Proceedings of Machine Learning Research 193:326–342, 2022 Machine Learning for Health (ML4H) 2022

Predicting Attrition Patterns from Pediatric Weight Management Programs

Hamed Fayyaz fayyaz@udel.edu
University of Delaware, Newark, DE, USA

Thao-Ly T. Phan ThaoLy.Phan@nemours.org
Nemours Children's Health, Wilmington, DE, USA

H. Timothy Bunnell Tim.Bunnell@nemours.org
Nemours Children's Health, Wilmington, DE, USA

Rahmatollah Beheshti rbi@udel.edu
University of Delaware, Newark, DE, USA

Abstract

Obesity is a major public health concern. Multidisciplinary pediatric weight management programs are considered standard treatment for children with obesity who are not able to be successfully managed in the primary care setting. Despite their great potential, high dropout rates (referred to as attrition) are a major hurdle in delivering successful interventions. Predicting attrition patterns can help providers reduce the alarmingly high rates of attrition (up to 80%) by engaging in earlier and more personalized interventions. Previous work has mainly focused on finding static predictors of attrition on smaller datasets and has achieved limited success in effective prediction. In this study, we have collected a five-year comprehensive dataset of 4,550 children from diverse backgrounds receiving treatment at four pediatric weight management programs in the US. We then developed a machine learning pipeline to predict (a) the likelihood of attrition, and (b) the change in body-mass index (BMI) percentile of children, at different time points after joining the weight management program. Our pipeline is greatly customized for this problem using advanced machine learning techniques to process longitudinal data, smaller-size data, and interrelated prediction tasks. The proposed method showed strong prediction performance as measured by AUROC scores (average AUROC of 0.77 for predicting attrition, and 0.78 for predicting weight outcomes).

Keywords: Childhood obesity, Attrition, Weight trajectories, Transfer learning, Multi-task learning, Deep learning

1. Introduction

Despite extensive efforts to fight childhood obesity, it remains a major public health concern worldwide. In the United States, the prevalence of obesity in children and adolescents aged 2-19 years in 2017-2020 was 19.7% and affected about 14.7 million and their families (Akinbami et al., 2022). Childhood obesity increases the risk of childhood morbidity, diabetes, cardiovascular disease, and cancer, predominantly as a result of a substantially greater risk of obesity in adulthood (Ogden et al., 2014). Multi-disciplinary clinical weight management programs (WMPs) are recommended for children with obesity who fail to improve with management in the primary care setting, but these programs often require a moderate to high intervention dose delivered over an extended period (Ball et al., 2021). Children and their families who attend more intervention sessions and remain enrolled in care for longer periods achieve the greatest improvements in weight and health (Wilfley et al., 2017). However, due to a variety of reasons, such as dissatisfaction with the intervention progress or logistical issues with attending the programs, many families (as much as 80% (Ball et al., 2021)) discontinue attending WMPs prematurely; a phenomenon referred to as "attrition." Other potential predictors of attrition include psychological (Altamura et al., 2018; Jiandani et al., 2016), sociodemographic and anthropometric (Ponzo et al., 2020) factors, as well as the initial weight-loss success (Perna et al., 2018), demonstrating a complex problem with many factors involved. Besides not treating the disease effectively, a failed weight loss attempt may also lead to frustration, discouragement, and learned helplessness (Ponzo et al., 2020). Attrition can also be challenging for healthcare systems needing to ensure the effective delivery of their services and efficient utilization of their often limited resources. Prior studies have explored the variables associated with attrition from pediatric WMPs (Jelalian et al., 2008; Pit-ten Cate et al., 2017; Skelton et al., 2011); however, most of them have used statistical analysis techniques with a lower prediction performance.

In this study, we present a customized predictive model to address the current limitations in the field and apply our method to a large dataset that we have collected. Our dataset represents four pediatric WMP sites within Nemours Children's Health (a large pediatric health system in the US). Specifically, we present a deep (neural network) model with two separate components for analyzing the static and dynamic input features, extracted from the electronic health records (EHRs) of children attending the WMPs. To improve the overall predictive performance of our models, we use a "multi-task learning" approach by combining two prediction tasks (predicting attrition and weight outcomes) such that one model predicts two values for these tasks. In our study, attrition prediction refers to predicting the time of the last visit (number of weeks after the baseline visit), and the weight outcome prediction task refers to predicting the change in the BMI percentile (BMI%) of patients at their last visit of the WMP. Besides multi-task learning, the presented model also follows a "transfer learning" design, by training on various lengths of observation and prediction windows (Ball et al., 2021), and then fine-tuning on the final target task to report the desired outcomes. The main contributions of our study are:

- We have collected one of the largest datasets dedicated to studying attrition from pediatric WMPs (4,550 children from four WMP sites, including EHR data linked to additional lifestyle and psychosocial factors).
- We present a deep model for predicting when attrition occurs and the patients' BMI% change at that time. We use a multi-task and transfer learning approach to improve the performance of our model, facilitating its deployment in real settings without large training data. Compared to the existing studies, one major advancement of our work is integrating the temporal information (including body weight trajectories) with other demographic and cross-sectional information for the patients, enabling our model to achieve better results.
- To make the findings more actionable, we identify the top predictors for attri-

tion patterns and study their clinical relevance.

