Extended Abstract Track

Graph Mixing Additive Networks

Maya Bechler-Speicher   MAYABS@META.COM
Meta

Andrea Zerio   ANZE@DCM.AAU.DK
Center of Excellence for Molecular Prediction of IBD, PREDICT, Department of Clinical Medicine, Aalborg University

Maor Huri   MAORHURY@MAIL.TAU.AC.IL
Sagol School of Neuroscience Tel Aviv University

Marie Vibeke Vestergaard   PEPIJNRH@KTH.SE
Center of Excellence for Molecular Prediction of IBD, PREDICT, Department of Clinical Medicine, Aalborg University

Ran Gilad-Bachrach   RGB@TAUEX.TAU.AC.IL
Department of Bio-Medical Engineering and Edmond J. Safra Center for Bioinformatics, Tel-Aviv University

Tine Jess   JESS@DCM.AAU.DK
Center of Excellence for Molecular Prediction of IBD, PREDICT, Department of Clinical Medicine, Aalborg University
Department of Gastroenterology and Hepatology, Aalborg University Hospital

Samir Bhatt   SAMIR.BHATT@SUND.KU.DK
University of Copenhagen
Imperial College London

Aleksejs Sazonovs   ALESAZ@DCM.AAU.DK
Center of Excellence for Molecular Prediction of IBD, PREDICT, Department of Clinical Medicine, Aalborg University

Abstract

We introduce GMAN, a flexible, interpretable, and expressive framework that extends Graph Neural Additive Networks (GNANs) to learn from sets of sparse time-series data. GMAN represents each time-dependent trajectory as a directed graph and applies an enriched, more expressive GNAN to each graph. It allows users to control the interpretability-expressivity trade-off by grouping features and graphs to encode priors, and it provides feature, node, and graph-level interpretability. On real-world datasets, including mortality prediction from blood tests and fake-news detection, GMAN outperforms strong non-interpretable black-box baselines while delivering actionable, domain-aligned explanations.

Keywords: Graph Learning, Graph Representation, Interpretaability.

Extended Abstract Track

1. Introduction

Many datasets contain heterogeneous measurements sampled at irregular and asynchronous times. For example in clinical data, A patient record, for example, may comprise multiple medical tests taken at task-dependent frequencies, which can be naturally viewed as a set of sparse temporal trajectories. As another example, news propagation unfolds along platform-specific trajectories through a social network, with timing and topology jointly shaping diffusion. Conventional pipelines regularize irregular data via imputation or grid alignment, using recurrent, attention, or diffusion-based imputers (Cao et al., 2018; Tashiro et al., 2021; Wu et al., 2022; Du et al., 2023). However, such preprocessing can distort dynamics, and obscure conditional dependencies.

To address this, we introduce Graph Mixing Additive Networks (GMAN), an interpretable framework for learning from sets of temporally sparse graphs. Concretely, GMAN treats each trajectory as a directed path graph (with edges parameterized by elapsed time) and each sample as a set of such graphs, preserving temporal distances and permutation invariance across channels/paths. Methodologically, GMAN (i) encodes each trajectory with an Extended GNAN (ExtGNAN) Bechler-Speicher et al. (2024) module that aggregates over temporal edges and applies univariate or multivariate shape functions to feature groups; and (ii) non-linearly combines trajectories via a partition of the set into subsets, enabling a tunable trade-off between interpretability and expressivity.

We evaluate GMAN on two structurally distinct problems central to learning representations for irregular signals: (1) In-hospital mortality prediction from routine blood-test trajectories, and (2) fake-news detection from social propagation paths (GossipCop). GMAN achieves state-of-the art performance while providing fine-grained temporal and graph-level attributions. Finally we prove that GMAN is strictly more expressive than GNAN, and that grouped (non-singleton) subset mixing strictly increases expressivity over singleton subsets.

2. Method

In this section, we present GMAN, an interpretable and flexible method for effectively learning over sets of trajectories of varying size. Let $S = \{G_1, \ldots, G_m\}$ be a *set* of m graphs. Each node v is associated with a feature vector $x_v \in \mathbb{R}^d$ and a time-stamp t_v. For instance, in the case of a patient's blood test data, each graph corresponds to a specific biomarker, and each node within the graph represents an individual measurement of that biomarker, annotated with its feature vector and time of collection. We denote the entry c of a vector $\mathbf{h}$ by $[\mathbf{h}]_c$, and the set of entries corresponding to a set of features S by $[\mathbf{h}]_S$.

