Enhancing Progressive Diagnosis Prediction in Healthcare with Continuous Normalizing Flows

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

Hengyu Zhang
College of Computer Science, Zhejiang University, China
zhyaq@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:

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 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 natively 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
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


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 $z_{i,0}$, such as a Gaussian, and needs to apply a differential function $\beta$ that is uniformly Lipschitz continuous in both $z_{i,0}$ and step $\psi$.

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:

$$\begin{align*}
\mu_i, \sigma_i & = \text{Encoder}(X_i; h_0), \\
z_{i,0} & = \mu_i + \epsilon \odot \sigma_i, \quad \epsilon \sim \mathcal{N}(0, I).
\end{align*}$$

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

$$\text{Self-Att}(E_i) = \text{softmax} \left( \frac{E_i W^Q (E_i W^K)'}{\sqrt{D}} \right) E_i W^V,$$

where the $D$ is the embedding dimension.

With the initial state $z_{i,0}$, we then apply a differential function $\beta$, 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 $q(z_{i,0}|X_i)$ in a continuous way as follows:

$$dz_{i,0}^\psi/d\psi = \beta(z_{i,0}^\psi, \psi),$$

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

$$d \log q(z_{i,0}^\psi|X_i)/d\psi = -\text{Tr} \left( \frac{\partial \beta(z_{i,0}^\psi, \psi)}{\partial z_{i,0}^\psi} \right),$$

where $\text{Tr}$ denotes the trace operation and can replace the intensive determinant computation in normalizing flows [21].

Then, the latent variables $z_{i,0}^\psi$ after step $\Psi$ can be computed as:

$$z_{i,0}^\psi = z_{i,0}^0 + \int_0^\Psi \beta(z_{i,0}^0, \psi) d\psi,$$

and its log densities are formulated as follows:

$$\log q_\psi \left( z_{i,0}^\psi | X_i \right) = \log q_\phi \left( z_{i,0}^0 | X_i \right) - \int_0^\Psi \text{Tr} \left( \frac{\partial \beta(z_{i,0}^\psi, \psi)}{\partial z_{i,0}^\psi} \right) d\psi.$$

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

$$\mathcal{L}_E = \mathbb{E}_{q_\phi} \left[ p(\hat{E}_{i,j}|z_{i,0}^\psi) - \log \frac{q_\phi (z_{i,0}^\psi|X_i)}{p (z_{i,0}^0|X_i)} \right] = \mathbb{E}_{q_\phi} \left[ \mathcal{L}_{\text{ELBO}}(z_{i,0}^\psi) \right] + \mathbb{E}_{q_\phi} \left[ \log p(z_{i,0}^\psi) - q_\phi (z_{i,0}^0|X_i) \right] + \int_0^\Psi \text{Tr} \left( \frac{\partial \beta(z_{i,0}^\psi, \psi)}{\partial z_{i,0}^\psi} \right) d\psi.$$

Figure 1: The overall framework of ProCNF.
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.

<table>
<thead>
<tr>
<th>Method</th>
<th>Recall@5</th>
<th>NDCG@5</th>
<th>Recall@10</th>
<th>NDCG@10</th>
<th>Recall@5</th>
<th>NDCG@5</th>
<th>Recall@10</th>
<th>NDCG@10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MIMIC-III</td>
<td></td>
<td>NELL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RATAIN</td>
<td>0.1510±0.18%</td>
<td>0.4188±0.16%</td>
<td>0.2134±0.13%</td>
<td>0.3537±0.12%</td>
<td>0.6272±0.21%</td>
<td>0.5974±0.16%</td>
<td>0.7535±0.16%</td>
<td>0.6227±0.13%</td>
</tr>
<tr>
<td>Dipole</td>
<td>0.1442±0.24%</td>
<td>0.3999±0.19%</td>
<td>0.2038±0.28%</td>
<td>0.3378±0.18%</td>
<td>0.5989±0.21%</td>
<td>0.5705±0.18%</td>
<td>0.7195±0.17%</td>
<td>0.5946±0.15%</td>
</tr>
<tr>
<td>GRAM</td>
<td>0.1429±0.13%</td>
<td>0.4059±0.10%</td>
<td>0.2112±0.14%</td>
<td>0.3510±0.12%</td>
<td>0.6394±0.15%</td>
<td>0.6118±0.12%</td>
<td>0.7277±0.16%</td>
<td>0.6325±0.13%</td>
</tr>
<tr>
<td>Timeline</td>
<td>0.1487±0.15%</td>
<td>0.4123±0.13%</td>
<td>0.2100±0.12%</td>
<td>0.3482±0.10%</td>
<td>0.6174±0.15%</td>
<td>0.5881±0.15%</td>
<td>0.7417±0.16%</td>
<td>0.6129±0.13%</td>
</tr>
<tr>
<td>KAME</td>
<td>0.1353±0.14%</td>
<td>0.3992±0.13%</td>
<td>0.2055±0.13%</td>
<td>0.3070±0.11%</td>
<td>0.5620±0.12%</td>
<td>0.5353±0.10%</td>
<td>0.6751±0.15%</td>
<td>0.5579±0.13%</td>
</tr>
<tr>
<td>HiTANet</td>
<td>0.1502±0.21%</td>
<td>0.4166±0.17%</td>
<td>0.2122±0.18%</td>
<td>0.3518±0.16%</td>
<td>0.6446±0.18%</td>
<td>0.6182±0.15%</td>
<td>0.7701±0.15%</td>
<td>0.6502±0.12%</td>
</tr>
<tr>
<td>CGL</td>
<td>0.1538±0.22%</td>
<td>0.4265±0.19%</td>
<td>0.2173±0.26%</td>
<td>0.3602±0.21%</td>
<td>0.6387±0.18%</td>
<td>0.6084±0.15%</td>
<td>0.7673±0.13%</td>
<td>0.6341±0.10%</td>
</tr>
<tr>
<td>LHM</td>
<td>0.1768±0.18%</td>
<td>0.4489±0.16%</td>
<td>0.2437±0.20%</td>
<td>0.3902±0.19%</td>
<td>0.6570±0.15%</td>
<td>0.6237±0.17%</td>
<td>0.7752±0.18%</td>
<td>0.6521±0.19%</td>
</tr>
<tr>
<td>Chet</td>
<td>0.1636±0.13%</td>
<td>0.4403±0.08%</td>
<td>0.2312±0.12%</td>
<td>0.3719±0.10%</td>
<td>0.6182±0.15%</td>
<td>0.5913±0.12%</td>
<td>0.7381±0.13%</td>
<td>0.6811±0.13%</td>
</tr>
<tr>
<td>ProCNF</td>
<td>0.1869±0.17%</td>
<td>0.4587±0.15%</td>
<td>0.2665±0.19%</td>
<td>0.4071±0.17%</td>
<td>0.6910±0.18%</td>
<td>0.6602±0.19%</td>
<td>0.7952±0.17%</td>
<td>0.6832±0.17%</td>
</tr>
</tbody>
</table>

