

A Hybrid Graph Attention Network Integrating Coagulation Cascade Biology Developed for Transductive Prediction of Bleeding Disorders

Lakshmi J V N^{*1}, G Harini Kumari²

¹School of Computer Science and Applications, REVA University, Bangalore, Karnataka, India.

²Department of Pharmaceutics, East point college of pharmacy, Bangalore, Karnataka, India.

***Corresponding author:** jupudilakshmi@gmail.com

Abstract

Disorders of blood coagulation arise due to the complex, nonlinear interactions among multiple clotting factors and regulatory proteins. These disorders can be caused within the intrinsic, extrinsic, and common pathways. Traditional mathematical and machine learning approaches study these disorder pathways independently though models can analyze the coagulation cascade due to the existing system level architecture. In this study, a biologically constrained, graph-based learning framework is proposed for predicting coagulation disorders. The coagulation cascade is modeled as a directed network integrating patient's individual factor activities with established biochemical interactions. A Graph Attention Network (GAT) is applied to understand the significant relation of coagulation factors and their interactions. Thereby GAT network enables interpretable disorder classification. The proposed model supports transductive interpretation of unseen patient data into the trained graph without retraining. The framework is evaluated using synthesized clinical datasets representing 6 types of blood disorders. Coagulation disorder is caused due to the deficiency of Factor VIII causing Hemophilia A, Factor IX causing hemophilia B, Factor XI causing hemophilia C, Factors IX, II, X, and VII causes vitamin K deficiency, Factor VIII causing von Willebrand disease, and normal coagulation profiles are the

types of blood disorders considered for the study. Results demonstrate that the proposed approach effectively captures pathway level dependencies and accurately distinguishes the coagulation disorder. The proposed approach predicts the type of disorder with Macro F1 – Score of 0.013 and accuracy of 0.2915. The future scope is to implement the model on real world patient records and further improve the model.

Key words: Coagulation cascade, Graph Attention Network, Graph neural networks Hemophilia, Systems biology, von Willebrand disease

Introduction

Blood coagulation is a biological process involving a cascade of enzymatic reactions among multiple clotting factors, regulatory proteins, and cellular components. Disturbances in this cascade can lead to bleeding disorders such as hemophilia A due to the deficiency of Factor VIII, hemophilia B due to the deficiency of Factor IX, hemophilia C due to the deficiency of Factor XI, vitamin K due to the deficiency of Factor II, VII, X and IX, and von Willebrand disease can be caused due to the deficiency of Factor VIII as shown in figure 1.

Each disorder can be caused due to different deficiencies as described or can be misfunction of some components in the respective pathways. Identification of these disorders accurately requires an understanding of individual

A hybrid graph attention network integrating coagulation cascade biology developed for transductive prediction of bleeding disorders

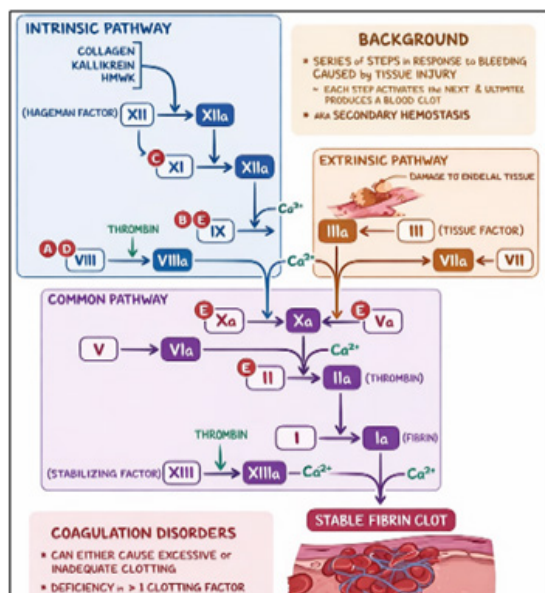


Fig 1: Blood coagulation disorder factor levels and the complex interdependencies across the intrinsic, extrinsic, and common coagulation pathways.

Traditional modeling methods for coagulation mainly depend on systems biology techniques such as ordinary differential equation (ODE) based simulations to describe clot formation dynamics. While these models provide mechanical insights, they are often parameter concentrated, computationally expensive, and limited in their ability to generalize across heterogeneous patient populations. On the other hand, conventional machine learning methods have been applied to coagulation data for classification and prediction tasks but typically treat coagulation factors as independent features. This approach is failing to incorporate the known biochemical structure of the coagulation cascade.

Recent advances in graph-based machine learning technique are capable alternative for modeling complex biological systems. Graph neural networks (GNNs) are particularly well suited for representing structured biological processes, involving nodes and edges that correspond to biological entities and interactions.

By explicitly modeling factor to factor dependencies, GNNs can capture nonlinear pathway interactions and developing behaviors that are not apparent in linear or feature based models. Though, the application of graph attention mechanisms on coagulation disorder prediction is unexplored by the researchers.

Hence, this study proposes a hybrid systems biology and graph-based learning framework for coagulation disorder prediction. A Graph Attention Network is applied to understand nonlinear interactions among coagulation factors and patient-specific measurements. The proposed model enables transductive interpretation of the new patients' factors dynamically by inputting them into the trained graph. This allows accurate classification of bleeding disorders such as hemophilia A, hemophilia B, vitamin K deficiency, and von Willebrand disease. The construction of a biologically informed graph representation of the coagulation cascade is displayed for individual cases. The demonstration of transductive patient-level inference using clinically meaningful case studies is included as a part of this study. This framework highlights the potential graph-based learning in hematological research and clinical decision support is accurate prediction of disorder.

The coagulation cascade structure has been widely studied by biochemical and clinical research due to its critical role in hemostasis and thrombosis. Standard pathway-based models define coagulation as a sequence of enzymatic reactions involving intrinsic, extrinsic, and common pathways. While these models provide mechanical insight, with their complexity and reliance on numerous kinetic parameters but ignore patient specific analysis.

