Neural CDEs as a Solution to Irregular EHR Data in 30-Day Unplanned Heart Failure Readmission Prediction*

Ryan Missel[1] [0000−0002−9509−6775]

College of Computing and Information Sciences, Rochester Institute of Technology, Rochester, NY, USA

Abstract. Unplanned intensive care unit readmissions of heart failure (HF) within 30 days contribute a large burden on hospitals and waste of medical resources yearly. The ability to accurately identify patients that are at risk for readmission is an open problem, to which many modern machine learning techniques have been applied. While these methods have had some success, they fail to represent Electronic Health Record (EHR) data well or provide sufficient model interpretability. The recent class of Neural Controlled Differential Equations (Neural CDEs) show promising ability to naturally incorporate the irregularity of EHR data while giving insight on the predictions of the network. In this work, we present a Neural CDE application on HF readmission prediction that demonstrates state-of-the-art results compared to previous applications and provide a method to interpret the decision process learned by Neural CDEs.

Keywords: Heart Failure Prediction · Neural CDEs · Electronic Health Record (EHR)

1 Introduction

Unplanned intensive care unit readmissions of heart failure (HF) within 30 days contribute a large burden on hospitals and waste of medical resources [9] [14]. The ability to accurately identify patients that are at risk for readmission is an open problem, one which would prevent avoidable suffering on individuals through early interventions and converse medical resources [14] [12]. Systemic advancements in the digitization of patient health recordings, called Electronic Health Records (EHR), allows for an interesting opportunity to construct large datasets used in building automated computational systems.

To this end, modern machine learning techniques have shown promise in tackling this problem due to their ability to handle immense quantity of messy data while learning deep representations of complex data distributions. However,

* Supported by Rochester Institute of Technology (RIT). **The results of this work are fake and are simply meant to be an exercise in research writing.**

previous techniques across the spectrum have fell short on reaching clinically acceptable sensitivity numbers and having a sufficient level of interpretability in their decision processing [5] [15]. This can be, in part, attributed to the complexity of EHR data, which consists of irregularly-sampled, partially-observed time series and heterogeneous features.

1.1 Classical Models

Standard regression models, such as Linear Regression (LR) and Support Vector Machines (SVM), are not inherently built to model time series and thus have to engineer static features to represent them [5]. Deep sequence models, such as Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks, are designed to model data series, but require interpolation and binning techniques to be able to handle time series that are irregularly-sampled due to their discrete hidden states [10] [6]. Both of these model types introduce levels of data representation error and cannot approximate the underlying process naturally.

One of the inherent problems with binning or imputation techniques on EHR time-series data is that false associations between specific events and recorded variables are introduced into the approximated data process [6]. For example, a timestep with a mix of imputed features alongside recorded features with sharp volatility contributes a potentially incorrect relationship between the imputed features and the sharp changes that were seen. The true underlying process may contain similarly volatile changes in these features that were imputed and thus the model may fail to capture the real relationship between predictors. This false association problem can be slightly handled through the incorporation of indicator variables specifying whether a certain feature was present or not in a timestep. However, this is a not a natural formulation for representing a time-series and introduces additional complexity to the model that may not be fully exploited by the models. It is this false association problem in the poor representation of the data process that we hypothesize has caused the limited performance of machine learning techniques in this domain.

1.2 Neural Controlled Differential Equations

The recently proposed class of Neural Controlled Differential Equations (Neural CDEs) allows us to solve this representation issue by modelling the hidden state as a differential equation whose dynamics are parameterized by a neural network [8]. Abstractly, it is the continuous-time formulation of a recurrent neural network and is built on the previous work "Neural Ordinary Differential Equations [3]," which are the continuous-time formulation of a residual network.

The Neural CDE model is defined as the solution to the CDE

$$z_t = z_0 + \int_{t_0}^{t} f_\theta(z_s) dX_s, \tag{1}$$

where z_0 represents an initial state defined by a learned linear map l_1 and the equation is driven by the continuous path X.

This hidden state formulation allows for the incorporation of observations and sampling of state at any time within the temporal bounds of the differential equation. Partially observed features are addressed through cubic spline interpolations on each data channel, which defines a continuous path X the hidden state depends on. Using cubic spline interpolations provide a minimum guarantee on complexity in capturing time series fluctuations, but different choices in spline order or continuous path interpolation, such as rough controls for discrete stochastic processes, may be used depending on the problem context.