We assume that the graphs in S are partitioned into $k, 1 \leq k \leq m$ disjoint subsets S_k such that $\bigcup_{i=1}^{k} S_i = S$. The partition $\{S_i\}_{i=1}^{k}$ provides a flexible trade-off between expressivity and interpretability. GMAN linearly aggregates representations of the subsets of S to form a final set representation, and then assigns a single label to S. The level of interpretability that GMAN provides for each graph depends on the size of the subset it belongs to. When a subset contains a single graph, GMAN offers fine-grained, node-level importance scores. In contrast, for larger subsets, it provides only set-level importance scores—trading interpretability for improved expressivity. This design enables a flexible trade-off between interpretability and expressive power, controlled by the chosen partitioning strategy.

2

Extended Abstract Track

$$\mathbf{h}_i = \Phi_i(S_i),$$

Then, it produce a representation for the whole set, $\mathbf{h}_S$ by summing the subsets' $\mathbf{h}_S = \sum_{i=1}^{k} h_i$. Finally, to produce the label, it sums over the d entries of $\mathbf{h}_S$. Overall:

$$\text{GMAN}(S) = \sum_{c=1}^{d} \sum_{i=1}^{k} [\Phi_i(S_i)]_c \tag{1}$$

Where $\Phi_i(S_i) = \mathbf{h}_{S_i}$ is a representation of the subset S_i.

For subsets of size one, $\Phi_i(S_i)$ applies an Extended GNAN (ExtGNAN), as described in Section 2.1. For subsets containing multiple graphs, a ExtGNAN is applied to each graph, followed by a DeepSet aggregation (Zaheer et al., 2018) over the resulting vectors. Importantly, each subset is assigned its own ExtGNAN, and all graphs within a subset share the same one. A DeepSet first applies a neural network $f : \mathbb{R}^d \to \mathbb{R}^d$ for each vector in the set $\{h_l\}_{G_l \in S_i}$, sums the results, and then applies another neural network $g : \mathbb{R}^d \to \mathbb{R}^d$.

$$g \left(\sum_{i \in S_2} f(h_i) \right)$$

Here, g and f are neural networks of arbitrary depth and width. We now turn to define ExtGNAN.

2.1. ExtGNAN

In GNAN Bechler-Speicher et al. (2024), univariate neural networks are applied to each feature of each node in isolation, to learn a representation for a graph. This has the benefit of generating interpretable models as features do not mix non-linearly. Nonetheless, when interactions between features are crucial for the task, it may result in sub-par performance. Therefore, ExtGNAN extends GNAN by allowing multivariate neural networks to operate on groups of features to gain accuracy at the cost of reducing the feature-level interpretability of the model.

Assume that the features are partitioned into K subsets $\{F_l\}_{l=1}^{K}$. For any subset of features greater than one, ExtGNAN applies a multivariate neural network for all the features in the subset together, instead of a univariate neural network for each one separately. To learn a representation of a graph G, ExtGNAN first computes representations for the nodes of G as follows.

ExtGNAN learns a distance function $\rho(x; \theta) : \mathbb{R} \to \mathbb{R}$ and a set of feature shape functions $\{\psi_l\}_{l=1}^{K}, \psi_l(X; \theta_k) : \mathbb{R}^{|F_l|} \to \mathbb{R}^{|F_l|}$. Each of these functions is a neural network of arbitrary depth. For brevity, we omit the parameterization θ and θ_k for the remainder of this section.

The entries of the representation of node j corresponding to the indices of the features in F_l, denoted as $[\mathbf{h}_j]_{F_l}$, is computed by summing the contributions of the features in the subset F_l from all nodes in the graph:

$$[\mathbf{h}_j]_{F_l} = \sum_{w \in V} \rho\left(\Delta(w, j)\right) \cdot \psi_l\left([\mathbf{X}_w]_{F_l}\right),$$

where $\Delta(w, j) = t_w - t_j$ and $[\mathbf{X}_w]_{F_l}$ are the features of node w corresponding to the subset F_l.