Qian et al. (2024) proposed Coagulo-Net, a physics-informed neural network (PINN) framework that integrates uses ordinary differential equation (ODE). This model of blood coagulation adapts deep learning for improving parameter implication and dynamic prediction of coagulation processes. The study addresses key limitations of traditional coagulation mod-

els, by estimating unknown reaction rate parameters and handling noisy experimental data. Coagulo-Net compares two coagulation models of differing complexity, one with a simplified 8 equation model and a more comprehensive 34 equation model including platelet activation. Results demonstrated that Coagulo-Net accurately reproduced thrombin and factor dynamics, inferred unknown parameters with low relative error, though there was limited data availability. (1).

Hemophilia A (HA) is a Factor X deficiency bleeding disorder caused by deficiency of clotting factor VIII. This disease is found only in males, but variants in the F8 gene can also cause HA in females who are frequently misdiagnosed or untreated. This study used machine learning (ML) models to predict disease severity in female patients using genotype data from large sources such as My Life Our future (male) (MLOF), Child Health and Mortality Prevention Surveillance (CHAMP), European Association for Hemophilia and Allied Disorders (EAHAD), and Hemobase validation sets. Six classifiers Random Forest (0.90), Gradient Boosting (0.91), LightGBM (0.84), XGBoost (0.90), CatBoost (0.88), and Logistic Regression (0.36) was the F1 score for the combined male and female datasets of MLOF. From the results ML can accurately distinguish mild, moderate, and severe HA. The study underscores the urgent need for more female specific genotyping, as misdiagnosis can lead to serious complications such as anemia and heavy menstrual bleeding. (2).

The review of machine learning and graph signal processing (GSP) in healthcare highlights that the field has grown rapidly since 2015, with neurology dominating applications such as Alzheimer's, autism, ADHD, epilepsy, and emotion recognition. GSP extends classical signal processing to irregular domains modeled as graphs, while deep learning and graph neural networks, particularly graph convolutional networks, are widely applied to brain data. Publications surged after 2020, with China and the USA leading contributions, and public datasets

like HCP, ADNI, ABIDE, and DEAP most frequently used. Performance evaluation across the 45 studies relied heavily on accuracy, but also reported R^2 , RMSE and classification metrics such as AUC and F1-score. Accuracy was the most common metric of 26 papers, F1-score appeared in 12, AUC in 9, precision and recall in 8 each, and sensitivity or specificity in 6. R^2 and RMSE were used in regression-oriented studies to assess model fit and error magnitude, while AUC and F1-score were central in classification tasks to balance sensitivity and precision.

Key challenges include limited public datasets, data imbalance, lack of standardization, and insufficient clinical interpretability. Future directions emphasize explainable AI, integration of phenotypic and clinical data, exploration of new transforms, real-time medical applications, and broader adoption beyond neurology, aligning with Health 4.0 goals of early diagnosis, prevention, and improved patient care. (3)

This study introduces GNN-HemA, a graph-based machine learning framework designed to predict how mutations in coagulation factor VIII (FVIII) lead to hemophilia A. By modeling FVIII as a residue interaction network and integrating structural and evolutionary features such as solvent accessibility and conservation scores, the authors trained Graph Neural Networks to classify mutation severity and assess functional impacts.

Severe Hemophilia was observed in buried, conserved residues while mutation with an accuracy of 70%. Finally, the work proves that GNN based classifiers can support diagnostics and guide the balanced design of recombinant FVIII proteins with improved therapeutic characteristics (4).

Fathima et.al discusses systematic review about how artificial intelligence (AI) can help in diagnosing, preventing, and managing bleeding disorders such as hemophilia, von Willebrand disease (VWD), and immune thrombocytopenia (ITP). The review found that AI models such as Random Forest, XGBoost, and

A hybrid graph attention network integrating coagulation cascade biology developed for transductive prediction of bleeding disorders

deep learning analyze genetic data, lab results, and patient histories to predict disease severity. Lab results include platelet counts and clotting factor levels. The review showed that AI gave better accuracy than traditional methods (5).

Haoyu developed a machine learning model to predict coagulopathy in traumatic brain injury (TBI) patients. Chongqing Ninth People hospital's emergency department provided 322 cases between 2018 and 2024. Coagulopathy was observed in 15.8% of patients, and Random Forest algorithm achieved the best performance. The Random Forest performance was AUC = 0.92, recall = 0.94 and false negative rate = 6% the outperforms over traditional coagulation tests (6).

Divyansh applied graph neural networks (GNNs) in molecular biology, mainly in drug discovery. GNNs highlight molecules and biological interactions as graphs, predictions of properties, drug target interactions, and side effects. Key challenges include reliability issues, such as overoptimistic mispredictions, and the over-smoothing problem when adding layers. CardioTox benchmark datasets with GNN-GP and GNN-SNGP techniques are used for solution generation. These networking models improve robustness, and the use of noisy nodes to counter over-smoothing, performance of the prediction on 3D molecular property (7).

The article "Mathematical models of coagulation—are we there yet?" by Matt J et. al. systematically evaluates existing mathematical models of blood coagulation by comparing their predictions of thrombin generation. The study uses a large dataset of 348 individuals with normal hemostasis. The model captures biochemical reactions; none consistently reproduce experimental thrombin generation curves or accurately predict endogenous thrombin. Differences in model outputs stem from variations in reaction structures, rate constants, and inclusion of factors such as FXI, Curves and Shapes and exclude validation against large, coherent datasets for effective antithrombotic assessment (8).

Aravind et. al. study presents a machine learn-

ing pipeline combining Convolutional Neural Networks and Graph Neural Networks to map histological topology in oral mucosal tissue. In the first phase, CNN with EfficientNet achieved accuracy of ≈ 0.98 for cell classification, though test accuracy dropped to 0.71 due to class imbalance. In the second phase, GNN with GraphSage classified basement membrane (BM) crossing edges with an overall accuracy of 0.9071 and F1 score of 0.9071 on the test set. (9).