These properties make Neural CDEs a natural fit to represent irregularly sampled and highly volatile EHR data, as no imputation techniques need to used that introduce mapping error to the approximated function. Since time is explicitly modelled into the ODE design, the quantity and frequency of observations are naturally learned as features. This is of particular importance to EHR data, as it has been shown that observational intensity for ICU patients has correlation with readmission.

An excellent and thorough introduction to this model class is [8], which is presented by the original authors. Though no knowledge is assumed on rough control theory for this work, a comprehensive overview is [11] and a strong introduction is [7].

1.3 Contributions

In this work, we apply the recent model class of Neural CDEs to the open problem of 30-day unplanned HF readmission prediction. We show how the inherent properties of its design are a natural fit to EHR data. We show through empirical tests the Neural CDE model demonstrates state-of-the-art performance (0.83 AUC, 95% CI: 0.81-0.85) when compared against previously proposed machine learning models (0.77 AUC, 95% CI: 0.75-0.79). Finally, we provide an analysis on the decision making process of the model through an ablation test over the features.

Due to a confidentiality agreement with our collaborators, we are unable to publicly share our data. However, the code used in our experiments is available at `www.github.com/qu-gg/ncde-hf-readmission-prediction`. The deep learning models are written in PyTorch v1.6 and our Neural CDE is built using the incredible torchcde library.

Our experiments are run on Windows 10 on a Nvidia RTX2060, Intel I7-10700k, and 16GB DDR6 RAM. In our repo, we provide an Anaconda environment file with the necessary libraries to run our code.

2 Related Work

There has been a number of similar works trying to apply a variety of machine learning architectures to the problem of ICU readmission prediction. Frizzell et al. [5] perform a survey study on a number of techniques for All-Cause 30-Day readmission prediction. Techniques such as Random Forest Regression to a Tree-Augmented Naive Bayes networks were compared against traditional statistical models (LASSO and Logistic Regression) and found that all predicted with a poor sensitivity. However, their work used limited time-series predictors that were modelled statically to fit into their linear models. Lin et al. [10] improve on the static modelling of these time series features for linear models by incorporating the learned parameters of regression lines fit to each time series as predictors in their feature vectors. This additional information on the trajectory of the series aided in the predictive capabilities of these models, but is limited in learning a true representation of the sequence as well as model order fitting to each individual feature.

Lin et al. [10] additionally propose a Bidirectional LSTM network that incorporates the time-series EHR features alongside static demographic features and a learned vector embedding of the ICD-9 codes a patient can have. They used an indicator variable to indicate missing observations in the time series features and imputed the features into hour intervals over the last 48 hours of EHR data for a patient. Compared to traditional methods and a Multi-filter CNN, their model performed significantly well on the prediction task and managed to predict a subset of cases with highly volatile features that the traditional methods could not. It provided promising results that more complex representation and incorporation of the time-series in EHR data are an important direction for improving the long stagnant predictive capabilities of models in this domain. However due to the discrete nature of LSTM networks in its hidden state, they were required to use imputation techniques with the limitations as discussed previously. Additionally, they used an immense number of redundant features in their feature vectors, carrying through static features into every interval. Learning a linear map from the static features to the hidden state initialization would reduce the computational required at each step and separate the static features from the time series ones.

While many techniques and architectures have been applied to this problem, all suffer from the inherent error that arises from limited representations of the underlying data process [5] [1]. Limited or fuzzy representations result in these models being unable to approximate the process sufficiently, which in turn results in limited predictive capabilities. The trend of better process representation techniques resulting in stronger predictive power is supported throughout the timeline of literature, starting from initial log regression models [13] [5] [15] to the most recent LSTM networks [10].

3 Methods

3.1 Data Cohort and Processing

The data for this study came from a cohort of 27,668 patients over a length of 20 years (2000-2020), gathered from 4 Pennsylvania hospitals under the Lewosky Administration. EHR data over the length of time a patient spent in the Intensive

Care Unit (ICU) and while hospitalized is included, as well as features such as medication orders and lab order tests.