2. Related Work

Attrition patterns have been studied in many healthcare domains. Studying attrition is related to studying other healthcare problems such as visit attendance and treatment adherence in clinical settings. Attendance prediction generally aims to predict whether a patient will show up for a scheduled appointment or not, and has been used to identify the important predictive factors in various fields such as rehabilitation (Sabit et al., 2008; Hayton et al., 2013), psychiatric care (Mitchell and Selmes, 2007), and primary care (Giunta et al., 2013; Kheirkhah et al., 2015). On the other hand, adherence prediction aims to predict whether a patient will be compliant with their treatment plan (e.g., taking the prescribed medicines). Treatment and medication adherence have been studied for conditions such as tuberculosis (Killian et al., 2019), heart failure (Son et al., 2010), and schizophrenia (Son et al., 2010). Another attrition study outside the healthcare domain is "churn" prediction, often used in economic domains to predict engagement patterns of individuals (such as customers and employees) to increase their retention. Churn prediction has been well researched in the fields of banking (Ali and Arıtürk, 2014), video games (Kawale et al., 2009; Hadiji et al., 2014), and telecommunication (Huang et al., 2012).

A major distinguishing aspect of different attrition studies discussed above is the "chronicity" of the conditions being treated which requires a long-term commitment to the intervention or treatment to succeed(Brown and Moore, 2019). Examples of conditions where this type of attrition pattern has been studied include mental health conditions (Linardon and Fuller-Tyszkiewicz, 2020), sleep disorders (Hebert et al., 2010), and addiction disorders (Murray et al., 2013).

Attrition prediction from WMPs, in particular, has been studied mainly using traditional methods such as linear and logistic regression (Jiandani et al., 2016; Altamura et al., 2018; Ponzo et al., 2020; Perna et al., 2018), assuming that the relationship between the independent and the dependent variables are linear. This category also includes a few studies dedicated to studying attrition and weight loss in pediatric WMPs (Cate et al., 2017; Zeller et al., 2004; Dhaliwal et al., 2014; Skelton et al., 2011). Among these studies, (Batterham et al., 2017) was the only group that used shallow decision trees to predict attrition in adult dietary weight loss trials (not exactly comparable to our focus). They used demographic information and early weight change characteristics but did not use weight trajectories or other temporal patterns. Some other work also have tried to predict children obesity trend using EHR data(Gupta et al., 2022a,b).

3. Dataset

The data collected by our team includes all children 0–21 years of age who visited any of the four WMPs of Nemours Children's Health System between 2007 and 2021. Only young adults with a first visit prior to age 18 years were allowed to be seen between 19 and 21 years of age. The four WMPs serve patients in the US states of Delaware, Florida, Maryland, New Jersey, and Pennsylvania. For each patient, data from an internal dataset collected by the providers inside the WMPs were linked to the system-wide EHR data, capturing their health records when interacting with the entire healthcare system. The internal dataset collected at the WMPs was specifically designed to capture important covariates generally missing in EHRs. Additional information about the system-wide EHR, WMP internal dataset, and data pre-processing steps is presented in Appendix A.

The final cohort included 4,550 children (with 26,895 total WMP visits) of diverse backgrounds (27% Black, 37% Hispanic, 48% with Medicaid) and a mean BMI% of 98. We use age- and sex-adjusted BMI% above the 95th percentile to define obesity as defined by CDC (Centers for Disease Control and Prevention, 2017). BMI% is recorded and fed to the model whenever available. The sequence of BMI% from baseline to the end time point of the particular prediction window (defined later) is included as a dynamic variable. For each child, we included 18 features capturing demographic, psychosocial, lifestyle, anthropometric, medical comorbidity, and visit variables (Table 1). Since height is in the equation for BMI, we exclude it from the features set.

