TIME SERIES FOR PATIENT ADHERENCE

Yin Li[1] *, **Yu Xiong**[2] *, **Wenxin Fan**[3], **Kai Wang**[1], **Qingqing Yu**[1], **Liping Si**[4], **Patrick van der Smagt**[5,6], **Jun Tang**[1], **Nutan Chen**[6]

[1] Department of Otorhinolaryngology, The First People's Hospital of Foshan, China

[2] Dept. of Otorhinolaryngology, The Second Affil. Hosp. of Guizhou Univ. of TCM, Guiyang, China

[3] Chinese Academy of Sciences Shenzhen Institutes of Advanced Technology, Shenzhen, China

[4] Department of Radiology, Zhongshan Hospital, Fudan University, Shanghai, China

[5] ELTE University, Budapest, Hungary

[6] Machine Learning Research Lab, Volkswagen Group, Munich, Germany

ABSTRACT

Subcutaneous Immunotherapy (SCIT) is the long-lasting causal treatment of allergic rhinitis (AR). How to predict and enhance the adherence of patients to maximize the benefit of allergen immunotherapy (AIT) plays a crucial role in improving the efficiency of AIT management. To address this challenge, this study explores the application of the sequential model of Stochastic Latent Actor-Critic (SLAC) and Long Short-Term Memory (LSTM) models in predicting patient adherence and symptom scores in AIT for allergic rhinitis. By developing and analyzing these models, we creatively apply sequential models in the long-term management of SCIT with promising accuracy in the prediction of SCIT adherence in AR patients.

1 INTRODUCTION

Allergic rhinitis (AR) is characterized by allergen-specific IgE-mediated inflammation in upper respiratory inflammation with a prevalence of up to 30% worldwide (Meltzer, 2016). Allergen-specific immunotherapy (AIT) aims to induce specific allergen immune tolerance, consequently achieving a status of clinical symptom remission. Among these approaches of AIT, SCIT is a clinic-dependent treatment in which the patient accepts an allergen extract injection subcutaneously.

Due to the long duration of SCIT, cumbersome process, slow onset, individual differences in treatment effect, and other factors fundamentally impact the completeness of therapeutics. From the reported studies on AIT, the rate of adherence ranged from around 25% to over 90% (Passalacqua et al., 2013). Multiple approaches were introduced into the field of improving adherence and supervising patient outcomes with systematic and technological interventions to prevent incomplete discontinuation of the treatment. With a multitude of personalized patient data, it is promising to employ a clinical prediction model to accurately identify and assess the risk of future non-adherent behavior. This approach enables targeted measures to enhance compliance among non-adherent patients, thereby improving efficiency.

Machine learning, particularly sequential models, is driving innovation in healthcare by analyzing medical data and enhancing patient treatments. These models are adept at handling time-sensitive data, crucial for forecasting patient compliance with treatments such as SCIT for AIT. Our study

*These authors contributed equally to this work.

evaluated two sequential models, showing their effectiveness in predicting treatment adherence and their significant impact on patient-focused healthcare. For those interested in a more comprehensive analysis and additional results, the full version of this paper can be found in (Li et al., 2024).

2 STUDY DESIGN

The study design is a critical component that shapes the direction and reliability of our research. It includes a systematic approach to selecting the study population, the treatment methods applied, and the evaluation criteria.

Population A retrospective analysis including 205 AR patients who started SCIT treatment between August 2018 and September 2019 in the Immunotherapy Center at the First People's Hospital of Foshan was performed. According to the Guidelines for the Diagnosis and Treatment of Allergic Rhinitis (2015 Edition), the recruit criteria (Li et al., 2024) were formulated: According to the guidelines for the diagnosis and treatment of allergic rhinitis (2015 Edition), the recruit criteria were formulated: patients who exposure to dust mites was confirmed as the major allergen by allergen tests with skin index (SI) of skin prick test (SPT) ++ or above or specific Immunoglobulin E(sIgE) level in serum to $\mathrm{Der\,p}/\mathrm{Der\,f} \geq 0.35$ kU/L.

SCIT treatment and evaluation Standard SCIT was performed according to the EAACI guideline including dose accumulation phase and maintenance phase(Roberts et al., 2018). Patients receive regular treatment evaluations, including symptom scores with visual analogue scale (VAS) and medication scores(Roberts et al., 2018). Medication score recorded the use of current adjuvant medication within 1 month to reach symptom relief. The use of oral antihistamines, antileukotrienes, and bronchodilators were recorded as one point, local glucocorticoids as two points, oral glucocorticoids or combined medication (hormones and $\beta 2$ receptor agonists) as three points, and the total cumulative score was the medication score. Symptom scores and medication scores were assessed once at registration of SCIT and then thereafter. All the chosen patients completed the four months of SCIT, and we chose the fourth month as the starting point of the observation. The data collection spans six time steps: at 0, 4, 12, 18, 24, and 36 months.

