Dynamic Spatial-Temporal Graph Model for Disease Prediction

Ashwin Senthilkumar[1]
School of Electronics Engineering
Vellore Institute of Technology
Chennai, 600127

Mihir Gupte[2]
School of Computer Science
and Engineering
Vellore Institute of Technology
Chennai, 600127

Shridevi S.[3]
Center for Advanced Data Science
Vellore Institute of Technology
Chennai, 600127

***Abstract*—Advances in the field of Neural Networks, especially Graph Neural Networks (GNNs) has helped in many fields, mainly in the areas of Chemistry and Biology where recognizing and utilising hidden patterns is of much importance. In Graph Neural Networks, the input graph structures are exploited by using the dependencies formed by the nodes. The data can also be transformed in the form of graphs which can then be used in such models. In this paper, a method is proposed to make appropriate transformations and then to use the structure to predict diseases. Current models in disease prediction do not fully use the temporal features that are associated with diseases, such as the order of the occurrence of symptoms and their significance. In the proposed work, the presented model takes into account the temporal features of a disease and represents it in terms of a graph to fully utilize the power of Graph Neural Networks and Spatial-Temporal models which take into consideration of the underlying structure that change over time. The model can be efficiently used to predict the most likely disease given a set of symptoms as input. The model exhibits the best algorithm based on its accuracy. The accuracy of the algorithm is determined by the performance on the given dataset. The proposed model is compared with the existing baseline models and proves to be outstanding and more promising in the disease prediction.**

***Keywords*—Spatial temporal graph convolution network; disease prediction; graph neural network; graph convolutional network; deep learning; knowledge graph**

I. Introduction

In recent times, Machine Learning has become very popular in the field of Medical Sciences, especially when it comes to detecting patterns that are associated with health and diseases. Disease prediction is one of the most sought after applications here, and the reason for that is the patterns that are associated with each disease, be it physical illness or mental well-being. Diseases are often associated and known for the symptoms that occur when a person is infected, therefore a study of the underlying structure of the symptoms, their behaviour, etc. is important for predicting them. These underlying structures usually have a temporal nature to them as the symptoms don't occur all at once, but they do in more or less sequential order.

It is common to see the use of ML models such as Naive-Bayes, Decision Trees and other classifiers to be used in these scenarios. The most common source for obtaining information about symptoms is by referring to Electronic Health Records (EHRs). Often these are used to extract features for input. In [1], the authors make a heart disease prediction model using ML techniques. They utilize features such as Age, Sex, Pain levels, etc of patients as input. In [2], it was found that most people tackled this problem using Support Vector Machines (SVM) and Naive-Bayes models. However, this did not prove to be efficient when dealing with huge volumes of Electronic Medical Records (EMRs) data without processing it and simply feeding it to these models, nor can these models take advantage of the temporal structure of the symptoms. To overcome these many researchers have started utilising Deep Learning, as a neural network can detect and represent the hidden and latent features more efficiently. In [3], the authors proposed a graph convolution network with mutual attention networks, to learn from EMR's directly and diagnose the patient. The author obtained an accuracy of 63.46% from the MIMIC-III dataset [4], which was better compared to all the other models in the paper like Convolutional Neural Networks (CNN), Graph Convolutional Network(GCN), etc. The authors in [5] use an Artificial Neural Network (ANN) to develop a model for Parkinson's disease and boast almost perfect accuracy. A method proposed in [6] uses Cascading Neural Networks in order to detect Melanoma, a form of skin cancer. In [7], the authors have developed a model called InceptionGCN for disease prediction and tested it out on various datasets. Thus it can be concluded that neural networks, especially Graph Neural Networks play an important role in the new and upcoming models of disease predictions.

However, most of the above-mentioned papers fail to utilise the temporal features of the symptoms that are associated with the diseases. Some papers [8],[9] make use of GNNs tailored for this specific purpose especially in [9],[10] where the authors use patient data to first construct a dependency of the patient with symptoms on each visit and a final diagnosis and then model a graph from which is fed to the network proposed by them. This model could achieve an accuracy of a little over 85% at the most favourable conditions.

So while the use of spatial-temporal networks is not new to this problem, a new architecture that models symptoms of a patient into a graph is quite novel. This paper proposes a method to model the input data by changing the underlying structure of the graph at each timestep, so as to utilise the temporal features of the input. Knowledge graph for diseases by [11] is utilized for proposed architecture. In this, the authors employ probabilistic models such as logistic regression, naive Bayes classifier, to derive features of various diseases from many EHRs. The authors feature a knowledge graph built on these features and so in the presented work a knowledge graph

based Spatial-Temporal GCN is proposed.

