World Insight Junction (WIJ)
Journal of Artificial Intelligence Applications

REVIEW PAPER

Machine Learning Methods for Gene Regulatory Network Reconstruction: A Short Review

Ebad Sirgy[1*]

[1]Applied Science Private University (ASU), Amman, Jordan.

*Corresponding author:
ebad12@gmail.com

Article History:
Received: 05.02.2026
Revised: 12.02.2026
Accepted: 22.02.2026
Published Online: 27.02.2026

Abstract

Gene regulatory networks (GRNs) capture how transcription factors and other regulators control gene expression, and their reconstruction from expression data remains a central problem in computational biology. This paper condenses a broader survey of machine-learning-based GRN inference methods published between 2019 and 2023 into a compact starting-point review. We summarize the main data types used for GRN inference, outline the dominant methodological families—network inference, predictive modeling, gene expression analysis, network topology analysis, and multi-omics integration—and briefly compare representative methods on standard benchmark metrics (AUROC, AUPR). We close with a discussion of open challenges, including nonlinearity, scalability, interpretability, and dynamic network inference.

Keywords: Gene regulatory networks, machine learning, network inference, multi-omics integration, single-cell RNA-seq

1. Introduction

Cells do not respond to their environment as isolated units executing fixed programs; they do so through webs of gene–gene interaction that continuously reshape which genes are switched on and off. These interactions underlie some of the most consequential aspects of biology, from susceptibility to disease and the orderly progression of cellular differentiation to the moment-to-moment response cells mount against environmental stimuli [1]. A gene regulatory network (GRN) is the formal abstraction typically used to capture this behavior: it describes how transcription factors bind regulatory regions of DNA to activate or repress the expression of target genes, and how the resulting web of activation and repression closes into feedback loops that allow a cell to sense and respond to both internal state and external signals [2]. Understood this way, a GRN is less a static wiring diagram than a dynamical system, and reconstructing it from data means recovering not just which genes interact but the logic that governs how those interactions play out over time.

For much of the history of molecular biology, this kind of reconstruction was constrained less by theory than by data availability. Mapping regulatory relationships required painstaking, gene-by-gene experimental work, and any network assembled this way necessarily remained partial. That constraint has loosened considerably over the past two decades. Advances in high-throughput sequencing have made large-scale gene expression measurement increasingly affordable and increasingly routine [3], to the point where expression profiles spanning thousands of genes across many samples are now a standard rather than an exceptional resource. This shift in data availability has, in turn, motivated a corresponding shift in method: rather than building networks primarily from targeted experiments, researchers increasingly rely on computational methods that infer regulatory structure directly from expression data itself, treating the expression matrix as the principal evidence from which regulatory relationships must be recovered.

It is within this computational turn that machine learning has come to occupy an increasingly central role. Where early inference methods leaned heavily on correlation and information-theoretic measures, more recent work draws on a much broader methodological toolkit — from ensemble tree methods to deep neural architectures — to model regulatory relationships that are often nonlinear, high-dimensional, and only partially observed. This growth in method diversity has been matched by a corresponding growth in the literature describing it, to the point where keeping track of the field's overall shape has become a task in its own right. This review responds to that difficulty. It focuses specifically

on machine-learning (ML) approaches to GRN reconstruction published between 2019 and 2023, and it does so by condensing a broader, more exhaustive survey into a shorter reference document intended as a starting point for readers who need an orientation to the field rather than an exhaustive account of it. In the sections that follow, we organize this literature around the data it draws on, the methodological families it falls into, and the benchmarks used to evaluate it, before closing with a discussion of the challenges that remain unresolved.

2. Modeling Approaches

GRN models are commonly grouped into topology, control-logic, and dynamic model classes [4]. Topological models represent genes as nodes and regulatory relationships as edges in a graph, useful for protein–protein interaction and coexpression networks. Control-logic models capture dependency structure without quantifying regulatory strength, while dynamic models—the earliest and most traditional approach—aim to reproduce the temporal behavior of gene expression under environmental perturbation. Beyond these, GRNs have been modeled through complex-network, biological-interaction, chemodynamic, and molecular-simulation formulations [5, 6], with machine learning methods (e.g., random forests, neural networks) increasingly used to forecast regulatory behavior directly from data. Figure 1 summarizes the five methodological families discussed in this review and how they relate to one another.