4.2 Experimental Results

We present the Recall and NDCG metrics values achieved by our 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).

<table>
<thead>
<tr>
<th>Visit $k$</th>
<th>Ground-truth diagnoses</th>
<th>Predictive diagnoses</th>
</tr>
</thead>
<tbody>
<tr>
<td>$250.00$ Diabetes mellitus without mention of complication</td>
<td>$250.00$ Diabetes mellitus without mention of complication</td>
<td></td>
</tr>
<tr>
<td>$401.9$ Unspecified essential hypertension</td>
<td>$401.9$ Unspecified essential hypertension</td>
<td></td>
</tr>
<tr>
<td>$272.4$ Other and unspecified hyperlipidemia</td>
<td>$272.4$ Other and unspecified hyperlipidemia</td>
<td></td>
</tr>
<tr>
<td>$300.00$ Anxiety state unspecified</td>
<td>$424.1$ Aortic valve disorders</td>
<td></td>
</tr>
<tr>
<td>$786.09$ Other respiratory abnormalities</td>
<td>$427.31$ Atrial fibrillation</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Visit $k+1$</th>
<th>Ground-truth diagnoses</th>
<th>Predictive diagnoses</th>
</tr>
</thead>
<tbody>
<tr>
<td>$250.00$ Diabetes mellitus without mention of complication</td>
<td>$250.00$ Diabetes mellitus without mention of complication</td>
<td></td>
</tr>
<tr>
<td>$401.9$ Unspecified essential hypertension</td>
<td>$401.9$ Unspecified essential hypertension</td>
<td></td>
</tr>
<tr>
<td>$272.4$ Other and unspecified hyperlipidemia</td>
<td>$272.4$ Other and unspecified hyperlipidemia</td>
<td></td>
</tr>
<tr>
<td>$300.00$ Anxiety state unspecified</td>
<td>$300.00$ Anxiety state unspecified</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Visit $k+2$</th>
<th>Ground-truth diagnoses</th>
<th>Predictive diagnoses</th>
</tr>
</thead>
<tbody>
<tr>
<td>$250.00$ Diabetes mellitus without mention of complication</td>
<td>$250.00$ Diabetes mellitus without mention of complication</td>
<td></td>
</tr>
<tr>
<td>$401.9$ Unspecified essential hypertension</td>
<td>$401.9$ Unspecified essential hypertension</td>
<td></td>
</tr>
<tr>
<td>$272.4$ Other and unspecified hyperlipidemia</td>
<td>$272.4$ Other and unspecified hyperlipidemia</td>
<td></td>
</tr>
<tr>
<td>$311$ Depressive disorder, not elsewhere</td>
<td>$427.31$ Atrial fibrillation</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Last Visit</th>
<th>Ground-truth diagnoses</th>
<th>Predictive diagnoses</th>
</tr>
</thead>
<tbody>
<tr>
<td>$427.31$ Atrial fibrillation</td>
<td>$427.31$ Atrial fibrillation</td>
<td></td>
</tr>
<tr>
<td>$414.01$ Coronary atherosclerosis of native coronary artery</td>
<td>$414.01$ Coronary atherosclerosis of native coronary artery</td>
<td></td>
</tr>
<tr>
<td>$250.00$ Diabetes mellitus without mention of complication</td>
<td>$272.4$ Other and unspecified hyperlipidemia</td>
<td></td>
</tr>
</tbody>
</table>

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