4. Problem setup

We study attrition patterns through two related predictive tasks: attrition prediction and weight outcome prediction. In the attrition prediction task, patients who dropped out before the prediction window were considered positive cases. In the weight outcome prediction task, any decrease in the child's BMI% was considered a success[1]. Accordingly, we defined the patients whose BMI% in the prediction window was lower than their BMI% at the time of their baseline visit to the WMP as negative cases. The rest of the patients (i.e., those whose BMI% remained the same or increased during the observation window) were considered positive cases.

1. This formulation captures the challenging nature of weight management interventions, as many patients join when they experience steady weight growth.

We considered these two (attrition and weight) prediction tasks in a binary classification framework and used flexible observation and prediction windows for both tasks. The start of the observation window was always fixed at the baseline WMP visit and its end was rolling (i.e., different time points). Additionally, the start of the prediction window was considered the end of the observation window and its end was rolling too. This type of flexibility in defining observation and prediction windows makes our models more practical in clinical settings, where a provider needs to know when a child (having various lengths of history) will drop.

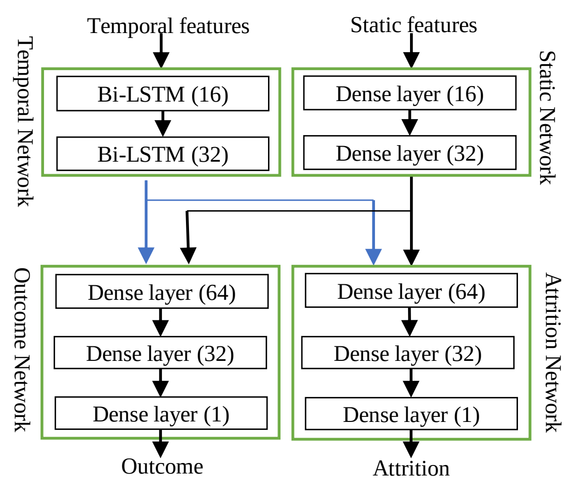


Figure 1: The proposed architecture for our model. The model receives static and temporal features and predicts the time of the last visit (attrition) and the BMI% change outcome at that time (weight outcome).

5. Predictive models

For implementing the two prediction tasks, we propose a deep neural network architecture, specifically designed to address the

Table 1: Characteristics of the study cohort.

Variable	Description
Sex	Male(2,122), Female(2,428)
Age	Range=(1–19), Mean=10.5
Race	Asian(57), White(1,767), Black(1,219), Other(1,462)
Ethnicity	Hispanic(1,680), Non-Hispanic(2,844)
Time btw visits	Mean=7 weeks
BMI% (per visit)	Mean BMI-for-age percentile at baseline visit=98
Insurance type	Medicaid(1,998), Private(1,570)
Food insecurity*	Often or Sometimes true Mean item-1(646), item-2(427)
Lifestyle score†	Range=(3–47), Mean=38.95
PSC-17‡	Range=(3–33), Mean=9
WMP visit type	Nutrition(4,293),medical(12,927),psychology(1,734),exercise(3,028)
Diagnosis codes	24 most commonly reported conditions (e.g., diabetes)

* as measured by the validated 2-item Hunger Vital Sign(Hager et al., 2010). † based on 12 evidence-based items about diet, activity, sleep, and hunger (each scored on a 4-point Likert scale with a total score of 12-48). ‡ Pediatric Symptom Checklist (Pagano et al., 1994), a validated 17-item screening tool (total score 0-34).

above problem. In order to fully utilize the static (e.g., demographics) and temporal (e.g., measurements) features in our data, our architecture is designed with the following four components as shown in Figure 1. Two initial components are shared between the two tasks of attrition and weight outcome predictions and the other two are task-specific components. The first component is a two-layer fully-connected network for processing the static features. The second is a two-layer bi-directional long short-term memory (Bi-LSTM) network for processing the temporal features. Bi-LSTM structures are similar to the common LSTMs, but they consist of two LSTMs units: one taking the input in a forward direction (left to right), and the other in a backward direction (right to left), thus improving the available context for the model (Goodfellow et al., 2016). The third component is a three-layer fully connected network for combining the extracted latent feature vectors from the first two parts and predicting the attrition time. Finally, the fourth component is a three-layer fully-connected network for combining the extracted latent feature vectors from the first two parts and predicting the weight outcome. We used dropout and batch normalization layers after all of the layers mentioned above and used binary cross-entropy as the loss function.

