Using Ubiquitous Mobile Sensing and Temporal Sensor-Relation Graph Neural Network to Predict Fluid Intake of End Stage Kidney Patients

Mingyue Tang, Guimin Dong, Jamie Zoellner, Brendan Bowman, Emaad Abel-Rahman, Mehdi Boukhechba
{utd8hj,gd5ss,jz9q,btb5k,ea6n,mob3f}@virginia.edu
University of Virginia
Charlottesville, USA

Abstract

End-Stage Kidney Disease (ESKD) patients on hemodialysis suffer from kidney failure, with the inability to remove excess fluid causing fluid overload. This can cause many morbidities, and is one of the most insidious and common risk factors for mortality in ESKD patients. Existing solutions for fluid intake monitoring such as self-report and weight gain monitoring are burdensome, non-continuous, and usually administered in clinics only. It is then critical to develop a ubiquitous fluid intake monitoring system to help ESKD patients better control their fluid consumption. In this study, we propose to leverage smartwatch sensor data (e.g., Photoplethysmography (PPG), Gyroscope, etc.) combined with a temporal sensor relation graph neural network (TSR-GNN) to predict fluid intake given past sensing data between two dialysis sessions. Our empirical experiments highlight promising findings about the feasibility of using ubiquitous sensing to predict fluid intake, and demonstrate that the proposed model TSR-GNN outperforms the selected baseline models in both accuracy and robustness. Additionally, an in-depth analysis of model interpretability by attention weights and GNNExplainer variant is conducted to better understand the inter-sensor interactions and sensor contributions to the fluid intake prediction results.

Keywords

On-body Sensor Network, Graph Neural Networks, Ubiquitous Computing, Cyber-Human System, Mobile Sensing

1 Introduction

Approximately 37 million Americans have chronic kidney disease (CKD) – a decrease in the ability of the kidneys to cleanse toxins from the blood and balance fluid volumes within the body [1]. Among CKD patients, 750,000 Americans suffer from the most severe form of CKD - end stage kidney disease (ESKD); generally diagnosed when kidney function falls below 15% of normal. While some receive optimal therapy in the form of kidney transplants, another 500,000 patients require mechanical blood cleansing known as dialysis [3]. In addition, ESKD patients must follow unique dietary restrictions to reduce the risk of kidney overworking. The most onerous of these is the need to restrict fluid/water intake. Normally, the kidneys and heart work in unison to keep only the necessary amount of water in the blood vessels. However, patients with ESKD suffer from kidney failure, with the inability to remove excess fluid causing symptoms of swelling, pulmonary edema, and refractory hypertension. Long-term consequences include markedly increased risks of mortality from cardiovascular disease (CVD).

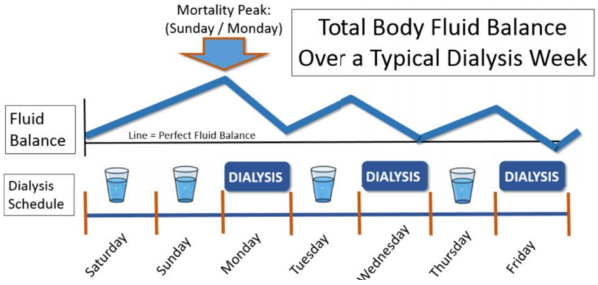


Figure 1: *Fluid Balance and the dialysis schedule.*

Dialysis therapy, substituting the natural work of the kidneys, can be a life-saving treatment for ESKD patients, nevertheless it's challenging for most patients. As shown in Figure 1, the dialysis schedule is fixed at four-hour treatment sessions delivered in a dialysis clinic three times a week. The "long break" of Saturday and Sunday, with significant fluid gains, is associated with the highest levels of mortality [17]. To reduce the dangers of rapid fluid gains and removal, patients are counseled to control their fluid intake (e.g. minimize water intake to 32 oz (1 liter) or less per day)[27]. However, in practice, this is nearly impossible, since thirst, a basic human need, can be driven by the high sodium western diets, dialysis sessions, and the attendant hypotension. Current fluid management methods estimate how much fluid patients have consumed by computing differences in weight between the last two dialysis sessions. The weight-monitoring based fluid management can be problematic, because patients are only monitored in the clinic, most deaths happen at home and mainly because of fluid overload (see mortality peak in Figure 1). It is then crucial to develop new ubiquitous methods to monitor patients' fluid consumption on a more granular scale to help clinicians intervene in time. To date, there has been no continuous method to monitor fluid intake of ESKD patients in their natural environment.

Recent advancements in mobile technology and computational methods for processing sensor signals provide novel opportunities to address these gaps. Wearable digital devices (e.g. smartwatches and smartphones) contain built-in motion and physiological sensors (e.g., heart rate, respiration, galvanic skin response [GSR][34]) can capture fine-grained human behavior information. In addition, a plethora of existing works has demonstrated how human behaviors,

978-1-6654-9624-7/22/$31.00 ©2022 IEEE
DOI 10.1109/IPSN54338.2022.00031

monitored through mobile sensing, can be related to health-related outcomes [7, 36]. Intuitively, the number of steps, location, hand gestures and other factors can have an impact on fluid consumption, these factors can also be captured by mobile sensing techniques.

In this work, we propose to leverage wearable sensing to design a multimodal sensor framework that extracts signatures from multiple sensors on wrist-worn watches [2] and predicts fluid intake of ESKD patients. The rational behind using multimodal sensing to estimate fluid intake of ESKD patients is the fact that it is known that fluid overload and over-hydration can manifest in patient's behaviors and physiology [42]. Several physiological biomarkers that can be captured through sensors such photoplethysmography (PPG to measure heart rate and heart rate variability), galvanic skin response (GSR to measure features from sweat), bioimpedance (to measure body composition), and pulse oximeter (SpO2 to measure respiration) can correlate with fluid overload symptoms [12]. Similarly, behavioral biomarkers such as mobility and physical activity patterns are also expected to correlate with fluid overload of ESKD patients.

There are multiple existing challenges when using multimodal mobile sensing data to capture complex human behavior dynamics such as fluids consumption. First, conventional sensor fusion techniques fall short in terms of capturing the dynamic correlation and complex interactions between heterogeneous sensor channels. Second, different modalities of human behaviors can have varying correlation levels with health outcomes. For example, patients' activity levels may have higher impact on fluid consumption then other modalities. This implies that graph representations must have different contribution levels to the learning targets, something that is not captured in current sensor network embedding methods. Third, static feature aggregation of human behaviors cannot capture temporal variations in human behavior dynamics. In this study, we propose a temporal sensor-relation graph neural network (TSR-GNN) to overcome the above challenges. TSR-GNN consists of the following components: 1) sensor-relation graph learning to automatically learn the dynamic relationships between different sensor channels; 2) attention-based sensor network embeddings, which can generate attention-weighted high-level topological sensor embeddings; and 3) temporal connection learning to capture the temporal dynamics within sensor interactions with respect to the corresponding human behaviors. Our contributions can be summarized as follows:

- We present a novel ubiquitous sensing method for personal-level fluid consumption estimation.
- We propose a Temporal Sensor Relation Graph Neural Network (TSR-GNN) with task-specific heterogeneous sensor processing and temporal learning modules to generate graph representation of on-body sensor network, capture multi-hop complex sensor interactions and produce robust sensor embeddings to predict fluid intake.
- We test our method in a real human-subject study involving 14 ESKD patients on hemodialysis, and demonstrate the superior performance of TSR-GNN (MAE:1.03, F1:0.93 in weight change prediction task, MAE:0.96, F1: 0.96 in self-reported fluid intake prediction task) when compared to other state-of-the-art models. We also conduct an interpretability analysis to unveil the working mechanism of TSR-GNN and discuss the clinical insights that can be extracted from this method.

Our paper is organized as follows: Section 2 highlights the existing works related to fluid intake monitoring, mobile sensing, and graph neural networks, explains why our solution is unique and how it is useful for fluid intake estimation. Section 3 proposes a temporal sensor relation based graph neural network approach to better interpret raw sensor data and to predict the fluid estimation accurately. Section 4 outlines the process of our ESKD fluid monitoring study and describes the multi-modal sensing data we collected from smart wearable devices. Section 5 provides comprehensive experiments and in-depth analysis of the effectiveness of our proposed approach. Section 6 discusses the limitation and the future directions of this research. Finally, Section 7 discusses the conclusion of our findings related to using mobile sensing and its potential for predicting fluid intake of ESKD patients.