Data Collection Data were collected from patient records in hospitals, and the following information was extracted for analysis: patient age, gender, distance to clinic, ratio of AIT cost to family income, allergen test results, etc., as well as patient VAS system score and medication score information, including baseline data of patients before injection therapy, adverse reactions to SCIT. For the descriptive analysis, categorical variables were given as numbers and percentages, and continuous variables were presented using mean, standard deviation, median, interquartile range (IQR), and minimum and maximum values. Further details about the data are available in Appendix A, and the dataset for this study can be found in a GitHub repository[1].

Survey methods Adherence was defined as the accomplishment of three years of AIT including the patients further received AIT. Non-adherence was defined as discontinuation of AIT at random time points during three years. The follow-up contents (see Appendix A) included (1) the main reasons for patients' discontinuation of treatment; (2) the duration of discontinuation of treatment, and (3) Allergic symptoms after discontinuation of treatment.

[1] `https://github.com/leexxe/Subcutaneous-Immunotherapy-Dataset`

3 SEQUENTIAL MODELS

Our study aims to develop sequential models for accurately predicting symptom progression and patient adherence during SCIT. We explore and compare two distinct sequential models, SLVM of Stochastic Latent Actor-Critic (SLAC) (Lee et al., 2020) and Long short-term memory (LSTM) (Hochreiter & Schmidhuber, 1997).

Data We have a dataset $D = \{[\mathbf{x}^{(i)}, \mathbf{y}^{(i)}, \mathbf{a}^{(i)}, s^{(i)}], i \in \{1, \ldots, N\}]\}$ consisting of states $\mathbf{x} = \{x_t \in \mathbb{R}^{11}\}$, $t = 1 \ldots T$; observations $\mathbf{y} = \{y_t \in \mathbb{R}\}$, $t = 1 \ldots T-1$; and actions $\mathbf{a} = \{a_t \in \mathbb{R}\}$, $t = 1 \ldots T-1$. The observations y_t indicate whether the patient will cease the treatment in the interval between the scoring measurements at x_t and x_{t+1}. The actions a_t represent the ongoing medical procedures in the period from t to $t+1$. The state x consists of nasal itching, sneezing, rhinorrhea, nasal congestion, ocular itching, lacrimination, shortness of breath, tightness in chest, perennial cough, wheezing, and rescue medication score. For each patient we have basic information $s \in \mathbb{R}^{14}$: age, gender, commute distance to clinic, ratio of cost to family income, eosinophils count and percentage, nasal allergen provocation test (change of nasal resistance, $\Delta\mathrm{NR}(\%)$), peak nasal inspiratory flow; $\Delta\mathrm{PNIF}(\%)$), serum total IgE level, sIgE of Dermatophagoides pteronyssinus (Derp), sIgE of Dermatophagoides farinae (Derf), skin prick test (Derp, Derf).

Sequential latent variable models In our research, we use the SLAC model. However, our application differs from the original use of SLAC which is typically associated with reinforcement learning. Instead, we only use its sequential latent-variable model (SLVM) without the Actor-Critic.

SLVM consists of an inference model and a generative model (see Fig. 1). The inference model in SLVM typically aims to approximate the posterior distribution of the latent variables z^1 and z^2 given the observed data, x and s. The generative model, on the other hand, describes how the observed data is generated from the latent variables. q_ϕ and p_ϕ indicate the parameterized distributions. We have the evidence lower bound (ELBO) in Eq. (14). We adapt it to a constraint optimization problem:

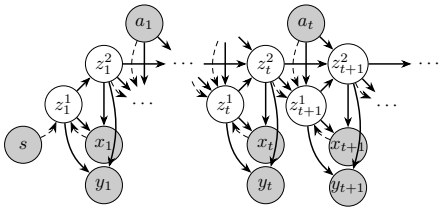

Figure 1: Schematic of the SLVM part of SLAC. Solid and dashed lines denote the generative and inference model pathways, respectively. The gray circles represent observed data, and the white circles denote latent variables. The figure is adapted from (Lee et al., 2020).