To filter out the intended patients for this study, first only the patients who had at least one emergency hospital encounter with a primary diagnosis of heart failure were kept, given by their diagnosed ICD-9 code. Patients that are under 21, left their admission against medical advice, transferred in from another hospital, or died during their ICU visit are excluded from the set. Additionally, patients that were only in the ICU are excluded, due to their time stays being significantly shorter on average and representing a different distribution of patients. Only patients that have at least 48 hours of EHR are considered for prediction, given the findings of [10] that the first 48 hours of EHR data are important for readmission prediction.

Following this filtering, the resulting patient cohort consisted of 7,407 encounters stemming from 5,190 patients. The number of cases with unplanned readmissions within a 30-day period is 441 cases for heart-failure primary readmissions and 1,996 cases for all-cause emergency readmissions, constituting 6.0% and 27% positive cases respectively. These numbers fall in line with cohorts from other work and represents a difficult highly imbalanced class problem [5].

3.2 Features

Given the heterogeneity of EHR data, different groups of features need to be handled independently to format them into a cohesive feature vector. In this section, we first note the distribution of the available features in our dataset before describing their implementation in each of the models being compared.

Demographics Data. Demographics data has been widely used as a feature set for readmission prediction [1]. The features for this set consists of age, gender, race, number of previous hospitalizations within the last year, and insurance type. Insurance type has been shown as correlated in readmission prediction, as cheaper options may be representative of a patient who either may be uninsured and unable to process payment, leading to an early discharge, or a patient of lower socioeconomic status. Lower economic status patients have been correlated with higher rates of readmission in other studies and while the ability to extract this information is unavailable in this cohort, insurance type may embed that information into it [10]. Table 1 shows the features and their associated dimensions, as well as potential values.

The distribution of race in this cohort is highly imbalanced, with 98.2% of patients being white, 1.4% being black, and all remaining groups under 0.4% combined. There is a relatively even balance between gender, with 53.2% male and 46.8% female. The average age in the cohort is 82.0 ± 12.4 years old. For insurance types, 50.9% are on Medicare, 27.1% are on Lewosky's Health Plan, 19.2% are on private plans, 1.8% self-paid, and the remaining 1% are classified under Other.

Table 1. Demographic Features

Name	Dim	Values
1. Age	1	25-106
2. Gender	2	Male/Female
3. Race	4	Asian, Black, Hispanic, White
4. # Prev Hosp.	1	0-5
5. Insurance	5	Medicare, Lewosky's, Private, Self-Paid, Other

Time-series Data. From our EHR data, we extract 15 chart events over the first 48 hours of EHR, which captures all of ICU into a portion In-Patient data for the majority of patients. This consists of categorical features (i.e. Glasgow scores) and numerical features (i.e. Pulse, BP Diastolic). We define the normal values for each of these features in 2, which are used in the event that an encounter has no readings present for any of the features.

Table 2. Time-Series Features

Name	Normal Value
1. Glasgow - Motor Response	6 Obeys Commands
2. Glasgow - Verbal Response	5 Oriented
3. Glasgow - Eyes Open	4 Spontaneously
4. Glasgow - Total Score	15
5. Diastolic Blood Pressure	64.2
6. Systolic Blood Pressure	121.3
7. Respiratory Rate	19.3
8. Mean Arterial Pressure	81.6
9. Pulse	80.2
10. Oxygen Saturation (SpO2)	95.9
11. Temperature	97.8
12. Height	65.8
13. Weight	197.7
14. Urine Output	310
15. Modified Early Warning Score (MEWS)	1.1

Chronic Features. There has been significant research into the correlations between varying chronic diseases and heart-failure admission, highlighting that certain diseases, such as diabetes; anemia; or renal disease, significantly heightened the risk of heart failure readmission [1]. Patients with diabetes had rehospitalizations rates up to 28% within 30-days while patients with anemia also had significantly increased readmission risk.

To incorporate this information into our models, we utilize a pre-trained embedding model on the ICD-9 codes related to the other diagnoses given to the patient during this encounter and all the previous encounters within the prior year [10]. These codes are individually embedded into a 250-dimension feature vector and then piece-wise summed together to produce a single 250-dimension feature vector, which represents the embeddings of all the chronic diseases a patient is diagnosed with [4].