2 Related Works

2.1 Fluid Intake Monitoring

Existing fluid intake monitoring research using ubiquitous devices can be categorized as the follows: smart tables, cameras, smart containers and wearable devices [10]. In the smart table-based solution, by leveraging embedded table sensors (e.g. pressure sensor), fluid instake can be estimated through measuring weight changes of objects on a table surface. Zhou et al. developed a smart table embedded with pressure sensors and force sensor resistors to monitor the weight changes in containers [44]. However, the design of smart tables to monitor fluid intake is dependent upon users' compliance and willingness to place food/waters in specific areas of the table. These directives can become burdensome, especially for a population such as ESKD patients which is already enduring a burdensome treatment. In the camera-based solution, fluid intake is estimated based on drinking activity detection using computer vision algorithms. Cippitelli et al. used a a depth and RGB camera, installed on the ceiling, to monitor eating and drinking actions [9]. Based of activity videos, Iosifidis et al. generated 3D volumes of human body poses to detect eating and drinking activities[22]. These methods can generate privacy concerns and are dependent upon the presence of cameras. The smart tables and camera-based solutions are unable to deliver continuous fluid intake monitoring because users can consume fluids where smart tables and cameras are not available. In the smart container-based solution, liquid levels are monitored through load and pressure measure, RFID, vibration measure and acoustic sensors. As signal strength of RFID can be impacted by liquid level, some researches have developed RFID-based fluid intake estimation methods for drinking event detection [23] and liquid level measurement [6]. Liquid resonance frequency is measured through Wi-Fi signal phase change detection and has been used to measure the liquid level in containers [32]. One of the limitations of smart container-based solution is that people usually do not consume fluids using only one specific container, which can underestimate the fluid consumption of ESKD patients causing life-threatening results. Other researchers have used wearable devices to estimate fluids. Mark et al. propose a body-worn audio and motion sensors system to estimate food type and amount consumed [30]. Other works weight fluctuation monitoring to detect

fluid intake [5, 33]. A hierarchical fluid intake monitoring system is proposed by Huang et al. to monitor fluid intake through drinking activity detection, drinking gesture recognition and intake amount estimation [20]. However, The existing wearable device-based methods are typically based on gesture recognition and rely on motion sensors only. They have high error rate (near 40%-50%) and usually evaluated in a controlled experimental environment. In this study, we deploy our Fluisense in real world situation and collect multi-modal sensor data (physiological, behavioral bio-markers) to provide accurate and robust fluid intake estimations. To the best of our knowledge, this is the first work deploying a ubiquitous sensing method to estimate fluid intake of ESKD patients.

2.2 Mobile Sensing

Sensor-rich smart devices such as smartphones and smartwatches enable unobtrusive monitoring of human behaviors and passive collection of behavioral and physiological signals in people's natural environment. The sensors embedded within those devices (e.g., Global Positioning System (GPS), Galvanic skin response (GSR) and accelerometer sensors) have been used to unveil bridges linking human behaviors and relevant health outcomes [7, 8]. Dong el at. demonstrate a sensor-based saliva cortisol level prediction framework by leveraging Actigraphy devices with embedded accelerometer and inclinometer [13]. By using the passively collected people's mobility information through GPS, Lin and Lyu el at. propose an interpretable machine learning model to predict health conditions by analyzing people's location traces with the associated points of interest [26]. Based on smartphone sensing, Huckins et al. evaluate people's travel pattern, phone usage, sleep behavior and physical activities and use the fined-grained human behavior features to estimate mental health metrics (e.g., depression and anxiety) [21]. In this study, our Fluisense contains a diverse set of sensors, such as Photoplethysmography (PPG), Accelerometer, and Magnetometer, to encode complex human behaviors from multiple perspectives and decode the high-level sensor embeddings to predict fluid intake.

2.3 Graph Neural Networks

Graphs are ubiquitous with respect to their capacity to represent multitudinous connected structures, such as social network, molecular structure, transportation network, and biological interaction graph [18]. As a consequence, graph neural networks (GNNs) have been evolving and attracting increasing attention in both industry and academia. In transportation, researches use graphs to represent transportation network and develop numerous spatio-temporal GNNs for traffic flow prediction, parking demand forecasting and travel time estimation [24, 37, 43]. In recommender systems, graphs represent the social network and interaction graph between users and items, and graph based recommender systems have been developed and deployed in real world application, achieving superior performance than traditional techniques in recommender systems [28, 29, 31]. In mobile health, researchers use GNNs to capture complex human behavior dynamics and translate the high level behavior embeddings to predict health symptoms such as mental health inference, biomedical marker prediction, and symptom detection [13–15]. However, previous works modeling multimodal mobile sensing data do not consider the complex interactions between different sensors. In this study, our proposed `TSR-GNN` generates graph representation of on-body sensor network and produce high-level sensor embeddings by capturing dynamic sensor interactions and temporal variations that can effectively characterize human behaviors.

3 Methodology

3.1 Problem Formulation

Given a set of on-body sensors, denoted as $\mathcal{S}$, our goal is to use the multivariate time series data collected by $\mathcal{S}$ to predict fluid intake values Y. In our proposed method, we generate a learnable sensor embedding $E_{S(i)}$ for each sensor $S(i) \in S$, where $i = \{1 \cdots N\}$ (N is the number of sensors), and we denote the raw multivariate time series sensor data as $R_{S(i)}$. To make predictions in real-time, we divide $R_{S(i)}$ into T time windows and set $R^t_{S(i)}, t \in T$ as the inputs to predict fluid intake. The raw heterogeneous sensor signals in $R^t_{S(i)}$ have different frequencies and cannot be fused naturally. To make $R^t_{S(i)}$ compatible with graph neural networks, we convert the $R^t_{S(i)}$ into the same dimensional feature space $F^t_{S(i)}$ with a transformation function f. Given the learnable sensor embeddings E and sensor data feature space F^t, the sensor relation graph structure G can be learned from an attention-based GNN, refereed as an attention-based Graph Isomorphic Network (GIN) [39]. The output of $\phi(G, F^t_{S(i)}, E)$ is a graph-level embedding GE^t for fused sensory data at time $t \in T$. We input the sequence graph embeddings GE into a recurrent neural network and produce the final prediction set Y.

3.2 Overview of `TSR-GNN`

The goal of our proposed prediction framework is to learn the relationship between different on-body sensors, and predict the interdialytic fluid intake by learning the sensor interactions indicative of fluid intake. This considers the inner connections between sensors and reduces the noise from possible sensor information loss comparing to traditional neural network methods. Our approach has four major components:

1. Heterogeneous Data Transformation, maps the original heterogeneous raw sensor data R into a sensor feature space F with a unified dimension.

2. Sensor Relation Graph Learning, learns the relationship between sensors and constructs a sensor relation structure graph G based on the correlation between learned sensor embedding vectors E.

3. Attention-Based GNN, generates a fused graph-level embedding GE given relation graph structure G and sensor feature F.

4. Temporal Connection Learning, learns sequential graph embedding GE within time domain $t \in T$.

The overall structure has been provided in figure 2. Our proposed TSR-GNN is inspired by Graph Deviation Network (GDN) [11] and Multivariate Time series Graph Neural Network (MTGNN) [38]. Both MTGNN and GDN can automatically learn inner time series interaction between two variables, and encode the learned between-variable interdependence to generate high-level node representations from multivariate time series input. However, in our

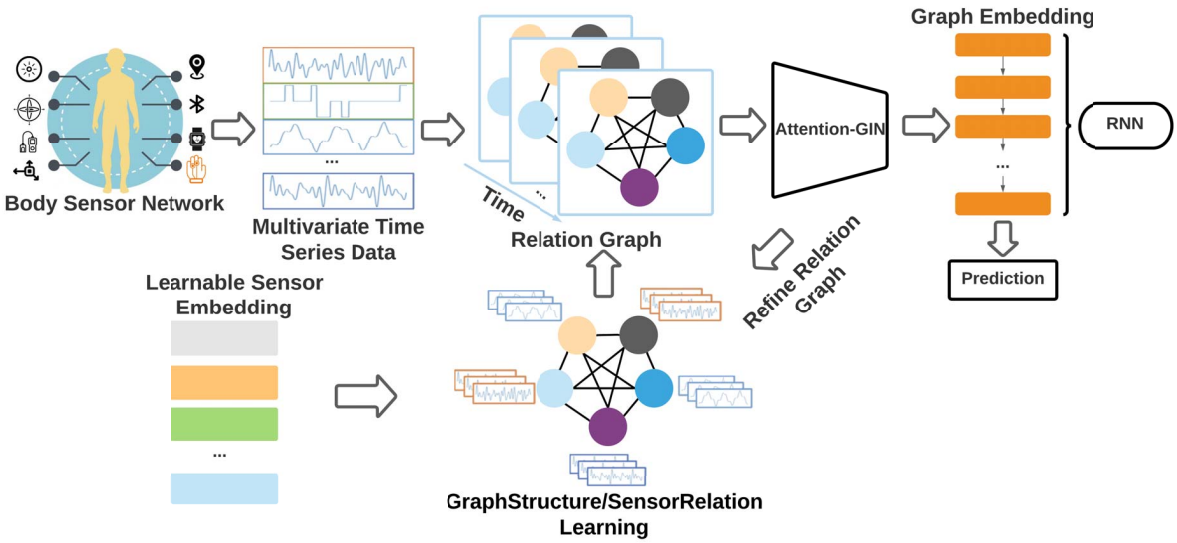


Figure 2: *The diagram of* TSR-GNN *model with multivariate time series sensor data.*