$$\min_{\phi} \mathop{\mathbb{E}}_{(x_{1:T}, a_{1:T-1}) \sim D} \left[\sum_{t=0}^{T-1} \left[D_{\mathrm{KL}} \left(q_\phi(z_{t+1} \mid x_{t+1}, z_t, a_t) \parallel p_\phi(z_{t+1} \mid z_t, a_t) \right) \right] \right] \tag{1}$$

$$\text{s.t.} \quad \mathop{\mathbb{E}}_{(x_{1:T}, a_{1:T-1}) \sim D} \left[\mathop{\mathbb{E}}_{z_{1:T} \sim q_\phi} \left[\sum_{t=0}^{T-1} -\log p_\phi(x_{t+1} \mid z_{t+1}) \right] \right] \leq \xi_{\text{score}} \tag{2}$$

$$\mathop{\mathbb{E}}_{(x_{1:T}, a_{1:T-1}, y_{1:T-1}) \sim D} \left[\mathop{\mathbb{E}}_{z_{1:T} \sim q_\phi} \left[\sum_{t=0}^{T-2} -\log p_\phi(y_{t+1} \mid z_{t+1}) \right] \right] \leq \xi_{\text{adherence}} \tag{3}$$

where ξ is a baseline error. We abbreviate $q(z_1 \mid x_1, z_0, a_0) := q(z_1 \mid x_1, s)$ and $p(z_1 \mid z_0, a_0) := p(z_1)$. In Eq. (2) we have regression with Gaussian distribution, and in Eq. (3) we use cross-binary entropy loss for classification. Eq. (1) is the Kullback-Leibler (KL) divergence between the variational distribution and the prior distribution of the latent variables. Implementation details are in App. B.

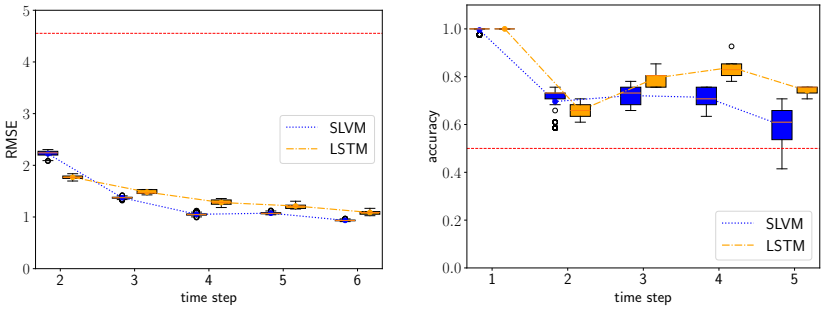

Figure 2: (left) RMSE and (right) Accuracy of the one-step prediction. The red dashed line is the RMSE or accuracy of random prediction with Uniform distribution.

4 RESULTS

We have a total of $N = 205$ samples, which we have randomly divided into a test dataset comprising $20\,\%$, i.e., 41 samples. For our analysis, we use a five-fold cross-validation approach. Additionally, we apply zero-mean and unit standard deviation (STD) normalization to the variables x and s. The Root Mean Square Error (RMSE) and classification accuracy metrics are used to evaluate the precision of our medical score and adherence predictions, respectively. The uncertainties for both models are calculated using five-fold cross-validation. In addition, as SLVM is a probabilistic model, we also perform 100 samples from the latent space to compute its uncertainty.

One-step prediction and rollout In Fig. 2 (left, middle), our focus is on predicting the immediate next step. As illustrated in Fig. 2 (left), SLVM surpasses LSTM in performance beginning at time step three. The figure indicates that with an increased amount of historical data (additional time steps), SLVM achieves better RMSE. Fig. 2 (right) demonstrates that from steps two to four, accuracy in adherence predictions improves with the inclusion of additional information. The first step shows a notable bias, since it only includes data from adherent patients, as detailed in (Li et al., 2024). Nonetheless, both models adeptly manage this bias and achieve high-accuracy predictions. Prediction for the sixth step is not conducted due to the cessation of treatment by the hospital. In the fifth step, there is a decline in accuracy, likely due to the extended time interval of 12 months. Table 1 illustrates details of the classification for one-step prediction. Both SLVM and LSTM perform considerably better over random prediction methods. Our findings, while not depicted in the figures, indicate comparable outcomes for the rollout process, i.e., when provided with one or more initial steps, we are able to predict the subsequent steps.

Table 1: Comparison of LSTM and SLVM over different time steps. The results are expressed as a mean ± standard deviation. The better results are highlighted in bold.

