Utilizing Neural Temporal Point Processes to Forecast Subsequent Events in Diabetic Retinopathy and Associated Medications

Neeraj Yadav
220708603
Dr. Michael Barnes
MSc.Big Data Science

***Abstract*— This study delves into the potential of Neural Temporal Point Processes (TPPs) for forecasting subsequent events related to Diabetic Retinopathy (DR) and associated medications, considering DR as a predominant vision impairment among diabetic patients. Utilizing an expansive simulated Electronic Health Records (EHR) dataset, we juxtapose the efficacy of Neural TPPs against other prevalent machine learning paradigms. Key performance indicators, such as the Area Under the Receiver Operating Characteristic Curve (ROC-AUC) and Negative Log Likelihood, serve as our evaluation benchmarks. Our analysis uncovers that the GRU-based Conditional Poisson process stands out in terms of performance, especially in dealing with time-sensitive data and predictive precision. Meanwhile, attention-driven Neural Temporal Point Processes are also highlighted for their interpretability, despite showcasing comparable results. This investigation augments the expanding domain of healthcare analytics, offering an in-depth appraisal of Neural TPPs in predicting events related to DR and its medications, thereby paving the way for future scholarly endeavors and prospective clinical integrations.**

Keywords*— *Temporal Point Processes, Diabetic Retinopathy, Associated Medications, Electronic Health Records, Predictive Forecasting, Machine Learning, Healthcare Analytics, Performance Metrics, Synthea Dataset, Clinical Prognostication, Medical Informatics.

I. INTRODUCTION

Diabetic Retinopathy (DR) is a significant microvascular complication of diabetes mellitus and a predominant cause of vision impairment and blindness among adults worldwide. Prompt and precise detection of DR is crucial. Early interventions can prevent or mitigate vision loss, enhancing patient outcomes and life quality.

The surge in healthcare data digitization has opened doors for the development of predictive models to aid clinical decisions. Electronic Health Records (EHRs) encapsulate extensive patient data, spanning diagnoses, treatments, and temporal patterns of medical encounters. The Synthea Dataset, a synthesized, detailed EHR dataset, paves the way to design and validate predictive models for numerous health conditions, DR being one of them.

Temporal Point Processes (TPPs) cater specifically to event data where event timings play a pivotal role. Although they've showcased potential in various fields, their application in DR prediction remains relatively uncharted.

This study embarks on a journey to comprehensively evaluate the potential of Neural TPPs in forecasting subsequent events related to DR and associated medications, utilizing EHR data from the Synthea Dataset. The core of our evaluation revolves around Neural TPPs models, gauged through metrics like the Receiver Operating Characteristic Curve's Area Under the Curve (ROC-AUC).

The study's objectives encompass:

1. Evaluating Neural TPPs' performance in predicting events related to DR and medications using the Synthea Dataset.
2. Delineating the distinctions in the performances of various Neural TPPs models for these predictions.
3. Elucidating the pros and cons of Neural TPPs as a predictive model for events related to DR, coupled with insights into their plausible clinical applications.

A. Neural Temporal Point Processes (Neural TPPs)

In recent years, a significant advancement in the field of Temporal Point Processes has emerged in the form of Neural Temporal Point Processes (Neural TPPs). These models integrate the strengths of traditional TPPs with the expressive power of neural network architectures. By combining these two components, Neural TPPs can model more complex, nonlinear temporal dependencies in event data, which is often a limitation of traditional TPP models.

Neural TPPs leverage the flexibility of neural networks to learn intricate patterns directly from the data. This ability to learn complex relationships is especially beneficial when dealing with high-dimensional and noisy data, such as EHRs. These models can capture the intricate temporal patterns and relationships among various clinical events, and effectively incorporate additional covariates that may influence the timing of future events.

B. Potential for Diabetic Retinopathy Prediction

Within the DR sphere, Neural TPPs emerge as an enticing research and application prospect. The chronology and sequence of clinical events preluding DR onset or progression, such as fluctuating blood glucose levels and eye check-ups, are paramount. With their intrinsic ability to manifest these temporal patterns, Neural TPPs stand as a potent tool for DR event prediction.

For instance, a Neural TPP could discern that a swift series of high blood glucose levels, combined with irregular eye check-ups, augments the DR onset risk for a specific patient. Such insights become a treasure trove for medical professionals, empowering them to take pre-emptive actions to halt or lessen DR's advance. Moreover, as the prevalence of diabetes continues to rise globally, the significance of effective predictive tools like Neural TPPs cannot be understated. Their

application in real-world clinical settings could revolutionize early intervention strategies, minimizing the devastating impacts of DR and optimizing treatment plans for affected patients.

