Enhancing Progressive Diagnosis Prediction in Healthcare with Continuous Normalizing Flows

Yanchao Tan
College of Computer and Data Science, Fuzhou University, China
yctan@fzu.edu.cn

Hengyu Zhang
College of Computer Science, Zhejiang University, China
zhyqaq@zju.edu.cn

Zihao Zhou
College of Computer and Data Science, Fuzhou University, China
reviverkey@gmail.com

Guofang Ma
School of Computer Science, Zhejiang Gongshang University, China
maguofang@zjgsu.edu.cn

Fan Wang
College of Computer Science, Zhejiang University, China
fanwang97@zju.edu.cn

Weiming Liu
College of Computer Science, Zhejiang University, China
21831010@zju.edu.cn

Xinting Liao
College of Computer Science, Zhejiang University, China
12021080@zju.edu.cn

Vicki S Hertzberg
Nell Hodgson Woodruff School of Nursing, Emory University, USA
vhertzb@emory.edu

Carl Yang*
Department of Computer Science, Emory University, USA
j.carlyang@emory.edu

ABSTRACT

Progressive diagnosis prediction in healthcare is a promising yet challenging task. Existing studies usually assume a pre-defined prior for generating patient distributions (e.g., Gaussian). However, the inferred approximate posterior can deviate from the real-world distribution, which further affects the modeling of continuous disease progression over time. To alleviate such inference bias, we propose an enhanced progressive diagnostic prediction model (i.e., `ProCNF`), which integrates continuous normalizing flows (CNF) and neural ordinary differential equations (ODEs) to achieve more accurate approximations of patient health trajectories while capturing the continuity underlying disease progression. We first learn patient embeddings with CNF to construct a complex posterior approximation of patient distributions. Then, we devise a CNF-enhanced neural ODE module for progressive diagnostic prediction, which aims to improve the modeling of disease progression for individual patients. Extensive experiments on two real-world longitudinal EHR datasets show significant performance gains brought by our method over state-of-the-art competitors.

KEYWORDS

Progressive diagnosis prediction, Continuous normalizing flows, Neural ordinary differential equations

ACM Reference Format:
Yanchao Tan, Hengyu Zhang, Zihao Zhou, Guofang Ma, Fan Wang, Weiming Liu, Xinting Liao, Vicki S Hertzberg, and Carl Yang. 2024. Enhancing Progressive Diagnosis Prediction in Healthcare with Continuous Normalizing Flows. In *Companion Proceedings of the ACM Web Conference 2024 (WWW '24 Companion), May 13–17, 2024, Singapore, Singapore.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3589335.3651457

*Carl Yang is the corresponding author.

WWW '24 Companion, May 13–17, 2024, Singapore, Singapore

ACM ISBN 979-8-4007-0172-6/24/05…$15.00
https://doi.org/10.1145/3589335.3651457

1 INTRODUCTION

Diagnosis prediction, utilizing electronic health records (EHRs), has become a pivotal area within AI-driven healthcare applications. Many studies leverage deep learning techniques [1, 9, 11] to model dynamic hospital visits, thereby facilitating clinical decision support. However, a challenge arises as most disease progressions and changes in patient health are inherently continuous, while clinical records are discrete due to irregular patient visits. To alleviate the challenges posed by irregular timestamps, some studies have adopted neural ordinary differential equations (ODEs) [15, 16], offering a more comprehensive understanding of continuous-time disease progression and enhancing diagnostic prediction.

However, existing studies usually assume a pre-defined prior for the patient, e.g., Gaussian, which might result in the inferred approximate posterior greatly deviating from the real-world distribution. Such biased inference gaps can affect the prediction of individual disease progression, and thus limit the performance of progressive diagnostic prediction models. Furthermore, without additional supervision, it remains unknown how to close the gap between approximate posterior and real posterior solely based on the patient's historical visits [20].