metric	model	time step 1	time step 2	time step 3	time step 4	time step 5
accuracy	LSTM	$\mathbf{1.00 \pm 0.00}$	0.66 ± 0.03	$\mathbf{0.80 \pm 0.04}$	$\mathbf{0.84 \pm 0.05}$	$\mathbf{0.74 \pm 0.02}$
	SLVM	1.00 ± 0.01	$\mathbf{0.70 \pm 0.06}$	0.72 ± 0.04	0.71 ± 0.04	0.60 ± 0.06
precision	LSTM	$\mathbf{1.00 \pm 0.00}$	0.72 ± 0.01	$\mathbf{0.86 \pm 0.06}$	$\mathbf{0.90 \pm 0.05}$	$\mathbf{0.62 \pm 0.03}$
	SLVM	$\mathbf{1.00 \pm 0.00}$	$\mathbf{0.75 \pm 0.03}$	0.74 ± 0.03	0.71 ± 0.03	0.44 ± 0.05
recall	LSTM	$\mathbf{1.00 \pm 0.00}$	0.86 ± 0.06	0.83 ± 0.05	0.82 ± 0.08	0.61 ± 0.03
	SLVM	1.00 ± 0.01	$\mathbf{0.87 \pm 0.06}$	$\mathbf{0.90 \pm 0.03}$	$\mathbf{0.86 \pm 0.04}$	$\mathbf{0.70 \pm 0.10}$
F1 score	LSTM	$\mathbf{1.00 \pm 0.00}$	0.79 ± 0.03	$\mathbf{0.84 \pm 0.03}$	$\mathbf{0.85 \pm 0.05}$	$\mathbf{0.62 \pm 0.03}$
	SLVM	1.00 ± 0.00	$\mathbf{0.81 \pm 0.04}$	0.81 ± 0.02	0.78 ± 0.03	0.54 ± 0.06

Interpretability We use Integrated Gradients of Captum (Kokhlikyan et al., 2020) for interpreting the model (see Fig. 3). The magnitude of features highlights the significance of the model's prediction for a specific class. The distance to the clinic significantly impacts patient adherence, especially if a patient is located far from the clinic or has relocated, as they are more likely to discontinue their visits. Following the distance, SPT of Der f and sIgE of Der f greatly influence the adherence. In contrast, ΔNR(%), EOS(%), and the cost/family income(%) have minimal impacts.

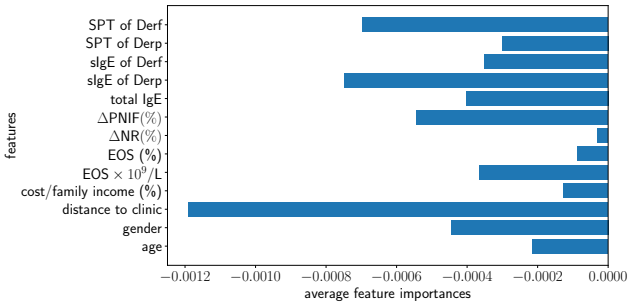

Figure 3: Importance of the factors using SLVM.

5 Discussion and Conclusion

The present study is the first research about the application of machine learning models in the adherence prediction of SCIT in AR patients. Previous research primarily concentrated on non-sequential prediction methods for adherence (Mousavi et al., 2022; Wang et al., 2020; Warren et al., 2022; Ruff et al., 2019). This approach presents a significant limitation in treatment processes, particularly for immunotherapy that often spans extended periods, such as three years. These non-sequential methods tend to predict only the overall outcome, overlooking the intricacies of intermediate time steps. To facilitate earlier intervention, a sequential model capable of making predictions at any given time step would be markedly more beneficial. While some subsequent studies have introduced sequential models (Hsu et al., 2022; Singh et al., 2022; Schleicher et al., 2023), their scope was restricted to predicting adherence alone. Our study enhances this approach by incorporating a state-action model, which can predict both adherence and score/state. This advancement allows for more precise and detailed analysis of patient cases by medical professionals.

The comparison of the SLVM part of the SLAC model with LSTM reveals the distinct strengths and limitations of each approach. This flexibility of SLVM is observed in its predictive capabilities. SLVM can predict y_t and use this prediction to influence the subsequent x_{t+1}. In contrast, standard LSTM only predicts a pair of y_t and x_{t+1} simultaneously, implying that we cannot use y_t to alter x_{t+1}. This advantage likely stems from its ability to efficiently learn and generalize in complex environments. Additionally, the SLVM outperforms the LSTM in score prediction. Conversely, the LSTM model shows better performance in predicting adherence, indicating its potential usage in scenarios. Both models demonstrate the capability to handle longer sequences, extending beyond one-step prediction. This ability is crucial in medical settings where long-term patient monitoring and prediction are essential for effective treatment planning.

Overall, the study underscores the importance of selecting the appropriate model based on the specific requirements of the task, whether it be flexibility, precision in score prediction, or adherence prediction. The findings contribute to the growing field of machine learning applications in healthcare, particularly in enhancing patient-centered treatment strategies through accurate and personalized predictions. Future research could focus on evaluating the SLVM model's performance in simulating various actions, further enriching its applicability in clinical settings.

REFERENCES

A. A. Alemi, B. Poole, I. Fischer, J. V. Dillon, R. A Saurous, and K. Murphy. Fixing a broken ELBO. *ICML*, 2018.

Nutan Chen, Alexej Klushyn, Francesco Ferroni, Justin Bayer, and Patrick Van Der Smagt. Learning flat latent manifolds with VAEs. *ICML*, 2020.

