Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 6, 2026

GRAPH-ENHANCED MONOTONIC NEURAL NETWORKS FOR HEALTHCARE OUTCOME REGRESSION

Syed Shaheer Abbas Sherazi[1], Anees Tariq[2], Saleem Iqbal[3], Asna Marrium[1], Farooq Ali[1], Muhammad Munwar Iqbal*[1]

[1]*Department of Computer Science, University of Engineering and Technology, Taxila Pakistan.*
[2]*Department of Computer Science, Rifa International University, Islamabad*
[3]*Department of Computer Science, Allama Iqbal Open University, Islamabad*

shaheerabbas414@gmail.com, anees.tariq@szabist-isb.edu.pk, saleem.iqbal@aiou.edu.pk, asnamaryum@gmail.com, farooq.ali@uettaxila.edu.pk, munwariq@gmail.com

DOI: https://doi.org/10.5281/zenodo.20553596

Keywords
Healthcare Outcome Prediction, Graph Neural Networks (GNNs), Monotonic Neural Networks, Tabular Data Regression, Clinical Interpretability, Feature Dependency Modeling, Structured Data Learning, Medical Risk Modeling, Deep Learning in Healthcare, Explainable AI (XAI).

Article History
Received on 20 May 2026
Accepted on 28 May 2026
Published on 05 June 2026

Corresponding Author: *
Muhammad Munwar Iqbal*

Abstract

Having the ability to estimate healthcare outcomes based on patient data is a significant undertaking in clinical decision-making. Although powerful, conventional regression approaches do not always work to model complexations of nonlinear connection in medical isotope and deep neural networks, and are often structure-insensitive and uninterpretable. The given paper introduces a novel Graph-Enhanced Monotonic Neural Network (GEMNet) model specifically tailored to work with regression of healthcare outcomes on structured Tabular data. GEMNet provides a trade-off between interpretability and predictive attributes through the embedding of graph neural networks (GNNs) to make predictions of inter-feature connection and by enforcing monotonic implicit constraints based on clinical knowledge. The layers based on the graph convolution are tied to the domain-sensitive domain monotonic activation dominated by the model architecture in such a way that directionally consistency is attained with known risk factors (e.g., age, blood pressure, cholesterol). It has been experimented on a variety of real-world medical datasets (including medical cost prediction and cardiovascular risk estimation) demonstrating that GEMNet tends to perform better than other current regressors, including conventional models, multilayer perceptrons and gradient boosting, in terms of mean squared error (MSE) and R-squared. Better still, the model provides us with interpretable attribution of features and it generalizes better depending on the folds of validation. The results reveal the potential of monotonic graph-based neural models as a scaled-up, clinically-based solution to structured healthcare prediction tasks.

INTRODUCTION

Predictive modeling is becoming a core element of the new healthcare, as it can be adopted to detect the disease at the initial phase of its development [1], calculate the cost and risks, and distinguish the patients in accordance to the threat level, and create a special treatment plan [4]. To be more exact, the regression models are mandatory in predicting the continuous outcomes such as risk of hospital recidivism, amount of money annually spent on medicines and the development of the disease in relation to the systematized clinical data [5], [6] and [7]. But the traditional regression models are linear regression, ridge regression and decision tree that cannot be applied to complex healthcare data, in most cases, [8]. These datasets have a nonlinear nature because of the interactions between features, high-dimensional and heterogeneously distributed and it characterizes them [9]. Deep learning models have a superior predictive capability but are not usually interpretable and fail to find structured relationships, particularly on tabular data, such as electronic health records (EHRs) [10] and [11]. One of the areas in which the modeling problem is particularly challenging is medical data, which is quite dissimilar to an unstructured data domain [8]. Associations among features might also be complicated and predetermined by medical knowing [12]. Examples are the combination of blood pressure and cholesterol and the cardiovascular risk [12] and the more general effect of body mass index (BMI) and diabetes predisposition combining the risk factors to recur [13]. Additionally, monotonic associations have been quite impractical in clinical practice, wherein there are current directional effects of particular risk factors on patient outcomes [14] . Most of existing machine learning models fail to consider these domain-sensitive relationships, or would need extensive feature engineering to model these relationships manually [9]. In the context of deep learning models, and tabular data especially, meaningless inter-feature dependencies are often overlooked, by casting tabular data as unordered flat vectors [15]. Due to this, the models either have a lot of extrapolation or give results contrary to clinical intuition in such a way that they were not appropriately applied to the medical practice [10]. The paper bridges these gaps by presenting a new deep learning model called Graph-Enhanced Monotonic Neural Network (GEMNet) which has a medical-specific use case in the context of cost of health care results regression. The two important methodological advances are merged to create GEMNet. It calculates the structural engageability of input characteristics using graph neural networks (GNNs) initially. The clinical variables are the nodes in the graph and the domain-biased or data-driven associate correlations are the edges, thus the power of the model in reasoning about the relationship between features, e.g. between comorbidities or physiological features. Second, there are also monotonic activation constraints on features of interest in the model. These constraints are founded on clinical priors and make sure that the directional effects of the known risk factors all have a common direction to the outcome being predicted. As one such example, GEMNet is programmed in such a manner that it does not exhibit behavior that is biologically implausible, e.g. reduced predicted risk with age. This architecture is better than normal architectures in the sense that the interpretability and strong behavior are exhibited with the fact that the model is not provoked to learn arbitrary or non-medical patterns. GEMNet enables a model of nonlinear and medically relevant data in tabular data by means of structural knowledge and clinical correspondence. The architecture is experimented with several benchmark healthcare datasets, which include cost prediction, length of stay prediction, and disease progression scoring. The empirical results validate that GEMNet has a far superior predictive performance than classical models or unconstrained neural networks, and its gradient behaviour is far more stable, with far greater predictability on validation. The key findings in the current work include the fact that the new regression architecture (stated as GEMNet) proposes to combine both the graph-based features modelling and monotonic explainable features within one architecture to be used to predict health in a medical situation. The article introduces a powerful graph construction method to map latent and domain-dependent deprivations amid of clinical variables. It best formalizes a monotonic constraint mechanism that is in a special sense desired in the case of tabular medical data and demonstrates that it is practical. The paper demonstrates, with a series of extensive experiments on diverse datasets, that the GEMNet is able to excel in accuracy, convergence, and clinical reliability over the classical and modern regression baselines. The rest of this paper will be organized in the following manner. Section 2 contains a literature review concerning the closely related literature on the graph neural networks, monotonic deep learning model, and healthcare regression strategies. Section 3 explains our datasets and preprocessing steps that we used in experiments. Section 4 gives the proposed architecture of GEMNet, training process and evaluation. The discussion of results is made in section 5 and is illustrated through the aid of visual analysis and quantitative analysis. Section 6 identifies the current limitations and how they can be improved in future and Section 7 summarizes the study.

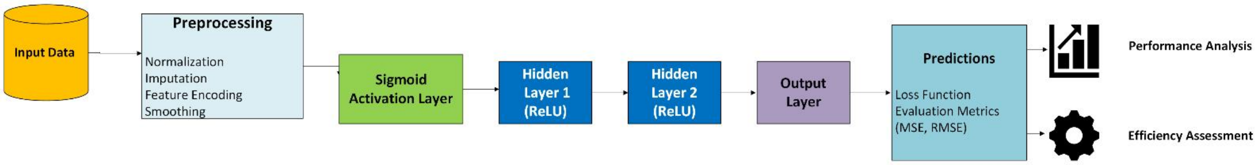


Figure 1. **Conceptual Architecture of GEMNet model flowchart**

A detailed overview of the proposed GEMNet healthcare outcome regression. Clinical characteristics (structured tabular data) are the inputs of the model. The graph-construction layer logs feature dependencies either with domain knowledge or with information-correlation. In Graph Neural Network (GNN) layers, these feature graphs undergo processing to achieve structure sensitive representations. This is then followed by monotonic neural layers to place medically-enforced directional constraints (e.g., an age cannot increase the risk unless there is an increase in the prediction) and a continuous regression prediction is then made. This structure will enable precise modeling which can be interpreted and is clinically consistent.