To this end, we propose a novel `ProCNF` framework, aimed at enhancing progressive diagnosis prediction in healthcare with continuous normalizing flows (CNF). By novelly integrating CNF with neural ODEs, we can effectively capture a more accurate posterior approximation of the patient distributions as well as the continuity underlying disease progression. Specifically, we first leverage CNF to transform a patient from a simple base distribution (e.g., Gaussian) into a more complex patient-specific distribution via a series of invertible mappings when the base distribution is reparametrizable. Then, based on the complex patient distributions, we propose a CNF-enhanced neural ODEs module for diagnosis prediction, where

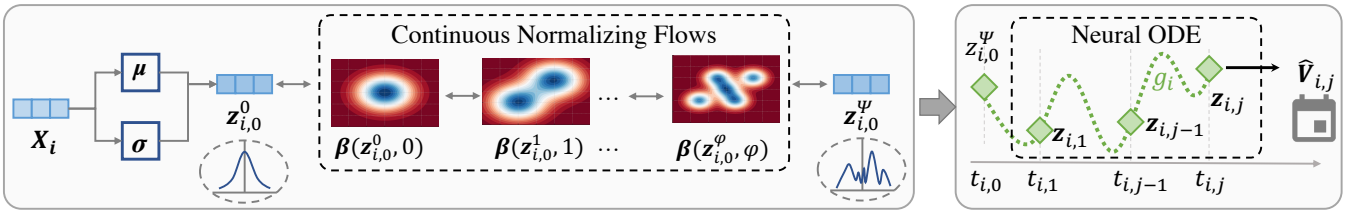

Figure 1: The overall framework of ProCNF.

we can continuously model disease progression under the accurate expressive patient embedding and discrete patient irregular visits. Finally, we evaluate the proposed ProCNF framework with extensive experiments on real benchmark electronic healthcare datasets for progressive diagnosis prediction tasks. Extensive experimental results show the great potential for our ProCNF over the nine representative state-of-the-art baselines.

2 RELATED WORK

Utilizing EHR data, deep learning methods have significantly advanced predictive modeling in diagnosis. For example, RETAIN [4] employed attention-based recurrent neural networks (RNNs), Dipole [12] integrated bidirectional long-short-term memory networks with attention, and Timeline [1] proposed time-aware attention in RNNs, demonstrated this progression. Chet [9] introduced dynamic graph learning for disease combinations, while GRAM [5], CGL [10], MedPath [19], and HealGCN [18] enhanced disease relation modeling using medical knowledge graphs.

Despite these advancements, irregular visit times and heterogeneous disease effects remain challenges. HiTANet [11] and Concare [14] alleviated these issues by incorporating time embeddings for temporal dependencies. As closest to us, Qian et. al [15] proposed LHM to describe continuous disease progression dynamics via neural ODEs, but they failed to capture complex distributions behind patients in EHR data.

3 THE PROCNF FRAMEWORK

3.1 Patient Embedding with CNF

Traditional approaches [7, 15] often rely on simple Gaussian assumptions for posterior distribution approximation, which are ill-equipped to model patient trajectories with underlying complex distributions and may lead to limited model performance. Addressing these constraints is crucial for developing more personalized and accurate healthcare predictive models.

To this end, we propose to learn patient embeddings based on CNF, which is promising in reversibly transforming a base distribution (e.g., Gaussian) to the complicated patient-specific patient distribution for accurate patient representations. In this way, we can reduce the gap between approximate posterior and true posterior by employing such richer posterior/prior distributions [2]. Specifically, the learning of CNF is initiated with a known probability distribution variable $z^0_{i,0}$, such as a Gaussian, and needs to apply a differential function β that is uniformly Lipschitz continuous in both $z_{i,0}$ and step ψ.

Firstly, we apply the widely adopted reparameterization trick [15] for the formulation of $z_{i,0}$ involves, where we utilize MLP as an encoder to sample a latent initial state $z_{i,0}$ for each patient as follows:

$$\begin{aligned}[\boldsymbol{\mu_i}, \boldsymbol{\sigma_i}] &= \text{Encoder}(\boldsymbol{X}_i; h_\theta),\\ \boldsymbol{z}_{i,0} &= \boldsymbol{\mu_i} + \epsilon \odot \boldsymbol{\sigma_i}, \quad \epsilon \sim \mathcal{N}(0, \boldsymbol{I}).\end{aligned} \tag{1}$$

Here, $\boldsymbol{X}_i$ represents the aggregated patient's historical visit embeddings processed through a self-attention mechanism [13, 17] (i.e., $\boldsymbol{X}_i = \text{Self-Att}(\boldsymbol{E}_i)$), with the visit embedding of patient i (i.e., $\boldsymbol{E}_i = (\boldsymbol{E}_{i,0}, \boldsymbol{E}_{i,1}, \ldots, \boldsymbol{E}_{i,j-1})$). The calculation of Self-Att with three learnable projection matrices (i.e., $\mathbf{W}^Q$, $\mathbf{W}^K$, and $\mathbf{W}^V$) is written as

$$\text{Self-Att}(\boldsymbol{E}_i) = \text{softmax}\left((\boldsymbol{E}_i\mathbf{W}^Q)(\boldsymbol{E}_i\mathbf{W}^K)^T/\sqrt{D}\right)\boldsymbol{E}_i\mathbf{W}^V, \tag{2}$$

where the D is the embedding dimension.