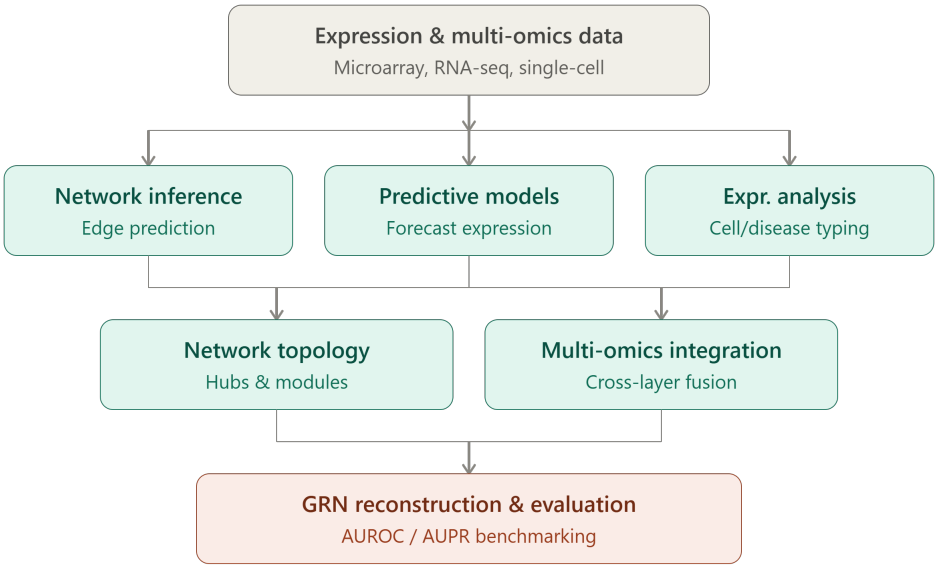

Figure 1. Taxonomy of machine-learning approaches to gene regulatory network (GRN).

3. Datasets

GRN inference relies on several data modalities (summarized in Table 1): microarray repositories such as GEO and ArrayExpress; RNA-seq resources such as TCGA, GTEx [7], and NCBI SRA [3]; and auxiliary resources such as dbGaP, BioSamples, and the Expression Atlas [8]. Time-series data from the DREAM Challenges [9] support dynamic GRN inference, while perturbation experiments (gene knockouts, drug treatments) provide causal information [10]. Multi-omics datasets integrating expression, protein interaction, genetic variation, and epigenetic data offer a more complete regulatory picture, and single-cell RNA-seq reveals cell-type-specific regulatory patterns.

Table 1. Representative GRN dataset types.

Type	Examples
Microarray	GEO, ArrayExpress
RNA-seq	TCGA, GTEx [7], NCBI SRA
Single-cell	scRNA-seq, scATAC-seq
Benchmark/time-series	DREAM Challenges [9]
Auxiliary	dbGaP, BioSamples, Expression Atlas [8]

4. Network Inference

Early GRN inference relied on Boolean networks, later supplemented by information-theoretic measures such as mutual information, the relevance network, CLR, and ARACNE. Link-prediction techniques [11] extend inference to missing or novel interactions using network topology, with applications spanning regulatory relations [12], protein–protein interactions [13], lncRNA–protein interactions [14], drug–target interactions [15], drug–drug interactions [16], gene–disease associations [17], and miRNA–disease associations [18].

Representative recent methods include D3GRN, which reformulates GRN inference as per-gene feature selection using functional decomposition and bootstrapped confidence scoring [19]; GNE, a deep-learning framework that jointly embeds gene expression and interaction-network topology [20]; RWRNET, which combines random-walk-with-restart topology encoding with Markov blanket discovery [21]; a nonlinear ODE/XGBoost framework for reconstructing dynamics from time-series and steady-state data [22]; DCNTC, which replaces mutual information with distance correlation and network centrality [23, 24]; GNNLink/GENELink, a graph neural network (GCN) approach for scRNA-seq-based inference [25]; CMIA, which combines a novel inference algorithm with the KSG mutual-information estimator [26]; GRINCD, a causal-discovery framework using GraphSAGE and additive noise models [27]; and scMTNI, which integrates scRNA-seq and scATAC-seq in a multi-task framework respecting cell lineage [28].

5. Predictive Modeling

Predictive modeling studies aim to build statistical or mathematical models that forecast gene expression or regulatory behavior. GC-MERGE integrates histone modifications and long-range spatial (Hi-C) interactions via graph convolutional networks with GNNExplainer-based interpretability [29]. The factor-graph network (FGN) framework combines probabilistic graphical modeling with loopy belief propagation to relate observed expression to regulatory structure [30]. DeepSEM generalizes structural equation modeling with a Beta-VAE architecture and an explicit GRN layer for scRNA-seq [31]. Other predictive frameworks target specific applications, such as gastric-cancer drug-response prediction [32], agent-based gene simulation with metaheuristically selected neural architectures [33], gene-interaction-graph-informed prediction via Neighbor Connection Neural Networks [34], and heterogeneous-graph gene–disease prioritization (HetIG-PreDiG/GDPS) [35].

6. Gene Expression Analysis

A distinct strand of the literature treats gene expression less as a substrate for reconstructing regulatory edges and more as a diagnostic signal in its own right. Rather than asking which gene regulates which, studies in this category ask what a given expression profile reveals about the state of a cell or tissue — typically by exploiting patterns of gene interaction to flag disease-associated genes or to classify samples into clinically meaningful categories [36]. This shift in emphasis, from network topology to sample-level prediction, brings the field closer to conventional supervised learning, and it is here that convolutional architectures have found some of their clearest applications in transcriptomics.