The major contributions of the paper are:

- A temporal approach to predicting diseases in order to make the models more robust and potentially exploiting such dependencies to enhance the quality of the proposed model.
- Graph Convolutional Network(GCN) model is proposed which exploits the spatial dependencies of the input first then uses these to explore the temporal dependencies of the input. This will help to understand how the symptoms relate to each other at each timestep and how that underlying relationship changes. The model will exploit this to predict and classify diseases.
- A graph model of the input data in the form of diseases and the symptoms that occur alongside proposed as well. Each symptom is associated with a probability which indicates how likely the symptom is to occur (as compared to all other symptoms for the same disease). Data is convert into a graph.

II. Related Works

Disease prediction has been difficult with a lack of patient history and noisy data. The work tries to overcome the difficulties and improve efficiency by using knowledge graphs obtained from medical ontology. Further, mechanism is also discussed to predict diseases using symptoms via a spatial-temporal Graph Convolutional Network.

Xuedong Li et al. [12] takes an innovative approach to overcome the lack of medical history by using medical knowledge. The lack of medical history of patients owing to the nature of rare medical diseases makes it a challenging process even for machine learning approaches to recognize the diseases. They have developed a text classification algorithm to create a bag of knowledge terms from the medical ontology to develop a knowledge graph that can be leveraged for the disease classification task. It works efficiently even if the knowledge graph of medical history is incomplete. The limitation of this approach is that the dataset being used is extremely imbalanced.

Rotmensch, M., Halpern, Y. et al. [9] has directed a review observational investigation utilizing recently gathered information from Electronic Medical Records (EMR) in order to develop a knowledge graph that relates indications to sicknesses and assessed competitor knowledge graphs against physically curated knowledge graph given by (Google well being knowledge graph, or GHKG) and also the expert opinion of physicians. However, the purpose of the knowledge graph was to test how efficiently a given algorithm could recover unknown causal relationships between the diseases and their symptoms. A drawback is that any approach that infers causal relations from observational data has major limitations inherently. The algorithms should be seen as a method of providing casual relations between the entities.

Zhenchao Sun, Hongzhi Yin et al. have introduced an innovative model [13] using GNNs for disease prediction. It uses multiple knowledge bases in order to obtain sufficient EMR data to learn highly representative node embeddings of medical concepts graph and the patient record graph (which include entities such as, the patients, diseases and symptoms), and are thus constructed from the medical knowledge base and EMRs. This results in accurate disease prediction for new patients under sparse data in an inductive manner.

Li, Y., Qian, B et al. have proposed GNDP, a disease prediction model which is based on a graph convolutional network. It exploits the spatial structure of the EHR data and the temporal dependencies of the entities to predict the patient's future diagnosis which is similar to [8,10]. Sun, Z., Dong, W. et al. propose a Reinforcement Learning mechanism that would take random walks over the knowledge graph with respect to the patient's symptoms and then propose the most likely disease. The authors have manually constructed the knowledge graph using the Mimic-III PLAGH dataset. A single knowledge graph has been made for all diseases. It is limited in terms of accuracy, as accuracy could be improved by using efficient methods. GNN models can be considered as the structure of the data is a graph and the performance can be improved using GNN algorithms.

Yuan, Q., Chen, J., et al. have constructed an elaborate GCN model [3] in which they have extracted symptoms from a given diagnosis in the form of a string. Then they embed the diseases first in the D-D GCN layer and attach symptoms (features) to the diseases in the next D-F GCN layer. Less important features have been pruned from the knowledge graph in this model. Finally, these features are given to a convolutional network modelled using attention, so that the most important disease corresponding to the symptoms is obtained. The accuracy obtained is only 63%. The limitation is that the model is not interpretable and complex to understand. Potential change or simplification of algorithms can lead to better accuracy. It only considers the features extracted from diagnosis and not the temporal relations between the features.

III. Methodology

In this section, the proposed architecture of spatial-temporal graph convolutional networks(STGCN) in elaborately discussed. STGCN is made up of spatial-temporal convolutional blocks that are arranged in a "sequential" structure with one sequential convolution layer and one spatial graph convolution layer, as shown in Fig. 1. The following sections go over the specifics of each module.

A. Data Collection and Preparation

One authentic medical dataset is experimented on to evaluate the proposed model. knowledge graph from electronic medical records [3], a public accessible benchmark dataset for diseases knowledge graphs with high-quality knowledge bases linking diseases and symptoms derived from the EMRs. These electronic records represent medical concepts collected from over 270,000 patient visits to the Emergency Department at Beth Israel Deaconess Medical Center (BIDMC), thus the knowledge graphs was automatically constructed using maximum likelihood estimation of three probabilistic models.