3.3 Regression Model Setup

As a baseline, we built and tested a number of static regression models on a flat feature vector. Demographic and chronic embedded features are static by nature and thus can be simply appending together in the feature vector. To incorporate the time-series features in a static context, we extract a number of statistical measurements from each feature. We take the max, min, and mean values that are observed throughout the time sequence of each feature. As well, we fit a linear regression model ($y = ax + b$) to each sequence and extract their parameters as a way to describe the general trajectory of each series. For series that lack any recordings for a patient, the normal value of that feature across all patients is included instead.

Figure 1 shows the layout of the feature vector with example values for a patient, before normalization.

13-Dim Demographic Features				250-Dim ICD-9 Embeddings				75-Dim Time-Series Statistics					
Age	Male	Female	...	Dim1	Dim2	...	Dim250	PulseMax	PulseMin	PulseMean	Pulse_a	Pulse_b	...
87	1	0		2.6	1.4		5.4	110.3	60.4	83.4	-1.3	15.3	

Fig. 1. Regression feature vector layout with example values.

To test a wide sweep of regression models, we performed a model selection loop over a list of models and hyperparameters to check if any were of a promising fit. Table 3 provides a list of the models searched over and Appendix X provides an in-depth list of the parameters searched over alongside the best parameters found over a 5-fold Grid Search Cross Validation pipeline.

3.4 LSTM Model Setup

The feature setup to the LSTM model is similar to the regression models, with the primary difference being the ability to take in time steps. Thus, the feature input now is instead a 2-D matrix with there being 48 timesteps representing the first 48 hours of EHR data. Static features (Demographics and ICD-9 embeddings) are carried through every timestep while the time-series features now becoming [1x48] columns. If a feature lacks an observation within that timestep, the Last Observation Carried Forward imputation method is employed. As before, if an encounter lacks a feature entirely, the normal value across the cohort is instead used. Figure 2 visualizes the format of the feature vector for the LSTM network.

Table 3. Regression Models Tested

Model Name
1. Linear Discriminant Analysis (LDA)
2. Quadratic Discriminant Analysis (QDA)
3. AdaBoost
4. Bagging
5. Extra Trees Ensemble
6. Random Forest
7. Ridge
8. Stochastic Gradient Descent (SGD)
9. Bernoulli Naive Bayes (BNB)
10. Gaussian Naive Bayes (GNB)
11. K-Nearest Neighbors (KNN)
12. Multi-layer Perceptron (MLP)
13. Decision Tree
14. Extra Tree

	13-Dim Demographic Features				250-Dim ICD-9 Embeddings				15-Dim Time-Series Features		
Hour	**Age**	**Male**	**Female**	**...**	**Dim1**	**Dim2**	**...**	**Dim250**	**Pulse**	**BP Dias**	**...**
1	87	1	0	...	2.6	1.4	...	5.4	80.2	65.3	...
2	87	1	0	...	2.6	1.4	...	5.4	85.2	67.2	...
...	87	1	0	...	2.6	1.4	...	5.4	89.3	63.7	...
48	87	1	0	...	2.6	1.4	...	5.4	81.5	62.1	...

Fig. 2. LSTM feature vector layout with example values.

To incorporate the values from the hidden state over the entire period rather than just the final output of the network, we take the output of the LSTM at every timestep and concatenate them together into a 48-dim vector that is passed to a 250-dim learned linear map to produce the final output via a sigmoid activation function.

Specifics on the size of the hidden states and hyperparameters throughout the LSTM network can be found in Appendix X. We initialize the model's weights using Xavier initialization. The optimization algorithm for the network is the Adam optimizer and we trained for at most 50 epochs over a 5-fold cross validation.

3.5 Neural CDE Setup

The feature vector of the Neural CDE varies quite a bit from those of the classical and LSTM models, due to its internal vector field formulation. The static features of the dataset are incorporated through a learned fully connected layer that initializes the hidden state of the model. The time series of the data are independently transformed into continuous paths through the cubic spline interpolations, where the coefficients of the splines over each observation are passed to the ODESolver.

An important additional factor that was found in [8] for EHR data is that of adding an observational intensity to the data points of each feature. This is simply a counter that increments at each observation of a given feature, which allows the model to exploit the frequency of observations in each feature. There have been studies that demonstrate the speed in observations for a patient encodes certain beliefs from the clinicians on the risk of a patient and can be useful for automated prediction [2].

