Patient-GAT: Sarcopenia Prediction using Multi-modal Data Fusion and Weighted Graph Attention Networks

Cary Xiao
Department of Computer Science
Stanford University
caryxiao@cs.stanford.edu

Nam Pham
McKelvey School of Engineering
Washington University in St. Louis
n.p.pham@wustl.edu

Erik A. Imel
School of Medicine
Indiana University
eimel@iu.edu

Xiao Luo
Purdue School of Engineering and Technology
IUPUI
luo25@iupui.edu

ABSTRACT

Graph Attention Networks (GAT) have been extensively used to perform node-level classification on data that can be represented as a graph. However, few papers have investigated the effectiveness of using GAT on graph representations of patient similarity networks. This paper proposes Patient-GAT, a novel method to predict chronic health conditions by first integrating multi-modal data fusion to generate patient vector representations using imputed lab variables with other structured data. This data representation is then used to construct a patient network by measuring patient similarity, finally applying GAT to the patient network for disease prediction. We demonstrated our framework by predicting sarcopenia using real-world El-IRs obtained from the Indiana Network for Patient Care. We evaluated the performance of our system by comparing it to other baseline models, showing that our model outperforms other methods. In addition, we studied the contribution of the temporal representation of the lab data and discussed the interpretability of this model by analyzing the attention coefficients of the trained Patient-GAT model. Our code can be found on Github[1]

CCS CONCEPTS

• **Applied computing** **Health informatics;**

KEYWORDS

Electronic Health Records, Graph Neural Networks, Data Fusion, Prediction, Model Interpretation

ACM Reference Format:
Cary Xiao, Nam Pham, Erik A Imel, and Xiao Luo. 2023. Patient-GAT: Sarcopenia Prediction using Multi-modal Data Fusion and Weighted Graph Attention Networks. In The *38th ACM/SIGAPP Symposium on Applied Computing (SAC '23),March 27 - March 31, 2023, Talltnn, Estonia.*ACM, New York, NY, USA,4 pages. https://doi.org/10.1145/3555776.3578731

[1] https://github.com/CaryXiaol/Patient-GAT

SAC '23,March 27 - March 31, 2023,Tallinn, Estonia

ACM ISBN 978-1-4503-9517-5/23/03.
https://doiorg/10.1145/3555776.3578731

1 INTRODUCTION

Electronic Health Records (EHR) data representation affects the accuracy of downstream tasks, such as disease prediction. Typically, when a machine learning model is used for clinical outcome or disease prediction using EHR data, each patient is represented as a vector using clinical variables or phenotypes. The vectors are then being fed into machine learning models for prediction. In recent years, graph neural network models have been applied in various application domains, including clinical domains, such as adverse drug reaction prediction [5] and many others [6, 11]. Graph representationsare also found in predictive analysis, where patients' medical information can be processed to find the similarity between given patients and define clusters or neighborhoods for patient similarity networks [3]. Recent research shows predictions of in-hospital mortality and length of stay can be improved through exploiting diagnoses as relational information by connecting similar patients in a network (10]. In the study [4], the authors have also shown that theycan achieve better results by utilizing clinical notes to build the patient network and using Graph Neural Networks (GNN) for predicting 30-day hospital re-admission.

In this paper, we propose Patient-GAT - a novel application of graph neural networks for sarcopenia prediction using EHR data. Sarcopenia is estimated to affect 10% of older adults across the world [8], where it contributes to a loss of independence, dysmobility, disability, hospitalizations, and heightened healthcare costs. However, the diagnosis of sarcopenia is often delayed by limited knowledge of sarcopenia among physicians, busy clinic time pressures, and the need for objective measures to establish its diagnosis. Our Patient-GAT model has a multi-model data representationcomponent that considers the temporal changes of the lab values, and it has a patient network built by measuring the similarities between the static features and the changing lab values.

The contribution of our research includes: (1) Wedevelop a multimodel data representation to integrate static EHR clinical variables with changing lab values. The multi-modal data representation includes an efficient data imputation part to deal with missing lab values. (2) We create a graph representation of patient networks by calculating similarity based on multi-modal data representation. The patient network takes advantage of similar patient data in the system to improve the performance and increase interpretability of the predictive model.