From the learned parameters, a graph of disease-symptom connections was elicited, and the developed knowledge graphs were assessed and approved, with consent, against Google's

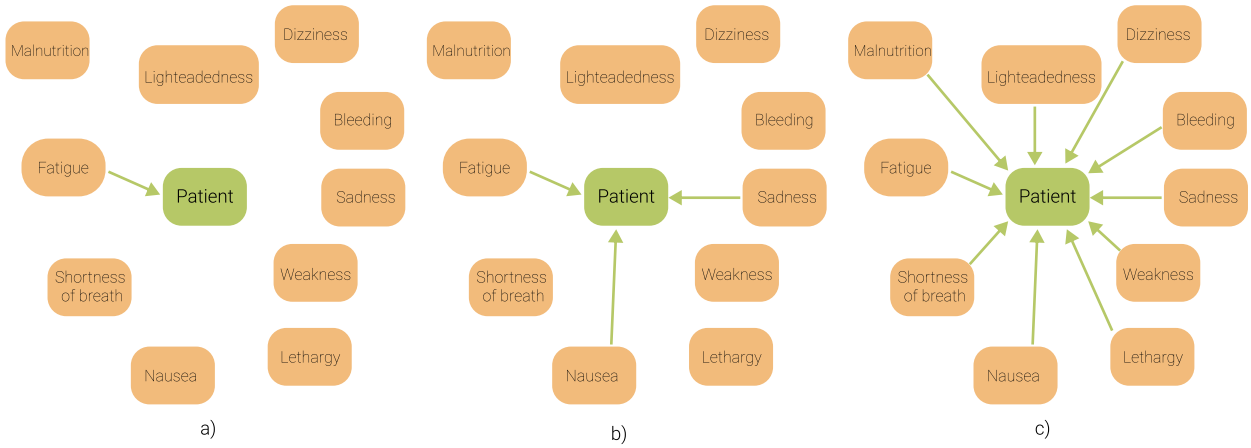

Fig. 1. Example of the Anemia Disease Dynamic Knowledge Graph that Generated in Preprocessing of the Dataset. a) Knowledge Graph at T = 0 b) Knowledge Graph at T = 3 c) Knowledge Graph at T = n, where n is Length of Disease Trajectory

manually built knowledge graph and master doctor suppositions. The graph records all 156 diseases and 491 symptoms, all edges between diseases and symptoms, and the significance scores given to each edge.

Temporal patterns in patient disease trajectories are either disregarded or only taken into account by assessing the temporal directionality of identified co-morbidity pairs [14], [15], [16], Which concludes naturally patients undergo different symptoms at different time instances on disease trajectories (stages of the disease). In this experiment, the temporal pattern for each disease as the preliminary basis of a disease prediction model is utilized.

Provided in Table I are the sample dataset values which includes a given disease and its symptoms along with the probability of a given symptom occurring. In Table II, important statistics of the dataset itself are given.

TABLE I. SAMPLE DATASET

Diseases	Symptoms
abscess	$pain(0.318), fever(0.119), ..$
anemia	$lethargy(0.096), weakness(0.087), ..$
common cold	$chills(0.083), sorethroat(0.075), ..$

Before applying the STGCN model to knowledge graphs extracted from electronic medical records [11], the existing knowledge graph must go through a series of steps to create a dynamic graph structure by leveraging temporal features for each disease knowledge graph.

As the dataset is not available with time element related with disease symptoms, a suitable time feature is created for the knowledge graph collected from electronic medical records [11]. This was done by generating all possible permutations of order symptoms that will occur in the patient's disease trajectories.

Then using the importance scores associated with each edge of symptoms, the top ten symptoms were selected with the highest probability of occurrence inpatient disease trajectories for all possible permutations. Since the disease prediction is a multi-class classification task, more than one data point is needed on each class but dataset have only one data point on each class. To overcome this top k permutations were selected with the highest importance scores associated with each edge of symptoms.

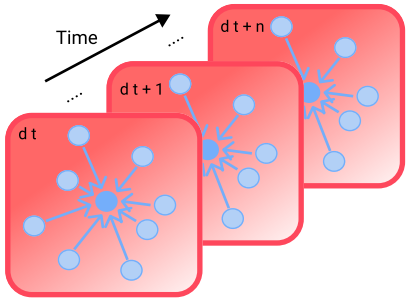

Fig. 2. Graph-Structured Disease Data. Each d_t Indicates a Frame of Current Disease Trajectory at Time Step t, which is Recorded in a Graph Structured Data Matrix.

To generate different graphs for each time instance with available data, by introducing a new initial node with the label as patient and linking new edges between each symptom node and patient node for each disease trajectory growth as shown in the figure.

This results in T unique graphs with the temporal patterns for each disease, where T represents the length of the disease trajectory. Nodal features of each node as initialized with a unique label code of symptoms and each edge weight of each edge assigned based on the importance score of symptoms associated with it.

