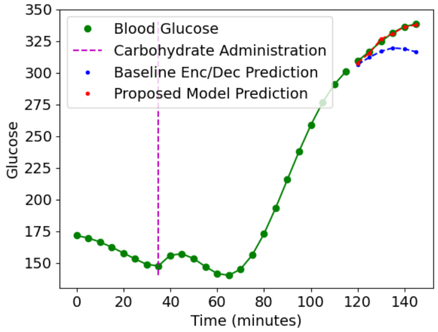


Figure 4: A sample prediction for a simulated individual (adult#004). Our model better accounts for the steep rise in the BG signal following a meal.

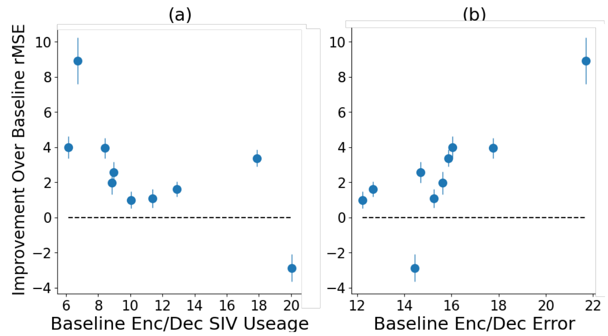


Figure 5: (a) Our architecture's improvement over baseline increases as baseline SIV usage decreases. (b) Improvement over baseline is positively correlated with baseline rMSE. Each point represents an individual in the simulated dataset, and errors bars represent standard error *i.e.*, the standard deviation of 100 bootstrapped samples.

baseline (rMSE 13.07 vs 14.14, paired t-test across individuals [t statistic/p-value]: 2.0/0.05). We also examine the clinical benefit of forecasting with our approach using a Clarke error grid in **Appendix C**.

| Model | rMSE [95%CI](Usage) | MAE [95%CI](Usage) |
|---|---|---|
| **BASELINES** | | |
| Enc/Dec | 15.63[14.1,16.9](11.13) | 12.42[11.1,13.6] (6.63) |
| SIV Fine-tune | 27.30[24.7,29.1](-0.54) | 22.22[19.9,23.9](-3.17) |
| SIV Initialize | 15.37[13.6,16.9](11.39) | 11.99[10.7,13.2] (7.06) |
| Full Capacity | 14.14[12.7,15.3](12.62) | 11.21 [9.9,12.2] (7.85) |
| **PROPOSED** | 13.07[11.8,14.2](13.69) | 10.45 [9.4,11.4] (8.61) |
| **ABLATIONS** | | |
| No Gating | 13.93[12.6,15.0](12.84) | 11.11 [9.9,12.1] (7.95) |
| No Restriction | 13.12[11.8,14.2](13.64) | 10.43 [9.3,11.4] (8.62) |
| No SIV Input | 14.20[12.7,15.4](12.56) | 11.18 [9.9,12.2] (7.87) |
| Only SIV Input | 13.97[12.6,15.1](12.79) | 11.12 [9.6,12.2] (7.94) |

Table 1: Mean forecasting error and SIV usage. Outcomes are reported as: Error [95% confidence interval] (SIV Usage). Our proposed approach outperforms baselines and ablations. CIs were calculated using 1,000 bootstrap samples.

**Individual Level Results.** In the simulated dataset, we consider ten individuals who differ in terms of simulated physiological parameters. Here, we investigate how model performance with respect to the baseline Encoder/Decoder varies across these individuals. In particular, we identify settings in which our approach is most beneficial.

Our model's benefit over baseline varies inversely with the extent to which the baseline approach relies on the SIV (*i.e.*, SIV usage) across individuals (Pearson r=-0.65, p=0.041, **Figure 5 (a)**). This supports the hypothesis that our model's improved performance over the baseline is due in part to the increased usage of the SIV. For individuals for whom the baseline model was able to achieve high usage, our model was not necessary, but individuals with low baseline usage stood to benefit. We also observe a strong correlation between baseline forecast error and our approach's improvement (r=0.80, p= 0.0056, **Figure 5 (b)**). This suggests that our approach addresses the deficits of the baseline at the individual level. There is higher rMSE variability across individuals for the baseline compared to the proposed approach (range: 9.5 vs 6.3), perhaps partially due to difficulties in SIV modeling, for which our model compensates.

**Ablations.** Ablation analyses reveal that, in general, our approach's strong SIV usage and forecast accuracy are a result of the combined effects of each implementation detail, rather than any one component. **Table 1** shows the results of our ablation study: removing any component results in a decrease in performance accuracy and SIV usage. Notably, removing the domain-guided restriction (*i.e.,* the ReLU functions) results in the smallest effect on performance. This is likely because, in our simulated dataset, the effect of the insulin boluses and carbohydrate administrations are strong enough that the model can learn them easily without supervision. When the restriction element is included, it seems to be utilized by the model: randomly switching the sign of the ReLU component at test time in a post-hoc analysis drastically hampers performance, more than doubling rMSE. We expect this component to have a greater effect in situations where the impact of the SIV is more subtle.

**Noise and Missingness.** In the above experiments, we assume that the SIV is accurately measured. To quantify the impact of measurement noise on our approach, we perturb the SIV as described in the experimental setup section. We find that as missingness and noise increase, our approach's performance degrades (**Figure 6**). Relative to the baseline, our approach is more impacted by corrupted SIV values, in part because of the increased dependence on the SIV (*i.e.*, greater SIV usage). As expected, completely omitting carbohydrates has a greater effect than simply corrupting the magnitude. Though performance decreases with increased noise/missingness, we are encouraged by the fact that our proposed approach remains competitive with the baseline.