C. Study Contribution

This research offers a novel exploration into the application of Neural TPPs to forecast subsequent events related to Diabetic Retinopathy and associated medications using the Simulated Synthea Dataset. By emphasizing on Neural TPPs, this study not only delves into the capabilities of advanced temporal models but also enriches the broader discourse on harnessing advanced machine learning methodologies to address pressing healthcare predicaments.

II. BACKGROUND

A. Temporal Point Processes

Temporal Point Processes (TPPs) are a class of statistical models that are designed to characterize the randomness in the timing of a sequence of events. TPPs have a long history of application in various fields such as seismology, telecommunications, and finance due to their ability to model complex temporal patterns. More recently, they have garnered attention in the context of healthcare analytics, particularly for modelling event data in Electronic Health Records.

B. Mathematical Formulation

Temporal Point Process is defined by its conditional intensity function $\lambda(t)$, which represents the instantaneous event rate at time t, given the history of events up to time t. Formally, the conditional intensity function $\lambda(t)$ is defined as:

$$\lambda(t)= \lim_{\Delta t \to 0} \frac{P(\text{Event in } [t, t+\Delta t) | \text{History up to } t)}{\Delta t}$$

where $P(\cdot)$ represents the probability, and the history up to time t is denoted by the sequence of past event times

$$H_t=\{\{t_i : t_i < t\}.$$

C. Modelling Capabilities

TPPs are particularly suited for modelling event data with complex temporal dependencies. Unlike traditional time series models that require data to be aggregated into fixed intervals, TPPs can naturally handle events that occur at irregular and unpredictable times. This makes TPPs highly flexible and capable of capturing intricate temporal patterns in the data.

D. Relevance in Healthcare Analytics

In healthcare, the timing of clinical events, such as diagnoses, treatments, and laboratory tests, is often irregular and can carry significant predictive information. For example, the timing pattern of blood glucose measurements for a diabetic patient can be highly informative for predicting the onset of complications such as DR. TPPs, with their ability to model such irregular event data, have emerged as a promising tool for healthcare analytics. They offer a principled way to leverage the rich temporal information contained in EHRs to predict future clinical events, including disease onset, progression, and treatment outcomes.

E. Application to Diabetic Retinopathy Prediction

In the context of Diabetic Retinopathy, TPPs have the potential to model the sequence of clinical events leading up to the onset or progression of the condition. For instance, TPPs can potentially capture the temporal patterns in a patient's history of blood glucose levels, eye examinations, and other relevant clinical events, and use this information to predict the patient's risk of developing DR or experiencing a significant progression of the condition.

This study aims to rigorously assess the efficacy of Neural TPPs as a predictive model for DR, providing a comprehensive comparison with other established machine learning models using EHR data from the Full Synthea Dataset.

D. Neural Temporal Point Processes (Neural TPPs)

Neural Temporal Point Processes (Neural TPPs) are a recent innovation that integrates the traditional TPP model with modern neural network architectures. This integration allows Neural TPPs to capture more complex, nonlinear temporal dependencies in event data.

In addition to healthcare, Neural TPPs have demonstrated effectiveness in the context of Natural Language Processing (NLP). The event sequence in NLP tasks can be viewed as a temporal sequence, where each event represents a linguistic unit, such as a word or a sentence, associated with a timestamp. Neural TPPs can capture the intricate temporal dependencies between these linguistic events, which is crucial for various NLP tasks, including machine translation, text generation, and dialogue systems.

For instance, in machine translation, the sequence of words in a sentence can be modelled as a temporal sequence, and Neural TPPs can capture the dependencies between words based on their occurrence times. This allows Neural TPPs to generate more contextually appropriate translations by understanding the temporal dynamics of a sentence.

Encoder Architecture

The encoder is designed to process an input sequence of events (e.g., words in a sentence) and compress this temporal information into a fixed-size latent representation, often referred to as a context vector.

Embedding Layer: This layer transforms discrete event types (e.g., word tokens) into continuous embeddings. These embeddings are then used as inputs for subsequent layers.

Recurrent Neural Network (RNN) Layer: This layer processes the sequence of event embeddings over time, capturing the temporal dependencies in the input sequence. The final hidden state of this RNN serves as the context vector, which encodes the salient information from the input event sequence.

Decoder Architecture

Conditioned on the context vector produced by the encoder, the decoder is responsible for generating an output sequence of events (e.g., translated words in the target language) that are temporally coherent and contextually appropriate.

Conditional Intensity Function Estimation Layer: This layer takes the context vector (output by the encoder) as input and uses it to parameterize the conditional intensity function $\lambda(t)$ of the point process. This function governs the timing of future events in the output sequence. Event Generation Mechanism: Based on the estimated $\lambda(t)$, the decoder generates predictions

for the timing of future events. This often involves sampling from a distribution determined by λ(t), thereby producing a sequence of output events (e.g., words in a translated sentence) that respect the temporal dynamics encoded in the context vector.

