Using Deep Learning for Individual-Level Predictions of Adherence with Growth Hormone Therapy

Matheus ARAUJO [a,1], Paula VAN DOMMELEN [b],
Ekaterina KOLEDOVA [c] and Jaideep SRIVASTAVA [a]

[a] *University of Minnesota, Minneapolis, MN, USA*
[b] *The Netherlands Organization for Applied Scientific Research
TNO, Leiden, The Netherlands*
[c] *Endocrinology Global Medical Affairs,
Merck KGaA, Darmstadt, Germany*

Abstract. The problem of consistent therapy adherence is a current challenge for health informatics, and its solution can increase the success rate of treatments. Here we show a methodology to predict, at individual-level, future therapy adherence for patients receiving daily injections of growth hormone (GH) therapy for GH deficiency. Our proposed model is able to generate predictions of future adherence using a recurrent neural network with adherence data recorded by easypod™, a connected autoinjection device. The model was trained with a multi-year long dataset with 2500 patients, from January 2007 to June 2019. When testing, the model reached an average sensitivity of 0.70 and a specificity of 0.88 per patient when predicting non-adherence (<85%) periods. When evaluated with thousands of therapy segments extracted from a test set, our model reached an AUC-PR score of 0.79 and AUC-ROC of 0.90; both metrics were consistently better than traditional approaches, such as simple average model. Using this model, we can perform precise early identification of patients who are likely to become non-adherent patients. This opens a path for healthcare practitioners to personalize GH therapy at any stage of the patients' journey and improve shared decision making with patients and caregivers to achieve optimal outcomes.

Keywords. Deep learning; growth hormone therapy; therapy adherence

1. Introduction

Growth hormone (GH) treatment for children is an efficacious treatment for various growth disorders but consistent adherence to the prescribed dose regimen is fundamental for achieving optimal growth [1,2]. The use of eHealth-based ecosystems provided a new mechanism to monitor adherence automatically and perform real-time data transmission to healthcare professionals (HCPs) [3]. This enables HCPs to detect low adherence and provide a personalized individual-patient intervention with the ultimate goal of positive population-based health outcomes. Data from clinical studies using an eHealth ecosystem including a connected injection device showed that most patients maintained

---

[1] Corresponding Author, Matheus Araujo, University of Minnesota, Minneapolis, MN, USA; E-mail: arauj021@umn.edu

an adherence to GH therapy of ≥85% of the prescribed dose for >3 years; however, the high number of non-adherent patients deserves careful investigation [2].

In this study we focus on individual- rather than group-level analysis for personalized interventions [4] to identify patients who are likely not to fully adhere to the treatment in the near future. We show how to train a recurrent neural network (RNN) with therapy segments from different time periods extracted from multiple patients. We also perform a feature importance analysis to understand how the information from each day contributes to the model's final decision.
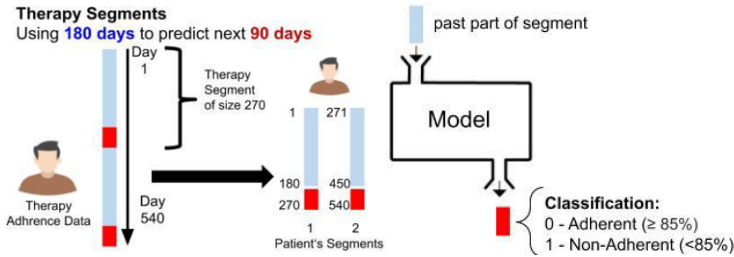
2. Methodology

Figure 1. Diagram showing how we extract therapy segments from the therapy adherence data of a patient.

Therapy segments allow us to specify different periods of the adherence data to be modeled by the deep learning algorithm and predict future adherence. Figure 1 shows how these segments are created, starting from the left, from top downwards, with the adherence 1d-vector of size 540, where each value represents a daily adherence ratio (100% x injected dose/prescribed dose). The therapy segment has two parts: the past (in blue), used as input for the model, and the future (in red), where we extracted the classification of the segment. We used 180 days of data to predict the next 90 days of therapy adherence. Thus, in the middle, we show the two therapy segments extracted from a patient with 540 days of therapy adherence data.