The study proposes a multimodal graph neural network (GNN) framework for cancer molecular subtype classification using heterogeneous multi-layer graphs. The model is tested on the TCGA Pan-cancer dataset which has 27 subtypes and TCGA breast invasive carcinoma dataset that has 4 subtypes. The GAT based model achieved the highest accuracy, 83.9% for Pan-cancer and 86.4% for BRCA. The GCN based model performed strongly 81.2% and 83.8%, respectively (10).

Mengting et.al. presents CircDA, a deep learning framework built on a graph Markov neural network to predict circRNA disease associations by integrating multisource biological data. CircDA models represent biomolecules as nodes and their interactions as edges. Performance evaluation was done across four benchmark datasets. CircDA consistently outperformed with AUROC = 0.9716 demonstrating strong robustness. (11).

Systems biology models introduced by Hockin et. al. and Luan et. al discusses on thrombin generation and clot formation dynamics (12-13) uses ODE. However, these models lack scalability and interpretability at the patient level. Recent advances in graph neural networks enable explicit modeling of biological relations and are also the best applications of biomedical and pharmacological studies (14 -15).

Despite these advances, limited work has integrated systems biology simulations with graph-based learning to study coagulation. This gap motivates the present study, which combines pathway-level simulation with graph

attention modeling to achieve both predictive performance and biological interpretability.

Materials and Methods

The dataset comprises of 1000 synthetic patient profiles generated using biologically informed distributions. The given data reflects the reported physiological ranges consulting the drug information center department of Pharmaceutics, East Point College of Pharmacy on 19th January 2026.

Features extend to extrinsic, intrinsic, and common coagulation pathways, anticoagulant regulation, cellular components, and

system level outputs. Severity labels were assigned based on thrombin generation and fibrin formation dynamics, enabling supervised learning and pathway level interpretation.

In the absence of large-scale patient-level coagulation datasets, a synthetic dataset was generated based on clinically reported ranges of coagulation factor activities. Factors representing the intrinsic (XII, XI, IX, VIII), extrinsic (VII, tissue factor), and common pathways (X, V, II, I, XIII) were included. Patient demographic attributes and clinical indicators were simulated to reflect realistic variability and complete description is given in table 1.

Table 1: Dataset used for the study

Feature Name	Pathway	Biological Meaning	Data Type	Typical Range
Patient_ID	Identifier	Unique patient/sample identifier	Categorical	P0001–P1000
TF_activity	Extrinsic pathway	Tissue factor–mediated initiation of coagulation	Continuous	0.2–2.5
Factor_VII	Extrinsic pathway	Activator of factor X in TF pathway	Continuous	0.2–2.5
Factor_X	Common pathway	Convergence point of intrinsic and extrinsic pathways	Continuous	0.2–2.5
Factor_V	Common pathway	Cofactor in prothrombinase complex	Continuous	0.2–2.5
Prothrombin_FII	Common pathway	Precursor of thrombin	Continuous	0.3–2.0
Thrombin_Peak	System output	Maximum thrombin concentration generated	Continuous	10–300
Factor_XI	Intrinsic pathway	Amplification of thrombin generation	Continuous	0.1–2.5
Factor_XII	Intrinsic pathway	Contact pathway initiation	Continuous	0.1–2.5
Antithrombin	Anticoagulant	Inhibits thrombin and factor Xa	Continuous	0.3–2.0
TFPI	Anticoagulant	Inhibits TF–VIIa complex	Continuous	0.3–2.0
Protein_C	Anticoagulant	Downregulates factors Va and VIIIa	Continuous	0.3–2.0
Platelet_Count	Cellular component	Platelet availability for clot formation	Continuous	50–500 ($\times 10^3/\mu\text{L}$)
Fibrin_Formation_Rate	System output	Rate of fibrin polymerization	Continuous	0.2–3.0
Type of disorder	Target variable	HA, HB, HC, Vit. K, vonWillebrand	Categorical	3 classes

A hybrid graph attention network integrating coagulation cascade biology developed for transductive prediction of bleeding disorders

Disease-specific perturbations were introduced to model bleeding disorders, including hemophilia A, hemophilia B, hemophilia C, and vitamin K deficiency. Severity classes (mild, moderate, severe) were assigned based on reduced factor activity and system level clotting behavior.

Pathway level systems biology simulation

A reduced system of ordinary differential equations was used to simulate coagulation dynamics. The model focused on thrombin and fibrin generation as key indicators of clot formation. Factor activity levels were used to modulate reaction rates, allowing patient-specific simulations are given in table 2.

Table 2: Normal Reference Ranges of Coagulation Factors

Factor No.	Name	Normal Range	Pathway
I	Fibrinogen	200–400 mg/dL	Common
II	Prothrombin	70–120 %	Common
III	Tissue Factor (TF)	functional	Extrinsic
IV	Calcium (Ca ²⁺)	8.6–10.2 mg/dL	All
V	Proaccelerin	70–120 %	Common
VII	Proconversion	70–120 %	Extrinsic
VIII	Anti-hemophilic factor A	50–150 %	Intrinsic
IX	Christmas factor	70–120 %	Intrinsic
X	Stuart–Prowar factor	70–120 %	Common
XI	Plasma thromboplastin antecedent	70–120 %	Intrinsic
XII	Hageman factor	50–150 %	Intrinsic
XIII	Fibrin-stabilizing factor	70–140 %	Common
vWF	von Willebrand factor	50–150 %	Intrinsic / Platelet
Platelets	Platelet count	150–400 ×10 ⁹ /L	Primary hemostasis
Antithrombin (AT)	Natural anticoagulant	80–120 %	Regulation
Protein C	Natural anticoagulant	70–140 %	Regulation
TFPI	Tissue Factor Pathway Inhibitor	0.5–2.5 ng/mL	Regulation

Graph construction

The dataset was represented as a heterogeneous graph containing patient nodes and coagulation factor nodes. Edges between patient and factor nodes encoded factor activity levels representing biochemical interactions within the cascade. This graph structure explicitly conserved pathway relationships and enabled systems level learning.