After preprocessing the knowledge graph extracted from electronic medical records [11], 'n' disease classes of 'k' top permutations were generated with the highest importance scores associated with each edge of symptoms and each permutation contains 'T' unique dynamic graphs with the temporal patterns for each disease. The dimension of each dynamic graph data will be [Number of disease classes 'x' number to permutations 'x' length of disease trajectory].

STGCN disease prediction was a spatial-temporal graph

classification task, hence the final dataset will contain, Adjacency matrix with the shape of (length of disease trajectory x number of nodes x number of nodes), Nodal features (length of disease trajectory x number of nodes), graph label (1 x 1).

TABLE II. IMPORTANT STATISTICS OF THE FINAL DATASET

Features	
Size of the Dataset	2000
Unique Diseases	100
Features per Disease	10
Unique Symptoms	308

IV. PROPOSED MODEL

In this section, the proposed architecture of Spatial-Temporal Graph Convolutional Networks (STGCN) is discussed in more detail. STGCN is made up of spatial-temporal convolutional blocks that are arranged in a "sequential" structure with one sequential convolution layer and one spatial graph convolution layer, as illustrated in Fig. 2. Each module's specifics are as follows.

Fig. 1 depicts an overview of knowledge-based STGCN and Fig. 2 depicts the convolution process. When importing disease data into STGCN in temporal graph format. Two convolution layers comprise the STGCN unit shown below. By broadcasting the features of each node along with the graph edge via the property of graph convolution, the first layer performs spatial graph convolutional operation parallelly on each dynamic graph of a different time instance, then node features vectors containing the aggregation label information of their neighbours can be extracted [17]. The next layer employs the temporal layer to capture the temporal relations of the resultant feature vectors of each entity in the graph.

The entire architecture of the model is depicted via Fig. 3 and Fig. 4. In the former it can be observed how samples of different timestamps are given as input to the network and how they are processed in the spatial layer, while in the latter the input obtained is processed by the temporal layer and the fully connected layer to lastly obtain the output.

A global attention pooling layer [18] is used after the STGCN unit and the output is reshaped in order to achieve proper feature dimension before the output is given to the Fully Connected unit. This unit consists of two linear layers. The first linear layer has an input channel of 55 and an out channel size of 128. The output of the first linear layer was passed through the batch normalization layer, activation function and dropout layer (p=0.3). The following second linear layer has an input channel of 128 and an output channel of n, where n is the number of diseases. Finally, a Softmax layer is applied to predict the final output which is then used for classification. This model is capable of being trained in an end-to-end scenario and the configuration is unified.

A. Spatial Graph Convolution Layer

The spatial graph convolution layer performs the first convolution operation on incoming data. Using the adjacency matrix A and the nodes feature vector F as inputs, the following function defined by [17] can perform an effective and effective convolution operation. Adjacency matrix A was converted into Edge Index E_i and Edge weight E_w for simplicity.

$$X' = (D'^{-1/2} A' D'^{-1/2}) \cdot X \tag{1}$$

Where A' = A + I denoted the adjacency matrix with inserted self-loop and $D'_{ii} = \sum A'_{ij}$ its diagonal degree matrix. The adjacency matrix can include other values than 1 representing edge weights via optional edge weight E_w tensor.

Its nodes wise formulation is given by:

$$X'_i = \sum_{j \in N(v) \bigcup i} \left(\frac{e_{i,j}}{\sqrt{d'_j \cdot d'_i}} \cdot X_j \right) \tag{2}$$

with $d'_i = 1 + \sum_{j \in N(i)} e_{i,j}$ denoted the edge weight from the source node j to target node i.

In this experiment, the temporal patterns for each disease knowledge graph simulate changing nodal features and adjacency matrix for each time instance by creating dynamic graph data for each disease. To perform spatial convolution operations for each time instance. The knowledge graph data is passed to T parallel spatial convolution layers, where T represents the length of disease trajectory (the number of time instances recorded in the disease knowledge graph).

A tensor of (Nf, Ei, Ew) can be used to represent the input feature of a spatial graph convolutional layer, where Nf represents node features (symptom labels) of the dynamic disease knowledge graph, Ei represents the edge index of the dynamic graph data, and Ew represents the edge weight of the dynamic graph data. A new tensor with the shape of (output Channel, Number of nodes, dimension of node features) is generated by using the Conv2D layer which is the standard 2D convolution layer. This is implemented using [1,1] kernel size and (4,4) stride as features on the input tensor, which is obtained by multiplying the input matrix with learnable weight matrix W and adding the bias b. The graph convolution is the result of the product of the normalised adjacency matrix A' and the new tensor's 2^{nd} dimension. Finally, a tensor with the dimensions (output Channel, Number of Nodes, Node Features Dimension) can be created.

$$F_{spatial} = GCN(F_{in}) = A' \times F'_{in} \tag{3}$$

$$F'_{in} = \sum F_{in} \cdot W + b \tag{4}$$

B. Temporal Convolutional Layer

The dynamic disease graph's temporal aspect is created by stacking the output of the spatial convolutional layer and generating a feature matrix. The temporal axis is well-ordered, with the duration of the disease trajectory limitation, allowing for a straightforward convolutional process to extract temporal information.