Mostavi et al. illustrate this well: by arranging expression values into one- and two-dimensional representations, they show that CNN architectures — spanning a straightforward 1D layout, a 2D-Vanilla arrangement, and a 2D-Hybrid variant that incorporates auxiliary structure — can distinguish cancer types and subtypes directly from expression data [37]. The appeal of this approach lies less in any single architectural choice than in the demonstration that image-style convolutional reasoning transfers reasonably well to tabular expression data once it is reshaped appropriately, a design decision that later work in the field continues to revisit.

Other methods stay closer to the network-based tradition while retaining a disease-classification objective. The Co-hub Differential Network (CDN) approach, for instance, does not simply build a static interaction network; it compares networks across conditions to isolate hub genes whose connectivity changes between disease states, applying this differential logic to contexts as varied as colorectal cancer and COVID-19 [38]. The underlying premise — that a gene's regulatory importance is best revealed by how its interactions shift under perturbation, rather than by its interactions in isolation — distinguishes this line of work from purely correlation-based hub detection. JDINAC pursues a related but genetically grounded strategy, constructing gene interaction networks anchored in SNP data to examine how inherited variation shapes breast cancer risk, thereby linking expression-level network structure back to germline genetics rather than treating expression as the sole source of signal [39].

Spatial and tissue-level context introduces a further layer of complexity that purely expression-based methods cannot capture on their own. CLARIFY addresses this by jointly inferring cell–cell interactions alongside cell-specific gene regulatory networks, using a multi-level graph autoencoder to exploit the spatial coordinates that accompany spatially resolved transcriptomic data [40]. In doing so, it treats the identity and behavior of a cell as inseparable from its neighborhood, an assumption that is increasingly common as spatial technologies mature and that marks a departure from earlier methods built for dissociated, non-spatial single-cell data.

Finally, weighted gene co-expression network analysis (WGCNA) remains a durable tool in this space despite its relative methodological simplicity. Its continued use is illustrated by work identifying hub gene modules that mediate therapeutic response in psoriasis, where co-expression modules — rather than individual genes — are treated as the functional unit of interest [41]. That such a comparatively established method continues to yield clinically interpretable modules alongside newer deep-learning approaches suggests that the choice of technique in this subfield is often shaped as much by interpretability requirements and sample size as by predictive performance alone.

Taken together, these studies indicate that gene expression analysis for disease and tissue classification has diversified along at least three axes: the representation of expression data (raw profiles versus interaction networks), the granularity of the biological unit under study (individual hub genes versus co-expression modules versus cell neighborhoods), and the source of structuring information (transcriptomic, genetic, or spatial). This diversification complicates direct comparison across methods, since a CNN trained on raw expression and a differential-network method built on hub connectivity are answering related but not identical questions — a tension that recurs throughout the broader literature reviewed here.

7. Network Topology Analysis

Where the methods discussed in the preceding sections are organized primarily around a predictive or integrative objective, a further strand of the literature approaches gene regulatory networks from the opposite direction: it starts from the network itself and asks what its structure — its hubs, its modules, and the way that structure evolves — reveals about regulation. This emphasis on topology as the primary object of study, rather than as an intermediate representation on the way to some other goal, distinguishes the methods below from much of the network inference work discussed earlier, even though the two strands frequently draw on the same underlying graph formalisms.

One recurring question in this space is how best to recover network structure from data that is simultaneously high-dimensional and comparatively scarce, a combination that plagues much of genomics. Graafland and Gutiérrez address this directly, showing that Gaussian Bayesian networks fitted with a Hill Climbing search procedure outperform sparsity-based undirected alternatives such as Glasso variants when reconstructing GRNs from high-dimensional microarray data [42]. The result is notable less for the specific algorithm than for what it implies about model class: a directed, score-based Bayesian formulation appears better suited to this regime than the undirected, penalty-based models that dominate much of the sparse-network literature, suggesting that the choice between directed and undirected representations is not merely a matter of convenience but has real consequences for reconstruction accuracy.

Temporal structure introduces a further axis along which topology-focused methods diverge from their static counterparts. Dual-stage attention recurrent neural networks (DA-RNN) tackle this by combining temporal prediction with an attention mechanism applied jointly over regulators and over time points, so that the resulting model is not only predictive but also interpretable in the specific sense of indicating which regulators mattered and when [43]. This pairing of prediction with interpretability is significant precisely because temporal models are often harder to inspect than their static counterparts; by making the attention weights themselves a source of biological insight, DA-RNN partially closes the gap between predictive accuracy and mechanistic understanding that recurs as a theme throughout this review.