In the attrition task, we train the model to distinguish between the patients who dropped out and patients who stayed in the WMP. In the outcome task, we train a model to distinguish between patients who successfully decreased their BMI% and those who had no change or a BMI% increase. To implement these two and to improve the overall performance of our models on the small size of training data, we used a "multi-task learning" approach for designing the model. In this design, we use a hard parameter-sharing approach, by sharing the first two components between the two tasks. This way, sharing the parameters of the two initial parts of the model improves the performance for both prediction tasks. Additionally, following a "transfer learning" theme, we iteratively trained our model on sliding observation and prediction windows. We initialized the weights of the model in each step with the previous model weights. We started from the smallest window and moved toward the largest window. No data used for prediction was used for training in previous iterations.Figure 5 shows the training process and the way that the four components are involved in our multi-task and transfer learning themes. Additionally, Algorithm 1 presents a high-level pseudocode of our training process. The model presented in this paper was implemented using Keras (Chollet et al., 2015) inside the TensorFlow (Abadi et al., 2016) framework. Our code is publicly available on GitHub[2].

6. Experiments

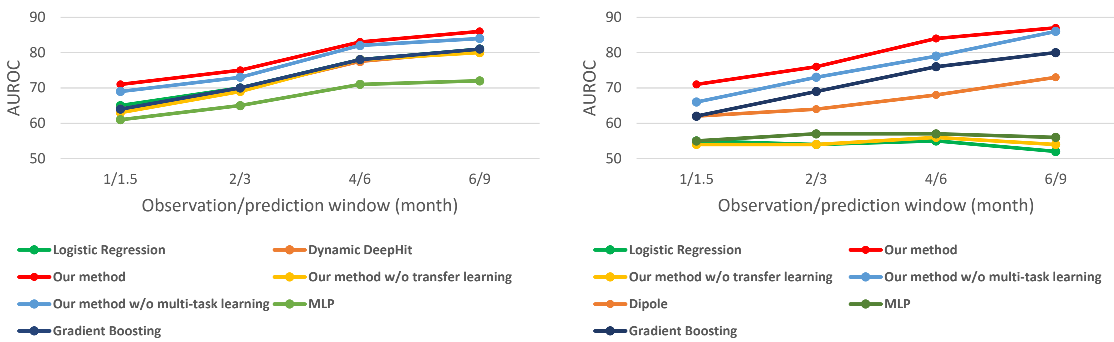
Baselines While a very large family of traditional and advanced machine learning methods could be used for our prediction tasks, we opted to use a few of the most relevant methods frequently used in the literature. One is the general logistic regression (LR) method, used by many other groups for studying attrition. Similar to the common practice in the field (Shipe et al., 2019), we aggregate the temporal variables by calculating the average values in the observation window. We train an LR model with L2 penalty and a BFGS solver using the scikit-learn library (Pedregosa et al., 2011) in Python.

For the second baseline, we use another method for studying similar problems in clinical domains, which takes a survival analysis approach. Survival analysis (time-to-event analysis) is widely used in many biomedical applications (Clark et al., 2003) to estimate the expected duration until the desired event occurs (e.g., dropout or death). We use a state-of-the-art method for studying survival analysis, called Dynamic DeepHit (Lee et al., 2020), which presents a deep neural network to learn the distribution of survival times. Dynamic DeepHit utilizes the available temporal data to issue dynamically updated survival predictions. This survival analysis method only fits our attrition prediction task (and not the weight outcome prediction task). We also use two other general baselines: a multilayer perceptron (MLP) and the Dipole method. The MLP baseline consists of three layers; each containing 100 cell. Dipole (Ma et al., 2017) is an end-to-end model for predicting patients' future health information and has been widely used in EHR applications. We used the PyHealth package (Zhao et al., 2021) implementation of Dipole for the experiments.

Performance measures To measure the prediction performance of the models, we report accuracy, precision, recall, area under the receiver operating characteristic (AUROC), and area under the precision-recall curve (AUPRC). The performance of the proposed model in the outcome and attrition prediction tasks for a series of observation and prediction windows are shown in Table 2 (observation window = 1, 2, 4, and 6 months; and prediction window = 1.5, 3, 6, 9 months). This specific set of observations and windows were selected based on prediction windows used in prior studies (Moran et al., 2019; Jiandani et al., 2016) and based on the Cox hazard model analysis of our data (Figure 4). Additionally, Figure 3 shows the results related to an alternative scenario where a fixed prediction window (at 6 months) is used, while the observation windows vary.

Ablation analysis We also study the effect of the extra components in our model added to implement multi-task and trans-

2. https://github.com/healthylaife/wm_attrition



(A) (B)

Figure 2: Performance of our method versus baselines, in predicting (A) attrition, and (B) weight outcome. Deephit and Dipole were not applicable to A and B, respectively.

fer learning themes in our design. To this end, we report the results of our method without (a) multi-task learning implementation (i.e., the two shared sub-networks and only one of the latter sub-networks for each task), and (b) transfer learning implementation (i.e., fine-tuning the models on the corresponding data to each observation-prediction window setting without pretraining). Figure 2 shows how the results obtained from our method compare to the baselines and the ablated versions of our model.