The input feature matrix $Fspatial$ is implemented as a tensor with the dimensions (T, Number of nodes, spatial Output

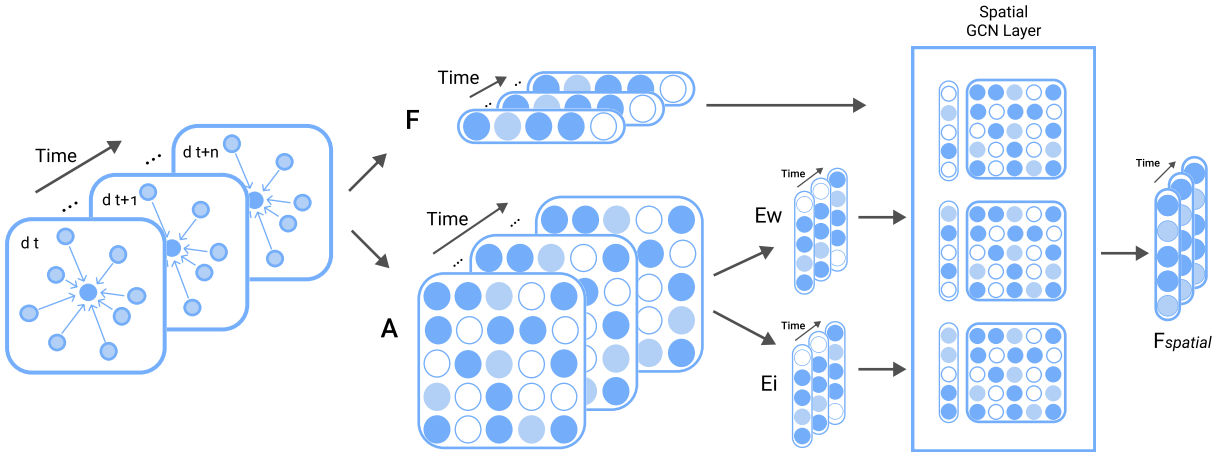

Fig. 3. F Includes the Input Node Feature for given Graph, A Indicates the Adjacency Matrix in Time Order, E_i Signifies Edge Index Matrix in Time Order, E_w Means Edge Weight in Time Order. $F_{spatial}$ is the Final Feature Vector which is Obtained as the Output.

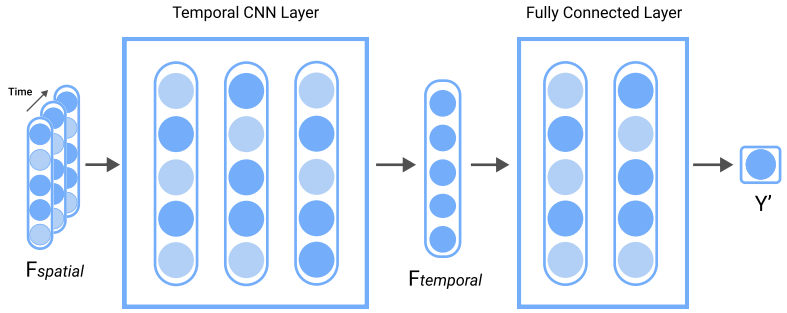

Fig. 4. The $F_{spatial}$ Denotes the Output Features of Spatial Convolutional Layer in Time Order, $F_{temporal}$ Denotes the Output Features of Temporal Convolutional Layer, Y' Denotes the Predicted Disease Class.

channel), where T is the length of the illness trajectory. The temporal kernel size is a parameter that controls how many timestamps are included in the disease graph sequence. The temporal convolutional layer's output channel dimension was determined by parameter γ [19].

$$Output\ channel = (Input\ Channel - y) + 1 \quad (5)$$

Thus, inspired by [20], [21], the temporal convolution operation can be defined as

$$F_{temporal} = TCN(F_{spatial}) = \sum_{i=0}^{\gamma-1} F_{spatial} \cdot W + b \quad (6)$$

V. Result

A. Objective Function

Cross-entropy loss is used as the objective function because disease prediction is a multi-class classification task. Using cross-entropy loss, the loss is quantified between the ground truth class d and the model output y', represented by the following formula:

$$loss(y', d) = \frac{1}{n} \sum d^T \cdot log(y') + (1 - d)^T \cdot log(1 - y') \quad (7)$$

where n is the total number of category classes.