case, as we have multi-modal sensor data as input, and patients fluid intake amount as output, both models cannot deal with heterogeneous sensory data and predict graph-level results. To address this, we propose **heterogeneous data transformation** to transform the sensory data that have varying dimensions in each type of sensor. Also, our method uses **Sensor Embeddings** inspired by BERT to produce parameterized sensor embeddings that can describe the unique characteristics of each senor. In addition, instead of targeting node forecasting tasks in GDN and MTGNN, we generate a sequence of synthesized **graph embeddings** that can capture and fuse dynamic sensor interactions to target the prediction of a instance/whole graph. Finally, we add a **temporal connection learning** module to learn the temporal relationship between instances which models the period change of ESKD patients behaviors.

3.3 Heterogeneous Data Transformation

Different sensors can have different frequencies and units of data. For example, the Photepletosmnography sensor (PPG) sampled at 100HZ can have significantly more observations than GPS sensor which is typically contains only few observations in a given hour. Therefore, the raw data $R^t_{S(i)}$ from different sensor $S(i)$ cannot be directly fed to our graph neural networks.

To address the above problem, we first use a transformation function f to transfer the heterogeneous raw sensor data into a homogeneous feature space, described as

$$F^t_{S(i)} = f(R^t_{S(i)}), i \in \{1 \cdots N\} \tag{1}$$

The transformation function of each sensor f can be a universal mapping (i.e. MLPs) that can be learned during the training process, or it can simply be some handcrafted features extracted from the original data, like Min, Max, Variance in statistic domain; Absolute energy, Entropy in temporal domain; or FFT mean coefficient, Fundamental frequency in spectral domain. The main purpose is to convert heterogeneous data into a homogeneous form. We use both methods in our experiments and present results with the best performance.

3.4 Sensor Relation Graph Learning

In multi-modal sensing applications, sensors can be related in different ways. For example, heart rate and step count sensors are intuitively connected because when an increase in activity levels is expected to cause an increase heart rate. Due to the neighborhood aggregation process of graph neural networks, connecting similar sensors in one graph can fuse sensor features together, reducing noise (e.g, missing values in data collecting) in single sensor and improve performance of the predictive model.

Sensor Embedding, As illustrated in section 3.3, different sensors can have various types of data, and these data can be related in a non-linear complex way. We adopt the sensor relationship learning method from [11, 38] with random initialized trainable sensor embedding vectors $E_{S(i)} \in \mathbb{R}^d, S(i) \in S$.

These embeddings are trained along with the attention-based GNN model, and can be gradually refined along with the relational structure and prediction result. E has the following properties: 1) similar embeddings indicate similar sensor behaviors, and 2) similar sensor behaviors have a high probability that are correlated together. We use the top k correlation of these embeddings to construct a relational graph and let E perform as attention weights in GNN neighborhood aggregation steps.

Relational Graph Construction, In order to represent the connection between sensors, we use directed graph $G = (S, Re)$, where sensors S are the vertex and relations Re are the edges. If the sensor $S(i)$ has proven to be helpful to model the behavior of $S(j)$, where $i, j \in \{1 \cdots N\}$, then a directed relational edge $Re_{i,j}$ will be formed

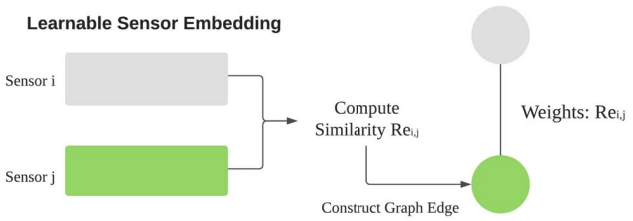


Figure 3: *Constructing graph edges according to sensor embedding similarity.*

(shown in figure 3), and the edge weight is

$$Re_{i,j} = sim(E_{S(i)}, E_{S(j)}) = \frac{E_{S(i)}^{\mathrm{T}} E_{S(j)}}{||E_{S(i)}|| \cdot ||E_{S(j)}||} \tag{2}$$

To construct the desired sensor relational graph G, the model selects the top K closest neighbors of a sensor node $S(i)$ according to the similarity equation 2, and k is a hyperparameter that can be tuned to control the sparsity of G.

$$Re_{i,j} = \emptyset, \{j \notin argtopk(Re_{i,n}, n \in \{1 \cdots N\})\} \tag{3}$$

3.5 Attention-Based Sensor Neural Networks

To generate an embedding vector which contains all nodes' information at time t, a graph-level representation GE^t of graph structure G^t, node features F^t, and sensor embeddings E is learned by a specially designed GNN which takes sensor interaction behaviors as an input feature.

The design of the desired attention-based GNN is described below:

Graph neural networks (GNNs), including graph convolutional networks (GCNs), Graph Isomorphic Networks (GINs), Graph Attention Networks (GATs), are all neural networks that perform graph convolution operations on graph structured inputs.

Then layer-wise propagation at layer l in GCN [25] can be expressed as

$$X^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}X^{(l)}W^{(l)}). \tag{4}$$

where W is a trainable weight matrix which applies a shared linear transformation to every node, and D, A are graph structure properties like node degrees and adjacent matrix. The propagation process aggregate 1-hop neighborhood information

GATs assign attention weights for each node v when doing the neighborhood aggregation.

$$z_v = \sigma(\sum_{u \in \mathcal{N}_{(v)}} \alpha_{v,u} W_t), \tag{5}$$

Where the $\alpha_{v,u}$ indicates the contribution coefficient for each sensor node to its connected sensors. The proposed attention-based GNN method adds sensor embeddings as graph structure controller and an augmented feature in on-body sensor network prediction.

Sensor embedding enabled attention aggregation, unlike the existing GNNs, the neighborhood aggregation process of `TSR-GNN` incorporates sensor embedding vectors E along with node features F^t. The attention coefficient between sensors is calculated by the feature-sensor combination.

First, we concatenate the node feature $F_{S(i)}^t$ and sensor embedding $E_{S(i)}$ of sensor $S(i)$ to generate a new variable $c_{S(i)}^t$,

$$c_{S(i)}^t = E_{S(i)} \oplus WF_{S(i)}^t \tag{6}$$

where $\oplus$ denotes a concatenation. Then, we assign a learned coefficients vector a for the attention mechanism and calculate attention $\alpha_{i,j}$ between $S(i)$ and its neighbor $S(j)$

$$\pi(i,j) = LeakyReLU(a^{\mathrm{T}}(c_{S(i)}^t \oplus c_{S(j)}^t)) \tag{7}$$

$$\alpha_{i,j} = softmax(\pi(i,j)) \tag{8}$$

where $LeakyReLU$ is a non-linear activation function used to compute the attention vectors, and the final attention coefficient α is normalized by softmax function in Eq. 7.

The new neighborhood aggregation/node embedding $NE_{S(i)}$ for sensor $S(i)$ with attention can be represented as

$$NE_{S(i)} = MLP(\alpha_{i,i}WF_{S(i)}^t + \sum_{j \in \mathcal{N}(i)} \alpha_{i,j}WF_{S(j)}^t) \tag{9}$$

where multi-layer perceptrons (MLP) is an aggregation function suggested by [39].

Output graph representation, After we have all node embeddings in Eq. 10, we interact the sensor embeddings E with node embeddings NE, in order to take sensor characteristic as an augmented feature for the final prediction. The interactive node-sensor embedding NSE^t at time t is

$$NSE_{S(i)}^t = NE_{S(i)}^t \odot E_{S(i)} \tag{10}$$

$\odot$ can be anything that interact two matrices (i.e. element-wise multiplication, concatenation), we use element-wise multiplication here.