Overall, the full representation of node j can be written as:

$$\mathbf{h}_j = \big([\mathbf{h}_j]_{F_1}, [\mathbf{h}_j]_{F_2}, \ldots, [\mathbf{h}_j]_{F_K}\big).$$

Then EXTGNAN produces a graph representation by summing the node representations,

$$\mathbf{h}_G = \sum_{i \in V} \mathbf{h}_i. \tag{2}$$

2.2. Expressivity

In this section we provide two theoretical results of the expressivity of GMAN. Proofs are in the Appendix.

Theorem 1 GMAN *is strictly more expressive than GNAN.*

Theorem 2 *Let S be a set of graphs $\{G_i\}_{j=1}^m$. Let $S_1 = \{S_i\}_{i=1}^m$ be a partition of S such that $|S_i| = 1$. Let $S_2 = \{S_i\}_{i=1}^k$ such that there exists k with $|S_k| > 1$. with a subset partition $\{S_i\}_{i=1}^k$. Then a GMAN trained over S_2 is strictly more expressive than a GMAN trained over S_1.*

3. Empirical Evaluation

We present preliminary results evaluating GMAN on two diverse tasks: (1) predicting mortality from biomarker (blood tests) trajectories, and (2) detecting fake news propagation patterns in social networks. These tasks vary significantly in domain and graph structure, allowing us to assess generalization and interpretability in real-world settings. Full experimental setup and dataset details are detailed in the Appendix.

In-hospital mortality prediction We use the PhysioNet2012 (P12) dataset, introduced by Goldberger et al. (2000), which contains records from 11,988 ICU patients, following the exclusion of 12 samples deemed inappropriate according to the criteria in Horn et al. (2020). For each patient, time series measurements from 36 physiological signals (excluding weight) were recorded over the initial 48 hours of ICU admission. Additionally, each patient has a static profile comprising 9 features, including demographic and clinical attributes such as age and gender. The dataset is labelled for a binary classification task: predicting in-hospital mortality.

We compare GMAN to other 9 baselines evaluated in Zhang et al. (2022) and use their splits, including: *Transformer* (Vaswani et al.,

Table 1: Evaluation of GMAN in-hospital mortality prediction (P12).

Method	AUROC
Transformer	65.1 ± 5.6
Trans-mean	66.8 ± 4.2
GRU-D	67.2 ± 3.6
SeFT	66.8 ± 0.8
mTAND	65.3 ± 1.7
IP-Net	72.5 ± 2.4
DGM^2	71.2 ± 2.5
MTGNN	67.5 ± 3.1
RAINDROP	72.1 ± 1.3
GMAN	$\mathbf{76.64 \pm 1.2}$

Extended Abstract Track

2017), *Trans-mean*, *GRU-D* (Che et al., 2016), *SeFT* (Horn et al., 2020), *mTAND* (Shukla and Marlin, 2021), IP-Net(Shukla and Marlin, 2019), DGM^2(Wu et al., 2021) and MT-GNN (Wu et al., 2020).

We conducted a grid search by training on the training set and evaluating on the validation set. We selected the best performing model over the validation set. We report the average AUROC score and standard-deviation of the selected configuration with 3 seeds.

The complete subsets information is provided in the Appendix.

Fake News Detection. The GossipCop (GOS) dataset contains 5,464 retweet cascades, each represented as a tree-structured graph. We decompose cascades into directed propagation paths and represent each as a graph. Graphs are grouped into subsets based on structure, enabling controlled non-linear aggregation. More details are provided in the Appendix. Results are presented in Table 2, showing GMAN outperforms all GNN baselines, including GATv2 (Brody et al., 2022), GraphConv (Morris et al., 2021), GraphSAGE (Hamilton et al., 2018) and GCNFN (Monti et al., 2019).

Table 2: Evaluation of GMAN on fake news detection (GOS).