This encoder-decoder architecture is particularly powerful for sequence-to-sequence tasks in NLP. It allows the Neural TPP model to learn complex temporal dependencies in an input sequence (e.g., a sentence in a source language) and use this knowledge to generate an output sequence (e.g., a translated sentence) that is both temporally coherent and contextually appropriate.

E. Advancements Over Traditional TPPs

Neural TPPs address a key limitation of traditional TPP models: their inability to capture complex, nonlinear temporal patterns. By leveraging the expressive power of neural networks, Neural TPPs can learn these complex patterns directly from the data, enabling more accurate and flexible modelling of event sequences.

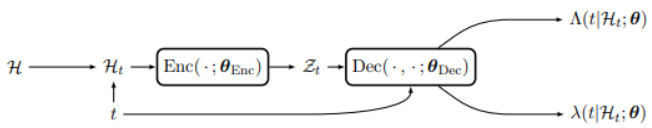


Figure 1: Encoder/decoder architecture of Neural TPPs. Given a query time t, the sequence H is filtered to the events H_t in the past of t. The encoder maps H_t to continuous representations $Zt = \{zi\}^{|Ht|}_{i=1} = Enc(H_t ; \theta_{Enc})$. Each zi can be considered as a contextualised representation for the event at ti . Given Zt and t, the decoder outputs $Dec(t, Z_t ; \theta Dec) \in R^M$ that the conditional intensity and conditional cumulative intensity are derived from without any learnable parameters.

III. Previous work

Temporal Point Processes (TPPs) have gained considerable traction in recent years, especially within the realm of healthcare analytics. Their suitability for modeling sequences of discrete events makes them particularly well-suited for analyzing Electronic Health Records (EHRs), which often comprise a series of time-stamped medical events. Such EHR data is rife with intricate temporal dynamics and dependencies, which traditional predictive models might overlook or inadequately capture. Neural TPPs, an evolution of traditional TPPs, have emerged at the forefront of this domain, offering a blend of temporal modeling with the flexibility and power of neural networks. Their ability to encapsulate intricate temporal patterns and dependencies allows them to predict future clinical events with a level of precision that was previously challenging to achieve. Neural TPPs' adaptability stems from their architecture, where the choice of encoders and decoders plays a pivotal role. The encoder captures the historical context, while the decoder predicts the timing and type of the next event, making the decoder's design a crucial determinant of the model's overall performance.Several researchers have delved into this area, exploring different encoder-decoder combinations, each with its strengths and limitations. The literature is replete with various architectures and methodologies, with each iteration aiming to refine the model's predictive accuracy, interpretability, and clinical relevance. This section aims to provide a comprehensive overview of these approaches, with a particular emphasis on the nuances of decoder selection, given its pronounced impact on model outcomes.

A. Closed Form Likelihood

The closed-form likelihood approach implies that each event contributes to the likelihood in a closed-form manner. A notable example is the Hawkes process, which models the conditional intensity in a way that results in a closed-form expression for the cumulative intensity function. Du et al. (2016) extended this approach by conditioning an exponential linear decoder on the output of a Recurrent Neural Network (RNN) encoder, termed RMTPP. While this model is limited in its ability to capture only exponential dependencies in time, it has been applied to various domains, notably in healthcare.

B. Analytic Conditional Intensity

In this approach, the conditional intensity is approximated by a neural network (NN) whose output is constrained to be positive. This approach does not require the cumulative intensity function to be in a closed form; instead, it can be approximated numerically. For example, Mei and Eisner (2017) employed Monte Carlo estimation given a sampling strategy. Zhu et al. (2020) also applied this strategy for spatio-temporal point processes, which jointly models times and labels, but only uses an embedding layer as an encoder.

C. Analytic Conditional Cumulative Intensity

This approach approximates the conditional cumulative intensity with a NN whose output is positive. The derivative of this function, computed using backpropagation, approximates the conditional intensity. Omi et al. (2019) used a Multi-Layer Perceptron (MLP) with positive weights to model a monotonic decoder and applied activation functions to ensure a positive output.

D. Attention based Neural TPPs for EHR Analysis

Enguehard et al. (2020) presented a comprehensive study on Neural Temporal Point Processes (Neural TPPs) for modeling Electronic Health Records (EHRs). They meticulously explored the choices of encoder and decoder architectures for Neural TPPs. While a Neural TPP encoder could be conveniently chosen from existing sequence models such as Recurrent Neural Networks (RNNs), they highlighted

that choosing a decoder was more challenging due to the integral calculations involved.

The authors categorized existing work based on the relationship between the conditional intensity λ (t) and the conditional cumulative intensity Λ(t), focusing on three approaches: Closed Form Likelihood, Analytic Conditional Intensity, and Analytic Conditional Cumulative Intensity.