A separate line of work asks whether the structural regularities that make link prediction possible can be learned directly from data, rather than engineered by hand. Devi and Singh propose a link-prediction framework of this kind, one that infers connectivity from learned representations of subgraph topology instead of relying on handcrafted structural features such as common-neighbor counts or preferential attachment scores [44]. Because such a model learns its own notion of which topological patterns are predictive, it is, in principle, more generalizable across networks with different structural properties than methods tied to a fixed set of hand-designed features — though this flexibility typically comes at the cost of reduced interpretability relative to the feature-based alternatives it replaces.

Higher-order interactions among genes pose a related but distinct challenge, since pairwise measures by construction cannot capture regulatory effects that only emerge from combinations of three or more genes acting jointly. Cui et al. address this gap with a deep-learning framework that incorporates structured sparsity constraints to detect such higher-order gene–gene interactions directly from SNP data [45], extending the reach of topology-focused inference beyond the pairwise interactions that dominate much of the earlier literature discussed in this review. Finally, at the level of the underlying similarity measure itself, part mutual information (PMI) offers an alternative to Pearson correlation for constructing coexpression networks in cancer transcriptomics, one designed to isolate direct associations between genes from those that arise only indirectly, through a shared third gene [46]. This distinction between direct and indirect association is not a minor technical refinement: as later sections of this review make clear, disentangling direct from indirect regulatory relationships remains one of the more persistent open problems in the field, and PMI represents one of several attempts, alongside the higher-order methods discussed above, to address it at the level of the similarity measure rather than the network architecture.

Across these five contributions, topology-focused analysis emerges as less a single method than a set of complementary strategies for making network structure itself the bearer of biological meaning — through model class (Bayesian versus sparsity-based), through temporal attention, through learned rather than handcrafted structural features, through explicit higher-order interaction modeling, and through refined similarity measures. Each strategy targets a different limitation of naive pairwise, static network construction, and together they anticipate several of the open challenges — interpretability, higher-order structure, and the direct/indirect distinction chief among them — discussed later in this review.

8. Multi-Omics Integration

No single omics layer tells the full regulatory story on its own. Expression data reveals which genes are active but says little about why, while proteomic, epigenomic, and genomic variation data each capture a different piece of the mechanism — protein-level interactions, chromatin state, and inherited variation, respectively. Methods in this category proceed from the premise that fusing these layers, rather than analyzing them in isolation, yields a more complete and more reliable picture of regulation than expression data alone can provide. What distinguishes the approaches below is less the goal, which is broadly shared, than the mathematical machinery used to reconcile data sources that differ in scale, noise structure, and biological meaning.

Matrix factorization has proven a natural fit for this fusion problem, since it offers a principled way to project heterogeneous data types into a shared latent space. iMTF-GRN exemplifies this strategy, applying non-negative matrix tri-factorization to simultaneously decompose expression data, protein–protein interaction networks, and Gene Ontology annotations, thereby allowing information from each source to constrain and inform the others during inference rather than being combined only after the fact [47]. A related but distinct question is addressed by ActivePathways, which sets aside network reconstruction in favor of pathway-level integration: it performs enrichment analysis across multiple omics layers jointly, asking not which genes interact but which biological pathways are consistently implicated once several data types are considered together [48]. The two methods illustrate a broader division within this subfield between approaches aimed at recovering fine-grained network structure and those aimed at coarser, pathway-level integration.

Kernel-based methods offer an alternative route to combining heterogeneous data, one that sidesteps some of the assumptions matrix factorization requires about how different data types relate to one another. MKL-GRNI adopts this approach, extending multiple-kernel-learning to a parallelized framework capable of incorporating DNA methylation alongside protein interaction data, and in doing so scales kernel-based integration to problem sizes that earlier formulations struggled to accommodate [49]. Statistical approaches, meanwhile, tend to favor a more conservative and interpretable route: Hybrid-Network combines several statistical tests with network-informed priors, using established hypothesis-testing machinery rather than a learned embedding to determine which associations across omics layers are likely to be genuine [50]. This preference for statistical rigor over representation learning reflects a persistent tension in the multi-omics literature between predictive flexibility and interpretability.

Bayesian approaches address the integration problem from yet another angle, treating prior biological knowledge not as a post-hoc validation step but as an explicit input to the inference process itself. IntOMICS follows this logic, using Bayesian networks together with adaptive Markov chain Monte Carlo sampling to integrate multi-omics measurements with existing biological knowledge, allowing what is already known about a

system to shape which regulatory structures the model considers plausible [51]. Graph-based embedding methods pursue a related aim through different means: BRANEnet embeds multilayer heterogeneous omics networks by combining a random-walk-based positive pointwise mutual information (PPMI) formulation with matrix factorization, producing a unified representation that preserves structure from each omics layer within a single embedding space [52].

Finally, the rise of single-cell technologies has pushed multi-omics integration toward finer resolution still. SMGR jointly models scRNA-seq and scATAC-seq data — transcriptomic and chromatin accessibility measurements taken, ideally, from the same cells — to discover co-regulatory programs that would be invisible to either modality examined alone [53]. This pairing of expression with chromatin accessibility marks a natural extension of multi-omics integration from the bulk-tissue level, where most of the methods above originate, down to the level of individual cells, and it foreshadows a broader trend toward single-cell multi-omics that recurs elsewhere in this review.