Nutan Chen, Patrick van der Smagt, and Botond Cseke. Local distance preserving auto-encoders using continuous knn graphs. In *Topological, Algebraic and Geometric Learning Workshops 2022*, pp. 55–66, 2022.

I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner. Beta-VAE: Learning basic visual concepts with a constrained variational framework. *ICLR*, 2017.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9: 1735–1780, 1997.

William Hsu, James R Warren, and Patricia J Riddle. Medication adherence prediction through temporal modelling in cardiovascular disease management. *BMC Medical Informatics and Decision Making*, 22(1):1–21, 2022.

D. P. Kingma and M. Welling. Auto-encoding variational Bayes. *ICML*, 2014.

Durk P Kingma, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling. Improved variational inference with inverse autoregressive flow. *Advances in neural information processing systems*, 29, 2016.

A. Klushyn, N. Chen, R. Kurle, B. Cseke, and P. van der Smagt. Learning hierarchical priors in VAEs. *Advances in Neural Information processing Systems*, 32, 2019.

Narine Kokhlikyan, Vivek Miglani, Miguel Martin, Edward Wang, Bilal Alsallakh, Jonathan Reynolds, Alexander Melnikov, Natalia Kliushkina, Carlos Araya, Siqi Yan, et al. Captum: A unified and generic model interpretability library for pytorch. *arXiv preprint arXiv:2009.07896*, 2020.

Alex X Lee, Anusha Nagabandi, Pieter Abbeel, and Sergey Levine. Stochastic latent actor-critic: Deep reinforcement learning with a latent variable model. *Advances in Neural Information Processing Systems*, 33:741–752, 2020.

Yin Li, Yu Xiong, Wenxin Fan, Kai Wang, Qingqing Yu, Liping Si, Patrick van der Smagt, Jun Tang, and Nutan Chen. Sequential model for predicting patient adherence in subcutaneous immunotherapy for allergic rhinitis. *arXiv preprint arXiv:2401.11447v2*, Jan 2024. Available: https://arxiv.org/abs/2401.11447v2.

Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. On the variance of the adaptive learning rate and beyond. *arXiv preprint arXiv:1908.03265*, 2019.

Eli O Meltzer. Allergic rhinitis: burden of illness, quality of life, comorbidities, and control. *Immunology and Allergy Clinics*, 36(2):235–248, 2016.

Hediye Mousavi, Majid Karandish, Amir Jamshidnezhad, and Ali Mohammad Hadianfard. Determining the effective factors in predicting diet adherence using an intelligent model. *Scientific Reports*, 12(1):12340, 2022.

G Passalacqua, I Baiardini, G Senna, and GW Canonica. Adherence to pharmacological treatment and specific immunotherapy in allergic rhinitis. *Clinical & Experimental Allergy*, 43(1):22–28, 2013.

D. J. Rezende and F. Viola. Taming VAEs. *CoRR*, 2018.

Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *ICML*, volume 32, pp. 1278–1286, 2014.

G Roberts, Oliver Pfaar, CA Akdis, IJ Ansotegui, SR Durham, Roy Gerth van Wijk, Susanne Halken, Désirée Larenas-Linnemann, Ruby Pawankar, C Pitsios, et al. Eaaci guidelines on allergen immunotherapy: allergic rhinoconjunctivitis. *Allergy*, 73(4):765–798, 2018.

Carmen Ruff, Ludmila Koukalova, Walter E Haefeli, and Andreas D Meid. The role of adherence thresholds for development and performance aspects of a prediction model for direct oral anticoagulation adherence. *Frontiers in Pharmacology*, 10:113, 2019.

Miro Schleicher, Vishnu Unnikrishnan, Rüdiger Pryss, Johannes Schobel, Winfried Schlee, and Myra Spiliopoulou. Prediction meets time series with gaps: User clusters with specific usage behavior patterns. *Artificial Intelligence in Medicine*, 142:102575, 2023.

Ankita Singh, Shayok Chakraborty, Zhe He, Shubo Tian, Shenghao Zhang, Mia Liza A Lustria, Neil Charness, Nelson A Roque, Erin R Harrell, and Walter R Boot. Deep learning-based predictions of older adults' adherence to cognitive training to support training efficacy. *Frontiers in Psychology*, 13:980778, 2022.

T. Sønderby, C. K.and Raiko, L. Maaløe, S. K. Sønderby, and O. Winther. Ladder variational autoencoders. *NeurIPS*, 2016.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

Xudong Sun, Nutan Chen, Alexej Gossmann, Yu Xing, Carla Feistner, Emilio Dorigatt, Felix Drost, Daniele Scarcella, Lisa Beer, and Carsten Marr. M-hof-opt: Multi-objective hierarchical output feedback optimization via multiplier induced loss landscape scheduling. *arXiv preprint arXiv:2403.13728*, 2024.