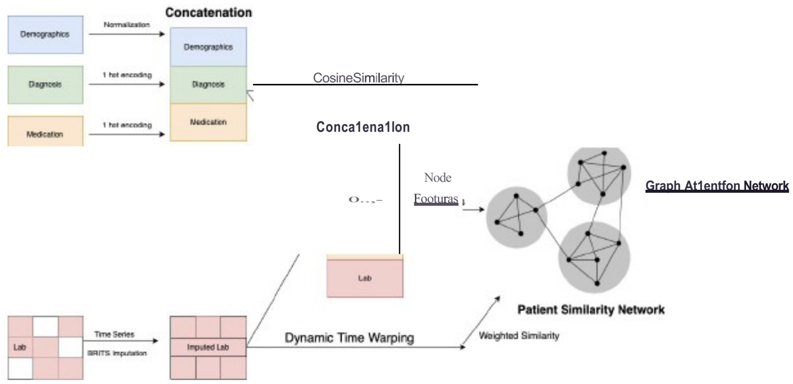

Figure 1: Framework of Patient-GAT

2 THE PROPOSED MODEL

Figure 1 shows the framework of Patient-GAT. Lab data is imputed using BRITS. Dynamic Time Warping (DTW) is used to measure the similarity between the temporal lab series. The static features, Demographics, Diagnosis, and Medication, are concatenated to build the non-lab vector representation of the patient. The patient network is then built based on the measured similarity between the vector representation and temporal lab series. Finally, GAT is performed on the patient network for disease prediction, which is represented by binary node classification.

2.1 Multi-Modal Data Fusion

2.1.1 Demographics, Diagnosis and Medication Embedding. For the Demographics data, we map the patient (all adults) age in our study cohort to age groups (18-29, 29-39, *etc.)*.The patient gender and race are mapped to a binary or integer encoding, respectively. The *BMI* is Z-Score normalized. The Diagnosis and Medication data are one hot encoded. To reduce the dimension, we represent the common groups of diagnosis and medication based on the International Statistical Classification of Diseases (ICD) hierarchies and the national drug code (NCD) directory, respectively. All less commonly occurring diagnoses and medications in our study cohort are grouped under "other", respectively.

2.1.2 Lab Data and BRITS Imputation. For each patient **Lab data,** we construct time series for each lab test as follows. First, we fix a length k for the time series sequence and a time step t_s between each collection in the sequence. In the experiment, based on the natural of the lab tests and data distribution in our study cohort, we set $k = 6$ and $ts = 180$ days. Then we map the earliest value v_0 of the lab to the 0^{th} index of the time series and use the time to as the anchor time. For eachsubsequent value v_i, we calculate the number of days between the time of measurement t; and the anchor time t_0 as $8d = t_i - t_0$. Then, we divide $8d$ by ts to find the index i to map v; to. It is possible that multiple values are mapped to i; in that case, we take the average of all the values.If i k, we ignore such values. We repeat the above procedure to build multiple time series for different lab tests of interest. The time series vectors for the lab data are likely to contain missing values. Therefore, we use a state-of-the-art imputation method- BRITS to fill in the missing values [1].

BRITS utilizes a bidirectional recurrent dynamical system that uses binary masking as well as a time gap matrix, where the imputed values are treated as variables of an RNN, which can be effectively updated during back propagation **[1].**

2.2 Patient Network with Weighted Similarity Metric

We propose a weighted similarity metric measure that incorporates both static features (Demographics, Diagnosis, and Medication) and temporal features (Lab data) to build a patient graph network. Given two patients p and q, for the static features of p and q, we concatenate De, D, M resulting in static feature vectors $Sp = D'j$, $||$ Dp $||$ Mp and $Sq = D;$ $||$ Dq $||$ Mq, where $||$ is the concatenation operation. We then compute cosine similarity between Sp and Sq, shown as Equation $ap,q = \frac{Sesq}{||Sp|| ||Sq||}$