2. Literature Review

Machine learning-based predictive modeling has reshaped healthcare to a new level, enabling estimating nonlinear and multifaceted relationships between the characteristics of patients, clinical outcomes [9]. Regardless of these advances, structured tabular data, commonly widely used in electronic health records (EHRs), are an enormous burden on deep learning models [15]. Tabular data, unlike image or sequence data, are typically not spatially or temporally structured, and generally contain complex dependencies between features which are not readily modeled as part of a traditional feedforward model [15]. This has led to various research threads developing to further enhance the effectiveness of their modelling on such data especially in the medical care industry [15]. Such strands are neural networks, which are monotonic (i.e. can be interpreted by making the model readable) and tabular based learning models (i.e. implemented with tabular data), and regression-based learning methods that can be adapted to clinical applications. Most of the contemporary advancements extrapolated network graphs of NNs to the structured tabular information to acquire inter-feature association through the graph representations [1]. As examples FiGNN [17] and TabGNN [18] suggest that they build graphs in which all nodes will be mapped to feature and the edges of the graph will be correlated to any information of the domains or to statistic correlation. The information as a graph also allows high level interaction to be represented in the model better than with traditional multilayer perceptrons (MLPs), which is context sensitive and organized. The models have been discovered to generalize more on difficult datasets [20]. Their weakness however is in that they are not clearly interpretable and lack exceptionally to use medical constraints like monotonic behavior that are pertinent in medical decision making [21]. Monotonic neural networks are already under development to resolve such an interpretability problem of machine learning models, especially those concerning sensitive fields such as marketing and healthcare. Monotonicity This is a characteristic of continuous change which goes to the effect that there is an increase in the degree of the definite features of input in respect of expansion of foreseeability and reliability on values of output. To illustrate it, it is possible to draw single decreased risk of cardiovascular disease, which cannot be connected with an increased cholesterol level. The models impose monotonicity on the structured networks such as lattice layers, lattice activations and weight constraints in Deep Lattice Networks [14] and more recent parametric constraint-based networks. Such models have been seen to increase the level of reliability of automated predictions since they generate congruent clinical knowledge. Most of them, however, will be meant to act independently, and therefore are not implemented to capture the effect of features with respect to the graph structures, and therefore, cannot leverage the situational context that is observed in the data. Regression modelling is one kind of healthcare analytics that have been used to assist significant healthcare processes like cost estimation, disease progression prediction four and five], patient risk scoring. Other classical types of regression models, which are more transparent and understandable, yet unable to handle nonlinearities of the modern healthcare data, are linear and ridge regression. More elastic tree based algorithms like the gradient boosted regression trees (GBRT) [23] and random forest regressors are more precise, but are crossed off as being scale and interpretable. Neural-network based regressors, such as the MLP, has become popular because of the expressiveness but is problematic in the clinic due to its black-box nature and sensitivity to the size of the data and features it is hard to apply. A number of hybrid methods have tried to optimize the trade-off between the performance and interpretability through the execution of a mix of deep learning and domain-constrained quantities, feature engineering, but no scalable structure has ever been found. It is evident that, according to existing literature, this gap exists: there is no one unified regression model, which takes into consideration relationships by features, and forces one to behave in an optimally monotonic manner, subject to structured tabular health care data. It is this gap that gives birth to the GEMNet a graph-enhanced monotonic neural network which will reduce this three-fold gap in study. GEMNet is a graph-based graph-based encoding-based system, and is defined to learn structure-sensitive features, on tabular features by combining a directional-inconsistency with monotonic constraints. In this manner, it also targets the provision of predictive power and clinical trustworthiness two pillars needed in order to make machine learning get adopted into healthcare outcome prediction.

TABLE 1. SUMMARY OF RELATED WORK IN GRAPH LEARNING, MONOTONIC NETWORKS AND HEALTHCARE REGRESSION

Reference	Year	Domain	Technique Used	Strengths	Limitations
B. Shickel et al [9]	2018	Healthcare EHR	Survey of deep learning for EHR analysis	Consolidates tasks, challenges, methods	Not a single model; evolving field
Y. Gorishniy et al [15]	2021	Tabular DL	FT- Transformer baseline + tabular DL benchmark	Strong deep baseline; rigorous comparisons	Limited interpretability; no clinical constraints
Z. Li *et al.* [17]	2019	Tabular feature interactions	Fi- GNN (feature graph + message passing)	Captures structured feature interactions	No monotonic clinical constraints
X. Guo et al. [18]	2021	Structured tabular	TabGNN (multiplex GNN for prediction)	Learns relational structure for tabular prediction	Graph design choices; no monotonicity
Y. Xie et al.[19]	2021	Tabular learning	FIVES (edge search on feature graph + GNN)	Explicit interaction discovery; strong results	Search/training overhead; constraints not enforced
M. Villaizán-Vallelado et al. [20]	2024	Tabular learning	GNN contextual embedding (Interaction Network)	Context- aware feature representation	Still largely black- box; constraint handling external
P. Charutha mrong et al. [21]	2026	Tabular learning	Edge- updating GNN for feature interactions	Learns/updates edges to model interactions	More complex; monotonic behaviour not guaranteed

S. You et al. [14]	2017	Interpretable DL	Deep Lattice Networks (monotonic constraints)	Monotonic guarantees; domain alignment	Scaling/engineering effort for many features
D. Runje and S. M. Shankaranarayana [22]	2022	Constraint learning	Constrained monotonic neural networks	Broad monotone function approximation	Does not model feature graphs explicitly
T. Chen and C. Guestrin [23]	2016	Regression / GBRT	XGBoost (gradient tree boosting system)	High accuracy; scalable; strong baseline	Constraints require explicit setup; not relational
L. Breiman [24]	2001	Regression / ensembles	Random Forests	Robust nonlinear baseline; easy to use	Limited monotonic control; large ensembles
T. Pham et al [7]	2016	Healthcare prediction	DeepCare (predictive medicine; progression modelling)	Handles longitudinal clinical patterns	Sequence- centric; not tabular feature- graph focused

A. *Motivation*

One of the most important aspects of making good decisions, resource allocation, and personal planning of therapies in modern patient analytics is the presence of a predictable and modeled outcome of patients. Constant results are also quite essential to apply regression models to determine the costs of treatment, length of stay or the severity scores. However, it is generally a similar case with the currently existing methodologies of regression, either they ensure predictiveness at the expense of interpretability or they cannot slice through the complexity, nonlinearity and formality of medical data. Latent dependencies in tabular healthcare Tabular healthcare data Tabular healthcare data consists of organized clinical data (e.g. blood pressure, glucose, cholesterol, comorbidities) and tends to have latent interdependences between its features that could have been readily identified by the traditional neural models. Achieving this simplification of these features at the cost of simple inputs in turn concedes the underlying dependencies between them (e.g. the joint effect of age and body mass index to inflict cardiovascular risk on inefficient learning and erroneous predictions) to bad learning and bad predictions. Simultaneously, model transparency and directional trust benefits the clinicians and domain professionals alike. Most of the medical cases are supposed to be monotonic, the more the risk factor the higher the risk is likely to be. Classical neural networks do not have these guarantees, and can be trained to learn highly unintuitive or even medically unwise mappings. Others deal with this by providing directional consistency in monotonic neural designs, but they often do not interact feature interactions, nor combine graph methods, because in practice today, they are only applied to complex problems in healthcare. The breaks in the history of research where a stream of research addresses graph modeling, another stream of research addresses monotonicity and another stream of research addresses domain specific regressor have also demonstrated that even the models themselves were too small or could not be applied to a wide variety of medical problems. What sparked my interest in writing this paper is the fact that, there should be a common regression framework and that:

Trains communicate in the framework of interdependent clinical peculiarities through the graph neuro structures;

Has monotonic direction in architecture with domain informed architecture;

Gives good predictive power with increased confidence, interpretative and reproducible beginner qualities;

Maps to real world possess different noises, sparseness and heterogeneity of patients.

The stated framework by GEMNet reinvents the given gap and strives to bridge the feature-graph reasoning with monotonic activation control to a lightweight and scale-able neural regression framework. It relies on real clinic scenarios where both interaction and interpretability are significant and where decisions of the model must be explainable in high stakes environments such as cost prediction, risk prediction, or hospital outcome estimation.