Methods	Accuracy
GATv2	96.10 ± 0.3
GraphConv	96.77 ± 0.1
GraphSage	94.45 ± 1.5
GCNFN	96.52 ± 0.2
GMAN	$\mathbf{97.34 \pm 0.2}$

4. Conclusion

In this extended abstract we introduced GMAN, a work-in-progress framework for learning from sets of sparse temporal graphs that achieves strong predictive performance while preserving multi-resolution interpretability. By combining distance-aware additive encoders with a flexible subset-mixing mechanism, GMAN enables fine-grained attributions when needed, and non-linear modelling when beneficial. Together, these features support actionable insight in high-stakes domains such as healthcare, while maintaining the flexibility to generalize across diverse data modalities.

Extended Abstract Track

References

Maya Bechler-Speicher, Amir Globerson, and Ran Gilad-Bachrach. The intelligible and effective graph neural additive network. *Advances in Neural Information Processing Systems*, 37:90552–90578, 2024.

Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks?, 2022. URL https://arxiv.org/abs/2105.14491.

Wei Cao, Dong Wang, Jian Li, Hao Zhou, Lei Li, and Yitan Li. Brits: Bidirectional recurrent imputation for time series. *Advances in neural information processing systems*, 31, 2018.

Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent neural networks for multivariate time series with missing values, 2016. URL https://arxiv.org/abs/1606.01865.

Wenjie Du, David Côté, and Yan Liu. Saits: Self-attention-based imputation for time series. *Expert Systems with Applications*, 219:119619, 2023.

Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220, 2000.

William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs, 2018. URL https://arxiv.org/abs/1706.02216.

Max Horn, Michael Moor, Christian Bock, Bastian Rieck, and Karsten Borgwardt. Set functions for time series, 2020. URL https://arxiv.org/abs/1909.12064.

Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, and Michael M. Bronstein. Fake news detection on social media using geometric deep learning, 2019. URL https://arxiv.org/abs/1902.06673.

Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks, 2021.

Satya Narayan Shukla and Benjamin M. Marlin. Interpolation-prediction networks for irregularly sampled time series, 2019. URL https://arxiv.org/abs/1909.07782.

Satya Narayan Shukla and Benjamin M. Marlin. Multi-time attention networks for irregularly sampled time series, 2021. URL https://arxiv.org/abs/2101.10318.

Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. Csdi: Conditional score-based diffusion models for probabilistic time series imputation. *Advances in neural information processing systems*, 34:24804–24816, 2021.

Extended Abstract Track

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186*, 2022.

Yinjun Wu, Jingchao Ni, Wei Cheng, Bo Zong, Dongjin Song, Zhengzhang Chen, Yanchi Liu, Xuchao Zhang, Haifeng Chen, and Susan Davidson. Dynamic gaussian mixture based deep generative model for robust forecasting on sparse multivariate time series, 2021. URL https://arxiv.org/abs/2103.02164.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojun Chang, and Chengqi Zhang. Connecting the dots: Multivariate time series forecasting with graph neural networks, 2020. URL https://arxiv.org/abs/2005.11650.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan Salakhutdinov, and Alexander Smola. Deep sets, 2018. URL https://arxiv.org/abs/1703.06114.

Xiang Zhang, Marko Zeman, Theodoros Tsiligkaridis, and Marinka Zitnik. Graph-guided network for irregularly sampled multivariate time series, 2022. URL https://arxiv.org/abs/2110.05357.

Appendix A. Node, graph and set importance

GMAN retains all interpretability properties of GNAN (Bechler-Speicher et al., 2024), including feature-level and node-level importance. However, it extends beyond GNAN by operating on sets of graphs rather than single graphs, enabling additional forms of interpretability such as graph-level and subset-level importance. Because GMAN allows a flexible trade-off between interpretability and expressivity, permitting non-linear mixing within graph subsets, some adaptations are required to extract meaningful attributions under this more expressive regime. We can extract the total contribution of each node j to the prediction by summing the contributions of the node across all feature sets. This is only valid when the node belongs to a graph that is not combined non-linearly with other graphs, i.e., it belongs to a subset of size one.