Across these seven methods, a consistent pattern emerges: as more omics layers are brought into a single analysis, the choice of integration strategy — factorization, kernel methods, statistical testing, Bayesian priors, or graph embedding — becomes less a matter of computational convenience and more a substantive modeling decision, since each strategy encodes different assumptions about how biological layers should be allowed to inform one another.

9. Comparative Evaluation

Table 2 summarizes representative benchmark performance (AUROC/AUPR) reported across studies. Comparisons on shared benchmarks (e.g., DREAM3/4/5) show that methods incorporating nonlinear modeling or multiple data sources (e.g., GNE [20], GENELink [25], CMIA [26]) tend to outperform simpler correlation-based approaches (e.g., D3GRN [19]), though performance is highly dataset- and metric-dependent. AUROC and AUPR remain the dominant evaluation metrics, valued for their robustness to class imbalance between true and false regulatory edges.

Table 2. Selected benchmark results (illustrative subset).

Method	Dataset	AUROC / AUPR
D3GRN [19]	DREAM4/5	–/0.175–0.373
GNE [20]	Yeast, *E. coli*	0.65–0.71 / 0.66–0.68
ODE+XGBoost [22]	Yeast, *E. coli*	0.55–0.96 / 0.02–0.85
GC-MERGE [29]	GM12878, K562, HUVEC	0.90–0.92 / 0.88
GENELink [25]	scRNA-seq (hHEP)	higher than 6 baselines
CLARIFY [40]	Mouse cortex/visual	AUPRC ratio 1.48
MKL-GRNI [49]	*E. coli*+TCGA	0.90 / 0.69
iMTF-GRN [47]	*E. coli*, yeast, human	0.73–0.97 / 0.37–0.46

10. Challenges and Future Directions

Read across the sections above, the literature on machine-learning approaches to GRN reconstruction gives an impression of a field that has grown rapidly in method diversity while leaving a comparatively stable set of underlying difficulties unresolved. Rather than being specific to any one methodological family, these difficulties tend to resurface across network inference, predictive modeling, expression analysis, topology analysis, and multi-omics integration alike, which suggests they are properties of the underlying problem rather than of any particular modeling choice. Five such challenges recur with enough regularity to merit closing discussion.

The first is a statistical one, and in some sense the most fundamental: high dimensionality paired with comparatively small sample sizes continues to complicate robust inference throughout this literature [42, 46]. Gene expression datasets routinely measure thousands of genes simultaneously while including only a modest number of samples, an imbalance that leaves many candidate regulatory relationships statistically underdetermined regardless of how sophisticated the inference method applied to them may be. This tension is visible in the very studies that motivate it: Graafland and Gutiérrez turn to Bayesian model structure specifically to cope with high-dimensional microarray data, while Jiang et al. turn to part mutual information specifically to obtain sparser, more direct associations in similarly high-dimensional settings. That two methodologically distinct responses converge on the same underlying constraint is itself evidence of how persistent the problem is.

A second, closely related challenge concerns interpretability. Deep learning models have delivered some of the strongest predictive results surveyed in this review, yet their accuracy is frequently purchased at the cost of transparency: it is often unclear which features or relationships a given model is actually relying on to make its predictions. Explainability techniques have emerged as a partial response to this difficulty rather than a complete solution — GNNExplainer offers one such route by identifying the subgraphs and features most responsible for a graph neural network's predictions [29], while Shapley-based interaction scores offer a complementary, game-theoretic route to attributing predictive weight across gene–gene interactions [45]. Both approaches help narrow the gap between predictive power and mechanistic understanding, but neither eliminates it outright, and interpretability remains more a matter of degree than a problem that has been solved.

Third, even where a regulatory relationship between two genes can be detected with confidence, it does not follow that the relationship is direct. Many statistically significant associations arise only indirectly, mediated through a shared regulator or a longer causal chain, and disentangling direct from indirect regulatory relationships remains an open problem across the literature. Part mutual information has been proposed as one way of addressing this at the level of the similarity measure itself, by explicitly discounting the influence of third-party genes when estimating association strength [46]; but as a general matter, distinguishing cause from correlation in regulatory networks continues to demand more than any single measure can currently provide.

Fourth, the majority of methods surveyed in this review infer networks that are implicitly static, treating regulatory structure as fixed at the

moment the data were collected. Biological regulation, however, is not static in this sense: relationships between genes shift as cells differentiate, respond to stimuli, or progress through disease states, and a network reconstructed from a single snapshot necessarily misses this dynamism. A smaller subset of methods has begun to address this gap directly. DA-RNN incorporates temporal prediction with attention over both regulators and time points, allowing the model to indicate not just which relationships hold but when they are most active [43], while ODE-based frameworks model regulatory dynamics more explicitly still, treating gene expression trajectories as the solution to a system of differential equations rather than as a sequence of independent observations [22]. That such dynamic approaches remain the exception rather than the rule, despite the clear biological motivation for them, suggests that capturing time-varying regulatory behavior is likely to remain an active area of methodological development rather than a solved problem.