With the initial state $z_{i,0}$, we then apply a differential function β, which enables the transformation of the initial Gaussian distribution into a more complex patient-specific distribution. In this way, we can simplify the computation of the change in $z_{i,0}$ and its log densities to transform $q(z_{i,0}|\boldsymbol{X}_i)$ in a continuous way as follows:

$$dz^\psi_{i,0}/d\psi = \beta(z^\psi_{i,0}, \psi), \tag{3}$$

which describes a continuous transformation of $z^\psi_{i,0}$. With the theorem of instantaneous change of variables [3], the change in log densities $\log q(z^\psi_{i,0}|\boldsymbol{X}_i)$ also follows a differential equation:

$$d\log q(z^\psi_{i,0}|\boldsymbol{X}_i)/d\psi = -\,\text{Tr}\left(\partial\beta(z^\psi_{i,0}, \psi)/\partial z^\psi_{i,0}\right), \tag{4}$$

where Tr denotes the trace operation and can replace the intensive determinant computation in normalizing flows [21].

Then, the latent variables $z^\Psi_{i,0}$ after step Ψ can be computed as:

$$z^\Psi_{i,0} = z^0_{i,0} + \int_0^\Psi \beta\left(z^\psi_{i,0}, \psi\right) d\psi, \tag{5}$$

and its log densities are formulated as follows:

$$\log q_\phi\left(z^\Psi_{i,0}|\boldsymbol{X}_i\right) = \log q_\phi\left(z^0_{i,0}|\boldsymbol{X}_i\right) - \int_0^\Psi \text{Tr}\left(\partial\beta/\partial z^\psi_{i,0}\right) d\psi, \tag{6}$$

where Ψ can be arbitrarily set for more transformations and we empirically set Ψ as 1 which is consistent with [20].

Finally, with the approximated posterior distribution q_ϕ, we have the evidence lower-bound (ELBO) objective based on CNF, where the formula is given as follows:

$$\begin{aligned}\mathcal{L}_E &= \mathbb{E}_{q_\phi}[p(\hat{\boldsymbol{E}}_{i,j}|z^\Psi_{i,0}) - \log\frac{q_\phi\left(z^0_{i,0}|\boldsymbol{X}_i\right)}{p\left(z^0_{i,0}\right)}] = p(\hat{\boldsymbol{E}}_{i,j}|z^\Psi_{i,0})\\ &+ \mathbb{E}_{q_\phi}[\log p\left(z^\Psi_{i,0}\right) - q_\phi\left(z^0_{i,0}|\boldsymbol{X}_i\right) + \int_0^\Psi \text{Tr}\left(\frac{\partial\beta(z^\psi_{i,0}, \psi)}{\partial z^\psi_{i,0}}\right) d\psi].\end{aligned} \tag{7}$$

Table 1: Experimental results on two benchmark EHR datasets. The best performances are highlighted in boldface and the second runners are underlined. ProCNF achieves the best performance on both datasets, where * denotes significant improvements based on the Wilcoxon signed-rank test.