B. Implementation Details

All the mentioned approaches are implemented using PyTorch 1.9.0, PyTorch-Geometric, PyTorch Geometric Temporal. All training processes are refined through Nvidia T4 GPU of 8.1 TFLOPS Performance and CUDA 11.1 with Intel(R) Xeon(R) processor. The dataset is then divided into various proportions to assess the performance of the model. It is haphazardly separated into training, validation, testing set in a 0.70 : 0.15: 0.15.

C. Baseline Methods

A comparison is established with the models given below in Table 3, but since the other models work on different datasets a direct comparison cannot be made. However their performance is highlighted with respect to the proposed model in terms of their accuracy, precision, recall, etc.. These models are:

- Graph Neural Disease Prediction model is proposed in [9] which also implements STGCN blocks on patient data for prediction.

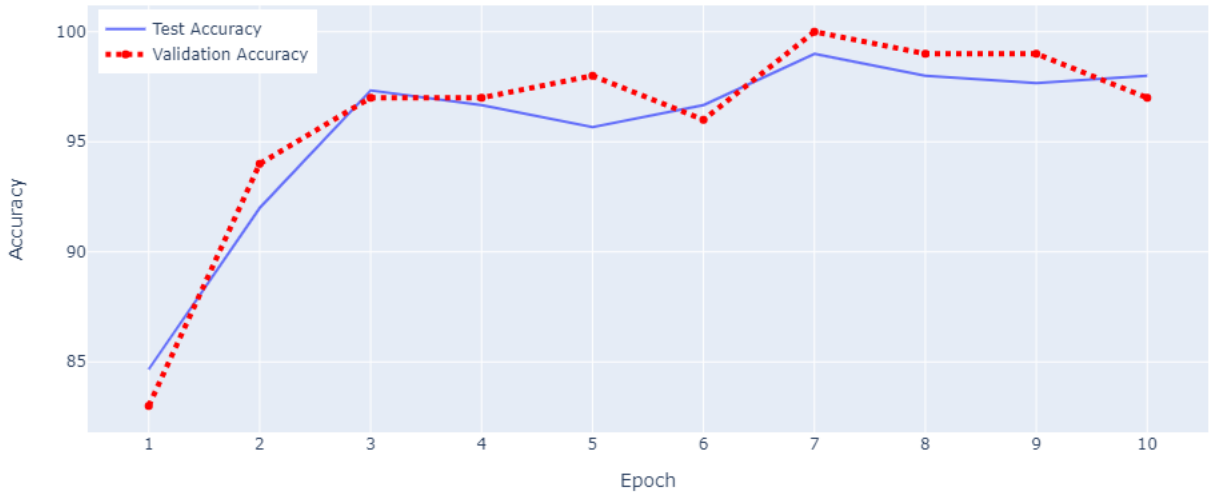

Fig. 5. Validation Accuracy and Test Accuracy Score for each Epoch

TABLE III. COMPARISON WITH OTHER DISEASE PREDICTION MODELS (CONSIDER METRIC AS ACCURACY IF NOT MENTIONED OTHERWISE)

Sr No.	Model	Dataset used	Number of Classes	Performance Metrics
1	**Proposed**	**EMR [11]**	**100**	**98.00%**
2	DP-GNN [13]	Proprietary EMR	71	Recall - 75.70% Precision - 22.40%
3	Inception GCN [7]	TADPOLE [22]	1	83.40%
2*4	2*RPL [23]	PLAGH [24]	2*65	AUC - 74.30%
		MIMIC [4]		AUC - 63.9%
2*5	2*GNDP [9]	MIMIC [4]	171	86.29%
		EHR Database	154	87.49%
2*6	2*LPNL [25]	TADPOLE [22]	2*1	91.85%
		UKBB		63.91%
7	HRP [26]	NHANES [27]	13	85.50%
8	GMAN [3]	MIMIC [4]	50	Recall - 62.13% Precision - 63.46%

- This is a GNN model for disease prediction proposed in [13] (DP-GNN). The authors map a graph of symptoms linked to the diseases and then use Graph Neural Networks to test their model.
- The model InceptionGCN is proposed in [7]. The author's utilise GCN layers in their model, however it is used to predict only one class of brain disease.
- The model proposed in [23] utilises Reinforcement Path Learning (RPL) over knowledge graph to predict diseases.
- This model, [25] makes use of simple Multilayer Perceptrons and Latent patient Network (LPNL) in order to predict diseases.
- The model given in [26] describes a method for representing symptoms in the form of knowledge graphs for health risk prediction (HRP).
- Graph Mutual Attention Network (GMAN) - This model [3] makes use of attention layers over graph convolution layers for disease prediction.

D. Evaluation Results

The model is trained to differentiate among 100 unique diseases by using 308 unique symptoms. Using LabelEncoder each disease is given a unique label id from 1-100, similarly the symptoms are labelled. For batching, 10 dynamic knowledge graphs are used per batch, and the $spatial_\gamma$ and $temporal_\gamma$ is set to 6. Thus, the channel size and the temporal size will be 5.