Lei Wang, Rong Fan, Chen Zhang, Liwen Hong, Tianyu Zhang, Ying Chen, Kai Liu, Zhengting Wang, and Jie Zhong. Applying machine learning models to predict medication nonadherence in Crohn's disease maintenance therapy. *Patient preference and adherence*, pp. 917–926, 2020.

David Warren, Amir Marashi, Arwa Siddiqui, Asim Adnan Eijaz, Pooja Pradhan, David Lim, Gary Call, and Mark Dras. Using machine learning to study the effect of medication adherence in opioid use disorder. *PLoS One*, 17(12):e0278988, 2022.

Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017.

A DATA STATISTICS

The statistics of our data are presented in Table 2, Table 3, and Figure 4.

variables		patients		
		total	adherent	non-adherent
age	≤ 12	96 (46.7)	40	56
	13–17	30 (14.6)	10	20
	≥ 18	79 (38.7)	23	56
gender	Female	62 (30.2)	22	40
	Male	143 (69.8)	51	92
distance to clinic(km)	≤ 10	136 (66.3)	56	80
	> 10	69 (33.7)	17	52
cost/family income(%)	< 30	107 (52.4)	37	70
	30–50	77 (37.4)	32	45
	> 50	21 (10.2)	4	17
EOS($\times 10^9$/L)		0.37; 0.41	0.36; 0.52	0.38; 0.36
EOS %		0.05; 0.04	0.05; 0.05	0.05; 0.05
ΔNR(%)		16.67; 59.70	30.00; 92.80	14.80; 50.00
ΔPNIF(%)		11.90; 34.50	12.70; 39.30	11.10; 28.80
total IgE (kU/L)		286; 543	340; 487	226; 555
sIgE of Der p (kU/L)		30.80; 68.480	31.30; 74.40	30.40; 67.80
sIgE of Der f (kU/L)		40.00; 68.20	40.60; 75.10	37.10; 65.70
Der p SPT SI		1.04; 0.58	1.00; 0.59	0.82; 0.55
Der f SPT SI		1.00; 0.50	0.82; 0.51	0.80; 0.45

Table 2: Demographic and clinical data of the patients under subcutaneous immunotherapy. In the rows from Age to Cost/Family income, values indicate the number of patients (percentage, if available). Other rows represent the median and IQR. P-values are omitted due to their large values.

reasons for SCIT withdrawal	number of non-adherent patients				
	5–12 mths	13–18 mths	19–24 mths	25–36 mths	total by reason
no clinical improvement	18	11	8	21	58
medical issue	3	1	2	0	6
improved efficacy	0	0	0	24	24
schooling	3	3	0	5	11
side effects	2	1	1	2	6
COVID-19	9	7	3	1	20
personal issue	0	3	0	4	7
total by time period	35	26	14	57	132

Table 3: Detailed reasons for withdrawal from SCIT at different time points.

B IMPLEMENTATION DETAILS

B.1 SEQUENTIAL LATENT VARIABLE MODEL

The sequential latent variable model of the SLAC consists of an inference model and a generative model (see Fig. 1). The inference model in a sequential latent-variable model typically aims to approximate the posterior distribution of the latent variables given the observed data. It tries to infer the hidden states z based on the observed inputs x and initial states s. The inference models the probability distributions of the latent variables z^1 and z^2 at different time steps. q_ϕ denotes the