Top predictors To study the degree to which each input feature contributes to predicting the final outputs in each task, we used the popular Shapley additive explanations (SHAP) method implemented in the SHAP toolbox (Lundberg and Lee, 2017b,a). Inspired by the Shapley values, this toolbox is a unified framework that assigns each input feature an importance value indicating its importance in the prediction task. A SHAP value (partially) indicates the degree to which a feature contributes to pushing the output from the base value (average model output) to the actual predicted value, so the higher value can be indicative of the higher importance. Here, we report the importance values (as indicated by SHAP) for the top five features predicting attrition and outcome patterns, for each prediction/observation window. The results for the attrition prediction are shown in Table 3 and for the outcome prediction in Table 4 (in Appendix C).

7. Discussion

The preliminary experiments in this study demonstrate that the presented prediction pipeline can achieve AUROCs of around 0.75 in most observation/prediction configurations. As expected, longer observation windows generally allowed the model to show better performance. While various observation and prediction combinations show the flexibility of our method in letting the end users decide about the desired length, in practical usages the providers we talked to indicated the sixth month as one of the critical points to focus on. Fixing our model on this point (shown in Figure 3) shows that our model can predict attrition patterns from early on.

Table 2: Results for the attrition and outcome prediction tasks (mean ± STD). Obr: Observation window (months), Prd: Prediction window (months), B.AUPRC: Baseline AUPRC which is equal to fraction of positives samples in the dataset.

	Attrition					
Obr/Prd	Precision	Recall	Specificity	AUROC	AUPRC	B.AUPRC
1/1.5	56 ± 3	48 ± 1	73 ± 0	69 ± 4	59 ± 6	0.4
2/3	66 ± 3	71 ± 5	68 ± 0	73 ± 4	71 ± 5	0.5
4/6	79 ± 2	84 ± 1	66 ± 2	82 ± 3	86 ± 3	0.63
6/9	84 ± 3	89 ± 1	64 ± 1	84 ± 3	91 ± 3	0.69
	Outcome					
Obr/Prd	Precision	Recall	Specificity	AUROC	AUPRC	B.AUPRC
1/1.5	30 ± 8	68 ± 4	61 ± 1	71 ± 2	38 ± 2	0.19
2/3	46 ± 4	67 ± 7	68 ± 0	73 ± 6	53 ± 9	0.27
4/6	63 ± 7	79 ± 2	71 ± 0	84 ± 4	71 ± 2	0.31
6/9	73 ± 3	70 ± 2	68 ± 1	86 ± 4	75 ± 4	0.31

Having a broad range of machine learning pipelines that could have been used for our model, we opted for a simple model consisting of basic LSTM and dense layers. This was mainly due to not observing any superior performance in more complex predictive deep models, such as recent transformer-based methods (Li et al., 2020; Poulain et al., 2021, 2022) and EHR-based time series prediction methods Gupta and Beheshti (2020). Comparing our model against several other baselines (including state-of-the-art deep models) also demonstrated that it can achieve superior prediction performance, as shown in Figure 2. Our ablation analysis also showed that incorporating transfer and multi-task learning themes was essential to enhance the performance of our model.

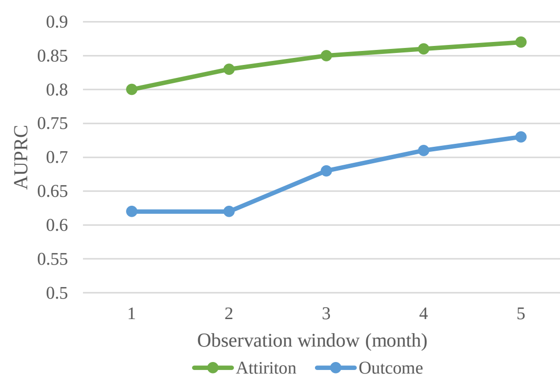


Figure 3: Performance of our method shown in AUPRC for a fixed prediction window (6 months) and different observation windows.

Studying the top predictor variables using SHAP values show that the top variables do vary across the tasks (attrition versus weight outcome) and across various lengths of the observation and prediction windows. This may indicate the importance of targeting different risk factors at different points in pediatric WMPs. One can also observe that age, the average time between the visits, and the patient's BMI% trajectory were important features for determining both attrition and weight outcomes. These top predictors are consistent with previous studies (Batterham

Table 3: The top five features predicting attrition, as determined by the scaled SHAP values (raw SHAP $\times 10^{-3}$) shown in parentheses. shown in parentheses. BMI% shows the BMI% trajectory recorded during the observation window. Food Ins: Food insecurity. Visits Int: Visits intervals.