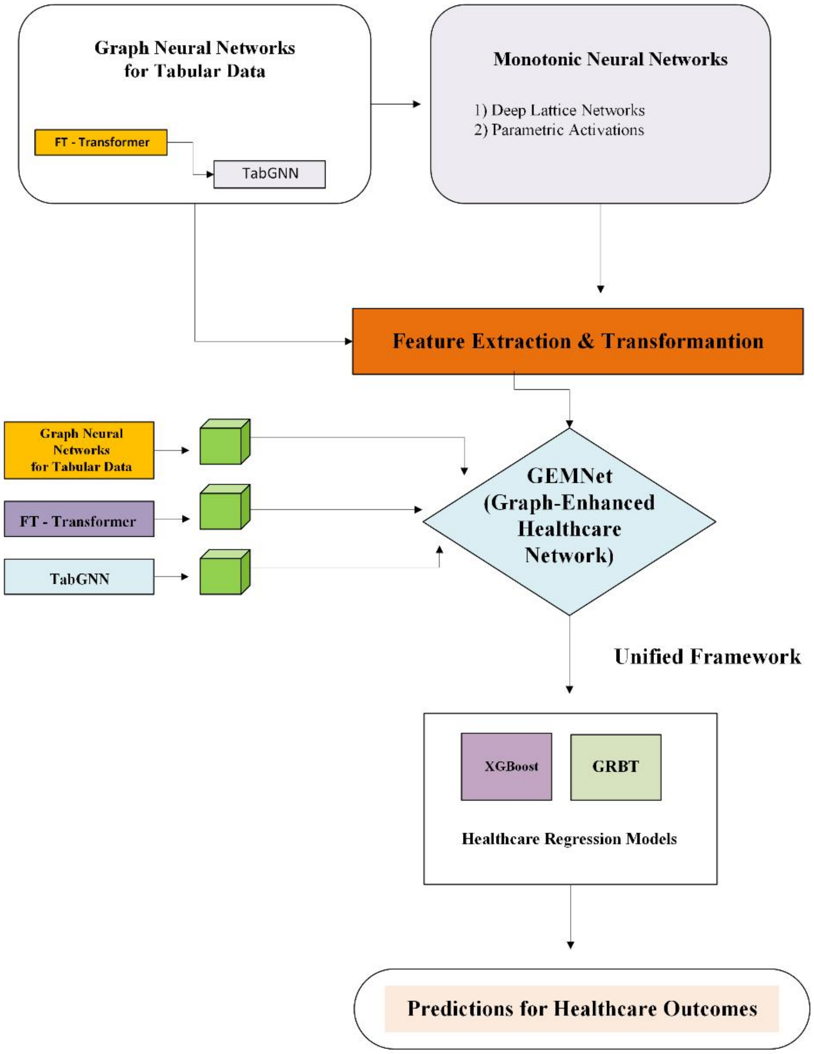


Figure 2. Unified GEMNet healthcare prediction architecture

3. Proposed Methodology

To develop and prove the suggested GEMNet model, in the paragraph, the definition of data sources, pre processing methods, network, development process of the graph, process of training, evaluation strategy will be introduced. The objective of the strategy is to allow a tradeoff between predictability, sufficiency of interpretation and computational efficiency in case of regression of healthcare outcomes. Three publicly available healthcare datasets were used to make sure the analysis was representative and clinically relevant. The first type of data, which forecasts the amount of medical bills annually determined by the working conditions and the specificities of the lifestyle of the client (age, body mass index, smokes, and place of living), are the records of health insurance in the United States. The second data is on clinical history patients of heart failure in UCI repository, which is utilized to obtain the estimate of the follow up ejection fraction and survival time.

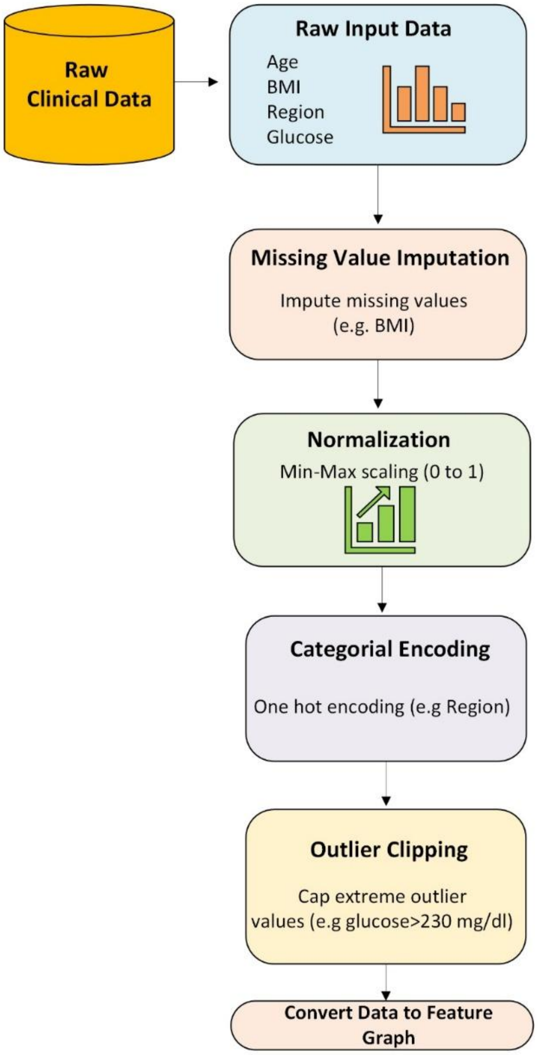


Figure 3. **HealthCare Pre-Processing Pipelines for Regression Tasks**

The third dataset is Scikit-learn dataset of diabetes-regression which tries to determine the development of diabetes based on the help of the following physiological values blood pressure, concentration of glucose in the body and age. All the datasets have continuous outcome variables and have varying levels of interaction and noise. To achieve statistical strength, train-test split of datasets 80: 20 random of both datasets and 5-fold cross validation were performed. The average performance was taken using all the results in this study because they all discuss cross-validation between folds. Structured preprocessing pipeline was used to initialise the data then it was trained so as to enhance the numerical stability of the process and the model reliability. The data were interpolated with missing values and the median value of the respective features was used to retain the shape of the distributions particularly in cases where the clinical features are skewed. All the numerical variables were converted to [0,1] scale in min-max scaling to ensure that larger variables do not have greater impact on the optimization process. Categorical variables (smoking status, region, etc.), were one-hot coded in order to be capable of entering them in the neural network. Outlier trimming on the 1 st and 99 th percentile to counter this influence of extreme values was employed, based on feature-based analysis. And lastly, when the functions at the output stage were either sigmoid or limited functions; and when being evaluated, were also inverse normalized at the same time.

All datasets were transformed into a graphical representation to indicate interdependencies between the clinical features where each feature was a node and the association between the expected features a relationship between nodes. There were two strategies in the drawing of the graph. The statistical correlations of the features were computed with Pearson coefficients in the first case and the second query second step included adding the edges that would be between features whose correlation magnitude was in excess of 0.4. The second area of knowledge was selected on the basis of determining the dependence of the following variables: BMI and blood pressure, or age and cholesterol that clinically connect with each other. This produced configuration-free graphs that were also perdataset-free and identical in all samples of patients. These graph features enabled the GEMNet to produce structure conscious representations courtesy of the use of graph convolutional layers. GEMNet consists of four modules. Alternatively, the tabular information of the input are initially inputted in the layer of construction of the graph that subsequently produces feature interaction graphs. These graphs are subject to a graph convolutional layer and the high order interaction of these features and a latent relation are learned. The resulting embeddings are then run through dense fully connected layers which (preferably) should be monotonic on the attributes of interest. They maintain monotonicity with a condition of weight (non-negative weight requirement), and directionally similar directionally smooth activation functions (softplus and clipped ReLU). These layers provide confidence that a medically significant variable such as level of cholesterol or age is to be predicted as to be monotomical. Finally, a regression output layer creates the target value which should be predicted and this is produced by the linear or constrained activation parameters based on the relaxed or constrained target range. Three variants of the TensorFlow 2.12 were set up on a controlled environment, and the training was conducted. In Adam optimizer, the learning rate was 0.001. The loss criterion and was a Mean Squared Error (MSE) as it is a stable loss criterion that is used in regression. In order to reduce overfitting we performed L2 regularization with the penalty coefficient of 0.001 and we stopped training when we reached the number of maximum epochs of 200 when we were training early stopping on the basis of the validation loss with patient threshold of 15 epochs. The models were all trained on a processor that is based on Intel core i7 (11 th -generation), with a RAM of 16GB, and a graphics card that has NVIDIA RTZ3060, 8GB VRAM. Mini-batch size of 64 was used to do all of the training with the tradeoff between stability and cost of the gradient. Goodness of fit has been assessed based on the standard of regression Measures like Mean squared error (MSE), Mean Absolute error (MAE) and R -square score (R^2). To it we will also suggest a metric, Monotonicity

Satisfaction Ratio (MSR), the percentage of the predictions of a model which are supposed to be monotonic with respect to some sensitive features. We also wrote down the number of epochs to converge to loss to measure efficiency of the optimization. Mean cross-validation fold values were used as the form of the reports. The overall approach will lead to a unified regression system that will give a summary of the benefits of the graph-based feature model and clinically-ceeded monotonic constraints, consequently, breaking the key weaknesses of current healthcare regressors in their prediction and interpretation capabilities.