Method	Recall@5	NDCG@5	Recall@10	NDCG@10	Recall@5	NDCG@5	Recall@10	NDCG@10
	MIMIC-III				NELL			
RATAIN	0.1510±0.18%	0.4188±0.16%	0.2134±0.13%	0.3537±0.12%	0.6272±0.21%	0.5974±0.16%	0.7535±0.16%	0.6227±0.13%
Dipole	0.1442±0.24%	0.3999±0.18%	0.2038±0.28%	0.3378±0.18%	0.5989±0.21%	0.5705±0.18%	0.7195±0.17%	0.5946±0.15%
GRAM	0.1429±0.13%	0.4059±0.10%	0.2112±0.14%	0.3510±0.12%	0.6394±0.15%	0.6118±0.12%	0.7277±0.16%	0.6325±0.13%
Timeline	0.1487±0.15%	0.4123±0.13%	0.2100±0.12%	0.3482±0.10%	0.6174±0.15%	0.5881±0.15%	0.7417±0.16%	0.6129±0.13%
KAME	0.1353±0.14%	0.3992±0.13%	0.2055±0.13%	0.3070±0.11%	0.5620±0.12%	0.5353±0.10%	0.6751±0.15%	0.5579±0.13%
HiTANet	0.1502±0.21%	0.4166±0.17%	0.2122±0.18%	0.3518±0.16%	0.6446±0.18%	0.6186±0.15%	0.7701±0.15%	0.6502±0.12%
CGL	0.1538±0.22%	0.4265±0.19%	0.2173±0.26%	0.3602±0.21%	0.6387±0.18%	0.6084±0.15%	0.7673±0.13%	0.6341±0.10%
LHM	0.1768±0.18%	0.4489±0.16%	0.2437±0.20%	0.3902±0.19%	0.6570±0.15%	0.6237±0.17%	0.7752±0.18%	0.6521±0.19%
Chet	0.1636±0.13%	0.4403±0.08%	0.2312±0.12%	0.3719±0.10%	0.6182±0.15%	0.5913±0.12%	0.7381±0.13%	0.6181±0.13%
ProCNF	**0.1869±0.17%***	**0.4587±0.15%***	**0.2665±0.19%***	**0.4071±0.17%***	**0.6910±0.18%***	**0.6602±0.19%***	**0.7952±0.17%***	**0.6832±0.17%***

3.2 Progressive Diagnosis Prediction

With the learned sophisticated patient embeddings (i.e., $z^{\Psi}_{i,0}$), we further design CNF-enhanced neural ODEs for progressive diagnosis prediction, so as to model continuous patient health status over time. In this way, the patient status representation $z_{i,j}$ at time $T_{i,j}$ can be inferred by solvin the neural ODE as follows:

$$[z_{i,1}, \ldots, z_{i,j}] = \text{ODESolve}\left(g_i, z^{\Psi}_{i,0}, [T_{i,1}, \ldots, T_{i,j}]\right), \quad (8)$$

where g_i is a learnable neural network for each patient u_i and we adopt the fourth-order Runge–Kutta for an accurate and efficient approximation [3, 8].

Since personalized diagnosis prediction is a multi-label classification task, we use a dense layer with a softmax function to calculate the predicted probability. Specifically, we fully leverage the historical visits to serve as supervision. The predicted visits $\hat{V}_i$ is based on the inferred patient status z_i and the predictive objective function $\mathcal{L}_P$ are listed as follows:

$$p([\hat{V}_{i,1}, \ldots, \hat{V}_{i,j}]|X_i, z^{\Psi}_{i,0}) = \text{softmax}(\text{MLP}([z_{i,1}, \ldots, z_{i,j}])), \quad (9)$$

$$\mathcal{L}_P = -\sum_{i=1}^{N}\sum_{j=1}^{M_i} V_{i,j} \log\left(\hat{V}_{i,j}\right) + \left(1 - V_{i,j}\right) \log\left(1 - \hat{V}_{i,j}\right). \quad (10)$$

where $V_{i,j}$ is the ground-truth of patient u_i's (j)-th diagnosis.

Finally, we have the overall objective function of our ProCNF:

$$\min \mathcal{L} = \mathcal{L}_P - \lambda \mathcal{L}_E, \quad (11)$$

where $\mathcal{L}_P$ is calculated in Eq. 10 and the ELBO loss $\mathcal{L}_E$ is calculated in Eq. 7. λ is a hyperparameter to adjust the weight of $\mathcal{L}_E$.

4 EXPERIMENT

4.1 Datasets

We use two real-world EHR datasets to verify the effectiveness of compared methods, i.e., MIMIC-III [6] and NELL. NELL is a large-scale real-world clinical data collected by the Nell Hodgson Woodruff School of Nursing at Emory University. Both datasets are fully anonymized and carefully sanitized before our access. We split dataset randomly according to patients into training/validation/test sets (i.e., 2100/61/210 on MIMIC-III and 3125/391/391 on NELL), which is consistent with [9, 10]. We chose patients who made at least three visits for both datasets. We use Recall@k and NDCG@k metrics that are consistent with [9, 10, 18].