We further explore the effects of a corrupted SIV signal using the 'Ohio' data, a real dataset generated by individuals with type 1 diabetes. While we cannot measure the amount of noise in the Ohio dataset, we expect it to be more in line with the simulated-noise setting than the noise-free setting. Somewhat reassuringly, even in this noisy setting, our approach performs no worse than existing approaches and even

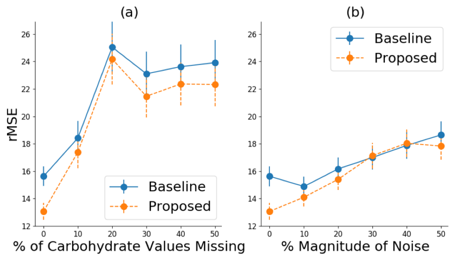


Figure 6: Mean model performance across all individuals as (a) simulated carbohydrate values are hidden and (b) noise is added to carbohydrates. As noise and missingness increase, our model fails to reliably outperform the baseline. Errors bars represent standard error *i.e.*, the standard deviation of 100 bootstrapped samples.

provides a small benefit over baselines. Specifically, our approach consistently leads to lower forecast rMSE compared to all baselines, though performance gains are modest (rMSE=20.16 vs. the strongest baseline rMSE=20.36). Furthermore, in ablation analyses we found that the restriction component was beneficial for this dataset, supporting the hypothesis that domain knowledge insertion is beneficial for more challenging tasks (see **Appendix A** for full results).

## Conclusions

The SIV problem arises in forecasting domains when the relative sparsity of an auxiliary signal makes learning its effect on a target signal challenging. We introduce the problem and propose a forecasting approach that leverages SIVs. Our approach isolates SIV dynamics and restricts them based on domain knowledge, achieving higher SIV-usage and stronger forecasting performance than baselines. While our approach assumes accurately measured SIVs, it performs no worse than baselines in the presence of missing or noisy SIV measurements. Though we focus on a specific use case for which we have a reliable simulator, we expect the SIV problem to arise frequently in healthcare. In such settings, SIVs are likely associated with time periods during which a patient is most vulnerable (*i.e.*, medication administration). Therefore, prediction models that address the SIV problem could lead to more accurate predictions during time periods that are critical for health outcomes.

We are the first to identify the SIV problem that arises when using RNNs for multi-input forecasting and the first to propose a solution. While sparsely sampled variables (SSVs) have been studied (**?**), the SIV problem is distinct. Interpolation approaches for addressing missingness and noise in SSVs are not directly applicable to the SIV setting. Although the SIV problem has not yet been addressed, several techniques have been proposed to learn inter-variable relationships in forecasting tasks, which in part inspire our approach. In the context of multi-input forecasting, Pantiskas et al. and Qin et al. use attention mechanisms to identify which variables to focus on (**??**). However, in contrast to our approach, these approaches do not account for signals that are mostly zero-valued, nor do they incorporate domain knowledge. Beyond attention based mechanisms, in a probabilistic setting, normalizing flows have been used to directly model the joint probabilities between variables (**??**). However, SIVs are often too sparse to accurately estimate a joint probability. Several other approaches have been proposed to explicitly model inter-variable relationships (**?????**). However, none explicitly addresses the sparsity issue. Moreover, while some of these architectures separate the effects of variables, none use this isolation to restrict the effects based on domain knowledge as we do. There has been other work in forecasting that combines deep learning with domain knowledge to reduce the hypothesis space. However, researchers have relied on strong assumptions, *e.g.*, structuring deep architectures to match clinical intuition (**?**), combining deep approaches with physiological-model-based simulators (**?**), and estimating expert judgements on model outputs via Monte-Carlo approximations (**?**). In contrast, we only restrict the sign of the SIV network's hidden state, a more flexible approach.

While there are many different ways to forecast signals, we focus on RNN-based techniques. Our primary contributions are the identification of the SIV problem in forecasting and noting the failure of common RNN-based approaches when addressing this problem. We demonstrate how addressing the SIV problem can lead to improvements over directly comparable baselines. We do not claim SOTA in forecasting, but our findings could apply to many settings in which variants of RNNs are applied to forecasting problems with SIVs, which could include vital sign forecasting with medication administration as an SIV, stock prices with quarterly reports or news articles as SIVs, and traffic forecasting with holidays or events as SIVs. Our approach is designed to utilize SIVs once an appropriate domain has been identified, but empirically identifying SIVs represents an interesting research direction. The two main ideas behind our approach include gating and output restriction. While neither of these methodological developments are unprecedented on their own, their combined application to the SIV problem poses a novel direction for forecasting in related domains.

## Acknowledgements

This work was supported by JDRF (award no. 1-SRA-2019-824-SB). The views and conclusions in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied of JDRF.

Porro assumenda molestias dolorem esse voluptatem cupiditate commodi, ad laborum quos, ratione aut praesentium aliquam neque pariatur, porro velit sit soluta quae amet harum quos impedit commodi?Et ratione voluptatem atque iure, inventore nostrum corporis aperiam nihil earum et corrupti, sequi asperiores placeat quaerat, porro at itaque a iusto reiciendis perspiciatis ea, alias voluptate numquam expedita accusamus hic?Esse dicta enim rerum minima possimus fugit corporis laborum doloribus, quia sit ratione tempore et cum repellat perferendis dolorum, minima necessitatibus voluptates consectetur quia omnis dolore libero velit pariatur, magnam magni alias tempore quaerat ullam perferendis quis maxime laudantium libero sint, repudiandae sint