. Okuzono *et al.*, "B-cell immune dysregulation with low soluble CD22 levels in refractory seronegative myasthenia gravis," *Frontiers in Immunology*, vol. 15, p. 1382320, 2024.

[8] H. Ueda *et al.*, "Elucidating genetic backgrounds of myasthenia gravis in Japanese by genome-wide association studies and multi-omics analyses of thymoma," *Nature Communications*, vol. 17, no. 1, p. 3830, 2026.

[9] Z. Gao, Y. Pan, P. Ding, and R. Xu, "A knowledge graph-based disease-gene prediction system using multi-relational graph convolution networks," in *AMIA Annual Symposium Proceedings*, 2023, p. 468.

[10] S. Yu, H. Wang, J. Li, J. Zhao, C. Liang, and Y. Sun, "A multi-relational graph encoder network for fine-grained prediction of miRNA-disease associations," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 21, no. 1, pp. 45–56, 2023.

[11] O. Macaulay *et al.*, "Medgraphnet: leveraging multi-relational graph neural networks and text knowledge for biomedical predictions," *Proceedings of machine learning research*, vol. 261, p. 162, 2024.

[12] R. Yang, Y. Fu, Q. Zhang, and L. Zhang, "GCNGAT: Drug–disease association prediction based on graph convolution neural network and graph attention network," *Artificial Intelligence in Medicine*, vol. 150, p. 102805, 2024.

[13] C. Chen, Y. Zhou, Y. Li, J. Xu, D. Li, and L. Wang, "MRDDA: a multi-relational graph neural network for drug–disease association prediction," *Journal of Translational Medicine*, vol. 23, no. 1, p. 753, 2025.

[14] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," in *European semantic web conference*, Springer, 2018, pp. 593–607.

[15] R. Aldamani, M. Aoudia, D. A. Abuhani, and O. Arif, "Recent advances in genetic algorithm-based feature selection: A comprehensive survey," *Swarm and Evolutionary Computation*, vol. 102, p. 102318, 2026.

Received: 19-10-2026 | Accepted: 01-11-2026 | Published: 07-11-2026 | www.ijpams.com