The final graph embedding GE^t at time t is pooled by all NSE^t.

$$GE^t = pooling(NSE_{S(i)}^t), pooling \in \{MLP, Concatenate\} \tag{11}$$

where $i \in \{1 \cdots N\}$.

3.6 Temporal Connection Learning

All of our previous steps are at a single timestamp t, the temporal and periodical information is also critical for a continuous fluid intake prediction. For example, people tend to drink more water in a specific range of time or after a dialysis session. Therefore, we introduce a recurrent neural network method to connect all graph-level embeddings $GE^t, t \in T$.

$$a_t = f_1(W_{a,a}a_{t-1} + W_{a,GE^t}GE^t), f_1 \in \{\text{sigm}, \tanh\} \tag{12}$$

$$Y_t = f_2(W_{y,a}a_t) \tag{13}$$

where, a is the hidden layer of RNN, f_2 the last layer of RNN. Given a ground truth fluid intake value $\hat{Y}_t$ at time t, the regression loss L for minimization can be calculated by Mean Squared Error (MSE):

$$L = \frac{1}{N}\sum_{i=1}^{N}(Y_i^t - \hat{Y}_i^t)^2 \tag{14}$$

The classification cross entropy loss can be represented as:

$$L = -\sum_{i}^{C} \hat{Y}_i^t log(softmax(Y_i^t)) \tag{15}$$

where C is the number of classes.

3.7 Training Process

The detailed training process of the proposed method is outlined in Algorithm 1. It is to be noted that our TSR-GNN framework does not rely on more hyper-parameters than the basic ones for classic GIN, LSTM, GAT, and ST-GIN models. In this work, our model takes the raw sensor data streams as input and generates fluid intake predictions.

Algorithm 1: Training Algorithm for TSR-GNN

Input: $\{R_S^t, t = 0, \dots T\}$, $epochs$
Initialize: $\{E_S(i)|i \in 1, \dots N \leftarrow Random\}$,
$\{Re^0 = 1|Re^0 \in G^0\}$, $W \leftarrow Random$,
$\alpha \leftarrow Random$, $\theta \leftarrow Random$, $h^0 \leftarrow Random$
Output: $\{G^t|t = 0, \dots, T\}$, $\{Y^t|t = 0, \dots, T\}$

for $epoch \in \{1 \dots iterations\}$ **do**
  for $t \leftarrow 0$ *to* T **do**
    for $i \in \{1 \dots N\}$ **do**
      $F^t_{S(i)} \leftarrow f(R^t_{S(i)})$
      $Re_i \leftarrow \{\mathrm{Re}^t_{i,j} = sim(E_{S(i)}, E_{S(j)})|j \in \{1 \dots N\}\}$
      $Re_i \leftarrow \{\{\mathrm{Re}_{i,j} = \emptyset, \{j \notin argtopk(Re_{i,n}, n \in \{1 \cdots N\})\}\}$ # k candidates neighbors selection, form graph structure
      $c^t_{S(i)} \leftarrow \mathrm{E}_{S(i)} \oplus WF^t_{S(i)}$
      $\pi(i,j) \leftarrow \{LeakyReLU(a^{\mathrm{T}}(c^t_{S(i)} \oplus c^t_{S(j)}))|j \in \{1 \dots N\}$
      $\alpha_{i,j} \leftarrow \{softmax(\pi(i,j))|j \in \{1 \dots N\}\}$
      $NE_{S(i)} = \{MLP(\alpha_{i,i} WF^t_{S(i)} + \sum_{j \in \mathcal{N}(i)} \alpha_{i,j} WF^t_{S(j)})|j \in \{1 \dots N\}\}$
    $GE^t \leftarrow$Concatenate($\{\mathrm{NE}_{S(i)}|i \in \{1 \dots N\}$)
    $Y^t, h^t \leftarrow$LSTM(h^{t-1}, GE^t)
    $loss \leftarrow$MSE($\mathrm{Y}^t, \hat{Y}^t$)
    Update the Sensor Embeddings, Attention-GNN, and LSTM model parameters E_S, W and θ according to the loss by gradient descent

Inputs. After the collection of N channels multi-modal sensory data from a smartwatch S, the system stores these streams R_S and divide them into time windows which fall in between every two dialysis sessions (typically 2 days worth of data). The divided stream data R^t_S presents as graph node attributes according to its modality(e.g., PPG node, GPS node, Step Count node, etc.).

Initialization. To capture the salient features of each sensor and to model the complex relationships between them, we introduce the sensor embeddings E^t_S, which are initialized randomly at $t = 0$ and are learned during training. The sensor relation graph $G^0 = (S, Re^0)$ is simply initialized as a fully connected graph because the sensor relation set Re^0 are initialized to 1. The setup of GNN weights W, attention weights α, LSTM parameter weights Θ and LSTM hidden state h^0 follow random initialization.

Overall algorithm. Algorithm 1 shows the overall proposed model training scheme. After initializing all the weights, hidden state, and sensor embeddings, the algorithm takes the raw sensor data taking form of time windows, transforms the raw input data into homogeneous form and feeds the data into a graph structure construction process. The graph is generated by calculating the cosine similarity between different sensor embeddings. Then, we feed both graph structure adjecent matrix and homogeneous sensor features into an attention-based graph neural network (GNN) to get the node embeddings for each sensor. After that, we interact the node embeddings with the sensor embeddings to get the final sensor representation, interactive node-sensor embeddings. Finally, we concatenate all node-sensor embedings to a graph embedding, and use the recurrent neural network (RNN) to connect all the graph embbedings in every timestamp and predict the fluid intake amount at that timestamp. The model parameters and learnable sensor embeddings are updated based on the MSE/Cross Entropy Loss.

The detailed hyper-parameter settings of TSR-GNN model training are discussed in Section 5.1.

4 Study and Data Description

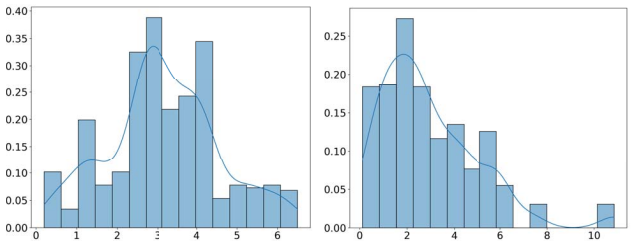
Data Collection After getting approval from the Institutional Review Board, N=14 ESKD patients were recruited from the Kidney Clinic to participate in a study investigating the use of wearable sensing to estimate fluid intake around 4-weeks. Participant's age was recorded as follows: 26-40: 21.42%, 41-55: 35.71%, 56-65: 28.57%, 66+: 14.29%, with 64.29% female, 28.57% male, and 7.14% transgender. Participants belong to 4 different ethnic groups: 57.14% are Caucasian, 14.29% are African, 21.43% are African American, and 7.14% are Asian. Majority of them (92.86%) are right-handed. Participants were given an Android smartwatch with an in-house developed app pre-installed (Fluisense, available on Android play store [4]). Patients were asked to log their fluid intake through the app by choosing from a list of predefined volumes each time they consume any liquid (e.g., 2oz, 4oz, 12oz, etc.). The app computed and displayed the self-reported daily volume intake to help patients monitor their own fluid consumption (Figure 5). The Fluisense application also stores all stream sensor data in real-time and uploads all recorded data to Amazon AWS S3 when internet is available. Participants were compensated up to $100 for completing the study.

Modality	Description
Heart Rate	The number of heartbeats per per minute
PPG	Detects volumetric changes in blood in peripheral circulation
Step Count	Measures the number steps and distance traveled
GPS	Longitude and latitude of locations
Compass	Captures the directions with respect to the North and South
Gyroscope	Senses the change in rotational angle per unit of time
Sound	Amplitude of sound waves
Battery	Monitors the smartwatch's battery level
Light	Measures ambient light in lux
Magnetometer	Measures the direction and strength of the magnetic field
Weight Changes	Measures of weight change before and after dialysis sessions
Self-Reported Intake	Self-reported fluids consumed in a given period

Table 1: *Multi-modal Sensory Data Description*

Patients received text messages twice a day (9am and 8pm) to remind them to use the watch app and the smartscale. We also

recorded patients' weights before and after each of the thrice-weekly dialysis sessions. Data was automatically and periodically synced to a secure amazon cloud storage. Participants were compensated with up to $100 in gift cards for completing the study.