Therefore, the contribution of node j is

$$\text{TotalContribution}(j) = \sum_{l=1}^{K} [\mathbf{h}_j]_{F_k} = \sum_{w \in V} \rho\left(\Delta(w,j)\right) \sum_{l=1}^{K} \psi_k\left([\mathbf{x}_w]_l, l \in F_k\right). \qquad (3)$$

The contribution of a graph G is then

$$\text{TotalContribution}(G) = \sum_{v \in G} \text{TotalContribution}(v).$$

For graphs that are mixed non-linearly, i.e., graphs that belong in subsets of size greater than one, interpretability is more limited, and we can only provide the total contribution of the set to the final prediction

$$\text{TotalContribution}(S) = \sum_{l=1}^{K} [\mathbf{S}]_{F_k}. \qquad (4)$$

Appendix B. Expressivity properties

In this section we provide a theoretical analysis of the expressivness of GMAN.

Theorem 3 GMAN *is strictly more expressive than GNAN.*

Proof of Theorem 1 We will prove that GMAN is strictly more expressive than GNAN. To prove this, we use a ground truth function that is a feature-level XOR. Let a single-node graph be endowed with binary features $x = (x_1, x_2) \in \{0,1\}^2$ and define the target $f_\oplus(x) = x_1 \oplus x_2$.

First we will show that GNAN cannot express $f_\oplus$. A GNAN scores the graph by $\hat{y} = \sigma\left(\phi_1(x_1) + \phi_2(x_2)\right)$, where each ϕ_i is univariate. Put $a = \phi_1(0)$, $b = \phi_1(1)$, $c = \phi_2(0)$, $d = \phi_2(1)$. To match the XOR truth-table there must exist a threshold τ such that

$$a + c < \tau, \quad b + c > \tau, \quad a + d > \tau, \quad b + d < \tau.$$

Summing the first and last inequalities yields $a+b+c+d < 2\tau$, while the middle pair gives $a+b+c+d > 2\tau$—a contradiction. Thus no GNAN realises $f_\oplus$.

8

Extended Abstract Track

Now we will show that GMAN can express $f_\oplus$. Place the two features in the same subset $F = \{x_1, x_2\}$ and choose the subset-network

$$\psi_F(x_1, x_2) = x_1 + x_2 - 2x_1 x_2.$$

For the four binary inputs this mapping returns $(0, 1, 1, 0)$, exactly $f_\oplus$. Hence GMAN represents a function unattainable by GNAN, proving that GMAN is strictly more expressive.

Theorem 4 *Let S be a set of graphs $\{G_i\}_{j=1}^m$. Let $S_1 = \{S_i\}_{i=1}^m$ be a partition of S such that $|S_i| = 1$. Let $S_2 = \{S_i\}_{i=1}^k$ such that there exists k with $|S_k| > 1$. with a subset partition $\{S_i\}_{i=1}^k$. Then a GMAN trained over S_2 is strictly more expressive than a GMAN trained over S_1.*

Proof of Theorem 2 Let S be a set of graphs $\{G_i\}_{j=1}^m$. Let $S_1 = \{S_i\}_{i=1}^m$ be a partition of S such that $|S_i| = 1$. Let $S_2 = \{S_i\}_{i=1}^k$ such that there exists k with $|S_k| > 1$. with a subset partition $\{S_i\}_{i=1}^k$. We will prove that a GMAN trained over S_2 is strictly more expressive than a GMAN trained over S_1.

To prove this, we use a ground truth function that is a set-level XOR. Let every graph G_i carry a single binary feature $x_i \in \{0, 1\}$ and let the ExtGNAN encoder return this feature unchanged, i.e. $h(G_i) = x_i$. Denote a set containing two graphs by $S = \{G_1, G_2\}$ and define the permutation-invariant target

$$f_\oplus(S) = x_1 \oplus x_2.$$

Singleton partition (S_1). If each graph is placed in its own subset, GMAN aggregates *additively*: the model output is

$$\hat{y} = \phi(x_1) + \phi(x_2),$$

because the final GMAN stage simply sums subset scores. Write $a = \phi(0)$ and $b = \phi(1)$. To realise $f_\oplus$ via a threshold τ we would need

$$a + a < \tau, \quad b + a > \tau, \quad a + b > \tau, \quad b + b < \tau.$$

Adding the first and last inequalities yields $a + b < \tau$, while the middle pair gives $a + b > \tau$—a contradiction. Hence GMAN$_{S_1}$ cannot represent $f_\oplus$.