In the Closed Form Likelihood approach, Enguehard et al. (2020) critically examined the RMTPP model by Du et al., which conditions an exponential linear decoder on the output of an RNN encoder, and they discussed its limitations, particularly its assumption that labels are conditionally independent of time given a history.

In the Analytic Conditional Intensity approach, they described how the conditional intensity can be approximated with a Neural Network (NN) whose output is positive, emphasizing the practical challenges of training such models.

In the Analytic Conditional Cumulative Intensity approach, they detailed how the conditional cumulative intensity can be approximated with a NN whose output and derivative are positive, referencing the work of Omi et al. (2019) who used a Multi-Layer Perceptron (MLP) with positive weights to model a monotonic decoder.

Enguehard et al. (2020) also made significant contributions to the field by addressing challenges related to the flexibility and interpretability of Neural TPPs in the context of EHRs. They proposed a novel Neural TPP model that aimed to overcome the limitations of existing models, especially when applied to the intricate and diverse temporal event data found in EHRs

In their results, Enguehard et al. (2020) demonstrated that their proposed Neural TPP model achieved superior performance in various metrics compared to traditional and existing Neural TPP models. They reported that their model was able to capture complex temporal dependencies more effectively and that it demonstrated improved predictive accuracy. Importantly, their model was designed to be highly interpretable, allowing for meaningful insights into the learned temporal dynamics, which is a critical aspect for applications in healthcare where understanding the model's decisions can be as important as the decisions themselves.

IV. MODELS

We delve into the architectures and components of the Neural Temporal Point Processes (Neural TPPs) employed in this study for the prediction of Diabetic Retinopathy using the Synthea dataset. Following the significant insights from Enguehard et al. (2020), we adapt and implement various Neural TPP models, emphasizing the encoder and decoder components.

A. Encoder Architecture

The encoder is responsible for processing the historical events up to a query time t, mapping these events to a continuous representation. In this study, we explore the use of Recurrent Neural Networks (RNNs) as the encoder, in line with the approach proposed by Enguehard et al. (2020). This enables us to capture and learn the intricate temporal dependencies present in EHR data.

1. Recurrent Neural Networks (RNNs)

RNNs are employed as encoders due to their ability to model sequential data effectively. They process the events in a patient's history sequentially, maintaining a hidden state that serves as the memory of the network. This hidden state is updated as each new event is processed, allowing the RNN to capture the temporal dependencies between events in a patient's EHR.

B. Decoder Architecture

The decoder, which is more challenging to design due to the integral calculations involved, plays a crucial role in the prediction task. It takes as input the continuous representations generated by the encoder, along with the query time t, and outputs parameters from which the conditional intensity and conditional cumulative intensity are derived.

1. Closed Form Likelihood Approach

In this approach, we employ models that have a closed-form expression for the likelihood. The Hawkes Process, as well as extensions such as the RMTPP model as explored by Enguehard et al. (2020), fall under this category.

2. Analytic Conditional Intensity Approach

Following Enguehard et al. (2020), we explore decoders that approximate the conditional intensity with a positive output Neural Network. This network is designed such that the time integral of its output needs to be approximated numerically. The final activation function is selected to be positive, in line with the methodologies suggested by Enguehard et al. (2020).

3. Analytic Conditional Cumulative Intensity Approach

In this approach, we design the decoder to approximate the conditional cumulative intensity with a Neural Network whose output is positive, as inspired by Enguehard et al. (2020). The derivative of this output approximates the conditional intensity and is ensured to be positive.

C. Base intensity

An integral component of the model architectures discussed is the base intensity. This fundamental element characterizes the inherent event rate for the point process. The base intensity isn't contingent on past events, making it a foundational aspect of the model, laying the groundwork for further refinements and adjustments based on historical data.

The base intensity can be constant, effectively serving as a baseline. Alternatively, it can be dynamic, adapting to the specific temporal context. In the models explored by Enguehard et al. (2020), the base intensity exhibits variations depending on the specific model architecture. Some models deploy a constant base intensity throughout, while others allow it to evolve based on the event history or other contextual factors.

In this study, we will be deploying and evaluating a suite of models, each tailored to discern the intricate patterns inherent in the realm of Diabetic Retinopathy. The models slated for examination include:

GRU-Cond-Poisson: A model that integrates the Gated Recurrent Unit (GRU) with a conditional Poisson process.

GRU-LNM: A model that joins the GRU with a Log-Normal Mixture.

GRU-MLP-CM: Incorporating a Multi-Layer Perceptron (MLP) with a conditional mean approach, this model leverages the GRU for sequence modelling.

GRU-RMTPP: This model couples the GRU with the RMTPP process.

GRU-RMTPP-Poisson: A fusion of the GRU-RMTPP model with a Poisson process.

GRU-SA-CM: An amalgamation of the GRU with a self-attention mechanism and a conditional mean approach.