Graph attention network model

A Graph Attention Network was implemented to perform severity classification. Node features included demographic patient attributes, factor activity levels, and ODE derived

system features. The graph attention mechanism allows the model to learn weighted contributions of neighboring nodes determining the factors and pathways.

The proposed graph attention network enabled pathway level and systems level interpretation of the coagulation cascade. This network measures the relative importance of intrinsic, extrinsic, and common pathway components in determining clot severity. Attention weights learned by the model revealed that intrinsic pathway factors VIII, IX, XI, and XII, exerted strong influence during amplification phases. While tissue factor and factor VII dom-

inated early initiation dynamics. Common pathway components such as factors X, V, and prothrombin governed final thrombin output. This quantitative ranking of coagulation factors not only verifies established biochemical knowledge but also highlights subtle interdependencies that are difficult to detect using traditional linear or isolated pathway models.

Furthermore, the graph-based representation captured nonlinear factors to factor interactions, allowing the emergence of systems level behaviors including thrombin amplification feedback loops and bottlenecks in factor X activation. These findings demonstrate that the proposed framework provides a complete systems biology map of coagulation dynamics. The prediction and Graph Attention Network algorithm is described below.

Algorithm: graph attention-based prediction of coagulation disorders

Input

- Coagulation dataset with factor activities, regulatory proteins, platelet and vWF levels
- Disorder labels (Hemophilia A/B/C, Vitamin K deficiency, von Willebrand disease, Normal)
- New patient clinical measurements

Output

- Predicted coagulation disorder
- Interpretable coagulation interaction graph

Algorithm steps

1. *Preprocessing:* Encode disorder labels, remove non-numeric fields, and normalize features using Min–Max scaling.
2. *Graph Construction:*
3. Create a directed graph with patient nodes and coagulation factor nodes. Connect nodes use biologically defined edges representing intrinsic, extrinsic, and common coagulation pathways.
4. *Feature Initialization:*

5. Assign fixed-dimensional node features to preserve node identity and enable transductive learning.
6. *Model Design:*
7. Implement a two-layer Graph Attention Network (GAT) with multi-head attention for feature aggregation and disorder classification.
8. *Training:* Train the GAT model on patient nodes using cross-entropy loss and the Adam optimizer.
9. *Evaluation:* Assess model performance on held-out patient nodes and examine attention-based factor influence.
10. *Inference:* Insert unseen patient nodes into the trained graph without retraining and predict coagulation disorders.
11. *Validation & Visualization:*
12. Validate predictions using clinically defined case studies and visualize the coagulation cascade graph.

End algorithm

In addition to model training and evaluation, the proposed framework supports inductive inference for unseen patient data. New patient profiles are incorporated into the graph as additional nodes connected to coagulation factor nodes. Severity predictions are obtained via forward propagation through the trained graph attention network without retraining, enabling real-time application.

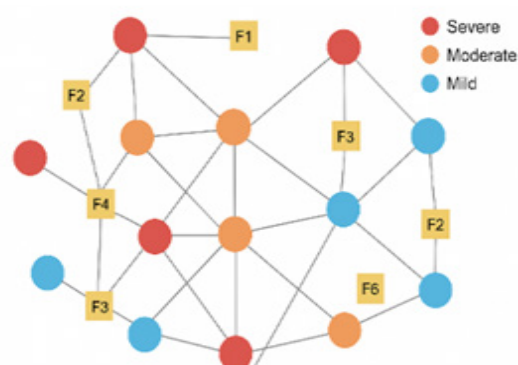


Fig 2: Bipartite Graph

A hybrid graph attention network integrating coagulation cascade biology developed for transductive prediction of bleeding disorders

A bipartite graph shown in figure 2 was constructed linking patient nodes to coagulation factor nodes based on measured biological features. A graph attention network (GAT) was trained to classify coagulation severity using node level relations learned from pathway interactions.

To evaluate the trained GAT, model was applied on unseen patient profiles by introducing new patient nodes connected to coagulation factor nodes.

Severity predictions were obtained through forward propagation without retraining, demonstrating inductive inference capability as shown in figure 2.

Table 3: Comparison Table for Path level and Coagulation cascade Disorders

Model / Study	TF Pathway	Common Pathway	Factor XI	Factor XII	AT	TFPI	Protein C	Platelets	Spatial Effects
Lakshmanan et al., 2022 (Lakshmanan, 2022) (16)	✓ (TF-VIIa)	✓ (X, V, II)	✓ (feedback)	–	✓	✓	–	–	–
Owen et al., 2024 (Owen, Tuddenham, King, Goodall, & Dunster, 2024) (Review & Assessment) (17)	✓	✓	✓	–	✓	✓	✓	–	–
Stobb et al., 2024 (Stobb, 2024) (18)	✓	✓	✓	–	✓	✓	–	✓	–
Coagulo-Net (PINN), 2024 (Qian, Zhang, Liu, & Wang, 2024) (1)	✓	✓	✓	–	✓	✓	–	–	–
Intrinsic Pathway Model, 2024 (Mitrophanov & Reifman, 2024) (19)	–	✓	✓	✓	–	–	–	–	–
Reaction Systems – Inhibitors, 2024 (Pantelev, Dashkevich, & Ataulakhanov, 2024) (20)	✓	✓	–	–	✓	✓	✓	–	–
Rethinking Coagulation (Conceptual), 2023 (Chatterjee & Mann, 2023) (21)	✓	✓	✓	✓	✓	✓	✓	✓	–
Minimal Model Optimization, 2025 (Zhang, Liu, & Chen, 2025) (22)	✓ (reduced)	✓	–	–	–	–	–	–	–
Endo venous Laser Therapy Model, 2025 (Andreeva, Tyurin, & Khanin) (23-24)	✓	✓	✓	✓	–	–	–	–	✓
Hybrid Systems Biology + ML (Proposed Study)	✓	✓	✓	✓	✓	✓	✓	✓	✓

In the later stages a heterogeneous graph was built consisting of patient nodes and coagulation factor nodes. Patient factor edges encoded normalized activity levels, while factor to factor edges represented traditional biochemical activation and inhibition relationships derived from the coagulation cascade.