Paired partition (S_2) Group the two graphs together and use a DeepSet $\Phi(S_2) = g\left(\sum_{i=1}^2 f(x_i)\right)$ with $f(x) = x$ and $g(s) = s(2 - s)$. Then

$$g(x_1 + x_2) = \begin{cases} 0 & (x_1, x_2) = (0, 0) \text{ or } (1, 1), \\ 1 & (x_1, x_2) = (0, 1) \text{ or } (1, 0), \end{cases}$$

exactly $f_\oplus$. The final GMAN sum over feature channels leaves this value unchanged, so GMAN$_{S_2}$ realises $f_\oplus$.

Strict separation. Because $f_\oplus$ is representable by GMAN$_{S_2}$ but not by GMAN$_{S_1}$, the former is strictly more expressive.

Appendix C. Dataset Details

C.1. PhysioNet P12

We provide the full list of the 36 physiological signals and 3 static patient features used in our experiments.

1. Alkaline phosphatase (ALP): A liver- and bone-derived enzyme; elevations suggest cholestasis, bone disease, or hepatic injury.

2. Alanine transaminase (ALT): Hepatocellular enzyme; increased values mark acute or chronic liver cell damage.

3. Aspartate transaminase (AST): Enzyme in liver, heart, and muscle; rises indicate hepatocellular or muscular injury.

4. Albumin: Major plasma protein maintaining oncotic pressure and transport; low levels reflect inflammation, malnutrition, or liver dysfunction.

5. Blood urea nitrogen (BUN): End-product of protein catabolism cleared by the kidneys; elevation signals renal impairment or high catabolic state.

6. Bilirubin: Hemoglobin breakdown product processed by the liver; accumulation indicates hepatobiliary disease or hemolysis.

7. Cholesterol: Circulating lipid essential for membranes and hormones; dysregulation is linked to cardiovascular risk.

8. Creatinine: Waste from muscle metabolism filtered by the kidneys; higher levels imply reduced glomerular filtration.

9. Invasive diastolic arterial blood pressure (DiasABP): Pressure during ventricular relaxation; low readings may reflect vasodilation or hypovolemia.

10. Fraction of inspired oxygen (FiO_2): Proportion of oxygen delivered; values above ambient air denote supplemental therapy.

11. Glasgow Coma Score (GCS): Composite neurologic score for eye, verbal, and motor responses; scores ≤ 8 indicate severe impairment.

Extended Abstract Track

12. Glucose: Principal blood sugar; hypo- or hyper-glycemia can cause neurologic compromise and metabolic instability.

13. Serum bicarbonate (HCO$_3$): Key extracellular buffer; low levels signal metabolic acidosis, high levels metabolic alkalosis or compensation.

14. Hematocrit (HCT): Percentage of blood volume occupied by red cells; reduced values denote anemia, elevated values hemoconcentration.

15. Heart rate (HR): Beats per minute reflecting cardiac demand; tachycardia indicates stress or shock, bradycardia conduction disorders.

16. Serum potassium (K): Crucial intracellular cation; deviations predispose to dangerous arrhythmias.

17. Lactate: By-product of anaerobic metabolism; elevation marks tissue hypoxia and shock severity.

18. Invasive mean arterial blood pressure (MAP): Time-weighted average arterial pressure; low values threaten organ perfusion.

19. Mechanical ventilation flag (MechVent): Binary indicator of ventilatory support; presence denotes respiratory failure or peri-operative care.

20. Serum magnesium (Mg): Cofactor for numerous enzymatic reactions; abnormalities contribute to arrhythmias and neuromuscular instability.

21. Non-invasive diastolic arterial blood pressure (NIDiasABP): Cuff-derived diastolic pressure; trends mirror vascular tone without an arterial line.

22. Non-invasive mean arterial blood pressure (NIMAP): Cuff-based mean pressure; used when invasive monitoring is unavailable.

23. Non-invasive systolic arterial blood pressure (NISysABP): Cuff-derived systolic pressure; elevations suggest hypertension or pain response.

24. Serum sodium (Na): Principal extracellular cation governing osmolality; dysnatremias cause neurologic symptoms and fluid shifts.

25. Partial pressure of arterial carbon dioxide ($PaCO_2$): Indicator of ventilatory status; hypercapnia implies hypoventilation, hypocapnia hyperventilation.