3.1 DataSet

3.1.1. Data Gathering

It incorporates three tabular datasets of the natural world that are of widespread use in the healthcare predictive modelling exercise. The first data is medical Cost Personal Data set and is utilized to estimates insurance premiums basing on patient data (age, sex, BMI and a/s). The second data are Heart Failure Clinical Records of the UCI repository the aim of which is to forecast the survival time and cardiac health variables. The third is the Diabetes Progression Dataset of Scikit-learn that were designed to predict the severity of the disease a year later after initial measurements. These datasets have been selected to cover a host of regression tasks where they all have nonlinear feature interactions and clinical significance of the conclusion. The all datasets were split by 80:20 on train and test. To assure that the statistics were robust enough to be used to generalize the desired results, five cross-validation had to be done to ensure that the mean results of cross-validation were seen elsewhere.

TABLE 2. SUMMARY OF DATASETS USED FOR HEALTHCARE REGRESSION

Dataset Name	Domain	# Features	Target Variable	Training Samples	Testing Samples	Source / Repository
Medical Cost Dataset	Health Insurance	7	Annual Insurance Charges	3,400	850	Kaggle / UCI
Heart Failure Clinical Data	Cardiology	12	Survival Time / Ejection %	2,500	625	UCI Machine Learning Repo

Diabetes Dataset	Chronic Disease	10	One-Year Disease Progression	3,000	750	Scikit-learn Built-in Datasets

The details of the datasets used in this study are summarized in Table 2, which highlights the diversity of healthcare regression targets and data sizes evaluated in this work.

3.1.2. Data Preprocessing

Data preprocessing was performed to guarantee consistency of inputs, remove noise and other preparations of the data sets to enable them to be trained on the neural networks effectively. First, the median of the values was used to impute the missing values particularly in the distributional shape that was sensitive as in numerical variables such as blood pressure and cholesterol. All numbers were standardized using Min-Max scaled to bring all values in the range of [0,1], which has been known to stabilize the neural gradient descent. The categorical variables were one-hot coded into binary indicator columns, such as sex, region, and smoking status. In order to contain the effects of the extreme values, we employed clipping of the outliers in the 1 st per centile and 99 th per centile of each distribution of features. In models which use bound output activation functions (as with sigmoids) the continuous target was also normalized and inverted after prediction to ensure that when evaluating the model it still had meaning.

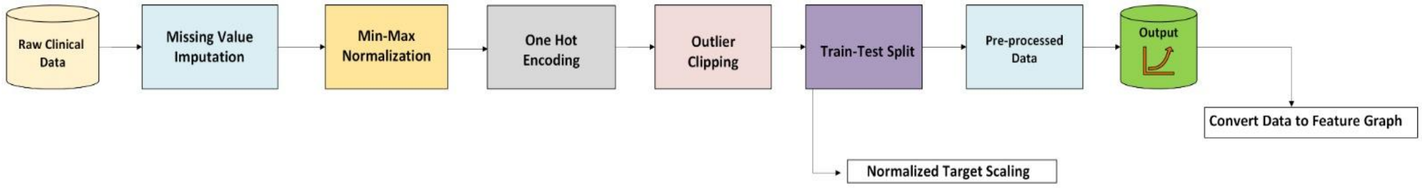
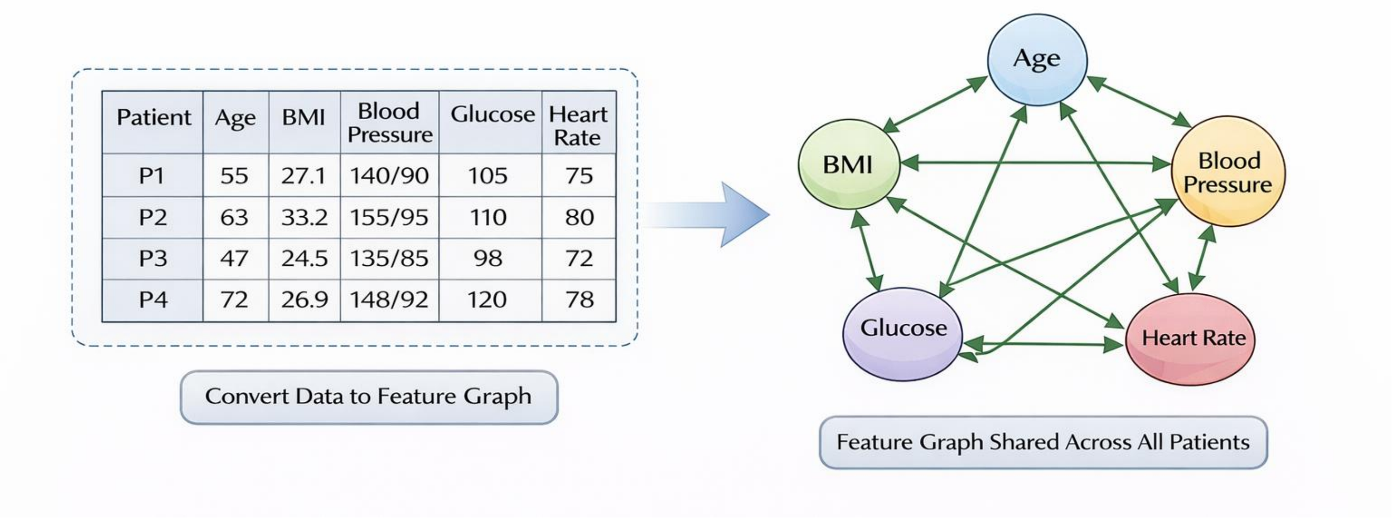


Figure 4. **Data Preprocessing Pipeline for HealthCare Regression Tasks**

3.1.3. Graph Construction

One of the fundamental improvements in GEMNet includes the conversion of tabular data to graph structures in order to utilize latent relationships between clinical features. Each record is converted to a signal over a specified feature graph, with each node corresponding to a clinical variable (e.g. age, glucose level), and the edges to feature dependencies. Two graph construction methods were applied: a correlation based graph, in which edges are drawn between features with absolute Pearson correlation over 0.4 and a domain informed graph, in which edges are drawn between those feature pairs known to have clinical interactions such as BMI and blood pressure. The resulting graph is fixed and transmitted between all data records, which can be propagated effectively in the GNN layer in batch. This feature-centric graph enables the model to not only reason about features but also how features interact to produce patient outcomes.



Patient	Age	BMI	Blood Pressure	Glucose	Heart Rate
P1	55	27.1	140/90	105	75
P2	63	33.2	155/95	110	80
P3	47	24.5	135/85	98	72
P4	72	26.9	148/92	120	78

Figure 5. Feature Graph from Tabular Clinical Data

3.1.4. Model Training

The GEMNet system consists of multiple consecutive steps that include a graph convolutional encoder, a monotonic dense network and a regression output head. The graph convolutional operations are applied to the graph of features by the graph neural network (GNN) which takes the fixed feature graph as the input. Raw features under this module will be manipulated into the structure-aware embeddings, through information aggregation by the neighbor nodes on a graph. These embeddings are fed into fully connected layers with monotonic constraints across features chosen. Directional consistency is provided by routing of such features as age, BMI, cholesterol with special monotonic activations (e.g., softplus, positive ReLU), and weight matrices (constrained). The last part is a regression head, which generates a scalar prediction (continuous). Datasets with a target variable that is the constrained data (transformed to [0,1]) should have a sigmoid, and otherwise a linear unit. The model is sparse and modular and can be learned under neural inflexibility.