Observation/Prediction window (months)			
1/1.5	2/3	3/4.5	6/9
Age(41)	Insurance(15)	Visits Int(23)	Age(63)
Ethnicity(6)	BMI%(6)	BMI%(20)	Insurance(9)
Food Ins(5)	Race(5)	Age(10)	Visits Int(3)
Sex(5)	Age(5)	Food Ins(7)	Sex(2)
Insurance(3)	Sex(4)	Insurance(5)	BMI %(2)

et al., 2017, 2016). Some of the top variables indicated by our models are not modifiable (such as sex and age). Knowing these non-modifiable factors can still inform early interventions targeting patients (Coleman et al., 2012).

Clinical relevance Focusing on the top variables that we have found in our analysis, we specifically study age, BMI%, and sociodemographic factors more closely. Other studies have also found that age is an important predictor of both attrition and weight outcomes, with children of younger ages having more success in WMPs (Jiandani et al., 2016; Batterham et al., 2016). Relatedly, we have also found that the average time between WMP visits is a primary predictor for both tasks, with worse outcomes and higher attrition rates for patients who had a prolonged time between visits. Besides age, a patient's early weight loss (identified by the BMI% trajectory) is shown to be predictive of overall weight loss progress and the patients who have success with early weight loss seem to have a lower risk of attrition (Batterham et al., 2016).

Finally, sociodemographics like race and ethnicity, and sex, as well as food security and insurance status, are important predictors of both attrition and weight outcomes. This is aligned with previous studies demonstrating inequities in health outcomes between subgroups (Martin and Ferris, 2007). Interestingly, other factors like medical diagnoses, medications, and lifestyle scores were less predictive of attrition and weight outcomes, which may highlight the importance of identifying and supporting key groups based on sociodemographics, as well as ensuring frequent visits and early success with weight outcomes during the treatment period, regardless of a patient's underlying conditions or lifestyle behaviors. We note that while we report the average SHAP values across all of the samples (children), our model is most helpful when the individual children are considered separately, and when a provider can see the predictors that can be targeted for a particular child to prevent attrition or increase the success with weight outcomes.

Censoring The way we define the positive and negative cases allows the effects of patient censoring to remain limited in our experiments. Specifically, for predicting attrition, we consider the time of the last visit as our target. No child in our dataset has returned to the program following a 6-month gap. 3–6 months is within the acceptable range for adhering to WMP interventions.

Clinical application A strength of our models is demonstrating the multi-factorial nature of obesity outcomes. The majority of clinicians providing care to children with obesity also understand that the etiology of obesity and the reasons for variable outcomes with treatment are complex. We would encourage users to interpret the data with this in mind and to use the data to tailor treatment for certain populations (e.g., providing culturally-competent care to certain racial and ethnic subgroups or tailoring treatment for adolescents to include more peer-based support). In addition to tailoring treatment, interventions can also address modifiable factors that predict outcomes (e.g., providing food resources to families who are food insecure or referring patients who endorse behavioral or mental health concerns to a psychologist) and ensure engagement of patients at risk for attrition with increased contact (e.g., texts or calls) in between clinic visits.

Limitations The current study is limited in several ways. First, our dataset only includes patients from one healthcare system. Still, our dataset is larger than similar ones used to study attrition and spans four geographically distinct sites in the Mid-Atlantic (Delaware, Pennsylvania, Maryland, New Jersey) and Southern (Florida) regions of the US. Additionally, our approach relies on discretizing future time into two-week windows. Considering attrition prediction as a regression task was an alternative natural choice. In our experiments, we have noticed that formulating our problem as a classification task can yield better results. Additionally, as most follow-up visits are not scheduled in shorter than two-week intervals, one can still use our approach for continuous (any time in the future) predictions. Lastly, our study focuses primarily on attrition from pediatric WMPs, our method should be applicable to adult obesity WMPs and other comparable problems such as mental health and addiction recovery programs.

Impact and future work As a publicly accessible tool that uses commonly available data on WMPs, our machine learning pipeline can be integrated into current clinical workflows to offer early and personalized insights assisting families and providers improve the success rates of weight management program interventions. Our team is currently working on deploying our predictive tool in the graphical dashboard that the providers in one of the WMPs of Nemours Children's Health use. As part of our future work, we plan to expand our model by including additional information from children's historical records (before joining a WMP) and also by including new data from other pediatric health systems. Moreover, we aim to explicitly identify the temporal phenotype (such as distinct shapes of the body-weight trajectory) that can predict (or stratify) attrition or weight outcome patterns.