(a) Weight change distribution (b) Self-reported fluid intake distribution

Figure 4: *Distribution of Interdialytic weight change (M=3.18, SD=1.38) and Interdialytic self-reported fluid intake (M=2.97, SD=2.12), both unit in kg.*

The goal of this study is to investigate how wearable sensors data can capture patient objective fluid intake biomarkers from the continuous sensor streams available on wearable devices. FluiSense can passively collect bio-behavioral features such as cardiovascular reactivity, activity levels, and mobility patterns that are hypothesized to correlate with fluid intake and motivate use to generate sensor network embeddings to encode the complex fluid-intake related human behaviors. The collected sensor data are summarized in Table 1.

Ground truth For ground truth $\hat{y}$, we use weight changes between dialysis sessions and interdialytic self-reported fluid intake data as our prediction labels. The rationale for using weight change and self-reported fluid consumption ground truth is because those two measures characterize both objective and subjective fluid consumption and because they represent the current clinical practice for measuring fluid consumption in the clinic. Because ESKD patients do not have the capability to remove fluid in their body, the weight change can reflect the real interdialytic fluid intake. Though the self-reported data may produce certain errors due to limitations in self-report methods (e.g., forgetfulness and personal perception), we found it to correlate with weight changes (with $P<0.01$, $r>0.25$), thus can provide more granular fluid intake data. Patients recorded fluids in total 496 times, Figure 4 shows the distribution of weight change and self-reported fluid intake data. The average interdialytic weight gain is 3.18 kg +/- 1.38 and the average self-reported interdyalic fluid consumption is 2.97 kg +/- 2.12.

5 Experiments

We design our experiments to evaluate TSR-GNN, focusing on the following research questions:

- **RQ1**: How do TSR-GNN and its ablations compare to the baseline methods?
- **RQ2**: How to interpret the node/sensor contribution and correlation in fluid intake predictions by using TSR-GNN?

Figure 5: *The user interface of the FluiSense smartwatch app*

- **RQ3**: What are the impacts of the major model hyper-parameters including training epoch, top-k, attention, and sensor embedding on TSR-GNN?

5.1 Experimental Setting

Datasets We summarize our fluid intake task into two major parts, one is a regression task, which estimates the precise quantity of fluid intake (in KG) between two dialysis sessions changeor on a daily basis. Another one is a classification task in which we divide the ground truth labels into 5 categories depending on their 5%, 25%, 50%, 75%, 90% percentiles, and predict the fluid intake level accordingly. The reason we choose the above 5 categories is because these are common percentiles used in statistical analysis. Furthermore, when fluid consumption is greater than 50% percentile, this can represent higher chances of fluid overload. Through these two tasks, a rigorous evaluation of the proposed method on different levels of granularity has been provided to study the feasibility of our proposed mobile sensing method and its effectiveness at detecting peaks in fluid accumulation that can indicate fluid overload. The Leave-one-subject out cross validation is one of the most common testing methods in this field. Thus, we split N=12 participants as training sets, N=1 as validation sets, and last N=1 as test sets. To test the robustness of our method, we perform 12-fold cross-validation over all collected data. Two time granularity levels have been tested: (1) training and testing performed on a two-days time granularity which represent the time between two consecutive dialysis sessions. This is driven by the fact that the weight pre/post dialysis has been measured during the thrice per weekly dialysis sessions typical two days apart. (2) training and testing on daily basis to test the performance of our model when making more fine-grained predictions.

Baselines, We compare TSR-GNN with 7 baselines (including one ablation), representing two types of regression methods, 1) Statistical machine learning (Mean, Support Vector Machine, Decision Tree), 2) Deep learning and GNNs (GIN, Long-Short Term Memory, Spatial-Temporal GIN).

- **Mean Model**, calculates the mean of training data, treat the mean as prediction result for test data.
- **SVR/SVM** [16], Support Vector Regression/Support Vector Machine, is a classic regression/classification method based on support vectors.
- **DTR** [35], Decision Tree Regression/Decision Tree, is a statistical machine learning method based on information gains.
- **GIN** [39], Graph Isomorphism Network (GIN) is a variant of GCN with MLP neighborhood aggregation, specifically

for graph-level representation, we trained this model with a strongly connected sensor graph.

- **LSTM** [19], Long-Short Term Memory (LSTM) is a type of recurrent neural network that learns the timeseries information between data.
- **ST-GIN** [40], Spatial-Temporal GIN (ST-GIN) is a variant of ST-GCN, specifically, use a GIN model connect to an LSTM model to reserve both temporal and spatial information.
- **SR-GNN**, is ablation of our approach, without learning temporal information from data (w/o LSTM).

Metrics, We choose Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and F1 measurements with standard errors as our evaluation metrics:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} y_i - \hat{y}$$

,

$$MSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2}$$

. Given the regression nature of our fluid intake monitoring tasks, the MAE, and RMSE are two most common metrics that evaluating the performance of the model in regression analysis and they are sensitive to different types of errors (e.g. RMSE is more sensitive to outliers).

$$F1score = \frac{2 \cdot Prec \cdot Rec}{Prec + Rec}$$

is a classification metrics which balanced the precision ($Prec$) = $TP/TP+FP$ and recall rate (Rec) = $TP/TP+FN$, where TP,TN, FP, FN are the numbers of true positives, true negatives, false positives, and false negatives.

Training Setting, we perform a cross-validation grid search to select the best hype-parameters for all models including baselines with the following default values, we set the number of GIN layers to 2, LSTM layers to 1, the embedding size of all layers in all models to 32. For TSR-GNN, the learning rate is select from 5e-2, 5e-3, 5e-4, 5e-5, epoch size 50, 100, 200, 300, 400, 500, we find $topk$ closest neighbors k in 3-10. We use the same GIN model for all baselines that involve GNN. All experiments are performed on an 8GB NVIDIA GeForce RTX 3070 GPU.

The main cross-validation experiment results are shown in table 2.

5.2 RQ1. Performance and Feasibility of TSR-GNN

Performance, All 7 baselines are included in the performance comparison presented in table 2. TSR-GNN achieves better results compared to all baselines on both **Weights Change** and **Self-reported** prediction tasks. Specifically,

(1) Almost all methods can significantly surpass the Mean Model in every aspects, which shows the potential prediction power of collected multi-modal sensor data. Specifically, in weights change estimation task, TSR-GNN MAE is improved 18.60%, RMSE improved 33.54%, and F1 score increased 60.57% compare to the mean model.

(2) Statistic methods (support vectors, decision trees) performed better in regression tasks, and deep learning approaches have higher F1 scores in classification tasks. TSR-GNN can beat both statistic methods and deep learning methods in regression and classification tasks. TSR-GNN MAE is improved 8.3%, RMSE improved 23.44%, and F1 score increased 3.88% compared to the second best model.

(3) Deep learning based models (including TSR-GNN), have better performance from a robustness perspective. They have relatively smaller standard deviation across all validation folds.

(4) Generally, methods without LSTM ablations (GIN, SR-GNN) are slightly inferior to models with LSTM (ST-GIN, TSR-GNN), which demonstrates that the temporal information between time windows are useful when predicting fluid intake.

(5) TSR-GNN and SR-GNN both overperformed other baseline models, especially on graph neural network models, which shows the potential of leveraging sensor embeddings and sensor interactions. Since, the baseline graph models GIN and ST-GIN do not consider the sensor relations.

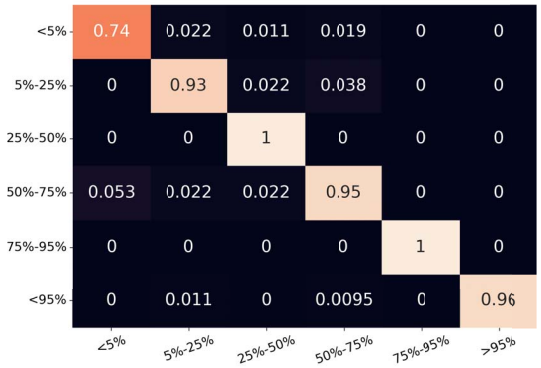


Figure 6: *Mean confusion matrix of our* TSR-GNN *method when predicting the interdialytic weight change.*

Feasibility, To prove the feasibility and effectiveness of our mobile sensing system, we select our best model TSR-GNN to analyze its prediction results. First, we calculate the error percentage of our regression tasks by

$$ER = \frac{100}{N}\sum_{i=1}^{N} |\frac{\hat{y_i} - y_i}{\hat{y_i}}|$$

The statistic of ground truth labels between two dialysis sessions is shown in table 3, and the average error percentage of our model for both weight changes and self-reported fluid intake is around 32%-33%, compared to the state-of-art wearable sensing solution based on motion sensors [20], the average error percentage reduced 8.68%.