After preprocessing, the data set is stratified based on diseases label, and then passed to the Spatial layer. The input here is of the size [10x10x11x5] (which is [batchsize x features x nodes x channel size]), the output of which is passed to the Temporal Layer which is of the size [10x10x11x5], where LeakyReLU is used as the threshold function (with a negative slope of 0.01). This is then passed through Global Pooling Layer where the output dimensions are [10x5x11x1] (which is [batch size x temporal size of features x nodes x new aggregated channel size]) and is then concatenated to shape [10x55].

The completely linked layer receives the output from here.

The first layer, which is a Linear layer, has an input of [10x55] and an output of [10x128]. This is normalised using the received output via a BatchNorm1D layer and then ReLU is used as the threshold function. Then a dropout layer is initialised, with a probability of 0.3, before passing it to the final Linear Layer which has an input of [10x128] and output of [10x100]. Then finally, ReLU is used as the threshold function so that whichever node has the highest value will be the predicted label, and hence the predicted disease.

The model was trained for 10 epochs, in which the validation accuracy obtained is 98.00%, and the test accuracy is 98.66% for $\gamma = 6$.

This accuracy indicates that there is a lot of merit in using temporal dependencies for disease prediction, especially by using the proposed method. This model can be also used practically in clinical sciences as a robust healthcare artificial intelligence system.

In Fig. 5, the validation accuracy and the test accuracy is plotted with respect to each epoch. It can be observed that there is an increasing trend and the model achieves high accuracy in a few epochs only.

E. Comparison with Baseline Models

A comparison is established with the given models in Table III, and their features are highlighted, namely the dataset they use, number of classes that were utilised and the accuracy (or the precision, recall, Area Under the Curve [AUC]) of their models.

The GNDP model [9] has been tested on the MIMIC-III dataset [4], so a direct comparison cannot be made, but the paper also makes use of Spatial Temporal Blocks to construct their architecture. In their network, they implement 5 STGCN units, pool the outputs at specific blocks, and then finally pass the output to the fully connected layer. Compared to that, the model only makes use of one STGCN unit before passing the output to the fully connected layer. It is also important to mention here that in GNDP, the input is processed differently from how the model processes the input, and hence the simplicity. GNDP model achieves a maximum accuracy of 86.29% on MIMIC.

VI. Conclusion

This paper proposes a novel deep learning framework STGCN for disease prediction, integrating graph convolution through Spatio-temporal convolutional blocks. GNDP solves the constraints of earlier techniques by using GNNs to learn spatial and temporal patterns from patients' sequential graph data, in which medical ontology knowledge and EMR information travel down distinct channels at different levels. The proposed model beats other state-of-the-art methods on datasets, demonstrating that it has a lot of potential in spatial-temporal structures.

These features are quite promising and practical for scholarly development. Moreover, the proposed framework can be applied to more general Spatio-temporal structured sequence prediction scenarios, such as evolving drug linkage, and preference prediction in diagnosis systems, etc.

VII. Future Work

In this paper, a template and a model is proposed that works well on that template, however the major challenge that was encountered was the absence of medical datasets in the structured format that is proposed. More work can be done to devise a model which converts EMR or EHR reports into a graph structure that [11] utilises and thus the temporal nature associated can be better exploited.

There is also a need of further structured data for disease prediction uses. The model can be better trained and would definitely give us more accuracy if the temporal as well as the sequential dependencies of the symptoms could be better utilised. For example, a fever is associated with a cold, which is associated with cough. These semantics and dependencies give the symptoms a structure which can be then utilised by the STGCN model directly.

Furthermore, methods could be devised to convert popular EMR databases such as [4] can be converted to the format that is proposed in order to establish better comparisons. This is challenging due to varied nature of each database to store information, and the diverse nature of writing reports.

References

[1] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE access*, vol. 7, pp. 81 542–81 554, 2019.

[2] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," *BMC medical informatics and decision making*, vol. 19, no. 1, pp. 1–16, 2019.

[3] Q. Yuan, J. Chen, C. Lu, and H. Huang, "The graph-based mutual attentive network for automatic diagnosis." in *IJCAI*, 2020, pp. 3393–3399.

[4] A. E. Johnson, T. J. Pollard, L. Shen, H. L. Li-Wei, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "Mimic-iii, a freely accessible critical care database," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[5] C. Castro, E. Vargas, A. Sánchez, E. Gutierrez Lopez, and D.-L. Flores, *Parkinson's Disease Classification Using Artificial Neural Networks*, 01 2020, pp. 1060–1065.