Finally, no single omics layer offers a complete account of regulation on its own, and the field continues to move toward methods that combine multiple layers with existing biological prior knowledge rather than treating each new dataset as an isolated inference problem. IntOMICS illustrates this trajectory by incorporating prior biological knowledge directly into a Bayesian inference procedure [51], while BRANEnet pursues a related goal through embedding-based fusion of heterogeneous omics networks [52]. Scaling such integrative approaches to the large, heterogeneous datasets that increasingly characterize the field — spanning multiple omics modalities, tissue types, and, increasingly, individual cells — remains a key direction for future work, one that several recent efforts have begun to take up explicitly [28].

Taken together, these five challenges — statistical robustness under high dimensionality, model interpretability, the direct/indirect distinction, dynamic versus static network representation, and multi-omics scalability — do not admit independent solutions. Progress on one frequently bears on the others: a more interpretable model, for instance, is often also better positioned to reveal whether a given relationship is direct or mediated, and a dynamic model capable of tracking regulatory change over time inevitably raises the dimensionality of the inference problem still further. It is this interdependence, more than any single unresolved question, that continues to motivate the hybrid statistical–deep-learning frameworks toward which much of the recent literature surveyed in this review appears to be converging.

11. Conclusion

This compact review distills recent (2019–2023) machine-learning approaches to GRN reconstruction across five methodological families: network inference, predictive modeling, gene expression analysis, network topology analysis, and multi-omics integration. While substantial progress has been made in accuracy and scalability, challenges around nonlinearity, interpretability, and dynamic modeling remain open, motivating continued development of hybrid statistical–deep-learning frameworks for gene regulatory network inference [28].

References

[1] Diane Gilbert-Diamond and Jason H Moore. Analysis of gene-gene interactions. *Current protocols in human genetics*, 70(1):1–14, 2011.

[2] Vân Anh Huynh-Thu and Guido Sanguinetti. Gene regulatory network inference: an introductory survey. In *Gene regulatory networks: Methods and protocols*, pages 1–23. Springer, 2018.

[3] Heba M AL-Bermany and Sura Z AL-Rashid. Microarray gene expression data for detection alzheimer's disease using k-means and deep learning. In *2021 7th International Engineering Conference "Research & Innovation amid Global Pandemic"(IEC)*, pages 13–19. IEEE, 2021.

[4] Guy Karlebach and Ron Shamir. Modelling and analysis of gene regulatory networks. *Nature reviews Molecular cell biology*, 9(10):770–780, 2008.

[5] Khalid Raza. Reconstruction, topological and gene ontology enrichment analysis of cancerous gene regulatory network modules. *Current Bioinformatics*, 11(2):243–258, 2016.

[6] Khalid Raza and Mansaf Alam. Recurrent neural network based hybrid model for reconstructing gene regulatory network. *Computational biology and chemistry*, 64:322–334, 2016.

[7] Marta Melé, Pedro G Ferreira, Ferran Reverter, David S DeLuca, Jean Monlong, Michael Sammeth, Taylor R Young, Jakob M Goldmann, Dmitri D Pervouchine, Timothy J Sullivan, et al. The human transcriptome across tissues and individuals. *Science*, 348(6235):660–665, 2015.

[8] Irene Papatheodorou, Nuno A Fonseca, Maria Keays, Y Amy Tang, Elisabet Barrera, Wojciech Bazant, Melissa Burke, Anja Füllgrabe, Alfonso Munoz-Pomer Fuentes, Nancy George, et al. Expression atlas: gene and protein expression across multiple studies and organisms. *Nucleic acids research*, 46(D1):D246–D251, 2018.

[9] Ziv Bar-Joseph. Analyzing time series gene expression data. *Bioinformatics*, 20(16):2493–2503, 2004.

[10] Frank Emmert-Streib, Matthias Dehmer, and Benjamin Haibe-Kains. Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Frontiers in cell and developmental biology*, 2:38, 2014.

[11] Linyuan Lü and Tao Zhou. Link prediction in complex networks: A survey. *Physica A: statistical mechanics and its applications*, 390(6):1150–1170, 2011.

[12] Baruch Barzel and Albert-László Barabási. Network link prediction by global silencing of indirect correlations. *Nature biotechnology*, 31(8):720–725, 2013.

[13] István A Kovács, Katja Luck, Kerstin Spirohn, Yang Wang, Carl Pollis, Sadie Schlabach, Wenting Bian, Dae-Kyum Kim, Nishka Kishore, Tong Hao, et al. Network-based prediction of protein interactions. *Nature communications*, 10(1):1240, 2019.