The classification tasks also demonstrates the potential of our method. We draw an average confusion matrix figure 6 over 9 folds of validations (the rest 3 folds do not have all five labels) on weight changes prediction. The confusion matrix shows that our system maintains high accurate fluid intake level prediction, even when the quality level is above 95% where we recorded a high 97.82% accuracy, similar result in regression task, as shown in red arrows in figure 7, our approach can capture the high peaks in continuous fluid intake estimation.

Model	Weights Change			Self Reported		
	MAE	RMSE	F1	MAE	RMSE	F1
Mean	1.2669 ±0.3172	2.1820 ±2.3353	0.3682 ±0.0790	1.5195 ±0.3645	3.3823 ±0.7793	0.6556 ±0.1408
SVR/SVM	1.1247 ±0.3340	1.8942 ±2.1731	0.3571 ±0.1653	1.1534 ±0.4454	2.3678 ±0.8630	0.5308 ±0.1282
DTR/DT	1.1584 ±0.4216	2.0132 ±1.3925	0.8510 ±0.2065	1.3231 ±1.1058	2.8329 ±1.7406	0.8887 ±0.2169
GIN	1.5688 ±0.3046	1.8571 ±0.2980	0.8571 ±0.1437	1.3519 ±0.3987	2.1335 ±0.7642	0.8066 ±0.0578
LSTM	2.5126 ±0.3613	3.1186 ±1.9921	0.8429 ±0.2108	1.6365 ±0.5047	2.7184 ±0.8616	0.9182 ±0.1627
ST-GIN	1.3244 ±0.2287	1.6337 ±0.2501	0.8621 ±0.1537	1.2762 ±0.3235	1.8495 ±0.4633	0.9161 ±0.1358
SR-GNN	1.0547 ±0.2331	1.4923 ±0.2776	0.9074 ±0.1879	1.0152 ±0.3051	1.6391 ±0.4180	0.9211 ±0.1107
TSR-GNN	**1.0313 ± 0.2372**	**1.4501 ± 0.2953**	**0.9337 ± 0.1816**	**0.9573 ± 0.2618**	**1.5523 ± 0.2514**	**0.9570 ± 0.0640**

Table 2: *The performance of Interdialytic fluid intake predictions.*

Label Type	Mean	Min	Max	EP
Weights Change	3.18	0.2	6.50	32.43%
Self-reported	2.97	0.12	10.85	32.25%

Table 3: *Ground truth label statistics. EP denotes the Absolute Error Percentage to the Mean value.*

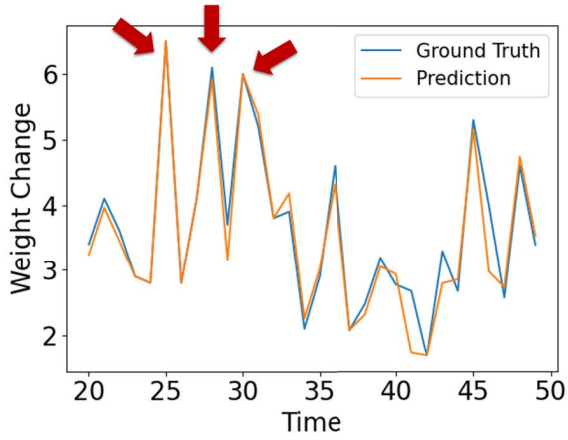


Figure 7: *An example of regression results generated by* TSR-GNN. *This example generated from one random participant illustrates how our method is able to accurately detect peaks in weight change. This is a crucial step to designing interventions in the future to help ESKD patients better anticipate and mitigate potential dangerous fluid overload.*

To further test the feasibility of using our method for monitoring fluid intake between clinical visits, we replicated the experiment on daily basis to predict the self-reported fluid intake . The reason why we only evaluate TSR-GNN on self-reported fluid intake is because the weight changes are only measured during the dialysis sessions. The results in table 4 shows similar trends in which we can see how TSR-GNN significantly outperformed all the other baselines in MAE, RMSE and F1 perspectives when predicting the daily fluid intake. This experiment demonstrates the prediction power of our model on a "between visits" granularity.

Model	Self Reported		
	MAE	RMSE	F1
Mean	0.9869 ±0.1439	1.1951 ±0.1822	0.6366 ±0.0688
SVR/SVM	0.5975 ±0.3019	1.3886 ±0.5425	0.7863 ±0.0594
DTR/DT	0.9278 ±0.2094	1.2237 ±0.2610	0.7844 ±0.1565
GIN	0.7972 ±0.2034	1.0580 ±0.2493	0.8657 ±0.1029
LSTM	1.0273 ±0.0273	1.3880 ±0.0395	0.8342 ±0.0472
ST-GIN	0.7155 ±0.3803	0.9109 ±0.4479	0.9078 ±0.0969
SR-GNN	0.4397 ±0.1054	0.5705 ±0.1317	0.9231 ±0.0522
TSR-GNN	**0.4183 ± 0.0866**	**0.5292 ± 0.0951**	**0.9497 ± 0.0325**

Table 4: *The performance of daily self-reported fluid intake predictions.*

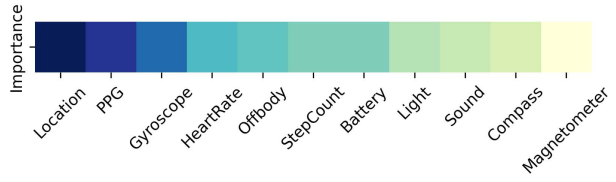


Figure 8: *Sensor importance 1D heatmap. Darker node indicate higher importance (dark blue to light green).*

5.3 RQ2. Interpretability of our TSR-GNN Model

Node Importance via GNNExplainer. To identity which nodes are more important than others at predicting fluid intake, we adopt GNNExplainer [41] on our converged model to create a node importance heatmap shown in figure 8. The original GNNExplainer learns an optimized X_s^F which maximizes the mutual information between GNN's prediction Y and the distribution of possible subgraph structures (G_s, X_s^F), where G is the subgraph structure, and X is the subgraph features. It uses a learnable feature mask F to select features from X_s, they redefine the subgraph features X_s^F as $X_s \odot F$. However in our case, we need node importance over feature importance, so we re-defined $X_s^F = X_s^N$ to $N \odot X_s$, where N acts as a learnable node mask, to highlight influential nodes.

The 1D heatmap identifies the most important nodes/sensors in the fluid intake prediction task. Location, photoplethysmography (PPG), and gyroscope sensors stood out as the most significant predictors of fluid intake. This can be justified by the followings: 1)

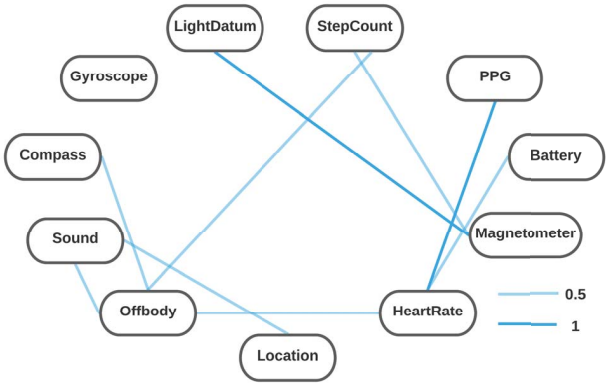


Figure 9: *Sensor relation graph according to edge attention weights.*

Mobility patterns captured from GPS locations can represent how active a particular patient is. We hypothesize that patients who are more physically active tend to drink more water to compensate for the lost water (through sweat) during physical activities. 2) PPG sensor, which generates an optically obtained plethysmogram that can be used to detect blood volume changes in the microvascular bed of tissue contains many physiological markers that can correlate with fluid overload (when ESKD patients drink water more than the standard). Heart rate variability, respiration rates, and blood oxygen levels are all indicative of fluid overload and are embedded within the PPG's plethysmogram. 3) the gyroscope sensor can characterize low-level motion data and is able to distinguish between different gestures indicative of drinking activities. We hypothesize that this sensor was less important than location and PPG because participants tend to wear the watch on their non dominant hand, so this sensor may have missed those instances when participants used their dominant hands when consuming fluids.

Sensor Relation/Interaction via Edge Attention, Edge weights learned by the attention network indicate the relation between sensor nodes. We draw a sensor relation plot depicted in figure 9 using $topk = 3$ to present the most reliable inner relationship between those multi-modal sensors.