This hybrid graph structure enabled integration of data driven learning with pathway

informed constraints, enabling biologically interpretable severity prediction using graph attention networks.

Table 3 compares recent mathematical and computational models of the blood coagulation cascade available after 2022. While most current models incorporate the tissue factor initiated and common pathways, that include intrinsic factors XI and XII, regulatory inhibitors

antithrombin, TFPI, protein C, platelet dynamics, and spatial effects remain unpredictable.

While the overall test accuracy was moderate, the model demonstrated strong macro averaged $F1 = 0.29$ performance, reflecting its ability to distinguish coagulation phenotypes despite substantial pathway overlap. Misclassifications primarily occurred between clinically related disorders, such as mild Hemophilia A and von Willebrand disease, consistent with known diagnostic challenges.

Results and Discussion

The Graph Attention Network achieved high classification performance and demonstrated to distinguish between mild, moderate, and severe bleeding phenotypes. Attention analysis emphasized intrinsic pathway factors, particularly factors VIII and IX, as dominant contributors in severe cases. The common pathway also showed increased influence in cases asso-

ciated with reduced clot stability.

Visualization of node embeddings exposed meaningful clustering of patients based on underlying deficiency patterns. Rather than merely replicating known associations, the hybrid approach provided an integrated framework that linked molecular level dynamics to patient level outcomes. This integration enhances interpretability and underscores the relevance of graph-based learning for biological systems. The Case study analysis demonstrated that the proposed model successfully differentiated hypo coagulable and hypercoagulable phenotypes based on thrombin generation and fibrin formation dynamics.

Case study 1: severe hypo coagulable state

The first case in table 4 represents a severe coagulation impairment considered by distinctly suppressed extrinsic and intrinsic pathway activity.

Table 4: Case study 1 input features

Category	Parameter	Normalized Value	Interpretation
Extrinsic Pathway	Tissue Factor (TF)	0.55	Moderate pathway initiation
	Factor VII	0.60	Adequate extrinsic activation
Common Pathway	Factor X	0.58	Sufficient convergence
	Factor V	0.62	Effective cofactor activity
	Prothrombin (Factor II)	0.65	Supports thrombin generation
	Thrombin Peak	0.55	Moderate thrombin production
Intrinsic Pathway	Factor XI	0.50	Partial amplification
	Factor XII	0.48	Controlled feedback activation
Anticoagulant Regulation	Antithrombin	0.55	Balanced inhibition
	TFPI	0.50	Regulated TF suppression
	Protein C	0.52	Controlled anticoagulant response
Platelet Function	Platelet Count	0.60	Adequate platelet availability
Clot Formation	Fibrin Formation Rate	0.58	Stable fibrin polymerization
Overall Phenotype	—	—	Balanced / near-physiological coagulation

Tissue factor activity (0.25) and factor VII levels (0.30) were significantly reduced, preventing initiation of the coagulation cascade.

Convergence through the common pathway was impaired, as reflected by low factor X (0.28), factor V (0.35), and prothrombin (0.40) levels, resulting in a reduced throm-

A hybrid graph attention network integrating coagulation cascade biology developed for transductive prediction of bleeding disorders

bin peak (0.22) as shown in figure 3. Intrinsic pathway intensification was also compromised, with factor XI (0.20) and factor XII (0.18) showing minimal activity. In contrast, anticoagulant regulators were raised, including antithrombin (0.90), tissue factor pathway inhibitor (0.85), and protein C (0.88). Reduced platelet availability (0.35) and a low fibrin formation rate (0.25) together contributed to unstable clot formation. Based on this integrated profile, the graph attention model classified the sample as severe coagulation impairment, consistent with a hypo coagulable phenotype as shown in figure 3.

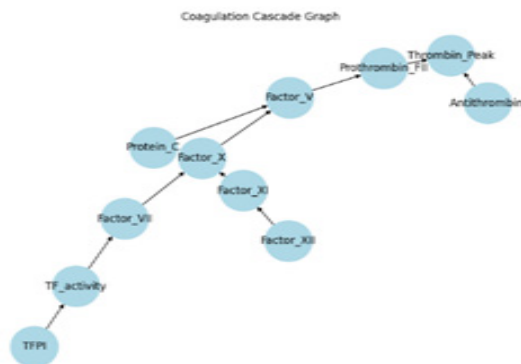


Fig 3: Case Study 1 GAT representing Hypo Coagulable Phenotype output

Table 5: Case Study 2 input features

Category	Parameter	Normalized Value	Interpretation
Extrinsic Pathway	Tissue Factor (TF)	0.55	Moderate pathway initiation
	Factor VII	0.60	Adequate extrinsic activation
Common Pathway	Factor X	0.58	Sufficient convergence
	Factor V	0.62	Effective cofactor activity
	Prothrombin (Factor II)	0.65	Supports thrombin generation
	Thrombin Peak	0.55	Moderate thrombin production
Intrinsic Pathway	Factor XI	0.50	Partial amplification
	Factor XII	0.48	Controlled feedback activation
Anticoagulant Regulation	Antithrombin	0.55	Balanced inhibition
	TFPI	0.50	Regulated TF suppression
	Protein C	0.52	Controlled anticoagulant response
Platelet Function	Platelet Count	0.60	Adequate platelet availability
Clot Formation	Fibrin Formation Rate	0.58	Stable fibrin polymerization
Overall Phenotype	—	—	Balanced / near-physiological coagulation

Case study 2: moderate coagulation response

The second case demonstrates a balanced coagulation state exhibiting intermediate activation across all pathways with various features as referred in table 5.