We adopt 9 representative state-of-the-art methods as baselines for the performance comparison with our ProCNF including RETAIN [4], Dipole [12], GRAM [5], Timeline [1], KAME [13], HiTANet [11], CGL [10], LHM [15], and Chet [9].

4.2 Experimental Results

We present the Recall and NDCG metrics values achieved by our proposed ProCNF and other nine baselines in Table 1. Overall, the ProCNF achieves the best performance on the MIMIC-III and NELL datasets, which constantly achieves an average of 4.99% improvement over the state-of-the-art baselines regarding both Recall and NDCG. These results affirm the effectiveness of ProCNF to model disease progression under complex patient distributions with discrete patient irregular visits.

Compared with the second-best model (i.e., LHM), the performance gains of ProCNF ranges from 2.18% with NDCG@5 to on MIMIC-III to 9.36% achieved with Recall@10 on MIMIC-III. Although LHM adopts neural ODEs for continuous disease progression, it fails to capture complex distributions behind patients. While other models like HiTANet and Chet show fluctuating ranks between datasets, ProCNF consistently demonstrates its robustness across datasets. This underlines the advantage of the CNF-enhanced neural ODEs module in ProCNF to accurately learn patient embeddings for progressive diagnosis prediction in healthcare.

4.3 Case Studies

To provide more insights into the advantages of ProCNF in modeling continuous disease progression, we provide the predictive diagnoses of one example patient (i.e., Jack) in Table 2. The analysis of the disease progression for the diabetic patient Jack in Table 2 showcases the strength of ProCNF in capturing continuous disease progression based on enhanced patient distributions. For instance, ProCNF successfully predicts the recurrence of "Atrial fibrillation" and "Other and unspecified hyperlipidemia" for visits $k+2$ and $k+1$ respectively, despite these diagnoses not being recorded in the preceding visits. This reflects our method's capability to understand and predict the ongoing nature of chronic conditions, ensuring a comprehensive and consistent monitoring of Jack's health status. It highlights how an enhanced patient distribution can lead to better health management by predicting potential complications.

Table 2: Predictive diagnoses for diabetic patient Jack (pseudonym) from the NELL dataset. Here "FN" in Red color refers to the diagnoses that are in the ground-truth diagnosis sets but are not predicted, while "FP" in Blue color denotes the diagnoses predicted but are not in ground-truth diagnosis sets (Best viewed in color).

	Ground-truth diagnoses	Predictive diagnoses
Visit k	250.00 Diabetes mellitus without mention of complication 401.9 Unspecified essential hypertension 272.4 Other and unspecified hyperlipidemia 300.00 Anxiety state unspecified	250.00 Diabetes mellitus without mention of complication 401.9 Unspecified essential hypertension 272.4 Other and unspecified hyperlipidemia 424.1 Aortic valve disorders 427.31 Atrial fibrillation
Visit k+1	250.00 Diabetes mellitus without mention of complication	250.00 Diabetes mellitus without mention of complication 427.31 Atrial fibrillation 401.9 Unspecified essential hypertension 272.4 Other and unspecified hyperlipidemia 414.01 Coronary atherosclerosis of native coronary artery
Visit k+2	250.00 Diabetes mellitus without mention of complication 401.9 Unspecified essential hypertension 272.4 Other and unspecified hyperlipidemia 300.00 Anxiety state unspecified 786.09 Other respiratory abnormalities 311 Depressive disorder, not elsewhere	250.00 Diabetes mellitus without mention of complication 401.9 Unspecified essential hypertension 272.4 Other and unspecified hyperlipidemia 300.00 Anxiety state unspecified 427.31 Atrial fibrillation
Last Visit	**427.31 Atrial fibrillation**	**427.31 Atrial fibrillation** 401.9 Unspecified essential hypertension 414.01 Coronary atherosclerosis of native coronary artery 250.00 Diabetes mellitus without mention of complication 272.4 Other and unspecified hyperlipidemia

5 CONCLUSION

In this paper, we propose to make progressive diagnosis predictions for patient visits with irregular intervals in healthcare. Specifically, we propose a novel progressive diagnostic prediction model (ProCNF) with two pivotal techniques, which jointly perform complex patient distributions and dynamic disease progression to achieve a more accurate approximation of patient health trajectories. Extensive quantitative experiments demonstrate the clear advantages of our ProCNF, which is consolidated with our real case study results.