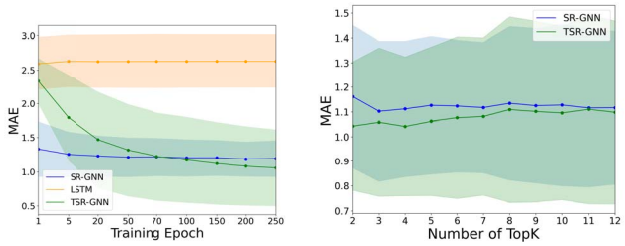
The sensor relation figure indicates that the importance of each node's neighbors in modelling the node's behavior (inner-interaction), and the relation between sensor embeddings. From figure 9, physiological biomarkers like PPG and heartrate have inner-connection indicating that they are similar and correlated. This finding proves how validated our approach is because heart rate is known to be extracted from PPG (by estimating the beat-to-beat intervals from the PPG signal). Some behavior biomarkers like step count and location are dependent on contextual sensors such off-body and sound. It's intuitively reasonable, since step count stops when wearable-devices are on off-body status and since noise levels vary based on the typle of location (e.g. indoor vs. outdoor). We only select top3 attention weights as edge between nodes, others still have other types of relationship with lower weights/weaker interaction.

In summary, 1) The sensor nodes in our model has interpretable contribution in the fluid intake estimation. The physiological and behavioral sensors contribute the most to the prediction result; 2) the attention weights help to find closely related sensors that have stronger inner-connection between each other.

5.4 RQ3. Hyperparameter Study

In table 2, we compare TSR-GNN with an ablation SR-GNN (without LSTM temporal information learning), the result shown SR-GNN lead to inferior performance on both regression and classification tasks. In this section, we present detailed results about hyper-parameter (learning curve, topk selection) and other ablation (w/o attention) analyses of TSR-GNN. We draw three figures to show the model sensitivity to hyperparameters and ablations.

As we observed from the learning curve figure 10 (a), TSR-GNN converges rapidly after 20-50 training epoch and achieve the best performance on test set after 100 epochs, other deep learning methods such as LSTM do not converge as the training epoch goes up. While SR-GNN converges quickly after 5 epochs, the regression error remain the same for the rest epochs, which indicates the proposed model w/o LSTM can converge quickly, and model with LSTM involves more parameters to train.

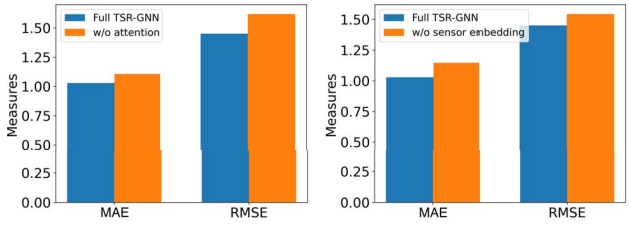


(a) Training curves vs. MAE (b) TopK vs. MAE

Figure 10: *Hyperparamter sensitivities.*

Moreover, as shown in figure 10(b), the performance of TSR-GNN and SR-GNN are robust to the neighbor size of $topk$, where $topk = 2, 3$ provide slight better performance than smaller and extremely big size.

Finally, to further understand the effectiveness of our designed attention-based model, we draw a bar plot in figure 11 to compare TSR-GNN with TSR-GNN w/o attention or w/o sensor embedding in node aggregation.



(a) Full vs. w/o Attention Ablation (b) Full vs. w/o sensor embedding

Figure 11: *Model ablations compression.*

To guarantee the fairness in model evaluation, we set the same set of parameters in both models including the same $topk$ and

learning rate. From figure 11 (a) and (b), the full TSR-GNN model achieves better result than TSR-GNN model without attention and sensor embeddings, indicating the effectiveness of the attention mechanism and learnable sensor embeddings applied here.

6 Discussion and Future Work

To the best of our knowledge, this is the first ubiquitous fluid intake monitoring method applied to ESKD patients using relational graph neural networks and on-body multi-modal mobile sensing. Instead of predicting the fluid intake behavior by inaccurate static single sensor estimation (e.g., gesture recognition), we introduce the importance of temporal information and multi-modal sensor interaction. Though the computational cost is higher than former methods, the proposed solution outperformed the-state-of-the-art predictive modeling models when predicting fluid intake of ESKD patients. The concept of sensor relationship is not only useful in fluid prediction task, but can also be applied to many other domains involving sensor networks.

Clinical implications 1) Our work provides evidence that mobile sensing techniques (e.g., GPS, PPG, Gyroscope, etc.) are potentially useful in designing accurate, ubiquitous, and unobtrusive fluid management systems and can perform as an additive system to the clinical examination, hemodynamic monitoring and weight assessment done three times weekly on dialysis. This can potentially help key stakeholders (patients, clinicians, providers, organizations) and represent a major component for next generation personalized intervention for dialysis patients. 2) Our work can provide rich information about patients' fluid intake patterns which can be leveraged to design adaptive personalized, context-relevant health services and treatment plans. 3) Although the time granularity tested in this work is in the order of one day and two days (time between two dialysis sessions), the same work can be used on varying time scales (e.g., hours, weeks, etc.).

Limitation We acknowledge that our current work is subject to the following limitations. 1) Our sample size is small. The collected data was generated from N=14 ESKD patients only. This is mainly due to the challenges associated with recruiting ESKD patients who tend to be less tech-savvy (due to age), have many commodities that interfere with the data collection (e.g., vision problem), and are busy with the burdensom dialysis treatement. Given all these challenges, the longitudinal dataset collected in this study is still unique and is still the first in its kind containing wearable sensing data from ESKD patients over multiple weeks. 2) There are multiple sensors that have not been investigated in this study (e.g., bio-impedance and sleep quality) that can have significant correlation with fluid overload. Those sensors were not present on the Fossil Gen 5 smartwatch used in this study.

In the future, we plan to replicate this study in a larger sample size, while including other sensing modalities and while collecting ground truth data on a more granular scale (e.g., every day). We will also develop forecast models to forecast the trajectories of fluid accumulation over time (e.g., in the next 24h, 48h, etc.) These additional and fully powered studies will allow us to better assess the clinical significance of our efforts, including establishing accuracy thresholds for mobile fluid sensing techniques (relative to and in combination with clinical examination, hemodynamic monitoring and in-clinic weight assessment), as well as determining minimally clinically important differences and effect sizes for different mobile fluid sensing modalities and modeling procedures. Likewise, the aforementioned studies are needed to determine the potential for mobile fluid sensing techniques to change fluid monitoring practices and recommendations in ESKD patients.

7 Conclusion

In this work, we present a novel wearable sensing fluid intake monitoring system, which combines multi-modal sensing and graph neural networks to collect physiological biomarkers, behavioral biomarkers, and contextual markers of fluid intake. To better leverage multi-modal sensor data, we propose a novel temporal sensor-relation graph neural network (TSR-GNN) to capture the inner relationship between different sensors and fuse sensor's data accordingly. The real-world empirical experiments demonstrate that our method overperformed all other baseline methods in both regression and classification tasks for interdialytic fluid intake estimation. Especially, TSR-GNN reduces the average error percentage by 8.68% compared to the state-of-arts fluid intake estimation methods, and achieve 97.82% f1 score on extreme fluid overload situation. The in-depth interpretability analysis provides a view of how physiological and behavioral sensors contribute to the prediction power.

Our proposed fluid intake monitoring solution can provide continuous and unobtrusive fluid manage solution to many health conditions in which fluid monitoring is critical. The interpretabilies of our model can also provide a fluid intake featurization cookbook to future fluid monitoring research.

References

[1] [n. d.]. CKD by the Numbers, howpublished = https://www.cdc.gov/kidneydisease/basics.html.

[2] [n. d.]. FluiSense, howpublished = https://play.google.com/store/apps/details?id=com.mob.fluisense&hl=en_US&gl=US.

[3] [n. d.]. UNITED STATES RENAL DATA SYSTEM. (n.d.)., howpublished = https://www.usrds.org/.

[4] 2021. Fluisense - Apps on Google Play. https://play.google.com/store/apps/details?id=com.mob.fluisense&hl=en_US&gl=US [Online; accessed 28. Oct. 2021].

[5] Jeremy D Akers, Rachel A Cornett, Jyoti S Savla, Kevin P Davy, and Brenda M Davy. 2012. Daily self-monitoring of body weight, step count, fruit/vegetable intake, and water consumption: a feasible and effective long-term weight loss maintenance approach. *Journal of the Academy of Nutrition and Dietetics* 112, 5 (2012), 685–692.

[6] Rahul Bhattacharyya, Christian Floerkemeier, and Sanjay Sarma. 2010. RFID tag antenna based sensing: Does your beverage glass need a refill?. In *2010 IEEE International Conference on RFID (IEEE RFID 2010)*. IEEE, 126–133.