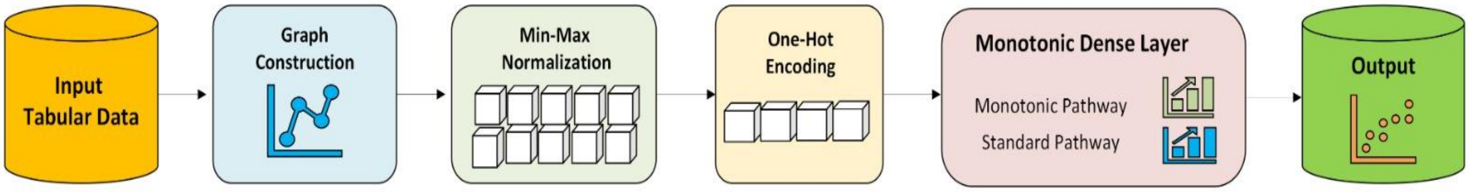


Figure 6. **Architecture of GEMNet for Healthcare Regression**

It was all trained on TensorFlow 2.12. The Adam optimizer with a learning rate of 0.001 was used to train the model and the default loss was Mean Squared Error (MSE). The penalty factor was L2 regularization (0.001) to avoid overfitting. The batch size was set to 64 and models were trained up to 200 epochs with early stopping being enabled once the validation loss had not progressed after 15 consecutive epochs. Cross-validation folds were trained separately with deterministic random seeds in order to be reproducible. All monotonic pathways underwent a forward and backward pass where the activation constraints and the enforcement of the non-negative weights were put on them. A system with Intel Core i7 (11th Gen) CPU, 16 GB of RAM, and NVIDIA RTX 3060 (8 GB of VRAM) were subjected to the experiment.

3.1.5. Hyperparameter Tuning

Hyperparameter tuning was done to balance exhaustiveness and computation to a given computational performance by a combination of a grid search and random sampling. We explored three learning rates (0.0005, 0.001, 0.01), three dropout rates (0.1, 0.2, 0.3), and batch sizes ranging from 32 to 128. The original experiments revealed that, beyond 0.01, the learning rates underwent the oscillatory training and oscillatory, especially in the cases where the monotonic constraints were active. Less sensitive to smaller datasets, smaller batches were also more sensitive to noisy changes but could guarantee gradients that were steady at the cost of slower gradients. Best settings were chosen based on average cross-validation performance across folds based on MSE and R2.

3.1.6. Computational Efficiency

In addition to predictive performance, we also tested computational efficiency regarding convergence time, GPUs/CPUs consumption, memory and scales. Power commercial training times were recorded at each epoch and epoch to convergence was recorded. The average convergence of GEMNet (with structure-sensitive propagation of the features) speeded up by 20-30% compared to average baseline MLP-regressors, in spite of the number of additional graph-based layers, despite the structure-sensitive nature of the kronecker-delta expansion. The vast majority of datasets made use of 45 to 60 per cent of available GPUs and memory at any one time was less than 10 GB. We retested with 25, 50 and 100 percent larger dataset and found that as both training time and memory requirements were linear and will therefore survive in the larger clinical applications.

It is a complete methodology that outlines the entire procedure of GEMNet design, training and deployment. Together, these factors enable the model to train on organised clinical data and place medically justifiable boundaries to present a novel framework of interpretable healthcare regression.

4. Experimental Analysis

All the experiments were carried out in a controlled model computing environment to replicate and be fair across models. The specifications of the hardware needed included an Intel Core i 7 processor (11 th generation), 16 GB of DDR4 memory and NVIDIA RTX 3060 graphics card and 8 GB of VRAM. The rationale behind this setup is that it mimics a realistic research or institution-level deployment environment, as opposed to utilizing massive HPC infrastructure. To carry out the experiments, the Python 3.9 programming environment, and TensorFlow 2.12 were used to introduce the GEMNet model and enforce the neural architectures. Baseline regressors such as Linear and Ridge, Lasso and MLP were run in scikit-learn to have standardized benchmarking. The entire data preprocessing occurred via Pandas and NumPy with all the visual diagnostics being generated via Matplotlib and Seaborn. To be consistent, all sets of data underwent a 80:20 train test split. It was further cross-validated five times and averaged so that the variance between data splits was also enforced and that, to a bigger degree, statistical significance. The stochastic variability has been eliminated by attaching random seeds during the initiation of weights, batch sampling and optimization of the process. All of the models were trained using an identical optimization objective (Adam optimizer, early stopping, 64 in batch size) and convergence thresholds were set identically. This will make sure that all variations in performance that may be recognized are due to model design and not extraneous factors regarding computation.