GRU-SA-MC: Incorporating the self-attention mechanism with a Monte Carlo approach, this model uses the GRU and a softmax layer.

SA-Cond-Poisson: A model that integrates the self-attention mechanism with a conditional Poisson process.

SA-LNM: This model synergizes the self-attention mechanism with a Log-Normal Mixture.

SA-MLP-CM: A fusion of the self-attention mechanism with an MLP, using a conditional mean approach.

SA-RMTPP: Pairing the self-attention mechanism with the RMTPP process.

SA-RMTPP-Poisson: This model combines the SA-RMTPP with a Poisson process.

SA-SA-CM: A model that employs dual self-attention mechanisms combined with a conditional mean approach.

SA-SA-MC: This model utilizes dual self-attention mechanisms with a Monte Carlo approach.

Each of these models has been meticulously crafted to capture the temporal dependencies and event patterns characteristic of Diabetic Retinopathy data. Our endeavour is to gauge their efficacy, understand their strengths and limitations, and ultimately discern the most potent model for predicting the progression of Diabetic Retinopathy.

1. Synthea Dataset Features

The Synthea dataset was developed using real-world medical data standards and modules that simulate medical conditions, thus ensuring that the synthetic data is realistic and representative of actual healthcare scenarios. Specifically, for our study's focus on Diabetic Retinopathy, the dataset includes:

Patient details: Patient age and other encounter information.

Clinical Events: Detailed records of clinical events related to diabetes and its complications, including blood glucose readings, eye examinations, and treatment interventions.

Temporal Patterns: Time stamps associated with each event, enabling us to model and analyse the sequence and timing of clinical events leading up to the onset or progression of Diabetic Retinopathy.

2. Data Pre-processing

Before feeding the data into our models, several pre-processing steps were undertaken:

Filtering: Given the comprehensive nature of the Synthea dataset, we filtered out data unrelated to diabetes and its complications, focusing solely on patients with a history of diabetes and associated clinical events.

Feature Engineering: Derived features were created based on domain expertise, such as the duration between consecutive blood glucose readings and the frequency of eye examinations.

3. Data Splitting

The dataset was divided into training, validation, and test sets. The training set was used to train the models, the validation set for hyperparameter tuning and model selection, and the test set for final evaluation. The split ensured that patient records were not overlapping across sets, maintaining the integrity and independence of the evaluation.

B. Hyperparameters

In our study, the model's performance was intricately tied to the careful selection and tuning of hyperparameters. It is crucial to understand the configuration parameters employed to comprehend the model's behaviour and outcomes fully. Below, we outline the specific hyperparameters and settings chosen for our model across different datasets:

Table 1. Properties of dataset used for evaluation.

Dataset	# classes	Task type	# events	Avg. length	Size			
					Train	Valid	Test	Batch
Synthea (Diabetic Retinopathy)	52	Multi-label	65,175	126	520	165	165	64

V. EVALUATION

A. Dataset overview

The data used in this study is sourced from the Synthea Dataset, a synthetic and comprehensive Electronic Health Records (EHR) dataset generated by human expert curated Markov process. This dataset provides a unique opportunity for researchers as it captures a wide range of information, including diagnoses, treatments, and temporal patterns of healthcare encounters.

Encoder and Decoder Structure:

- Utilized a single layer for both.
- Layer dimensions:
- 64 units for the Synthea dataset.

Batch Size:

- Adjusted based on dataset.

- Modifications made due to GPU memory constraints.

Optimization and Learning Rate:

- Used the Adam optimization technique (Kingma and Ba, 2017).
- Incorporated the Noam learning rate adjustment method (Vaswani et al., 2017).

Table 2: Evaluation on task

Encoder GRU	ROC-AUC	NLL/time
Decoder		
CP	**0.859**	90.252
RMTPP	0.759	126.499
LNM	0.768	**25.644**
MLP-CM	0.504	86.553
MLP-MC	0.702	46.67
ATTN-CM	0.506	83.111
ATTN-MC	0.793	36.704
Encoder SA		
CP	0.828	92.936
RMTPP	0.728	113.703
LNM	0.752	32.566
MLP-CM	0.504	86.135
MLP-MC	0.697	48.90
ATTN-CM	0.506	83.849
ATTN-MC	**0.776**	**43.491**

- Scheduler specifications:
- Peak learning rate of 0.01.
- 10 epochs for the warm-up phase.

Training Approach:

- All models trained using the Maximum Likelihood Estimation (MLE) method.
- Early termination applied to prevent overfitting.
- Termination triggered after 100 epochs without improvement.

C. Metrics

In our study, a combination of the ROC-AUC score and NLL was utilized to holistically evaluate the model's performance.