On the other hand, we adopt DTW to calculate the similarity between the imputed lab time series [9].DTW returnsa realnumber r 0 that represents how similar two times series are (larger r means more dissimilarity)[9]. For each lab l_i, we compute a DTW distance matrix DD_i by computing the pair-wise DTW distance of all the time series associated with I;. So, the entry (n, m) in DD_i is $DTW(lf, l;")$, where If and I;" represents the time series of lab I; for patient n and m, respectively. We then apply min-max normalization DD_i to bound the distance between 0 and 1. We subtract all entries in DD_i from 1 to produce a DTW similarity matrix DS_i. The reason we subtract the normalized values from 1 is that a larger distance r means more dissimilarity; thus, when we subtract from 1, this results in a smaller similarity score. Hence, $dd(p,q)$ is the DTW similarity between patients p and q for lab I;. If we have multiple labs of interest $\{lo, l1, \cdots, lk\}$, we take the mean of all DTW similarity metrics to produce a lab similarity coefficient between patient p and q shown as Equation $/Jp,q = \frac{\sum^m_{i} dd_{p,q}}{\cdots}$

To integrate the similarity value calculated based on the static features - ap,q, with the similarity value calculated based on the temporal lab data - $/Jp,q$, we take a simple and flexible weighted scheme to make a comprehensive similarity metric $8p,q$ between patients p and q, shown as Equation 1.

$$op,q = k * «p,q + (1 - k) * /Jp,q, \quad O \quad k \quad 1. \tag{1}$$

The patient graph network is constructed based on the siini-larities between patients. From each patient node p, we set it to connect to the top m nodes with the highest similarity values that are over a threshold w. Both m and w can be tuned based on the data set.

2.3 Weight Graph Attention Networks for Prediction

After we have built a patient similarity network, we leverage GAT to perform node-level binary classification [12]. GAT is a novel neural network model that operates on graph-structured data and uses attention mechanisms to achieve state-of-the-art results on both inductive and transductive learning tasks [12].

Consider that we have a set of node features $h = \{h1, hz, \cdots hN\}$, $h;$ E RF, where N is the number of node in the graph and Fis the dimension of the node features, the layer outputs a new set of node features (of potentially different node feature dimension F') $h' = \{h;,h;,\cdots h\}, h$ E RF'_

We apply a linear transformation W E $IRF'xF$ to every node to obtain sufficient expressive power. Then, we perform self-attention on the nodes using a shared attention mechanism, shown as Equation a : RP X RF' -+ R to compute attention coefficient (shown as Equation $e;j = a(Wh;,Whj))$ Theattention coefficient indicates the relative importance of node j's features to node i. These attention coefficients are used to explain the relative importance of each node (patient) in predicting the label of another node. Note that this attention coefficient is not necessarily symmetric.

These attention coefficients are normalized across all choices of j, where j lies in some neighborhood $N;$ of some fixed i using the softmax function, shown as Equation 2

$$«i,j = \text{softmax}(eij) = \frac{exp(eij)}{\sum_{kEN;} exp(e;k)} \tag{2}$$

In our model implementation, $N;$ is the first-order neighborhood of i. In our proposed model, the attention mechanism a is a single-layer forward neural network on which we apply the LeakyReLU non-linearity. The normalized attention coefficients are then used to compute a linear combination of the features corresponding to them, to serve as the final output features for every node, shown as Equation $h; = a(I,jEN; a;jWhj)$-

In the final prediction layer, we first take the average of the previous layers and apply a final nonlinearity (in our binary classification problem, this is the Sigmoid function). The equation for the final layer is 3

$$h_i' = \sigma\left(\frac{1}{K}\sum_{l=1}^{K}\sum_{j\in N_i}\alpha_{ij}^l W^l h_j\right). \tag{3}$$

3 EXPERIMENTAL SETTING AND RESULTS

3.1 Dataset, Baselines, and HyperParameters

We used a proprietary dataset obtained from the Indiana Network for Patient Care (INPC). The objective is to investigate whether we can use the structured data in the EHR for sarcopenia identification. Our initial cohort contains the medical histories of 1,304 patients. Toprove the robustness of our model, we excluded patients without any information on diagnoses, medications, or lab variables. At the end, our dataset included 884 patients, with 174 patients diagnosed with sarcopenia. The input feature setting and representation is determined based on our previous experiments using the baseline models along with SHAP [7] interpretation to identify the most important features for prediction. For diagnosis, instead of using the full ICD code, we used the third level to the leaf nodes of the ICD-10 code hierarchy, which groups diagnoses intocategories. For medication, instead of using the original drug name with dosage, we used the drug group specified in the NDC directory. For lab data, we considered the following four major lab tests: glucose, Low-density lipoprotein (LDL), High-density lipoprotein (HDL), and hemoglobin.