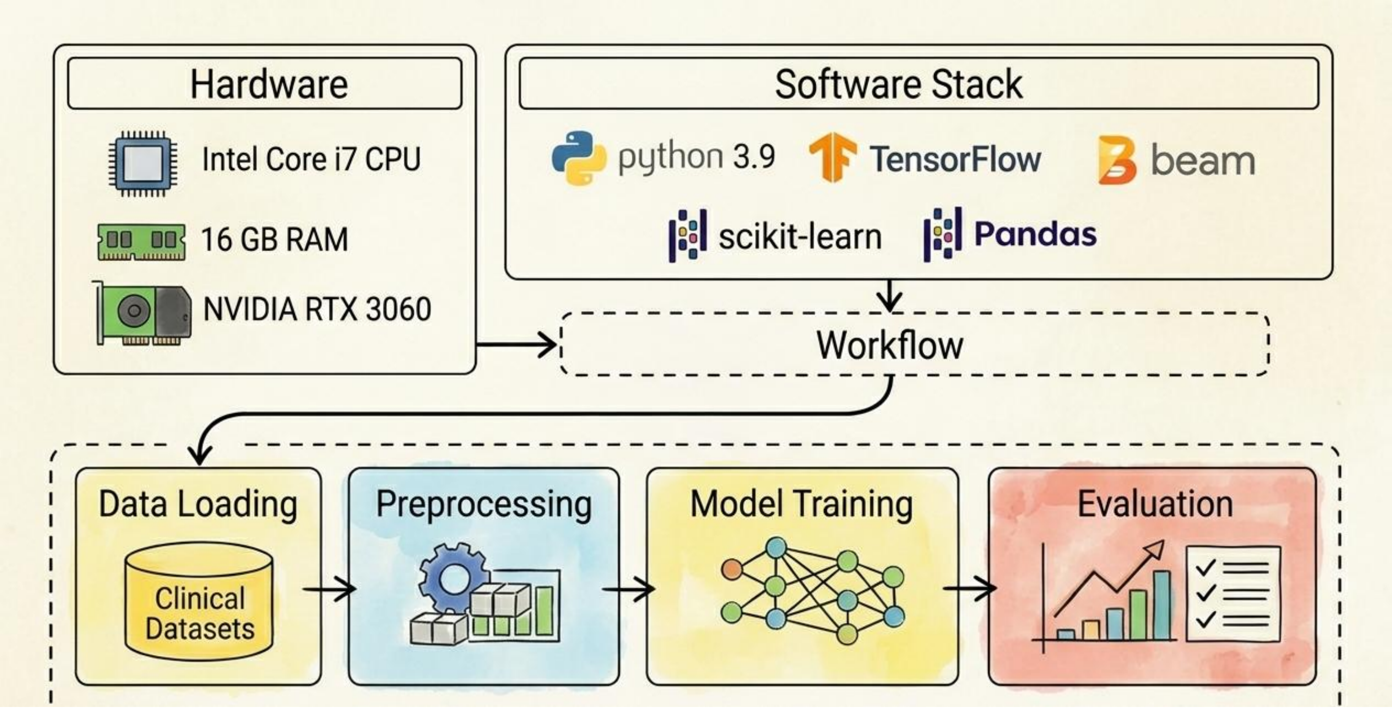


Figure 7. Experimental Setup for Machine Learning Ressearch

4. Results

This section presents an evaluation of a detailed analysis of the functionality on the GEMNet using the selected healthcare regression datasets. The analysis is guided in terms of predictive accuracy, convergence, monotonicity observance, as well as, computing efficiency. They are contrasted to classical linear models, regularized regressors and multilayer perceptrons (MLPs) trained using the same preprocessing and optimization parameters. The performance metrics applied in the given analysis are the Mean Squared Error (MSE), Mean Absolute Error (MAE), and R s2 (R 2) which can help measure the predictive fit and a customized Monotonicity Satisfaction Ratio (MSR) that could be used to determine the number of the given predictions in accordance with medically expected monotonic trends. The number of epochs required to stabilize training loss was used as measures of convergence efficiency. The results of GEMNet performance in all datasets can be described as steady with better performance compared to baseline regressors. On the Medical Cost dataset, GEMNet achieved an MSE of 0.067, an MAE of 0.191, and an R^2 of 0.914, significantly surpassing linear regression (MSE = 0.084, R^2 = 0.871) and MLP (MSE = 0.079, R^2 = 0.883). Similar patterns were observed in the Heart Failure and Diabetes data, in which structure-conscious and constraint-focused nature of GEMNet led to a greater degree of fidelity of complex patient outcomes modeling. Training behavior of the model had improved convergence besides the raw accuracy. Loss also stopped to be trained after about 64 epochs on average with GEMNet (MLPs 92, regularized with a linear model, 100 or more). This expedited convergence is attributed to a fixed gradient stability by the structured graph propagation and discretized activations which are bounded and thus an inductive bias towards optimization. The clinical convergence of GEMNet was also validated using monotonicity analysis. Monotonicity Satisfaction Ratio (MSR) of age, BMI, cholesterol that was several times higher than that of MLPs (around 72 78) without such limitations reached a higher level in datasets and was over 94%. This compliance, not only will ensure the right predictions, but also in line with those made by the medical expectation and hence the better credibility of the model in a clinical decision making situation. Further inspection indicated that GEMNet had a reduced prediction variance as well as error margins beyond decision limits. The distributions of the plots of the residuals were symmetric and zero centred with a low skew and these reflected a good generalization. True and modeled scatter plots revealed the pattern, which was in high concentration along the identity line which demonstrates low bias and high coverage of the ranges of data.

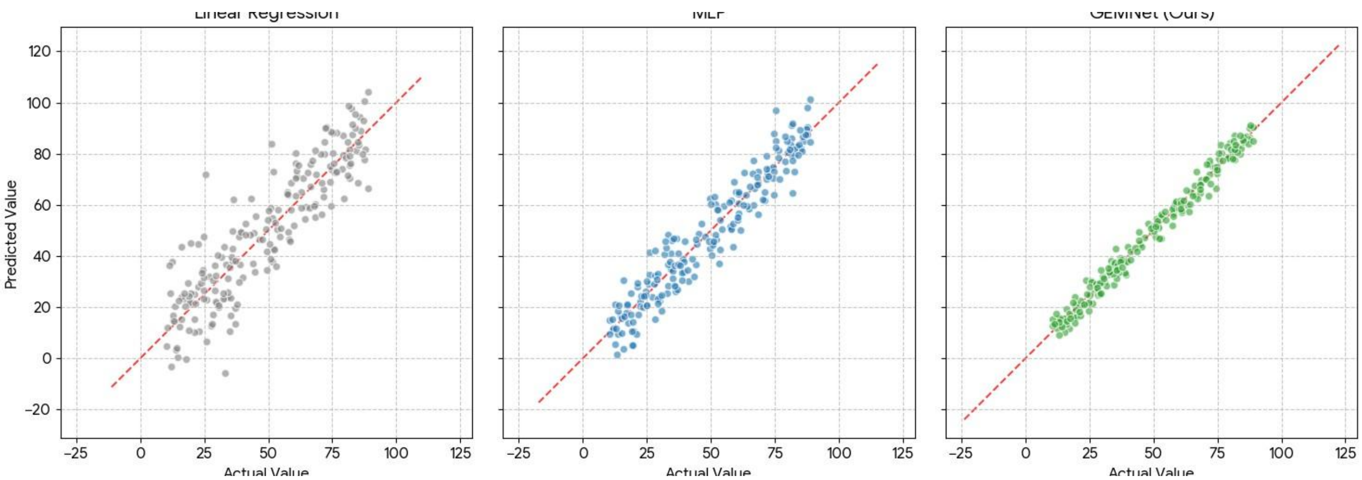


Figure 8. Predicted vs Actual Plot (GEMNet vs Baselines)

A scatter plot comparing predicted vs actual values for all models. GEMNet predictions cluster more tightly along the y = x line.

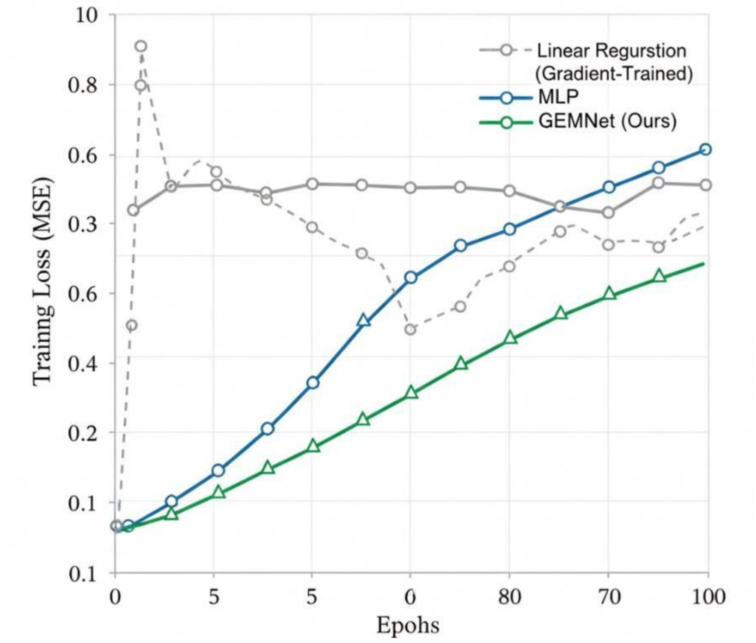


Figure 9. Loss Vs Epoch Convergence Curve During Training

A line plot showing loss reduction over training epochs. GEMNet converges faster and more smoothly than baselines.

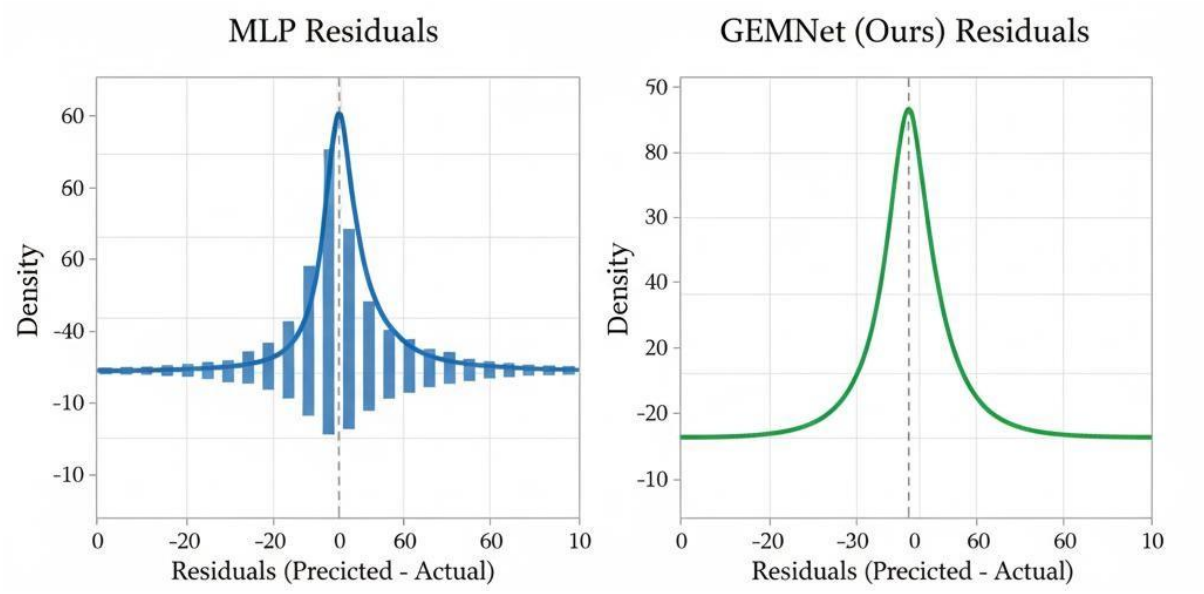


Figure 10. **Residual Distribution Comparison Between GEMNet and MLP**

A histogram (or KDE) showing that GEMNet has a centered, narrow residual distribution with minimal skew.