On the right of Figure 1, we illustrate how the model works. As input, it receives the patient's past information and outputs the classification of the future part of the therapy segment. Suppose the patient has an average adherence ≥85% (missed at most 1 injection/week [5]) in the future region of the segment. In that case, the perfect model should always output 0 (adherence) given the past region as input. Otherwise, if the patient has an average adherence <85% in the future region, the perfect model should always predict 1. To train and evaluate our model, we used data from 2500 patients divided randomly into three sets: train, validation, and test. The reported results are only based on the patients' data in the test set that was never looked at by the model during the training phase. While training, we used the train and validation sets to find the best parameters for our deep learning RNNs. The best configuration found was to use a single layer RNN using 16 gated recurrent unit cells followed by an output dense layer with two neurons with softmax activation. Due to class imbalance towards adherence, we used cost-sensitive learning to force the model to give more attention to non-adherence segments [6]. This is possible by setting the class weight learning parameter. The adherence class received weight 1, and the non-adherence class received weight 3.31, which is the imbalance ratio on the training set. We use Adam optimizer with a learning

rate of 0.0001 and batch size 16. We used the AUC-PR (area under the curve-precision recall) score on the validation set to be a reference for our early stopping mechanism, which allowed our model to be trained over 31 epochs. To implement the model, we used the Keras framework[2] backed with Python's Tensorflow library version 2.2.0[3].

To test the model performance, we compared our deep learning model with a simple average approach. This comparison reinforces a non-trivial pattern finding in the model, leveraged by a more complex artificial intelligence (AI) approach. The simple average computes the mean over therapy adherence in the past part of the segment and uses the value to predict future adherence. To compute the metrics of the result section, we must order therapy segments based on how far they are from the therapy classification threshold (e.g. segments close to 0% of adherence have close to 100% probability to non-adhere). Thus, we apply a sigmoid function on the absolute difference between mean adherence and the classification threshold of 85%.

3. Results

We analyzed a sample of 2500 patients (56% male), from which we extracted 5199 segments where the patients adhered to the therapy and 1555 segments in which the patients did not adhere to the therapy. We randomly assigned 1361 patients (3643 segments) for the training set, 166 patients (517 segments) for the validation set, and 973 (2594 segments) for the test set. The mean starting age of the therapy was 10.2 years, with a standard deviation of 3.1 years (min. 4 and max. 18 years).
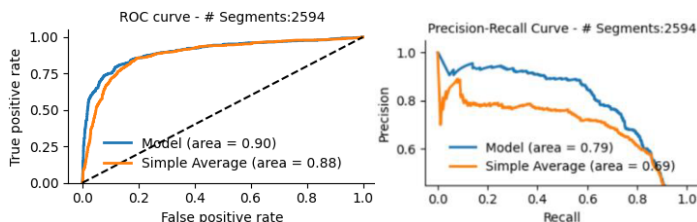

Figure 2. ROC curve (left) and PR-curve (right) when evaluating therapy segments of the test set.

We evaluated the predictions of our deep learning model on the 2594 therapy segments extracted from the patients in the test set. On the left of Figure 2, we show the receiver operating characteristic (ROC) curves for both deep learning model and simple average. The deep learning model had an AUC score of 0.90 versus 0.88 from the simple average. On the right of Figure 2, we show the PR curve, presenting the trade-off between recall (sensitivity) and precision (positive predictive value). The PR-curve is considered to be the more suitable metric for imbalanced datasets [7]. Our model had an AUC-PR score of 0.79 versus 0.69 from the simple average.

---

[2] Keras Framework. https://keras.io.
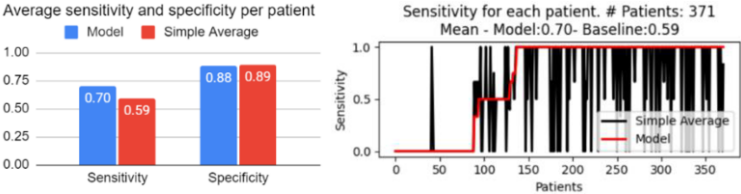[3] Tensorflow. https://tensorflow.org.

Figure 3. Aggregated average sensitivity and specificity per patient (left) and sensitivity per patient (right).