[14] Huan Hu, Chunyu Zhu, Haixin Ai, Li Zhang, Jian Zhao, Qi Zhao, and Hongsheng Liu. Lpi-etslp: lncrna–protein interaction prediction using eigenvalue transformation-based semi-supervised link prediction. *Molecular BioSystems*, 13(9):1781–1787, 2017.

[15] Feixiong Cheng, Chuang Liu, Jing Jiang, Weiqiang Lu, Weihua Li, Guixia Liu, Weixing Zhou, Jin Huang, and Yun Tang. Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS computational biology*, 8(5):e1002503, 2012.

[16] Wen Zhang, Yanlin Chen, Feng Liu, Fei Luo, Gang Tian, and Xiaohong Li. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. *BMC bioinformatics*, 18(1):18, 2017.

[17] Limei Lin, Tinghong Yang, Ling Fang, Jian Yang, Fan Yang, and Jing Zhao. Gene gravity-like algorithm for disease gene prediction based on phenotype-specific network. *BMC Systems Biology*, 11(1):121, 2017.

[18] Xing Chen, Zhihan Zhou, and Yan Zhao. Ellpmda: ensemble learning and link prediction for mirna-disease association prediction. *RNA biology*, 15(6):807–818, 2018.

[19] Xiang Chen, Min Li, Ruiqing Zheng, Fang-Xiang Wu, and Jianxin Wang. D3grn: a data driven dynamic network construction method to infer gene regulatory networks. *BMC genomics*, 20(Suppl 13):929, 2019.

[20] Kishan Kc, Rui Li, Feng Cui, Qi Yu, and Anne R Haake. Gne: a deep learning framework for gene network inference by aggregating biological information. *BMC systems biology*, 13(Suppl 2):38, 2019.

[21] Wei Liu, Xingen Sun, Li Peng, Lili Zhou, Hui Lin, and Yi Jiang. Rwrnet: a gene regulatory network inference algorithm using random walk with restart. *Frontiers in Genetics*, 11:591461, 2020.

[22] Baoshan Ma, Mingkun Fang, and Xiangtian Jiao. Inference of gene regulatory networks based on nonlinear ordinary differential equations. *Bioinformatics*, 36(19):4885–4893, 2020.

[23] Kuan Liu, Haiyuan Liu, Dongyan Sun, and Lei Zhang. Network inference from gene expression data with distance correlation and network topology centrality. *Algorithms*, 14(2):61, 2021.

[24] Kun Yu, Weidong Xie, Linjie Wang, Shoujia Zhang, and Wei Li. Determination of biomarkers from microarray data using graph neural network and spectral clustering. *Scientific reports*, 11(1):23828, 2021.

[25] Guo Mao, Zhengbin Pang, Ke Zuo, Qinglin Wang, Xiangdong Pei, Xinhai Chen, and Jie Liu. Predicting gene regulatory links from single-cell rna-seq data using graph neural networks. *Briefings in bioinformatics*, 24(6):bbad414, 2023.

[26] Lior I Shachaf, Elijah Roberts, Patrick Cahan, and Jie Xiao. Gene regulation network inference using k-nearest neighbor-based mutual information estimation: revisiting an old dream. *BMC bioinformatics*, 24(1):84, 2023.

[27] Ke Feng, Hongyang Jiang, Chaoyi Yin, and Huiyan Sun. Gene regulatory network inference based on causal discovery integrating with graph neural network. *Quantitative Biology*, 11(4):434–450, 2023.

[28] Shilu Zhang, Saptarshi Pyne, Stefan Pietrzak, Spencer Halberg, Sunnie Grace McCalla, Alireza Fotuhi Siahpirani, Rupa Sridharan, and Sushmita Roy. Inference of cell type-specific gene regulatory networks on cell lineages from single cell omic datasets. *Nature Communications*, 14(1):3064, 2023.

[29] Ellen V Rothenberg. Causal gene regulatory network modeling and genomics: second-generation challenges. *Journal of Computational Biology*, 26(7):703–718, 2019.

[30] Stephen Kotiang and Ali Eslami. A probabilistic graphical model for system-wide analysis of gene regulatory networks. *Bioinformatics*, 36(10):3192–3199, 2020.

[31] Hantao Shu, Jingtian Zhou, Qiuyu Lian, Han Li, Dan Zhao, Jianyang Zeng, and Jianzhu Ma. Modeling gene regulatory networks using neural network architectures. *Nature Computational Science*, 1(7):491–501, 2021.

[32] Heewon Park, Seiya Imoto, and Satoru Miyano. Predictivenetwork: predictive gene network estimation with application to gastric cancer drug response-predictive network analysis. *BMC bioinformatics*, 23(1):342, 2022.

[33] Francesco Zito, Vincenzo Cutello, and Mario Pavone. A machine learning approach to simulate gene expression and infer gene regulatory networks. *Entropy*, 25(8):1214, 2023.