[7] Mehdi Boukhechba, Anna N Baglione, and Laura E Barnes. 2020. Leveraging Mobile Sensing and Machine Learning for Personalized Mental Health Care. *Ergonomics in Design* 28, 4 (2020), 18–23.

[8] Mehdi Boukhechba and Laura E Barnes. 2020. Swear: Sensing using wearables. Generalized human crowdsensing on smartwatches. In *International Conference on Applied Human Factors and Ergonomics*. Springer, 510–516.

[9] Enea Cippitelli, Samuele Gasparrini, Ennio Gambi, and Susanna Spinsante. 2016. Unobtrusive intake actions monitoring through RGB and depth information fusion. In *2016 IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP)*. IEEE, 19–26.

[10] Rachel Cohen, Geoff Fernie, and Atena Roshan Fekr. 2021. Fluid Intake Monitoring Systems for the Elderly: A Review of the Literature. *Nutrients* 13, 6 (2021), 2092.

[11] Ailin Deng and Bryan Hooi. 2021. Graph neural network-based anomaly detection in multivariate time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 4027–4035.

[12] Duarte Dias and João Paulo Silva Cunha. 2018. Wearable health devices—vital sign monitoring, systems and technologies. *Sensors* 18, 8 (2018), 2414.

[13] Guimin Dong, Mehdi Boukhechba, Kelly M Shaffer, Lee M Ritterband, Daniel G Gioeli, Matthew J Reilley, Tri M Le, Paul R Kunk, Todd W Bauer, and Philip I Chow. 2021. Using Graph Representation Learning to Predict Salivary Cortisol

Levels in Pancreatic Cancer Patients. *Journal of Healthcare Informatics Research* (2021), 1–19.

[14] Guimin Dong, Lihua Cai, Debajyoti Datta, Shashwat Kumar, Laura E Barnes, and Mehdi Boukhechba. 2021. Influenza-like symptom recognition using mobile sensing and graph neural networks. In *Proceedings of the Conference on Health, Inference, and Learning*. 291–300.

[15] Guimin Dong, Mingyue Tang, Lihua Cai, Laura E. Barnes, and Mehdi Boukhechba. 2021. Semi-supervised Graph Instance Transformer for Mental Health Inference. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*. 1221–1228. https://doi.org/10.1109/ICMLA52953.2021.00198

[16] Harris Drucker, Chris JC Burges, Linda Kaufman, Alex Smola, Vladimir Vapnik, et al. 1997. Support vector regression machines. *Advances in neural information processing systems* 9 (1997), 155–161.

[17] Jennifer E Flythe and Eduardo Lacson Jr. 2012. Outcomes after the long interdialytic break: implications for the dialytic prescription. In *Seminars in dialysis*, Vol. 25. Wiley Online Library, 1–8.

[18] William L Hamilton, Rex Ying, and Jure Leskovec. 2017. Representation learning on graphs: Methods and applications. *arXiv preprint arXiv:1709.05584* (2017).

[19] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[20] Hsiang-Yun Huang, Chia-Yeh Hsieh, Kai-Chun Liu, Steen Jun-Ping Hsu, and Chia-Tai Chan. 2020. Fluid Intake Monitoring System Using a Wearable Inertial Sensor for Fluid Intake Management. *Sensors* 20, 22 (2020), 6682.

[21] Jeremy F Huckins, Alex W DaSilva, Weichen Wang, Elin Hedlund, Courtney Rogers, Subigya K Nepal, Jialing Wu, Mikio Obuchi, Eilis I Murphy, Meghan L Meyer, et al. 2020. Mental health and behavior of college students during the early phases of the COVID-19 pandemic: Longitudinal smartphone and ecological momentary assessment study. *Journal of medical Internet research* 22, 6 (2020), e20185.

[22] Alexandros Iosifidis, Ermioni Marami, Anastasios Tefas, and Ioannis Pitas. 2012. Eating and drinking activity recognition based on discriminant analysis of fuzzy distances and activity volumes. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2201–2204.

[23] Asangi Jayatilaka and Damith C Ranasinghe. 2017. Real-time fluid intake gesture recognition based on batteryless UHF RFID technology. *Pervasive and Mobile Computing* 34 (2017), 146–156.

[24] Guangyin Jin, Min Wang, Jinlei Zhang, Hengyu Sha, and Jincai Huang. 2022. STGNN-TTE: Travel time estimation via spatial–temporal graph neural network. *Future Generation Computer Systems* 126 (2022), 70–81.

[25] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).

[26] Zongyu Lin, Shiqing Lyu, Hancheng Cao, Fengli Xu, Yuqiong Wei, Hanan Samet, and Yong Li. 2020. HealthWalks: Sensing Fine-grained Individual Health Condition via Mobility Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–26.

[27] DeAnn Liska, Eunice Mah, Tristin Brisbois, Pamela L Barrios, Lindsay B Baker, and Lawrence L Spriet. 2019. Narrative review of hydration and selected health outcomes in the general population. *Nutrients* 11, 1 (2019), 70.

[28] Qi Liu, Ruobing Xie, Lei Chen, Shukai Liu, Ke Tu, Peng Cui, Bo Zhang, and Leyu Lin. 2020. Graph neural network for tag ranking in tag-enhanced video recommendation. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2613–2620.

[29] Chen Ma, Liheng Ma, Yingxue Zhang, Jianing Sun, Xue Liu, and Mark Coates. 2020. Memory augmented graph neural networks for sequential recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 5045–5052.

[30] Mark Mirtchouk, Christopher Merck, and Samantha Kleinberg. 2016. Automated estimation of food type and amount consumed from body-worn audio and motion sensors. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 451–462.

[31] Ruihong Qiu, Hongzhi Yin, Zi Huang, and Tong Chen. 2020. Gag: Global attributed graph neural network for streaming session-based recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 669–678.

[32] Yili Ren, Sheng Tan, Linghan Zhang, Zi Wang, Zhi Wang, and Jie Yang. 2020. Liquid Level Sensing Using Commodity WiFi in a Smart Home Environment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–30.

[33] Simone Stickel, Wendy Gin-Sing, Martha Wagenaar, and J Simon R Gibbs. 2019. The practical management of fluid retention in adults with right heart failure due to pulmonary arterial hypertension. *European Heart Journal Supplements* 21, Supplement_K (2019), K46–K53.

[34] Imam Yogie Susanto, Tse-Yu Pan, Chien-Wen Chen, Min-Chun Hu, and Wen-Huang Cheng. 2020. Emotion recognition from galvanic skin response signal based on deep hybrid neural networks. In *Proceedings of the 2020 International Conference on Multimedia Retrieval*. 341–345.

[35] Geoffrey KF Tso and Kelvin KW Yau. 2007. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy* 32, 9 (2007), 1761–1768.

[36] Fabian Wahle, Tobias Kowatsch, Elgar Fleisch, Michael Rufer, Steffi Weidt, et al. 2016. Mobile sensing and support for people with depression: a pilot trial in the wild. *JMIR mHealth and uHealth* 4, 3 (2016), e5960.

[37] Xiaoyang Wang, Yao Ma, Yiqi Wang, Wei Jin, Xin Wang, Jiliang Tang, Caiyan Jia, and Jian Yu. 2020. Traffic flow prediction via spatial temporal graph neural network. In *Proceedings of The Web Conference 2020*. 1082–1092.

[38] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojun Chang, and Chengqi Zhang. 2020. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 753–763.

[39] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2018. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826* (2018).

[40] Sijie Yan, Yuanjun Xiong, and Dahua Lin. 2018. Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Thirty-second AAAI conference on artificial intelligence*.

[41] Rex Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. 2019. Gnnexplainer: Generating explanations for graph neural networks. *Advances in neural information processing systems* 32 (2019), 9240.

[42] Julian Zelingher, Chaim Putterman, Yaron Ilan, Eldad J Dann, Fabio Zveibil, Yigal Shvil, and Eithan Galun. 1996. Case series: hyponatremia associated with moderate exercise. *The American journal of the medical sciences* 311, 2 (1996), 86–91.

[43] Weijia Zhang, Hao Liu, Yanchi Liu, Jingbo Zhou, and Hui Xiong. 2020. Semi-supervised hierarchical recurrent graph neural network for city-wide parking availability prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 1186–1193.

[44] Bo Zhou, Jingyuan Cheng, Mathias Sundholm, Attila Reiss, Wuhuang Huang, Oliver Amft, and Paul Lukowicz. 2015. Smart table surface: A novel approach to pervasive dining monitoring. In *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 155–162.