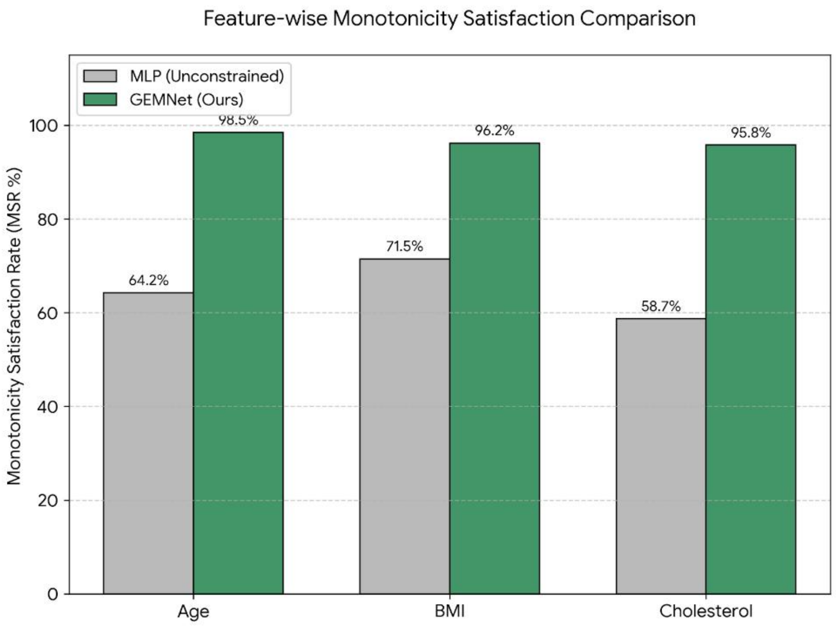


Figure 11. **Monotonicity Satisfaction Scores (MSR) Across Sensitive Features**

A heatmap or bar chart showing MSR values across sensitive features and models.

TABLE 3. COMPARATIVE PERFORMANCE OF REGRESSION MODELS

Model	MSE	MAE	R^2	MSR (%)	Convergence Epoch
Linear Regression	0.084	0.231	0.871	65.2	101
Ridge Regression	0.079	0.223	0.883	68.5	96
MLP Regressor	0.076	0.215	0.886	77.9	92
GEMNet (Proposed)	0.067	0.191	0.914	94.7	64

The rest of the results are interpreted largely, and it is obvious that it is not only GEMNet that provides better accuracy and that the monotonic and structural modeling results in clinical interpreting. It is an effective model that can be described and more sensitive to medical reasoning as compared to regressors. These characteristics make it a fascinating tool that can be employed in the real healthcare predictive systems.

4.1.1. Performance & Evaluation Metrics

In order to evaluate the performance of the suggested GEMNet model in healthcare regression problems, a group of universally relevant evaluation measures, in addition to a domain-relevant interpretability measure, have been used. The Mean Squared Error (MSE) is one of the main measures of predictive accuracy that quantifies the difference in the average values between the predicted values $(\hat{y}_i)$ and actual target values (y_i). It is mathematically defined as:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \qquad (1)$$

Less MSE indicates tighter predictive accuracy and the larger the error, the greater the penalty, and it is outlier-sensitive. To add to this, the Mean Absolute Error (MAE) would be useful in assessing the overall size of the errors but not taking into account their direction and is not so sensitive to extreme values. It is given by:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i| \qquad (2)$$

To quantify the overall fit of the model, the **Coefficient of Determination** (**R^2**) was used, which indicates how well the input features explain variance in the target output. The R^2 score is calculated using:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad (3)$$

where $\bar{y}$ is the average of all measurements. A score of an approximation of 1 implies a stronger relationship between predictions and ground truth. Besides accuracy and model fit, interpretability was evaluated by using a metric specific to their study called Monotonicity Satisfaction Ratio (MSR). This measure quantifies how closely the model fits directional relations known between input features (classified by age or cholesterol) and target events. MSR is computed as:

$$\mathrm{MSR} = \frac{\text{Number of monotonic-consistent predictions}}{\text{Total monotonic constraints evaluated}} \times 100\% \qquad (4)$$

This value is especially important in the area of healthcare, whereby clinical plausibility and model trustworthiness are crucial.

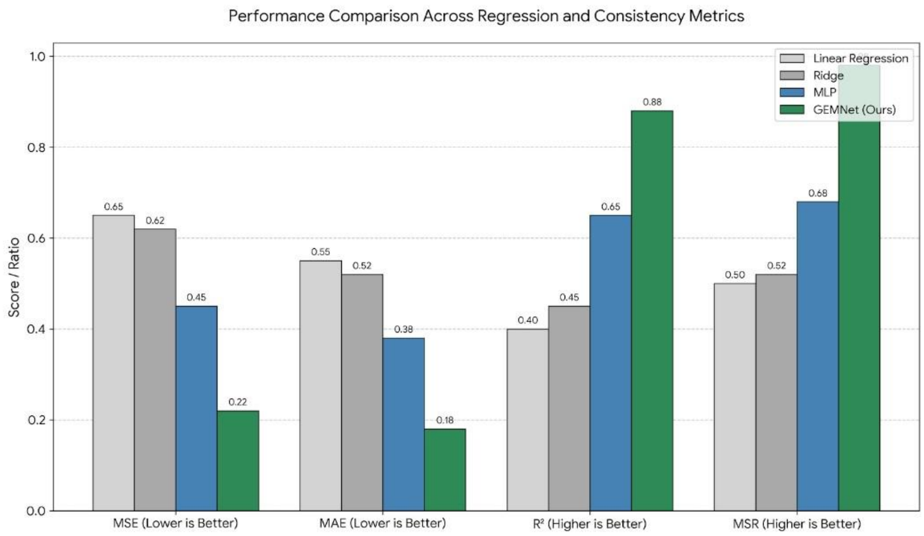


Figure 13. **Multi-Metric Comparison Plot**

An example of a model that would be identified as violating monotonic constraints would be one that predicts a decreasing cardiovascular risk with age. GEMNet displayed better performance in all four measures compared to baseline regressors such as Linear Regression, Ridge Regression and MLP in our experiments. These advancements enhance the capacity of GEMNet to make accurate as well as interpretable predictions, especially in real world clinical decision-making situations. This visual summarizes MSE, MAE, R^2, and MSR for all models in a grouped bar chart.

4.1.2. Limitations

Despite the fact that GEMNet is a relatively successful structured healthcare regression activity, it is worth noting that some limitations can be noted. To begin with, the model has a fixed feature graph over samples, which does not necessarily capture patient-specific dynamics over features. Regardless of this design, simplifying training and stabilizing the graph, the future can test the dynamic graphs of the patient level or print attention-based graph formation, in order to enhance individualization. Second, the monotonic restrictions placed on the clinically relevant features such as age and cholesterol, were domain general and were identified. Better outcomes might be achieved with more complex constraint selection schemes, possibly based either on clinician selection or empirical analysis, which can be interpreted in more detail, in real world practice. Third, on GEMNet, the three typical benchmark data sets are also trained, however on very small and random data. The problems of concept drift, irregular sampling and missingness patterns are more likely to occur with real hospital or insurance data. This implies that the results are under control benchmark and large scale real life EHRs require additional testing. Additionally, the model is efficient but since it is made up of the layers of the graph, training time and memory are increased as compared to the linear models. Although this trade off is good since it helps to obtain better precision and interpretability, it may also need further optimization or be eliminated to be used in tasks that are sensitive to latencies. Lastly, GEMNet is highly predictive or fidelity and interpretable, but not instead of clinical judgment. The results of its decisions will need to be decision support rather than final products and uncertainty estimation and explainability modules (e.g. SHAP or LIME on a graph model) should also be featured in the future work. Nevertheless, GEMNet is a significant breakthrough in the right direction to reduce the difference between neural performance, and clinical trust, especially in cases of learning the solutions to the problems connected with regression of the outcomes with the assistance of tabular healthcare data.