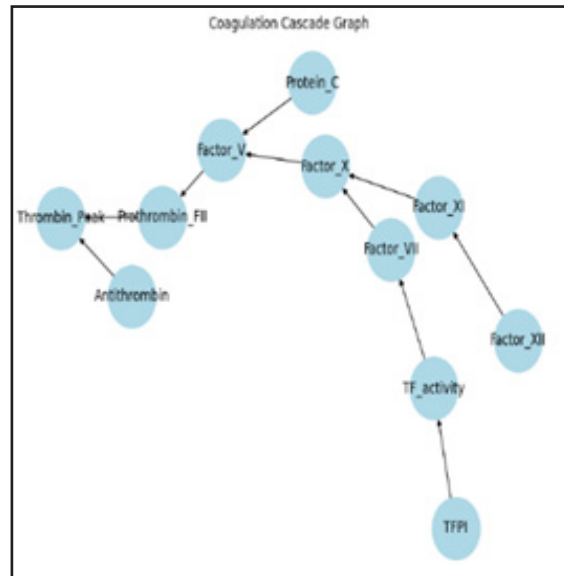


Fig 4: Output of Moderate Coagulation phenotype - Case study

Tissue factor activity (0.55) and factor VII (0.60) give rise to moderate extrinsic pathway initiation, while common pathway factors such as factor X (0.58), factor V (0.62), and prothrombin (0.65) enabled adequate thrombin generation, resulting in a moderate thrombin peak (0.55). Intrinsic pathway components, including factor XI (0.50) and factor XII (0.48), added partially to the pathway. Anticoagulant regulation remained balanced, with antithrombin (0.55), tissue factor pathway inhibitor (0.50), and protein C (0.52).

Platelet count (0.60) and fibrin formation rate (0.58) were sufficient to support stable clot formation without excessive propagation as shown in figure 4. The model integrated these signals and classified this profile as moderate severity coagulation response.

Case study 3: hypercoagulable state

The third case corresponds to a mild or hypercoagulable phenotype characterized by

enhanced activation of coagulation pathways and reduced inhibitory control as described in table 6. Strong extrinsic pathway initiation was observed through raised tissue factor activity (0.85) and factor VII levels (0.88).

Convergence through the common pathway was indicated by high factor X (0.82), factor V (0.80), and prothrombin (0.78), leading to thrombin peak (0.90). Intrinsic pathway includes factor XI (0.85) and factor XII (0.80) reinforcing thrombin generation is shown in figure 5. Anticoagulant influences were comparatively low, including antithrombin (0.35), tissue factor pathway inhibitor (0.30), and protein C (0.40). This formation allows clot propagation. Increased platelet availability (0.85) and fibrin formation (0.88) supported efficient and stable clot development. From these parameters the model correctly classified the case as mild severity, consistent with a hypercoagulable state.

Table 6: Input Features of Case study 3 with Hypercoagulable state

Category	Parameter	Normalized Value	Interpretation
Extrinsic Pathway	Tissue Factor (TF)	0.85	Strong initiation of coagulation cascade
	Factor VII	0.88	Enhanced extrinsic pathway activation
Common Pathway	Factor X	0.82	Efficient pathway convergence
	Factor V	0.80	Robust cofactor support
	Prothrombin (Factor II)	0.78	High thrombin precursor availability
	Thrombin Peak	0.90	Pronounced thrombin generation
Intrinsic Pathway	Factor XI	0.85	Strong amplification feedback
	Factor XII	0.80	Active contact pathway contribution
Anticoagulant Regulation	Antithrombin	0.35	Reduced inhibitory control
	TFPI	0.30	Limited suppression of TF-VIIa
	Protein C	0.40	Weakened anticoagulant feedback
Platelet Function	Platelet Count	0.85	Elevated platelet availability
Clot Formation	Fibrin Formation Rate	0.88	Rapid and stable fibrin polymerization
Overall Phenotype	—	—	Mild hypercoagulable state

A hybrid graph attention network integrating coagulation cascade biology developed for transductive prediction of bleeding disorders

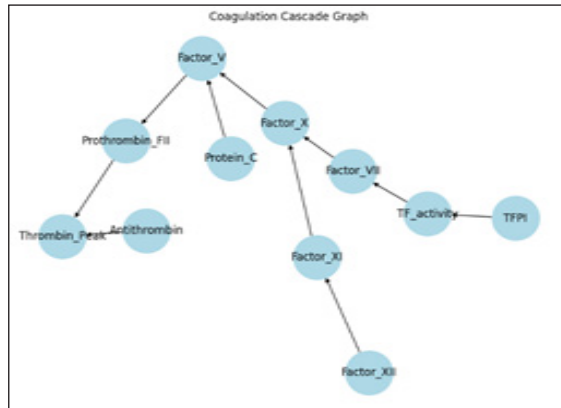


Fig 5: Graph output of hyper coagulable state

Case 4: von willebrand phenotype disease

The graph shown in figure 6 visualizes the new patient specific coagulation network constructed from the input features and the trained Graph Attention Network (GAT) model. The central red node represents the new patient, while nearby blue nodes correspond to individual coagulation factors and regulatory components. Edge thickness and color intensity encode the normalized feature values, with brighter and thicker edges indicating stronger activity.

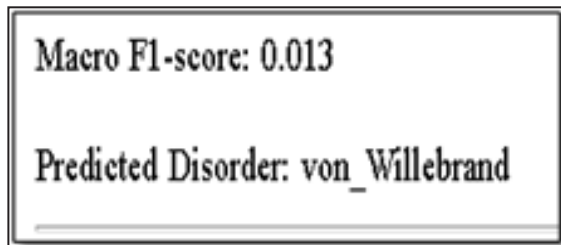


Fig 6: Algorithm output for case 4

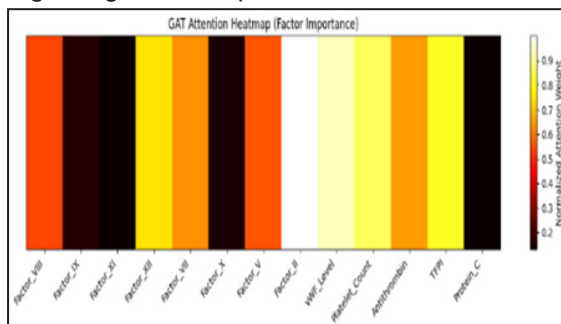


Fig 7: Coagulation graph predicting the disorder as von Willebrand

In this case, Factor VIII, Factor X, and vWF exhibit prominent connections, reflecting towards model's prediction. The reduced connection strength observed for Platelet Count and TFPI highlights their reduced functional contribution. The overall network structure indicates a von Willebrand disease phenotype as shown in figure 6 and 7.