NLL (Negative Log Likelihood) is a metric used to evaluate how well a probabilistic model predicts a given outcome. A lower NLL indicates that the model's predicted probabilities are closer to the true outcomes, making it a more reliable model. In the context of Temporal Point Processes, the NLL measures the model's ability to predict both the timing and type of the next event. When normalized by time (NLL/time), it provides a continuous measure of the model's prediction quality over time, enabling us to evaluate its proficiency across different time intervals.

The ROC-AUC score is a performance measurement for classification problems. It evaluates a model's ability to distinguish between classes, with a focus on its discrimination threshold. An AUC score of 1 indicates perfect classification, while an AUC score of 0.5 suggests that the model is no better than random guessing. In multi-label datasets, the weighted ROC-AUC score provides a consolidated metric, giving us insights into how well the model predicts multiple labels simultaneously.

The emphasis was on the weighted ROC-AUC score, which gauged the model's ability to predict the imminent event's label(s) based on its timing and the preceding events' sequence. Furthermore, the Negative Log Likelihood (NLL) normalized by time, or NLL/time, was employed. This metric offered insights into the model's capacity to predict not only the label(s) of the succeeding event but its precise timing as well, using the context of prior events.

Both metrics are crucial, with the ROC-AUC score focusing on label prediction accuracy and the NLL/time addressing both the event's label and its timing. A robust model should ideally excel in both metrics for a comprehensive and effective evaluation.

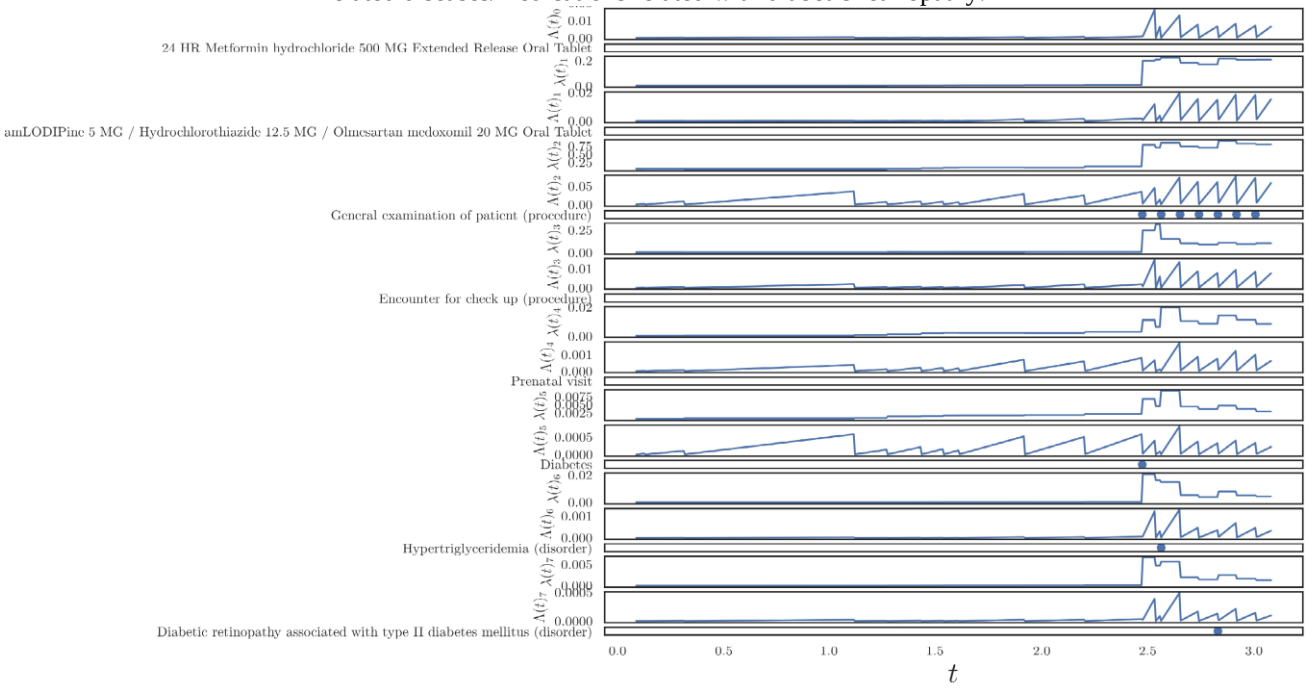
VI. Results

The performance of several models on the Synthea dataset is presented in this section. Each model was evaluated using two primary metrics: the Receiver Operating Characteristic Area Under the Curve (ROC-AUC) and the Negative Log Likelihood (NLL).