We compared the performance of our model against five baselines: k-Nearest Neighbors, Logistic Regression, support Vector classification (SVC), Random Forest, and Multi-layer Perceptron (MLP).

HyperParameter Setting: For Patient-GAT, the k **in** Equation 1 is set to be 0.25, m is set to be 5, and w is set to be 0.65. Both MLP and Patient-GAT were standardized to all have two hidden layers with 16 nodes per layer, a dropout rate of 0.1, and were all trained with 10,000 maximum epochs. The Adam optimizers were used with a learning rate of 0.0005 and a weight decay of 0.05. The Binary Cross Entropy (BCE) with Logits Loss using mean reduction is used.We used 70%of the study cohort for training, 20% for testing, and 10%for validation. Since our dataset is unbalanced, SMOTE oversampling [2] is used on positive sarcopenia nodes in each training split to balance the classifications.The baselines are fine tuned to gain the best performances for comparison.

3.2 Performance and Comparison

Table 1 summarizes the mean and standard deviation AUC of Patient-GAT and the baseline models. The results show that the performances of all models except SVC drop slightly without demographics. The performance ofMLP dropped significantly (5%), whereas the kNN, Logistic Regression, Random Forest, and our Patient-GAT model dropped by only approximately 1%. However, Patient-GAT outperformed all other baseline modelswithout demographics. This allows for real-world deployments of Patient-GAT to potentially ignore the direct metrics of patient age, gender, race, or BMI, leading to a more equitable model in clinical settings. In addition to our primary graph representation that uses all clinical variables, we compared Patient-GAT and the baseline models on data without demographics to investigate whether the model has any biases on populations with certain demographics.

Table 1: Overall **Model** Performance Comparison

Model Type	Without Demographics	Demographics
k-Nearest Neighbors	0.70 ± 0.03	0.71 ± 0.03
Logistic Regression	0.70 ± 0.03	0.71 ± 0.04
SVC	0.66 ± 0.03	0.66±0.03
Random Forest	0.70 ± 0.04	0.71 ± 0.04
MLP	0.55 ± 0.09	0.70±0.03
Patient-GAT	0.71 ±0.04	0.72±0.04

3.3 Model Analysis and Interpretation

To investigate the contribution of the lab time series similarity measure using DTW, we perform a model analysis by comparing Patient-GAT to a variant of Patient-GAT by creating the patient network with edges solely based on the cosine similarity of the vector representation through concatenating the imputed lab variables with other staticvariables, then calculate the cosine similarity between patients to build the patient network. Figure 2 shows that without using DTW to measure the lab tests' similarity, the performance of Patient-GAT dropped by about 4%. This demonstrates that DTW is effectively used to measure the similarities between the temporal lab test values based on the changes over time. Like other diseases, the changes in the lab values often indicate the progression of diseases.

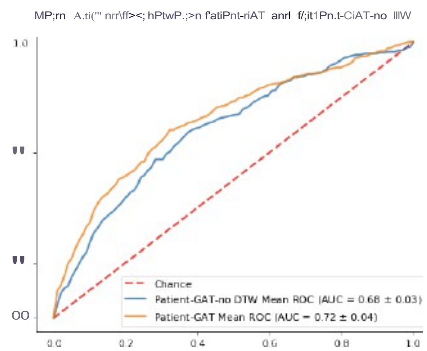

Figure 2: AUC-ROC Curves for Ablation Study on DTW

A notable feature of Patient-GAT is that it allows for implementations of Patient-GAT to achieve a certain level of model interpretability. We used attention weights of the trained model to identify the similar patients on the graph. Given patient with ID 66739 in the training data, the model correctly classified 66739 as not having sarcopenia. This patient has two other neighbors patient 40532 and 3077510 in the testing set, shown as Figure 3. The self attention of patient 66739 is the highest. The attention weight between patient 66739 and 40532 is higher than that between patient 66739 and 3077510. After further investigation we found that patient 40532 has a much closer BMI, age, and number of shared diagnoses with 66739 than patient 3077510 does. And patient 40532 shares thirteen different medications with patient 66739, while patient 3077510 shares only eight medications with patient 66739.