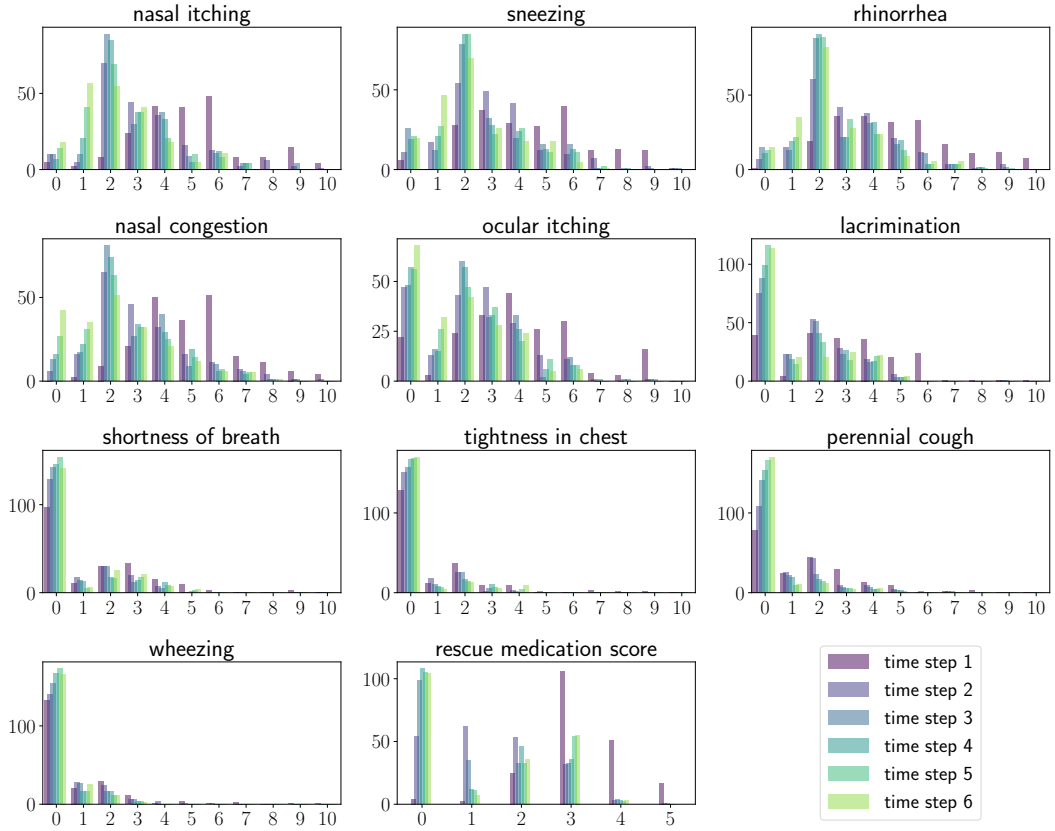

Figure 4: Histogram of scores across six-time steps. Score value (horizontal axis) vs. count (vertical axis).

variational distribution parameterized by ϕ,

$$z_1^1 \sim q_\phi(z_1^1 \mid x_1, s) \tag{4}$$
$$z_1^2 \sim p_\phi(z_1^2 \mid z_1^1) \tag{5}$$
$$z_{t+1}^1 \sim q_\phi(z_{t+1}^1 \mid x_{t+1}, z_t^2, a_t) \tag{6}$$
$$z_{t+1}^2 \sim p_\phi(z_{t+1}^2 \mid z_{t+1}^1, z_t^2, a_t). \tag{7}$$

The generative model, on the other hand, describes how the observed data is generated from the latent variables. The generative model is the probability distribution of both the initial latent states and their transitions over time, as well as the likelihood of the observations given the latent states, with p_ϕ indicating the parameterized generative distribution.

$$z_1^1 \sim p(z_1^1) \tag{8}$$
$$z_1^2 \sim p_\phi(z_1^2 \mid z_1^1) \tag{9}$$
$$z_{t+1}^1 \sim p_\phi(z_{t+1}^1 \mid z_t^2, a_t) \tag{10}$$
$$z_{t+1}^2 \sim p_\phi(z_{t+1}^2 \mid z_{t+1}^1, z_t^2, a_t) \tag{11}$$
$$x_t \sim p_\phi(x_t \mid z_t^1, z_t^2) \tag{12}$$
$$y_t \sim p_\phi(y_t \mid z_t^1, z_t^2). \tag{13}$$

We have the evidence lower bound (ELBO):

$$\log p_\phi(x_{1:t+1}|a_{1:t}) \geq \Bigg[\mathop{\mathbb{E}}_{(x_{1:T}, a_{1:T-1})\sim D} \Bigg[\mathop{\mathbb{E}}_{z_{1:T}\sim q_\phi} \sum_{t=0}^{T-1} \Big(\log p_\phi(x_{t+1} \mid z_{t+1}) - D_{KL}\big(q_\phi(z_{t+1} \mid x_{t+1}, z_t, a_t) \parallel p_\phi(z_{t+1} \mid z_t, a_t)\big)\Big)\Bigg]\Bigg]. \tag{14}$$

For ease of notation, we have $q(z_1 \mid x_1, z_0, a_0) := q(z_1 \mid x_1, s)$ and $p(z_1 \mid z_0, a_0) := p(z_1)$. The ELBO provides a lower bound to the log-likelihood of the observed data, which is computationally intractable to compute directly. It is composed of two terms: the expected log-likelihood of the observed data given the latent variables, and the Kullback-Leibler (KL) divergence between the variational distribution and the prior distribution of the latent variables. Minimizing the KL divergence can be interpreted as enforcing the variational distribution to be as close as possible to the prior, while maximizing the expected log-likelihood ensures that the model accurately captures the distribution of the observed data. To predict the adherence, we have $\log p_\phi(y_{t+1}|z_{t+1})$ as a regularisor in the loss function.

The objective is to compute the parameters ϕ that minimize the KL divergence between the variational and prior distributions of the latent variables, subject to certain constraints. These constraints are related to the expected log-likelihood of the data under the model and are represented by the inequalities with thresholds ξ. These thresholds ensure that while minimizing the losses, the model also satisfies a minimum standard for score prediction and adherence classification performances.