6 ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China under Grants (No.6230071268), the Natural Science Foundation of Zhejiang Province (LQ23F020007), Zhejiang Gongshang University "Digital+" Disciplinary Construction Management Project (SZJ2022B001). Carl Yang was not supported by any fund from China.

REFERENCES

[1] T. Bai, S. Zhang, B. L. Egleston, and S. Vucetic. Interpretable representation learning for healthcare via capturing disease progression through time. In *SIGKDD*, pages 43–51, 2018.
[2] J. Bose, A. Smofsky, R. Liao, P. Panangaden, and W. Hamilton. Latent variable modelling with hyperbolic normalizing flows. In *International Conference on Machine Learning*, pages 1045–1055. PMLR, 2020.
[3] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud. Neural ordinary differential equations. In *NeurIPS*, 2018.
[4] E. Choi, M. T. Bahadori, J. A. Kulas, A. Schuetz, W. F. Stewart, and J. Sun. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. In *NeurIPS*, 2016.
[5] E. Choi, M. T. Bahadori, L. Song, W. F. Stewart, and J. Sun. Gram: graph-based attention model for healthcare representation learning. In *SIGKDD*, pages 787–795, 2017.
[6] 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*, 2016.
[7] D. P. Kingma, T. Salimans, and M. Welling. Variational dropout and the local reparameterization trick. *NeurIPS*, 2015.
[8] Y. Li, H. Yi, C. Bender, S. Shan, and J. B. Oliva. Exchangeable neural ode for set modeling. *NeurIPS*, 33:6936–6946, 2020.
[9] C. Lu, T. Han, and Y. Ning. Context-aware health event prediction via transition functions on dynamic disease graphs. In *AAAI*, volume 36, pages 4567–4574, 2022.
[10] C. Lu, C. K. Reddy, P. Chakraborty, S. Kleinberg, and Y. Ning. Collaborative graph learning with auxiliary text for temporal event prediction in healthcare. In *IJCAI*, 2021.
[11] J. Luo, M. Ye, C. Xiao, and F. Ma. Hitanet: Hierarchical time-aware attention networks for risk prediction on electronic health records. In *SIGKDD*, pages 647–656, 2020.
[12] F. Ma, R. Chitta, J. Zhou, Q. You, T. Sun, and J. Gao. Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In *SIGKDD*, pages 1903–1911, 2017.
[13] F. Ma, Q. You, H. Xiao, R. Chitta, J. Zhou, and J. Gao. Kame: Knowledge-based attention model for diagnosis prediction in healthcare. In *CIKM*, 2018.
[14] L. Ma, C. Zhang, Y. Wang, W. Ruan, J. Wang, W. Tang, X. Ma, X. Gao, and J. Gao. Concare: Personalized clinical feature embedding via capturing the healthcare context. In *AAAI*, 2020.
[15] Z. Qian, W. R. Zame, M. van der Schaar, L. M. Fleuren, and P. Elbers. Integrating expert odes into neural odes: Pharmacology and disease progression. In *NeurIPS*, 2021.
[16] Y. Tan, Z. Zhou, L. Yu, W. Liu, C. Chen, G. Ma, X. Hu, V. S. Hertzberg, and C. Yang. Enhancing personalized healthcare via capturing disease severity, interaction, and progression. *ICDM*, 2023.
[17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
[18] Z. Wang, R. Wen, X. Chen, S. Cao, S.-L. Huang, B. Qian, and Y. Zheng. Online disease diagnosis with inductive heterogeneous graph convolutional networks. In *WWW*, pages 3349–3358, 2021.
[19] M. Ye, S. Cui, Y. Wang, J. Luo, C. Xiao, and F. Ma. Medpath: Augmenting health risk prediction via medical knowledge paths. In *WWW*, 2021.
[20] F. Zhou, L. Li, K. Zhang, G. Trajcevski, F. Yao, Y. Huang, T. Zhong, J. Wang, and Q. Liu. Forecasting the evolution of hydropower generation. In *SIGKDD*, pages 2861–2870, 2020.
[21] F. Zhou, Z. Wen, K. Zhang, G. Trajcevski, and T. Zhong. Variational session-based recommendation using normalizing flows. In *WWW*, 2019.