Similar to the LSTM network above, the state of the CDE is input to a linear map that outputs the final prediction through a sigmoid activation. Since we cannot just take the outputs at every function evaluation of the ODESolver, as that would have a variable number of inputs to the map, we specifically evaluate the ODE at an hourly interval in the vector field to get its output. Note that this doesn't affect the dynamics of the system and is just a means of getting the trajectory of the hidden state at a standard interval. Additional work in this setting would be testing different formulations on the CDE output stream to investigate which is the most natural model fit.

4 Results

In this section, we compare the results of the different models on a couple of tasks, from heart-failure specific readmission to all-cause readmission prediction. As well, we perform an ablation test on the features of the CDE to investigate which features are the most influential on identifying readmissions. In order to preserve the output of the experiments, only the top-3 regression models are shown alongside the LSTM and Neural CDE models.

Throughout all of these experiments, the number of parameters available to the LSTM and Neural CDE models are kept the same, as to provide a fair comparison between them. Precise experiment details can be found in Appendix X.

4.1 30-Day Heart-Failure Specific Readmission

Table 5 shows the results for a 5-fold cross validation run on the heart-failure specific readmission task, of which there were 6.0% positive cases within the encounter set. There was a significant difference between the top-2 regression models when compared to the remaining regression models. SVM achieved the best results with 0.703 Recall and 0.779 A.R.; AdaBoost achieved 0.670 Recall and 0.770 A.R.; and DTC achieved 0.565 Recall and 0.712 A.R. The remaining regression models were all above 0.50 Recall and 0.6 A.R. in their evaluations. The LSTM model achieves a better Recall of 0.742 and an A.R. of 0.791, demonstrating the strengths of time-series incorporation over static feature vectors.

Our Neural CDE model significantly outperforms the previous models with a Recall of 0.812 and an A.R. of 0.831. These results support our hypothesis that the better process approximation techniques granted by the CDE formulation leads to stronger predictive ability.

Table 4. Model Results on 30-Day Heart-Failure Specific Readmission Prediction

Model	**Recall (95% CI)**	**A.R. (95% CI)**
DTC	0.565 (0.550–0.580)	0.712 (0.693–0.730)
AdaBoost	0.670 (0.647–0.694)	0.770 (0.758–0.782)
SVM	0.703 (0.685–0.720)	0.779 (0.768–0.789)
LSTM	0.742 (0.718–0.766)	0.791 (0.782–0.800)
Neural CDE	**0.812 (0.789-0.824)**	**0.831 (0.815-0.854)**

4.2 30-Day All-Cause Readmission

Alongside the heart-failure specific readmission test, we evaluated the models on the task of predicting all-cause readmission within a 30-day window. The task of all-cause prediction is a more balanced problem with 27% of cases having readmissions within the window. Similar to before, the SVM and AdaBoost models were significantly above the other regression models with DTC being slightly ahead of the remaining set. The LSTM model performed again performed well above the conventional models but lagged behind the proposed Neural CDE model.

Table 5. Model Results on 30-Day All-Cause Readmission Prediction

Model	**Recall (95% CI)**	**A.R. (95% CI)**
DTC	0.655 (0.635–0.678)	0.758 (0.720–0.776)
AdaBoost	0.701 (0.683–0.722)	0.812 (0.795–0.823)
SVM	0.721 (0.685–0.720)	0.824 (0.805–0.844)
LSTM	0.821 (0.801–0.839)	0.876 (0.855–0.904)
Neural CDE	**0.862 (0.849-0.883)**	**0.912 (0.901-0.923)**

5 Discussion

In this section, we investigate into visualizing the decision process of the Neural CDE model and discuss the current limitations of our study and the CDE model class in this setting.