Latent variable models, such as Variational Autoencoders (VAEs) (Kingma & Welling, 2014; Rezende et al., 2014) and their variants (e.g., SLAC), often encounter challenges (Sønderby et al., 2016; Kingma et al., 2016). Furthermore, a higher ELBO does not always lead to enhanced predictive performance, as discussed by Alemi et al. (2018); Higgins et al. (2017). However, the integration of scheduling strategies inspired by constrained optimization methods has been shown to significantly improve the training of latent variable models (Rezende & Viola, 2018; Klushyn et al., 2019; Sun et al., 2024). Consequently, we formulate the training of our model into an optimization problem

$$\min_{\phi} \mathop{\mathbb{E}}_{(x_{1:T}, a_{1:T-1})\sim D} \left[\sum_{t=0}^{T-1} \left[D_{KL}\left(q_\phi(z_{t+1} \mid x_{t+1}, z_t, a_t) \parallel p_\phi(z_{t+1} \mid z_t, a_t) \right) \right] \right] \tag{1}$$

$$\text{s.t.} \quad \mathop{\mathbb{E}}_{(x_{1:T}, a_{1:T-1})\sim D} \left[\mathop{\mathbb{E}}_{z_{1:T}\sim q_\phi} \left[\sum_{t=0}^{T-1} -\log p_\phi(x_{t+1} \mid z_{t+1}) \right] \right] \leq \xi_{\text{score}} \tag{2}$$

$$\mathop{\mathbb{E}}_{(x_{1:T}, a_{1:T-1}, y_{1:T-1})\sim D} \left[\mathop{\mathbb{E}}_{z_{1:T}\sim q_\phi} \left[\sum_{t=0}^{T-2} -\log p_\phi(y_{t+1} \mid z_{t+1}) \right] \right] \leq \xi_{\text{adherence}} \tag{3}$$

where in Eq. (2) we have regression with Gaussian distribution, and in Eq. (3) we use cross-binary entropy loss for classification. To solve the optimization problem, we incorporate the constraints into the objective function using Lagrange multipliers λ. We apply methods from (Chen et al., 2022) to adapt λ. This allows the model to balance the importance of the constraints relative to the divergence terms, which can help in avoiding common pitfalls in training such as suboptimal local minima and posterior collapse.

To avoid over-fitting, we incorporate dropout (Srivastava et al., 2014) and Mixup (Zhang et al., 2017). Subsequent research has extended the application of Mixup to latent variable models, specifically within the latent space (e.g., (Chen et al., 2020)). However, considering our need for data augmentation

across all data dimensions, not limited to latent variables, we have selected to implement the original Mixup method in our experiments.

B.2 LSTM

The primary objective of this study is to forecast y_t from historical data, formulated as $y_t = f(x_{1:t}, y_{1:t-1}, s)$. To align this approach with the SLVM of SLAC for score prediction, an additional term x_{t+1} is also predicted,

$$(x_{t+1}, y_t) = f(x_{1:t}, y_{1:t-1}, s) \tag{15}$$

where f is a function represented by an LSTM. The loss consists of the cross entropy for adherence classification and the Normalized Mean Squared Error Loss (NMSE) for score prediction.

In our scenarios, SLVM stands out due to its inherent flexibility over traditional sequential models like LSTM. This flexibility is primarily observed in its predictive capabilities. SLVM can predict y_t and use this prediction to influence the subsequent x_{t+1}. In contrast, LSTM only predicts a pair of y_t and x_{t+1} simultaneously, implying that we cannot use y_t to alter x_{t+1}. Although it is possible to modify the LSTM model to predict a pair of y_t and x_t, this approach encounters a similar issue for y_t: it cannot predict y_t using the information from x_t.

B.3 Architecture and computation

In this study, computational experiments were performed using an NVIDIA GeForce GTX 1080 Ti GPU, with the implementation done in PyTorch, version 2.1.0.

The SLVM model's architecture featured 32 hidden dimensions each for variables z_1 and z_2. Its encoder and decoder were symmetrically structured, each comprising five layers with 128 units. The primary activation function was LeakyReLU, set with a negative slope coefficient of 0.2. Both the encoder and decoder's mean output layers were linear, while the STD layer utilized a Softplus activation. For binary classification tasks, a Sigmoid activation was used for output.

The LSTM architecture included a hidden dimension size of 128, with two LSTM layers. The output activation function for score prediction was linear, and as in the SLVM model, a Sigmoid function was used for binary classification outputs.

Both models shared the same optimization settings. They used the RAdam (Liu et al., 2019) optimizer with a learning rate of 0.001. The batch size was set at 64, and a gradient clipping value of 0.8 was applied to ensure training stability. To prevent overfitting and enhance model generalization, a dropout rate of 0.05 was introduced. Additionally, both models incorporated Mixup as a data augmentation during training.