Based on the table presented above, we can derive insights into the performance of different models on the Synthea dataset on the ROC-AUC Weighted score and the NLL/Time score. The former provides insights into a model's classification capabilities, especially when considering the prevalence of each class in the dataset. The latter, on the other hand, sheds light on the model's efficiency in predicting the timing of the next event while maintaining a low log-likelihood loss. The plots show the intensity functions on several labels applied to the same EHR. We can see the intensity functions in different models affecting the NLL and spikes on the related diseases. Leading the ROC-AUC Weighted metric is the GRU-Cond-Poisson model, registering an impressive score of 0.859. This score not only underscores the model's superior ability to classify but also

its prowess in prioritizing the importance of different classes. Following closely are the SA-Cond-Poisson and GRU-SA-MC models, with scores of 0.828 and 0.793, respectively. These models, too, exhibit the classification capabilities. However, it's worth noting that certain models like GRU-MLP-CM, GRU-MLP-CM, and SA-MLP-CM have scores hovering around the 0.50 mark. This suggests there's ample room for improvement in their weighted classification capabilities.

Figure 2: Intensity functions on several labels applied to the same EHR. We can see the intensity functions and spikes on related diseases/medications related with diabetic retinopathy.



Switching our attention to the NLL/Time metric, the GRU-LNM model stands out distinctly with a score of 25.644. This model's efficiency in predicting the timing of the next event, coupled with its ability to maintain a low log-likelihood loss, makes it a top contender. The SA-LNM model, with a score of 32.566, is another model that mirrors the efficiency of the GRU-LNM model. However, not all models fared well on this front. The GRU-RMTPP, for instance, recorded the highest NLL/Time score of 126.499. This elevated score suggests potential challenges this model might face in predicting event timings with accuracy, especially when compared to its counterparts.

In conclusion, while several models showcased satisfactory performances on specific metrics, the ideal model for practical applications would be one that strikes a balance between classification accuracy and prediction efficiency. In this context, both the GRU-Cond-Poisson and GRU-LNM models emerge as strong contenders, each excelling in one of the two key metrics. As advancements in the field of healthcare analytics continue, refining these models further could pave the way for more accurate and efficient predictive tools in clinical settings. Also, the plot represents the predicted intensities of various events (diseases/medications) related to diabetic retinopathy over time and we can observe:

Disease Intensities Over Time: The rising and falling curves indicate the predicted intensity or likelihood of each disease/medication event occurring at a given time. Peaks in the plot suggest periods of high intensity or risk for a particular event.

Comparison with Ground Truth: The plot also contains actual observed events, and we can compare the predicted intensities from the Neural TPPs model with the ground truth to assess the model's accuracy.

Interplay of Events: The concurrent rise in intensity of multiple events suggests a correlation or causative relationship between them. For instance, if the intensity of a medication rises shortly after the intensity of a disease peaks, it might suggest that the medication is commonly prescribed following the onset or exacerbation of that disease.

Clinical Implications: From a clinical perspective, periods of high predicted intensity for adverse events (e.g., severe stages of diabetic retinopathy) would be of utmost concern. These periods might indicate the need for increased monitoring, early interventions, or changes in treatment strategy.

Model Insights: Peaks or troughs in the plot might also offer insights into the model's behavior. For instance, unexpected peaks might suggest areas where the model sees patterns in the data that aren't immediately obvious to human observers. Conversely, discrepancies between the model's predictions and actual events could highlight areas for model refinement.

The visual representation of our model's predictions, as presented in the plot, offers profound insights into the progression and interplay of conditions associated with diabetic retinopathy. A clear trajectory can be obtained from the onset of diabetes, progressing to the development of

Hypertriglyceridemia, and culminating in the manifestation of Diabetic Retinopathy. This sequential progression underscores the intertwined nature of these conditions and reaffirms the importance of early interventions at the diabetes stage to potentially prevent or delay subsequent complications. Interestingly, the plot also hints at a notable correlation between neonatal visits and diabetes. While this might seem counterintuitive at first glance, it suggests that events or interventions during the neonatal period could have long-term implications on an individual's predisposition to diabetes. This observation warrants further investigation, potentially opening new avenues for early diabetes risk assessment and preventive strategies. Furthermore, the spikes in the intensities of medications associated with diabetes highlight the recurrent nature of treatment interventions. These peaks might correspond to periods of exacerbated symptoms or the progression of the disease, indicating the cyclical need for medical interventions.

In conclusion, the plot not only validates some established medical understandings but also brings forth new observations that could be instrumental in reshaping diabetic care strategies. It underscores the value of utilizing advanced predictive models like Neural Temporal Point Processes in healthcare, offering a holistic view of patient trajectories and enabling proactive medical interventions.

In summary, the plot serves as a visual representation of the Neural TPPs model's predictions regarding the progression of diabetic retinopathy and related events. By interpreting these results in the context of the broader project and clinical knowledge, one can gain valuable insights into the nature and progression of diabetic retinopathy and its associated conditions and treatments.