[6] T. Vaiyapuri, P. Balaji, S. S, H. Alaskar, and Z. Sbai, "Computational intelligence-based melanoma detection and classification using dermoscopic images," *Computational Intelligence and Neuroscience*, vol. 2022, p. 2370190, May 2022. [Online]. Available: https://doi.org/10.1155/2022/2370190

[7] A. Kazi, S. Shekarforoush, S. A. Krishna, H. Burwinkel, G. Vivar, K. Kortüm, S.-A. Ahmadi, S. Albarqouni, and N. Navab, "Inceptiongcn: receptive field aware graph convolutional network for disease prediction," in *International Conference on Information Processing in Medical Imaging*. Springer, 2019, pp. 73–85.

[8] D. Ahmedt-Aristizabal, A. Armin, S. Denman, C. Fookes, and L. Petersson, "Graph-based deep learning for medical diagnosis and analysis: Past, present and future," *Sensors*, vol. 21, p. 4758, 07 2021.

[9] Y. Li, B. Qian, X. Zhang, and H. Liu, "Knowledge guided diagnosis prediction via graph spatial-temporal network," in *Proceedings of the 2020 SIAM International Conference on Data Mining*. SIAM, 2020, pp. 19–27.

[10] T. Vaiyapuri, Liyakathunisa, H. Alaskar, E. Aljohani, S. Shridevi, and A. Hussain, "Red fox optimizer with data-science-enabled microarray gene expression classification model," *Applied Sciences*, vol. 12, no. 9, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/9/4172

[11] M. Rotmensch, Y. Halpern, A. Tlimat, S. Horng, and D. Sontag, "Learning a health knowledge graph from electronic medical records," *Scientific reports*, vol. 7, no. 1, pp. 1–11, 2017.

[12] X. Li, Y. Wang, D. Wang, W. Yuan, D. Peng, and Q. Mei, "Improving rare disease classification using imperfect knowledge graph," *BMC medical informatics and decision making*, vol. 19, no. 5, pp. 1–10, 2019.

[13] Z. Sun, H. Yin, H. Chen, T. Chen, L. Cui, and F. Yang, "Disease prediction via graph neural networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 3, pp. 818–826, 2020.

[14] A. Giannoula, A. Gutierrez-Sacristán, Á. Bravo, F. Sanz, and L. I. Furlong, "Identifying temporal patterns in patient disease trajectories using dynamic time warping: A population-based study," *Scientific reports*, vol. 8, no. 1, pp. 1–14, 2018.

[15] A. B. Jensen, P. Moseley, T. Oprea, S. Ellesøe, R. Eriksson, H. Schmock, P. Jensen, L. Jensen, and S. Brunak, "Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients," *Nature communications*, vol. 5, p. 4022, 06 2014.

[16] T. Ploner, S. Heß, M. Grum, P. Drewe-Boss, and J. Walker, "Using gradient boosting with stability selection on health insurance claims data to identify disease trajectories in chronic obstructive pulmonary disease," *Statistical Methods in Medical Research*, vol. 29, 07 2020.

[17] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[18] Y. Li, D. Tarlow, M. Brockschmidt, and R. S. Zemel, "Gated graph sequence neural networks," *CoRR*, vol. abs/1511.05493, 2016.

[19] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *IJCAI*, 07 2018, pp. 3634–3640.

[20] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[21] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 922–929, 07 2019.

[22] R. V. Marinescu, N. P. Oxtoby, A. L. Young, Bron *et al.*, "Tadpole challenge: Accurate alzheimer's disease prediction through crowdsourced forecasting of future data," in *International Workshop on PRedictive Intelligence In MEdicine*. Springer, 2019, pp. 1–10.

[23] Z. Sun, W. Dong, J. Shi, and Z. Huang, "Interpretable disease prediction based on reinforcement path reasoning over knowledge graphs," *arXiv preprint arXiv:2010.08300*, 10 2020.

[24] P. Li, C. Xie, T. Pollard, Johnson *et al.*, "Promoting secondary analysis of electronic medical records in china: summary of the plagh-mit critical data conference and health datathon," *JMIR medical informatics*, vol. 5, no. 4, p. e43, 2017.

[25] L. Cosmo, A. Kazi, S.-A. Ahmadi, N. Navab, and M. Bronstein, "Latent patient network learning for automatic diagnosis," *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*.

[26] X. Tao, T. Pham, J. Zhang, J. Yong, W. P. Goh, W. Zhang, and Y. Cai, "Mining health knowledge graph for health risk prediction," *World Wide Web*, vol. 23, no. 4, pp. 2341–2362, 2020.

[27] S. Kranz, L. J. Mahood, and D. A. Wagstaff, "Diagnostic criteria patterns of us children with metabolic syndrome: Nhanes 1999–2002," *Nutrition Journal*, vol. 6, no. 1, pp. 1–9, 2007.