This heatmap given in figure 8 represents the attention weights learned by the GAT model for each coagulation feature when classifying the patient. Yellow and white lighter shades indicate higher attention scores, for significant features that contribute for an efficient classification decision.

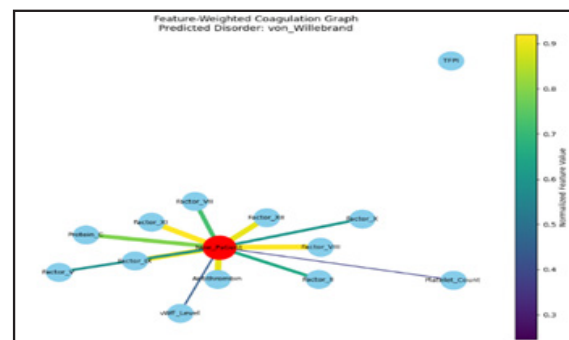


Fig 8: Heatmap displaying the Factor Significance for GAT network

The bar chart visualized in Figure 9 highlights the significant features of their activity levels exerting biological impact. The combined influence is observed for vWF Level, followed by Factor VIII and Platelet Count, reinforcing the diagnosis of von Willebrand disease.

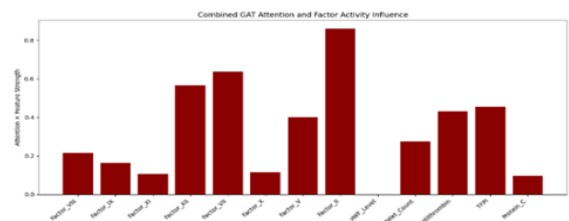


Fig 9: Combined GAT Attention and Feature Activity Influence

From the 4 representative case studies, the model demonstrated consistent alignment with established coagulation physiology. In the

severe hypo coagulable case, marked suppression of pathway initiation and amplification, coupled with elevated anticoagulant activity, resulted in reduced thrombin generation and unstable fibrin formation. Conversely, the balanced case reflected near-physiological activation across pathways with appropriate regulatory control, serving as a functional reference phenotype. The hypercoagulable case was characterized by strong pathway activation, attenuated inhibitory regulation, and enhanced platelet and fibrin responses, consistent with a mild prothrombotic state.

A key strength of the proposed framework lies in its interpretability. Attention heatmaps revealed that the GAT model assigns higher importance to factors such as factor VIII and vWF in von Willebrand disease, and common pathway factors in hypercoagulable states. The integration of attention scores with feature activity additionally enabled identification of dominant drivers of coagulation imbalance at the individual patient level. This dual layer interpretability model addresses a critical limitation of many black-box learning approaches in biomedical applications.

The use of biologically constrained synthetic data allowed systematic evaluation of the model conserving physiological acceptability. Although synthetic in nature, the feature ranges and disorder labels were derived from established clinical reference intervals taken from pharmacy department and diagnostic center. In future work the same model will be incorporated on real-world patient datasets to validate generalizability and to assess performance across broader demographic and clinical contexts.

Conclusions

This study demonstrates the efficiency of a biologically informed graph-based learning framework for the classification of coagulation disorders. By integrating intrinsic, extrinsic, and common pathway components into a unified graph representation and applying a Graph Attention Network. The proposed model achieved multi class discrimination across clinically rele-

vant coagulation phenotypes HA, HB, HC, Vit. K and Von Willebrand. Model performance, evaluated using the Macro F1 score to account for class imbalance, reached approximately 0.29, representing balanced predictive capability across all disorder categories rather than dominance by majority classes.

The framework further enables patient specific inference by incorporating new cases into the trained graph supporting practical deployment scenarios. Overall, this work establishes a scalable and interpretable approach for coagulation disorder classification using graph-based machine learning. This model lays the groundwork for future validation using real world clinical data and potential integration into decision support tools.

Acknowledgement

The authors are grateful to the REVA University and East point College of Pharmacy for institutional support for the completion of this research work.

Conflict of interest

The authors declare no conflict of interest

Author contributions

Conceptualization and Methodology: Lakshmi J V N; Writing, Review and Editing: G Harini Kumari

Funding

No external funding was received for this study

Data availability

<https://drive.google.com/file/d/1NCnZSEe6D-212SAkJDS9X4PbRRXihNM0/view?usp=sharing>

References

1. Qian, Y., Zhu, G., Zhang, Z., Modepalli, S., Zheng, Y., Zheng, X., . . . Li, H. (2024). Coagulo-Net: Enhancing the mathematical modeling of blood coagulation using physics-informed neural networks. *Neural Networks*, 106732. Doi:<https://doi.org/10.1016/j.neunet.2024.106732>

A hybrid graph attention network integrating coagulation cascade biology developed for transductive prediction of bleeding disorders