Patients can have one or more therapy segments to be evaluated. Thus, we also computed the aggregated sensitivity and specificity to understand our approach's performance per patient (instead of per segment) (Figure 3). On the left, we show that our model's sensitivity was 0.70, higher than the 0.59 of a simple average, maintaining almost the same specificity. On the right, we investigate the differences between both models, comparing their sensitivity score per patient. Since we can compute sensitivity only for patients with at least one therapy segment, these values are based on 371 patients from the test set. When the red line is above the black line, our model was better than the simple average baseline on sensitivity for the patient along the x-axis. Overall, our model had a higher or equal sensitivity for 359 out of the 371 patients.
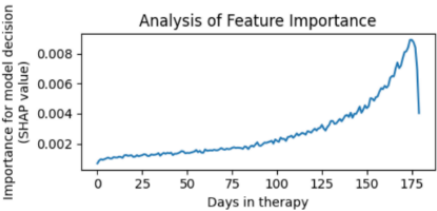

Figure 4. Analysis of absolute SHAP values for each day in the therapy segment.

Figure 4 shows how each of the 180 values used as input to the model impacts the model decision. These are the average absolute SHAP (SHapley Additive exPlanations) values for each day in the segment [8]; SHAP values are higher approaching Day 180 (more recent) in the segment, with a sharp drop on the last two days.

4. Discussion

The comparison between our deep learning model and a simple average model justifies the use of AI for the proposed task of predicting, at any time in the therapy, the future adherence of a patient. When we compare each model's ROC curve, we see the AI model consistently outperforms a simple average and the PR-curve, in which a 0.70 value on recall gives observed precision of 0.77 in our model, is better than 0.69 from the simple average with the same 0.70 recall. When looking at per patient performance, we observe that the deep learning model is better in terms of sensitivity. Our model has 0.11 higher sensitivity, at the cost of only 0.01 of specificity, versus the simple average approach.

As observed by the analysis of the importance of the values used as input to the model, the information that dictates the model output is mostly, but not entirely, defined by the last 30 days of the segment. Further investigation is needed to understand why

there is a sharp drop in importance on the two most recent days. It might suggest that the last day(s) of use may not be that important for future prediction, but the pattern of routine, i.e. the combination of various days, is more relevant. Furthermore, the model does not ignore any of the 180 values used as input, considering all information given. The application of this model is straightforward. We can collect the last 180 days of a patient's therapy history and predict how their coming 90 days will be classified at any moment in time. By sorting the patients based on the probability of non-adherence generated by the model, we can build a priority list for earlier and precise intervention.

5. Conclusions

This work presents a novel application that uses a deep learning approach evaluating a series of 180 daily recorded values of adherence to predict, at the individual level, the probability of a patient not adhering to their GH therapy during the next 90 days. We compared our approach to a simple average model and showed that we could have a higher sensitivity at very little cost of specificity. Overall, using both AUC-ROC and AUC-PR metrics, our proposed model is better than the simple average. We aim to integrate the proposed model into the physician interface for the easypod™ device, which may improve engagement and adherence to optimize treatment outcomes.

Conflicts of interest

MA and PvD have consultancy agreements with Merck KGaA. EK is an employee of Merck KGaA and holds shares in the company.

References

[1] Child J, Davies C, Frost K, et al. Managing paediatric growth disorders: integrating technology into a personalised approach. J Clin Res Pediatr Endocrinol. 2020 Sep;12(3):225-32.
[2] van Dommelen P, Koledova E, Wit JM. Effect of adherence to growth hormone treatment on 0-2 year catch-up growth in children with growth hormone deficiency. PLoS One. 2018 Oct;13(10):e0206009.
[3] Koledova E, Stoyanov G, Ovbude L, et al. Adherence and long-term growth outcomes: results from the easypod™ connect observational study (ECOS) in paediatric patients with growth disorders. Endocr Connect. 2018 Aug;7(8):914-23.
[4] Malik S, Moloney C, Koledova E, et al. Designing a personalized digital patient support program for patients treated with growth hormone: key design considerations. J Med Internet Res. 2020 Jul;22(7):e18157.
[5] Cutfield WS, Derraik JG, Gunn AJ, et al. Non-compliance with growth hormone treatment in children is common and impairs linear growth. PLoS One. 2011 Jan;6(1):e16223.
[6] Elkan C. The foundations of cost-sensitive learning. In: IJCAI'01: Proceedings of the 17th International Joint Conference on Artificial Intelligence – Volume 2. Morgan Kaufmann Publishers Inc. 2001 Aug;973-8.
[7] Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One. 2015 Mar;10(3):e0118432.
[8] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Advances in neural information processing systems. 2017 p. 4765-74.