26. Partial pressure of arterial oxygen (PaO_2): Measure of oxygenation efficiency; low values denote hypoxemia.

27. Arterial pH: Measure of hydrogen-ion concentration; deviations from normal reflect systemic acid–base disorders.

28. Platelet count (Platelets): Thrombocyte concentration essential for hemostasis; low counts increase bleeding risk, high counts thrombosis risk.

29. Respiration rate (RespRate): Breaths per minute; tachypnea signals metabolic acidosis or hypoxia, bradypnea central depression.

30. Hemoglobin oxygen saturation (SaO_2): Percentage of hemoglobin bound to oxygen; values below normal indicate significant hypoxemia.

31. Invasive systolic arterial blood pressure (SysABP): Peak pressure during ventricular ejection; extremes compromise end-organ perfusion.

32. Body temperature: Core temperature; fever suggests infection, hypothermia exposure or metabolic dysfunction.

33. Troponin I: Cardiac-specific regulatory protein; elevation confirms myocardial injury.

34. Troponin T: Isoform of cardiac troponin complex; rise parallels Troponin I in detecting myocardial necrosis.

35. Urine: Hourly urine volume as a gauge of renal perfusion; oliguria signals kidney hypoperfusion or failure.

36. White blood cell count (WBC): Reflects immune activity; leukocytosis suggests infection or stress, leukopenia marrow suppression or severe sepsis.

Static patient features: Age; Gender; *ICUType* – categorical code for the admitting intensive care unit (1 = Coronary Care, 2 = Cardiac Surgery Recovery, 3 = Medical ICU, 4 = Surgical ICU), capturing differences in case mix and treatment environment.

C.2. Biomarker Subsets

In the PhysioNet P12 task, we grouped the 36 physiological signals into one multivariate subset and 29 singleton subsets. Domain knowledge showed that only the respiratory and gas-exchange variables shared sufficiently strong, coherent dynamics to benefit from joint modeling. All other signals were physiologically diverse, so they were left as singletons to retain their unique predictive information.

1. **Respiratory gas exchange and ventilation**
   [FiO$_2$, PaO$_2$, PaCO$_2$, SaO$_2$, RespRate, pH, MechVent]
   These variables collectively describe oxygen delivery (FiO$_2$), pulmonary gas exchange efficiency (PaO$_2$, SaO$_2$), ventilatory adequacy (PaCO$_2$, RespRate, MechVent), and the resulting systemic acid–base balance (pH). Grouping them lets the model learn the tightly coupled patterns that arise during hypoxemia, hypercapnia, mechanical ventilation adjustments, and respiratory failure—yielding a more coherent representation of a patient's real-time respiratory status.

2. **Singleton biomarkers**
   Each remaining signal represents a distinct physiological domain (hepatic, renal, hematologic, hemodynamic, neurologic, metabolic, or cardiac). Their organ-specific pathophysiology favored treating them individually, preserving granular patterns while keeping the grouping scheme simple and interpretable.

Appendix D. Experimental Setup and Hyperparameter Choices

In-Hospital Mortality (P12) Experiments   We trained all PhysioNet12 models for a maximum of 500 epochs using the Adam optimizer with weight decay of 1e-4. We used a ReduceLROnPlateau scheduler with a max learning rate in the {1e-3, 1e-5} range, min learning rate of 1e-5, factor of 0.5 and patience=20.

We trained GMAN models with batch size of range {16, 32}, dropout of 0.2, n_layers in the {3, 5} range, hidden_channels in the {32, 64} range, num_lab_ids_embed in the {5, 8} range, num_biom_embed in the {3, 5} range, num_units_embed in the {3, 5} range.

Fake-News Detection (GosspiCop) Experiments   We trained all GNAM GosspiCop models for a maximum of 400 epochs using the Adam optimizer with weight decay of 1e-4. We used a ReduceLROnPlateau scheduler with a max learning rate in the {1e-3, 5e-5} range, min learning rate of 1e-8, factor of 0.5 and patience=20.
We trained GMAN models with batch size of 16, dropout in {0.0, 0.5} range, n_layers in the {3, 5} range, hidden_channels in the {16, 128} range.
Random seeds were fixed for reproducibility, and results are reported across three independent runs. All models were trained on a single NVIDIA Quadro RTX 8000 (48GB) GPU.