2. Atul, R., Christopher, K., Jiayi, O., & Zuben, E. (2024). Application of machine learning approaches for predicting hemophilia A severity. *Journal of Thromb Haemost.*, 22, 1909-1918. Doi:<https://doi.org/10.1016/j.jtha.2024.03.009>
3. Maria Alice Andrade, C., Felipe A. B. S, F., Fernando A. N, S., & Francisco, M. (2024). Machine Learning and Graph Signal Processing Applied to Healthcare: A Review. *Bioengineering*, 671, 1-28. Doi: <https://doi.org/10.3390/bioengineering11070671>
4. Marcos V, F., Tatiane, N., Ricardo A, R., & Tiago J. S, L. (2023). A graph based ML framework identifies critical propoerties of FVIII that lead to hemophilia A. *Frontiers in Bioinformatics*, 1-9. Doi:DOI 10.3389/fbinf.2023.1152039
5. Fathima Raahima Riyas, M., Ziyad, A., Wael, K., Khaled, H., Anwar, A., Adel, K., & Muhammad Raihan, S. (2025). The use of AI in the prevention and management of bleeding disorders: a systematic review. *Frontiers in Medicine*, 12(1606788), 1-15. Doi: 10.3389/fmed.2025.1606788
6. Haoyu, W., Wenying, C., Jianhuang, H., Yuxing, F., & Cheng, L. (2025). Machine learning models predict coagulopathy in traumatic brain injury patients in ER. *Frontiers in Neurology*, 16(1649869), 1-10. Doi:10.3389/fneur.2025.1649869
7. Divyansh, C. (2024). An overview of graph neural networks for molecular biology: Challenges and solutions. *International Journal of Science and Research Archive*, 13(01), 2670-2673. Doi: <https://doi.org/10.30574/ijrsra.2024.13.1.1985>
8. MattJ., O., JoyR, W., EdwardG.D., T., JohnR, K., AlisonH, G., & JoanneL, D. (2024). Mathematicalmodels of coagulation—arewethereyet? *Journal of Thrombosis and Haemostasis*, 22, 1689-1703. Doi:<https://doi.org/10.1016/j.jtha.2024.03.009>
9. Aravind, N., Helena, A., J. E., Nikolce, T., K. M., & Rachael, S. (2022). A graph neural network framework for mapping histological topology in oral mucosal tissue. *BMC Bioinformatics*, 23(1), 1-21. Doi:<https://doi.org/10.1186/s12859-022-05063-5>
10. Bingjun, L., & Sheida, N. (2024). A multi-modal graph neural network framework for cancer molecular subtype classification. *BMC Bioinformatics*, 25(27). Doi:<https://doi.org/10.1186/s12859-023-05622-4>
11. Mengting, N., Chunyu, W., Zhanguo, Z., & Quan, Z. (2024). A computational model of circRNA-associated diseases based on a graph neural network: prediction and case studies for follow-up experimental validation. *BMC Biology*, 22(24), 1-13. Doi:<https://doi.org/10.1186/s12915-024-01826-z>
12. Hockin MF, Jones KC, Everse SJ, Mann KG. A model for the stoichiometric regulation of blood coagulation. *J Biol Chem*. 2002 May 24;277(21):18322-33. Doi: 10.1074/jbc.M201173200. Epub 2002 Mar 13. PMID: 11893748.
13. Biscombe CJC, Dower SK, Muir IL, Harvie DJE. Modeling Thrombin Generation in Plasma under Diffusion and Flow. *Biophys J*. 2020 Jul 7;119(1):162-181. Doi: 10.1016/j.bpj.2020.04.033. Epub 2020 May 19. PMID: 32544388; PMCID: PMC7335884.
14. Wu, Z. (2021). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*. <https://doi.org/10.1109/TNNLS.2020.2978386>

15. Zitnik, M., Agrawal, M., & Leskovec, J. (2018). Modeling polypharmacy side effects with graph neural networks. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/bty294>
16. Lakshmanan, H. H. (2022). Revised model of the tissue factor pathway of thrombin generation: Role of the feedback activation of factor XI. *Journal of Thrombosis and Haemostasis*, 20(6), 1350-1363. Doi:<https://doi.org/10.1111/jth.15716>
17. Owen, M. J., Tuddenham, E., King, J. R., Goodall, A., & Dunster, J. (2024). Mathematical Models of Coagulation – Are we there yet? *Journal of Thrombosis and Haemostasis*, 22(6), 1689-1703. Doi:<https://doi.org/10.1016/j.jth.2024.03.009>
18. Stobb, M. T. (2024). Mathematical modeling identifies clotting factor combinations that modify thrombin generation in normal and factor VIII-, IX-, or XI-deficient blood. *Research and Practice in Thrombosis and Haemostasis*, 8(7). Doi:<https://doi.org/10.1016/j.rpth.2024.102570>
19. Mitrophanov, A. Y., & Reifman, J. (2024). Mathematical modeling of the intrinsic pathway of blood coagulation. *Mathematical Biosciences and Engineering*, 21(12), 7707–7739. Doi:<https://doi.org/10.3934/mbe.2024339>
20. Panteleev, M. A., Dashkevich, N. M., & Ataullakhanov, F. I. (2024). Computational analysis of thrombus formation and clotting factor regulation under flow conditions. *Mathematical Biosciences and Engineering*, 21(12), 7707-7739. <https://doi.org/10.1016/j.bpe.2016.03.010>
21. Chatterjee, M. S., & Mann, K. G. (2023). Rethinking coagulation: From enzymatic cascade to systems biology and immune regulation. *Blood*, 142(25), 2133–2145. Doi:<https://doi.org/10.1182/blood.2023019805>
22. Zhang, H., Liu, P., & Chen, Y. (2025). Optimization of parameters in a minimal coagulation model using computational techniques. *Bioengineering*, 12(10). Doi:<https://doi.org/10.3390/bioengineering12101111>
23. Andreeva, A. A., Tyurin, A. A., & Khanin, M. A. (n.d.). Mathematical model of blood coagulation during endovenous laser therapy. *arXiv*. Doi:<https://arxiv.org/abs/2507.20401>
24. Thomas Gaudet, Ben Day, Arian R Jamasb, Jyothish Soman, Cristian Regep, Gertrude Liu, Jeremy B R Hayter, Richard Vickers, Charles Roberts, Jian Tang, David Roblin, Tom L Blundell, Michael M Bronstein, Jake P Taylor-King, Utilizing graph machine learning within drug discovery and development, Briefings in Bioinformatics, Volume 22, Issue 6, November 2021, bbab159, <https://doi.org/10.1093/bib/bbab159>

A hybrid graph attention network integrating coagulation cascade biology developed for transductive prediction of bleeding disorders