VII. Limitations of study

The study, while pioneering in its approach, does come with specific limitations, particularly concerning the utilization of the Synthea dataset. Although Synthea provides a robust synthetic patient dataset, capturing the complexities and nuances of real-world patient data is inherently challenging. The generated synthetic data may not encapsulate the full range of variabilities and intricacies witnessed in genuine patient histories, potentially leading to an over-optimistic assessment of our model's efficacy. Temporal dynamics, pivotal in healthcare predictions, may manifest differently in synthetic datasets compared to real-world patient timelines. This disparity raises concerns about how well our model might generalize when confronted with genuine data. Another significant concern is the potential biases introduced during synthetic data generation. Such biases, whether subtle or pronounced, can skew our model's predictive capabilities and lead to inaccurate interpretations. Moreover, the application of our model to real-time health data, which is continuously evolving and rich in its informational content, presents an entirely different set of challenges. The dynamism and intricacies of real-time health data might expose certain limitations in our model, not evident when tested against static synthetic datasets. Furthermore, validation concerns arise when relying heavily on synthetic datasets. For our model to be truly considered effective and clinically relevant, rigorous validation against actual patient data sets is indispensable. This validation will ensure that our model's predictions remain consistent and accurate, irrespective of the data's origin.

Considering these limitations, while synthetic data offers invaluable insights and a safer environment for preliminary research, it's paramount to approach results with a balanced perspective. Emphasis should be placed on subsequent studies that integrate and validate our findings against real-world datasets, ensuring the model's adaptability and reliability in varied scenarios.

VIII. Conclusion

In our exploration to leverage Neural Temporal Point Processes for forecasting events related to Diabetic Retinopathy and its associated medications, we rigorously assessed a myriad of models on the Synthea simulated dataset. The results yielded diverse performances, with some models demonstrating unparalleled prowess in classification and others excelling in timing predictions. Specifically, the GRU-Cond-Poisson and GRU-LNM models emerged as front-runners in their respective metrics, thus presenting themselves as viable candidates for future applications. The visualized intensity functions, derived from our models, offer a panoramic view of the dynamic landscape of events associated with diabetic retinopathy. This visual analysis not only corroborates existing medical knowledge—like the progression from diabetes to Hypertriglyceridemia, followed by Diabetic Retinopathy—but also unveils novel observations. The intriguing correlation between neonatal visits and diabetes, for instance, introduces a compelling avenue for future research, highlighting the potential long-term ramifications of neonatal interventions on diabetes susceptibility. Moreover, the recurrent peaks in medication intensities underscore the cyclical nature of diabetic treatment, possibly reflecting periods of symptom exacerbation or disease progression. This insight is invaluable, emphasizing the importance of continuous monitoring and timely interventions to pre-empt complications. However, as with all research, our study is not without limitations. The utilization of the Synthea simulated dataset, while invaluable for its comprehensive nature, introduces concerns of validation and potential biases. Real-world data, with its inherent complexities and nuances, may yield different results. Furthermore, while our models have showcased efficacy in a simulated environment, their robustness in real-world settings remains to be tested. Future endeavours should focus on juxtaposing our findings with real-world data, ensuring a holistic validation and mitigating biases.

In summation, our research offers a promising step forward in the realm of predictive analytics for Diabetic Retinopathy and its associated events. The insights gleaned not only fortify existing knowledge but also pave the way for novel preventive and intervention strategies. As we continue to navigate the intricate corridors of healthcare analytics, studies like ours underscore the importance of precision, adaptability, and foresight in shaping a proactive and patient-centric healthcare paradigm.

ACKNOWLEDGMENT

I extend my heartfelt thanks to Dr. Michael Barnes for his profound guidance and thorough critique of this document. I owe a deep sense of gratitude to Alisha Angdembe for her steadfast support and specialized expertise that significantly influenced this research. My sincere appreciation is also directed towards the dedicated team at AI Multiply for their insights and guidance.

REFERENCES

[1] Enguehard, J., Busbridge, D., Bozson, A.J., Woodcock, C. and Hammerla, N.Y. (2020). Neural Temporal Point Processes For Modelling Electronic Health Records. arXiv (Cornell University). doi:https://doi.org/10.48550/arxiv.2007.13794.

[2] Hongyuan Mei and Jason M Eisner. The Neural Hawkes Process: A Neurally SelfModulating Multivariate Point Process. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, Advances in Neural Information Processing Systems 30, pages 6754–6764. Curran Associates, Inc., 2017.

[3] Shixiang Zhu, Minghe Zhang, Ruyi Ding, and Yao Xie. Deep Attention Spatio-Temporal Point Processes. arXiv:2002.07281 [cs, stat], February 2020.

[4] Hongyuan Mei and Jason M Eisner. The Neural Hawkes Process: A Neurally SelfModulating Multivariate Point Process. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, Advances in Neural Information Processing Systems 30, pages 6754–6764. Curran Associates, Inc., 2017.

[5] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. arXiv:1412.6980 [cs], January 2017.