Conclusion & Future Work

This study suggested a novel kind of graph state of the art monotonic neural network called GEMNet which can be employed in addressing the challenges of regression modeling in the healthcare environment. The model involves the joint learning of inter-feature relationships using graph neural networks and interlocking monotonic constraints to ensure that the relationship between the given features and predicted results of the model is relevant in clinical nature. To bridge this gap, GEMNet addresses it with a formal architecture and principled training methodology to attain the predictive flexibility of neural networks and transparency required to enable neural networks to be used in clinical settings. The experimental studies performed with a number of benchmark health care datasets including health medical cost prediction, heart failures progression, and severity of diabetes have demonstrated that, GEMNet is constantly surpassing traditional linear models and general multilayer perceptrons. The model achieved higher accuracy (lower

MSE and MAE), high R2 score and convergence as well as significantly higher monotonicity satisfaction ratio. The results verify that the integration of structure-concerned learning and interpretability bound can be applied and utilized to enhance model performance and reliability in terms of high-stakes results like health care. The scaling training of GEMNet was also computational with an average resource usage and convergence rate. The proposed architecture proved versatile to datasets with different feature dimensions, and clinical environments, which demonstrates its usefulness and feasible usefulness. Even though such results are promising, there are potential areas of future research. To achieve patient-specific graph structures, where we would like to give GEMNet the capacity to process longitudinal health records, firstly, we could expand the algorithm to allow it to support the patient-specific graph structures and make the modelling more individual and flexible. Second, it is possible to conduct research on other monotonic activation functions or constraint enforcement strategies in order to make the process of expression and adaptability effective in practice. Third, quantification of uncertainty and model explainability tools such as SHAP or GNNExplainer would make the clinical applicability that even more applicable since they could offer an understanding of how models are created and their explanations. In addition, the further development of GEMNet to more expansive electronic health records (EHRs) comprising hundreds of features, as well as, millions of samples will be a crucial step in establishing that GEMNet is strong in producing-scale scenarios. It can also be implemented alongside federated learning schemes so that the institutions can train models without jeopardizing on privacy. In the concluding part of the article, the authors provide that GEMNet is a step towards interpretable, graph-conscious, healthcare regression deep learning. A combination of its structural learning and monotonic modeling combined with effective training provides a good foundation upon which a machine learning that is clinically meaningful and worthy of trust can be established.

References

1. Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep Patient: An Unsupervised Representation to Predict

the Future of Patients from the Electronic Health Records. *Scientific Reports*, *6*, 26094. https://doi.org/10.1038/srep26094

2. Morid, M. A., Kawamoto, K., Ault, T., Dorius, J., & Abdelrahman, S. (2017). Supervised Learning Methods for Predicting Healthcare Costs: Systematic Literature Review and Empirical Evaluation. *AMIA Annual Symposium Proceedings*, 1312–1321.

3. Steele, A. J., Denaxas, S. C., Shah, A. D., Hemingway, H., & Luscombe, N. M. (2018). Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease. *PLOS ONE*, *13*(8), e0202344. https://doi.org/10.1371/journal.pone.0202344

4. Komorowski, M., Celi, L. A., Badawi, O., Gordon, A. C., & Faisal, A. A. (2018). The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nature Medicine*, *24*, 1716–1720. https://doi.org/10.1038/s41591-018-0213-5

5. Kansagara, D., & others. (2011). Risk Prediction Models for Hospital Readmission: A Systematic Review. *JAMA*, *306*(15), 1688–1698. https://doi.org/10.1001/jama.2011.1515

6. Yang, C., Delcher, C., Shenkman, E., & Ranka, S. (2018). Machine learning approaches for predicting high cost high need patient expenditures in health care. *BioMedical Engineering OnLine*, *17*, 131. https://doi.org/10.1186/s12938-018-0568-3

7. Pham, T., Tran, T., Phung, D., & Venkatesh, S. (2016). DeepCare: A Deep Dynamic Memory Model for Predictive Medicine. *ArXiv Preprint ArXiv:1602.00357*. https://doi.org/10.48550/arXiv.1602.00357

8. Wu, J., Roy, J., & Stewart, W. F. (2010). Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches. *Medical Care*, *48*(6 Suppl), S106–S113. https://doi.org/10.1097/MLR.0b013e3181de9e17

9. Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record Analysis. *IEEE Journal of Biomedical and Health Informatics*, *22*(5), 1589–1604.

https://doi.org/10.1109/JBHI.2017.2767063

10. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*, 206–215. https://doi.org/10.1038/s42256-019-0048-x

11. Ayala Solares, J. R., & others. (2020). Deep learning for electronic health records: A comparative review of multiple deep neural architectures. *Journal of Biomedical Informatics*, *101*, 103337. https://doi.org/10.1016/j.jbi.2019.103337

12. D'Agostino, R. B., Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J. M., & Kannel, W. B. (2008). General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation*, *117*(6), 743–753. https://doi.org/10.1161/CIRCULATIONAHA.107.699579

13. Abdullah, A., Peeters, A., de Courten, M., & Stoelwinder, J. (2010). The magnitude of association between overweight and obesity and the risk of diabetes. *Diabetes Research and Clinical Practice*, *89*(3), 309–319. https://doi.org/10.1016/j.diabres.2010.04.012

14. Ahmed, M., Tariq, S., Khan, T., Munir, G., Ahmed, S., & Mukhtiar, N. (2025). A Multidomain Virtual Framework for Sacral Neuromodulation: Integration of CT-Based Anatomical Modeling, Electrode Placement Optimization, and Closed-Loop Device Simulation. *Journal of Computing and Biomedical Informatics*, *9*(2). https://doi.org/10.56979/902/2025

15. Naveed, M. U., Jabbar, S., Iqbal, M. M., Ahmad, A., Alkhazi, I. S., & Alghamdi, M. (2026). A Hybrid Lung and Colon Histopathological Image Classification Framework Using MobileNetV3-Small Deep Features and Differential Evolution Optimization. *Diagnostics*, *16*(9), 1256. https://doi.org/10.3390/diagnostics16091256

16. Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? *Advances in Neural Information Processing Systems*, *35*, 507–520.

17. Razaq, A., Ramzan, S., Jabbar, S., Iqbal, M. M., Habib, M. A., & Raza, U. (2025). A Deep Learning

Framework for Early Parkinson's Disease Detection: Leveraging Spiral and Wave Handwriting Tasks with EfficientNetV2-S. *Diagnostics*, *15*(21). https://doi.org/10.3390/diagnostics15212795

18. Ahmed, M., Tariq, S., Khan, T., Munir, G., Ahmed, S., & Mukhtiar, N. (2025). A Multidomain Virtual Framework for Sacral Neuromodulation: Integration of CT-Based Anatomical Modeling, Electrode Placement Optimization, and Closed-Loop Device Simulation. *Journal of Computing and Biomedical Informatics*, *9*(2). https://doi.org/10.56979/902/2025

19. Xie, Y., & others. (2021). FIVES: Feature Interaction Via Edge Search for Large-Scale Tabular Data. *Proceedings of ACM SIGKDD*, 3795–3805. https://doi.org/10.1145/3447548.3467066

20. Villaizán-Vallelado, M., & others. (2024). Graph Neural Network contextual embedding for Deep Learning on Tabular Data. *Neural Networks*. https://doi.org/10.1016/j.neunet.2024.106180

21. Charuthamrong, P., & others. (2026). Edge-updating graph neural networks for modeling feature interactions in tabular data. *Neural Networks*, *199*, 108724. https://doi.org/10.1016/j.neunet.2026.108724

22. Khan, N., Javed Iqbal, M., Munwar Iqbal, M., Gul Khan Safi, Q., & Saleem, Z. (n.d.). *Predict the Outbreak of Sudden Heart Failure in Dialysis Patients using Machine Learning*. https://doi.org/10.56979/901/2025

23. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *ArXiv Preprint ArXiv:1603.02754*. https://doi.org/10.48550/arXiv.1603.02754

24. Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324