5.1 Feature Ablation Test

In order to better understand the decision process of our model and the influence of certain features on the prediction task, we performed a feature ablation test. Simply put, a loop over each feature was performed, setting that feature to be the normal value of the cohort for all cases in order to see the difference in prediction with and without that feature being present. This was done over all 5-folds trained for all features. The ratio of the test set that were wrongly predicted after the modification was recorded for each feature, which is what signifies the effect this feature has on the prediction task. Figure 3 illustrates the features that were most impactful on the predictions, from highest to lowest. We include the ICD-9 embedded feature in this experiment, as it has been shown to have a significant impact on readmission prediction in previous works. [1]

The ICD-9 embedding feature vector was the most important for readmission prediction, influencing up to 10% of cases on whether or not they are accurately predicted. This validates previous work in their findings of a similar phenomena [1] [10]. Past the embedding vector, features mean arterial pressure, pulse, and SpO2 were the most influential in the model's predictions, aiding in up to a combined 4.5% of correct predictions. There are a number of additional features that are important to the readmission prediction, which highlights the model's understanding of the setting.

5.2 Limitations

One of the current limitations of the Neural CDE model in this specific setting is the need to define a bounded time interval upon which to parameterize the flow [8]. A trade-off needs to be made between capturing the most amount of data possible while not leaving a significant portion of samples with a relatively short timeframe compared to the time interval. Having a longer time interval in order to include only a small subset of long samples leads to a number of inaccurate interpolations from the shorter samples past their ending. These inaccurate interpolations means the learned flow in the end time interval is trained only on a small portion of real data samples, leading to a poor fit of the true underlying data process throughout that section.

Since this study was conducted retrospectively, additional research would be need to be conducted in a variety of real clinical settings in order to validate this procedure as a potential application. Additionally, this method would need to be applied to a combined variety of regions and their hospital records in order to test its generality on different demographics.

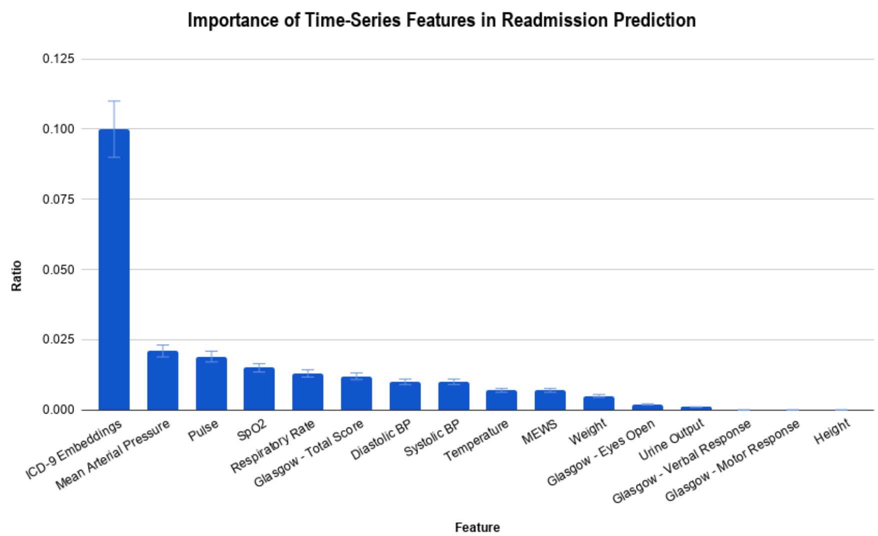

Fig. 3. Feature ablation test, where each feature is modified to be the normal value of the population in order to test its effect on the predictive capabilities of a model.

6 Conclusion

In this work, we demonstrated the state-of-the-art predictive capabilities of the recent Neural CDE model when compared to recent methodology in 30-day ICU readmission prediction for heart failure patients. It achieved a sensitivity of 0.812 and an AUC of 0.831, both of which are significantly higher than the most recently proposed LSTM-CNN model (0.782/0.791, p=0.005/0.004, paired t-test). The CDE model was able to confidently predict a subset of cases with uniquely volatile features that had high partial observance through the time window. These results support our hypothesis that the intricate representation of the underlying data process Neural CDEs provide is a direct benefit applications on EHR data.

Future work in this domain entails a deeper look into the rough path controls that can be utilized in generating the data controls used by the CDE models. While cubic splines provide the minimum guarantee for fitting, it's possible that there exist more appropriate techniques to apply to EHR data. Additionally, there is potential to better represent each time-series feature using different time intervals between them. By currently taking a bounded and uniform time boundary over all the features, there is the potential that some information is not being captured. Some features change on a hour by hour basis and these fluctuations are meaningful while other features may change slowly over a day by day window yet are just as important to the readmission prediction. Too long of an interval and there is the risk that the features with fast changes are overlooked while too short of an interval and the long-term changes aren't captured at all within the ODE vector field. Looking into different ways to represent these features individually or modify the ODE methodology to exploit these temporal differences may lead to even stronger results.