[34] Xuanyu Li, Xuan Zhang, Wenduo He, Deliang Bu, and Sanguo Zhang. Gene expression prediction based on neighbour connection neural network utilizing gene interaction graphs. *PLoS One*, 18(2):e0281286, 2023.

[35] Kathleen M Jagodnik, Yael Shvili, and Alon Bartal. Hetig-predig: a heterogeneous integrated graph model for predicting human disease genes based on gene expression. *Plos one*, 18(2):e0280839, 2023.

[36] EH Hadi, HA Lafta, and SZ Alrashid. Classification of covid-19 disease using genes expression and deep learning technique. *Mathemat Stat Eng Appl*, 71(4):3721–3731, 2022.

[37] Milad Mostavi, Yu-Chiao Chiu, Yufei Huang, and Yidong Chen. Convolutional neural network models for cancer type prediction based on gene expression. *BMC medical genomics*, 13(Suppl 5):44, 2020.

[38] Jiacheng Leng and Ling-Yun Wu. Interaction-based transcriptome analysis via differential network inference. *Briefings in Bioinformatics*, 23(6):bbac466, 2022.

[39] Liyuan Liu, Wenli Zhai, Fei Wang, Lixiang Yu, Fei Zhou, Yujuan Xiang, Shuya Huang, Chao Zheng, Zhongshang Yuan, Yong He, et al. Using machine learning to identify gene interaction networks associated with breast cancer. *BMC cancer*, 22(1):1070, 2022.

[40] Mihir Bafna, Hechen Li, and Xiuwei Zhang. Clarify: cell–cell interaction and gene regulatory network refinement from spatially resolved transcriptomics. *Bioinformatics*, 39(Supplement_1):i484–i493, 2023.

[41] Xiaoxiao Xue, Yatao Guo, Qianying Zhao, Yongwen Li, Mi Rao, Wenjing Qi, and Huijuan Shi. Weighted gene co-expression network analysis of oxymatrine in psoriasis treatment. *Journal of inflammation research*, pages 845–859, 2023.

[42] Catharina E Graafland and José M Gutiérrez. Learning complex dependency structure of gene regulatory networks from high dimensional microarray data with gaussian bayesian networks. *Scientific reports*, 12(1):18704, 2022.

[43] Michele Monti, Jonathan Fiorentino, Edoardo Milanetti, Giorgio Gosti, and Gian Gaetano Tartaglia. Prediction of time series gene expression and structural analysis of gene regulatory networks using recurrent neural networks. *Entropy*, 24(2):141, 2022.

[44] Salam Jayachitra Devi and Buddha Singh. Link prediction model based on the topological feature learning for complex networks. *Arabian Journal for Science and Engineering*, 45(12):10051–10065, 2020.

[45] Tianyu Cui, Khaoula El Mekkaoui, Jaakko Reinvall, Aki S Havulinna, Pekka Marttinen, and Samuel Kaski. Gene–gene interaction detection with deep learning. *Communications Biology*, 5(1):1238, 2022.

[46] Yi-Hua Jiang, Jie Long, Zhi-Bin Zhao, Liang Li, Zhe-Xiong Lian, Zhi Liang, and Jia-Rui Wu. Gene co-expression network based on part mutual information for gene-to-gene relationship and gene-cancer correlation analysis. *BMC bioinformatics*, 23(1):194, 2022.

[47] Nisar Wani and Khalid Raza. Imtf-grn: integrative matrix tri-factorization for inference of gene regulatory networks. *IEEE Access*, 7:126154–126163, 2019.

[48] M Paczkowska, J Barenboim, N Sintupisut, NS Fox, H Zhu, D Abd-Rabbo, MW Mee, PC Boutros, F Abascal, SB Amin, et al. Integrative pathway enrichment analysis of multivariate omics data. nat commun 11 (1): 735, 2020.

[49] Nisar Wani and Khalid Raza. Mkl-grni: A parallel multiple kernel learning approach for supervised inference of large-scale gene regulatory networks. *PeerJ Computer Science*, 7:e363, 2021.

[50] Demba Fofana, EO George, and Dale Bowman. Combining assumptions and graphical network into gene expression data analysis. *Journal of Statistical Distributions and Applications*, 8(1):9, 2021.

[51] Anna Pačínková and Vlad Popovici. Using empirical biological knowledge to infer regulatory networks from multi-omics data. *BMC bioinformatics*, 23(1):351, 2022.

[52] Surabhi Jagtap, Aurélie Pirayre, Frédérique Bidard, Laurent Duval, and Fragkiskos D Malliaros. Branenet: embedding multilayer networks for omics data integration. *BMC bioinformatics*, 23(1):429, 2022.

[53] Qianqian Song, Xuewei Zhu, Lingtao Jin, Minghan Chen, Wei Zhang, and Jing Su. Smgr: a joint statistical method for integrative analysis of single-cell multi-omics data. *NAR genomics and bioinformatics*, 4(3):lqac056, 2022.