Figure 3: Similarity-based neighbors of Patient 66739

4 CONCLUSIONS AND FUTURE WORK

We have proposed Patient-GAT, a novel framework for disease prediction and applied to a sarcopenia dataset. The model leverages EHR data with imputed lab variables to generate patient similarity networks, allowing for a Graph Attention Network to perform prediction. Our model outperforms the compared baseline models. The imputed lab variables and DTW similarity measurement capture the changes in the lab tests and contribute to the performance of the proposed model. The proposed model can be adapted to other clinical outcomes or disease predictions.

In future work, we will (1) extensively study the contribution of each modality, (2) utilize rich data representation to measure the semantic associations between the structured data elements including diagnoses, medication and lab tests, and (3) include more lab tests with various lengths of history. We plan to compare our model against more baselines and test our proposed model on other diseases. We will also want to investigate the heterogeneous graph representation to measure patient similarity.

ACKNOWLEDGEMENT

This research was supported by the National Science Foundation REU program NSF CNS-1852105 and partially supported by the National Institute of Health (grants 1R01AR077273 and P30AR072581). The authors wouldlike to thank Professor Feng Li and Sheila Walter for their support.

REFERENCES

[1] Wei Cao, Dong Wang, Jian Li, Hao Zhou, Lei Li, and Yitan Li. 2018. Brits: Bidirectional recurrent imputation for timeseries. *Advancesin neural information processing systems* 31 (2018).
[2] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. SM01E: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16 (2002), 321-357.
[3] Leyu Dai, He Zhu,and Dianbo Liu. 2020. Patient similarity: methods and applications. *arXlv preprint arXiv:2012.01976* (2020).
[4] Sara Nouri Golmaei and Xiao Luo. 2021. DeepNote-GNN: predicting hospital readmission using clinical notes and patient network. In *Proceedings of the 12th ACM Omference on Bioinformalics,Computational Biology,and Health Informatics.* 1-9.
[5] Heeyoung Kwak, MinwooLee, Seunghyun Yoon,Jooyoung Chang, Sangmin Park, and Kyomin Jung. 2020. Drug-disease graph: predicting adverse drug reaction signals via graph neural network with clinical data. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining.* Springer, 633-644.
[6] Yang Li, Buyue Qian, XianJi Zhang, and Hui Liu. 2020. Graph neural network-based diagnosis prediction. *Big Data* 8, 5 (2020), 379-390.
[7] Scott MLundberg and Su-In Lee. 2017. A unifiedapproach to interpreting model predictions. *Advances In neural information processing systems* 30 (2017).
[8] AJMayhew, KAmog,S Phillips,GParise, PDMcNicholas,RJDeSouza, LThabane, and P Raina. 2019. The prevalence ofsarcopenia in community-dwelling older adults, an exploration of differences between studies and within definitions: a systematic review and meta-analyses, *Age and ageing* 48, I (2019), 48-56,
[9] Meinard Miiller. 2007. Dynamic time warping. *Information retrieval for music and motion* (2007), 69-84.
[to] Emma Rocheleau, Catherine Tong, Petar Velickovii:, Nicholas Lane, and Pietro Lio. 2021. Predicting Patient Outcomes with Graph Representation Leaming. *arXiv preprint arXiv:2101.0394()*(2021).
[11] ZhenchaoSun, Hongzhi Yin, Hongxu Chen,TongChen,Lizhen Cui, and FanYang. 2020. Disease prediction via graph neural networks. *IEEE Journal of Biomedical and HeaUh Informatics* 25, 3 (2020), 818-826.
[12] Petar Velickovii:, Guillem Cucurull,Arantxa Casanova.Adriana Romern, Pietro Lio, and Yoshua Bengfo. 2017. Graph attention networks. *arXlv preprint* arXiv:*1710.10903* (2017).