Acknowledgments

Our study was supported by NIH awards, P20GM103446 and P20GM113125.

References

Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pages 265–283, Savannah, GA, USA, 2016. ACM.

Lara J Akinbami, Te-Ching Chen, Orlando Davy, Cynthia L Ogden, Steven Fink, Jason Clark, Minsun K Riddles, and Leyla K Mohadjer. National health and nutrition examination survey, 2017-march 2020 prepandemic file: Sample design, estimation, and analytic guidelines. *Vital and Health statistics. Ser. 1, Programs and Collection Procedures*, (190):1–36, 2022.

Özden Gür Ali and Umut Arıtürk. Dynamic churn prediction framework with more effective use of rare event data: The case of private banking. *Expert Systems with Applications*, 41(17):7889–7903, 2014.

Mario Altamura, Piero Porcelli, Beth Fairfield, Stefania Malerba, Raffaella Carnevale, Angela Balzotti, Giuseppe Rossi, Gianluigi Vendemiale, and Antonello Bellomo. Alexithymia predicts attrition and outcome in weight-loss obesity treatment. *Frontiers in Psychology*, 9:2432, 2018.

Geoff D. C. Ball, Meghan Sebastianski, Jessica Wijesundera, Diana Keto-Lambert, Josephine Ho, Ian Zenlea, Arnaldo Perez, James Nobles, and Joseph A. Skelton. Strategies to reduce attrition in managing paediatric obesity: A systematic review. *Pediatric Obesity*, 16(4):e12733, 2021.

M. Batterham, L. Tapsell, K. Charlton, J. O'Shea, and R. Thorne. Using data mining to predict success in a weight loss trial. *Journal of Human Nutrition and Dietetics*, 30(4):471–478, 2017.

Marijka Batterham, Linda C Tapsell, and Karen E Charlton. Predicting dropout in dietary weight loss trials using demographic and early weight change characteristics: implications for trial design. *Obesity research & clinical practice*, 10(2):189–196, 2016.

Tamara Brown and Theresa HM Moore. Interventions for preventing obesity in children. *Cochrane Database of Systematic Reviews*, (7), 2019.

Pit-Ten Cate, Hanen Samouda, Ulrike Schierloh, Julien Jacobs, Jean Francois Vervier, Saverio Stranges, Marie Lise Lair, Carine De Beaufort, et al. Can health indicators and psychosocial characteristics predict attrition in youth with overweight and obesity seeking ambulatory treatment? data from a retrospective longitudinal study in a paediatric clinic in luxembourg. *BMJ Open*, 7(9), 2017.

Centers for Disease Control and Prevention. Growth charts - clinical growth charts, Jun 2017. URL https://www.cdc.gov/growthcharts/clinical_charts.htm. Accessed: Feb 2022.

François Chollet et al. Keras. https://github.com/fchollet/keras, 2015.

Taane G Clark, Michael J Bradburn, Sharon B Love, and Douglas G Altman. Survival analysis part i: basic concepts and first analyses. *British journal of cancer*, 89(2):232–238, 2003.

Karen J Coleman, Anne C Hsii, Corinna Koebnick, Ana F Alpern, Brenna Bley, Marianne Yousef, Erin M Shih, Keila J Trimble-Cox, Ning Smith, Amy H Porter, et al. Implementation of clinical practice guidelines for pediatric weight management. *The Journal of pediatrics*, 160 (6):918–922, 2012.

Jasmine Dhaliwal, Nicole MI Nosworthy, Nicholas L Holt, Lonnie Zwaigenbaum, Jillian LS Avis, Allison Rasquinha, and Geoff DC Ball. Attrition and the management of pediatric obesity: an integrative review. *Childhood Obesity*, 10(6):461–473, 2014.

C. B. Forrest, P. A. Margolis, L. C. Bailey, K. Marsolo, et al. Pedsnet: a national pediatric learning health system. *Journal of the American Medical Informatics Association*, 21(4):602–606, 2014.

Diego Giunta, Agustina Briatore, Analía Baum, Daniel Luna, Gabriel Waisman, and Fernán Gonzalez Bernaldo de Quiros. Factors associated with nonattendance at clinical medicine scheduled outpatient appointments in a university general hospital. *Patient preference and adherence*, 7: 1163, 2013.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

Mehak Gupta and Rahmatollah Beheshti. Time-series imputation and prediction with bi-directional generative adversarial networks. *arXiv preprint arXiv:2009.08900*, 2020.

Mehak Gupta, Thao-Ly T Phan, H Timothy Bunnell, and Rahmatollah Beheshti. Obesity prediction with ehr data: A deep learning approach with interpretable elements. *ACM Transactions on Computing for Healthcare (HEALTH)*, 3(3):1–19, 2022a.