Finally, a full clinical pipeline may be constructed using a Neural CDE, given the potential that the CDE can be transformed into a probabilistic model. A patient's risk at each stage of admission - initial observations plus demographics, data throughout ICU, and all data up to discharge - can be evaluated through learned linear maps on the output of the CDE at each step. Thus, there is potential for CDE architectures to help guide clinician's treatment for a patient over their stay.

References

1. Alzaghari, O.: Factors influencing 30-day unplanned rehospitalization in adults with heart failure: data from an urban hospital in north carolina (2016)
2. Che, Z., Purushotham, S., Cho, K., Sontag, D., Liu, Y.: Recurrent neural networks for multivariate time series with missing values. Scientific Reports **8**(1) (Dec 2018). https://doi.org/10.1038/s41598-018-24271-9
3. Chen, R.T.Q., Rubanova, Y., Bettencourt, J., Duvenaud, D.: Neural ordinary differential equations (2019)
4. Choi, Y., Chiu, C., Sontag, D.: Learning low-dimensional representations of medical concepts. AMIA Joint Summits on Translational Science proceedings. AMIA Summit on Translational Science **2016**, 41–50 (07 2016)
5. Frizzell, J.D., Liang, L., Schulte, P.J., Yancy, C.W., Heidenreich, P.A., Hernandez, A.F., Bhatt, D.L., Fonarow, G.C., Laskey, W.K.: Prediction of 30-Day All-Cause Readmissions in Patients Hospitalized for Heart Failure: Comparison of Machine Learning and Other Statistical Approaches. JAMA Cardiology **2**(2), 204–209 (02 2017). https://doi.org/10.1001/jamacardio.2016.3956, `https://doi.org/10.1001/jamacardio.2016.3956`
6. Gelman, A.B., Hill, J.: Data analysis using regression and multilevel/hierarchical models. Cambridge University Press (2018)
7. Hodgkinson, L., van der Heide, C., Roosta, F., Mahoney, M.W.: Stochastic normalizing flows (2020)
8. Kidger, P., Morrill, J., Foster, J., Lyons, T.: Neural controlled differential equations for irregular time series (2020)
9. Kramer, A.A., Higgins, T.L., Zimmerman, J.E.: The association between icu readmission rate and patient outcomes*. Critical Care Medicine **41**(1), 24–33 (2013). https://doi.org/10.1097/ccm.0b013e3182657b8a
10. Lin, Y.W., Zhou, Y., Faghri, F., Shaw, M., Campbell, R.: Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. PLOS ONE **14**, e0218942 (07 2019). https://doi.org/10.1371/journal.pone.0218942
11. Lyons, T.J., Caruana, M.J., Thierry, L.: Differential Equations Driven by Rough Paths Ecole dEte de Probabilites de Saint-Flour XXXIV-2004. Springer Berlin Heidelberg (2007)
12. Mcilvennan, C.K., Eapen, Z.J., Allen, L.A.: Hospital readmissions reduction program. Circulation **131**(20), 1796–1803 (2015). https://doi.org/10.1161/circulationaha.114.010270

13. Nguyen, O.K., Makam, A.N., Clark, C., Zhang, S., Xie, B., Velasco, F., Amarasingham, R., Halm, E.: Predicting all-cause readmissions using electronic health record data from the entire hospitalization: Model development and comparison. Journal of hospital medicine **11 7**, 473–80 (2016)
14. Ponzoni, C.R., Corrêa, T.D., Filho, R.R., Neto, A.S., Assunção, M.S.C., Pardini, A., Schettino, G.P.P.: Readmission to the intensive care unit: Incidence, risk factors, resource use, and outcomes. a retrospective cohort study. Annals of the American Thoracic Society **14**(8), 1312–1319 (2017). https://doi.org/10.1513/annalsats.201611-851oc
15. Ross, J., Mulvey, G.K., Stauffer, B.D., Patlolla, V., Bernheim, S., Keenan, P., Krumholz, H.: Statistical models and patient predictors of readmission for heart failure (2016)