Mehak Gupta, Raphael Poulain, Thao-Ly T. Phan, H. Timothy Bunnell, and Rahmatollah Beheshti. Flexible-window predictions on electronic health records. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(11):12510–12516, Jun. 2022b.

Fabian Hadiji, Rafet Sifa, Anders Drachen, Christian Thurau, Kristian Kersting, and Christian Bauckhage. Predicting player churn in the wild. In *2014 IEEE Conference on Computational Intelligence and Games*, pages 1–8, Dortmund, Germany, 2014. IEEE.

Erin R. Hager, Maureen M. Quigg, Anna M. Black, Sharon M. Coleman, Timothy Heeren, Ruth Rose-Jacobs, John T. Cook, Stephanie A. Ettinger de Cuba, Patrick H. Casey, Mariana Chilton, Diana B. Cutts, Alan F. Meyers, and Deborah A. Frank. Development and Validity of a 2-Item Screen to Identify Families at Risk for Food Insecurity. *Pediatrics*, 126 (1):e26–e32, 07 2010. ISSN 0031-4005. doi: 10.1542/peds.2009-3146.

Conal Hayton, Allan Clark, Sandra Olive, Paula Browne, Penny Galey, Emma Knights, Lindi Staunton, Andrew Jones, Emma Coombes, and Andrew M. Wilson. Barriers to pulmonary rehabilitation: Characteristics that predict patient attendance and adherence. *Respiratory Medicine*, 107(3):401–407, 2013.

Elizabeth A. Hebert, Norah Vincent, Samantha Lewycky, and Kaitlyn Walsh. Attrition and adherence in the online treatment of chronic insomnia. *Behavioral Sleep Medicine*, 8(3):141–150, 2010.

Bingquan Huang, Mohand Tahar Kechadi, and Brian Buckley. Customer churn prediction in telecommunications. *Expert Systems with Applications*, 39(1):1414–1425, 2012.

Elissa Jelalian, Chantelle N Hart, Robyn S Mehlenbeck, Elizabeth E Lloyd-Richardson, Jamie D Kaplan, Katherine T Flynn-O'Brien, and Rena R Wing. Predictors of attrition and weight loss in an adolescent weight control program. *Obesity*, 16(6):1318–1323, 2008.

Dishay Jiandani, Sean Wharton, Michael A. Rotondi, Chris I. Ardern, and Jennifer L. Kuk. Predictors of early attrition and successful weight loss in patients attending an obesity management program. *BMC Obesity*, 3(1):1–9, 2016.

Jaya Kawale, Aditya Pal, and Jaideep Srivastava. Churn prediction in mmorpgs: A social influence based approach. In *2009 International Conference on Computational Science and Engineering*, volume 4, pages 423–428, Vancouver, BC, Canada, 2009. IEEE, IEEE.

Parviz Kheirkhah, Qianmei Feng, Lauren M Travis, Shahriar Tavakoli-Tabasi, and Amir Sharafkhaneh. Prevalence, predictors and economic consequences of noshows. *BMC health services research*, 16 (1):1–6, 2015.

Jackson A Killian, Bryan Wilder, Amit Sharma, Vinod Choudhary, Bistra Dilkina, and Milind Tambe. Learning to prescribe interventions for tuberculosis patients using digital adherence data. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2430–2438, Anchorage, AK, USA, 2019. ACM.

Changhee Lee, Jinsung Yoon, and Mihaela van der Schaar. Dynamic-deephit: A deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. *IEEE Transactions on Biomedical Engineering*, 67(1): 122–133, 2020.

Yikuan Li, Shishir Rao, José Roberto Ayala Solares, Abdelaali Hassaine, Rema Ramakrishnan, Dexter Canoy, Yajie Zhu, Kazem Rahimi, and Gholamreza Salimi-Khorshidi. Behrt: transformer for electronic health records. *Scientific reports*, 10(1):1–12, 2020.

Jake Linardon and Matthew Fuller-Tyszkiewicz. Attrition and adherence in smartphone-delivered interventions for mental health problems: A systematic and meta-analytic review. *Journal of consulting and clinical psychology*, 88(1): 1, 2020.

Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions, 2017a.

Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. 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 4765–4774. Curran Associates, Inc., 2017b.

Fenglong Ma, Radha Chitta, Jing Zhou, Quanzeng You, Tong Sun, and Jing Gao. Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1903–1911, 2017.

Katie S. Martin and Ann M. Ferris. Food insecurity and gender are risk factors for obesity. *Journal of Nutrition Education and Behavior*, 39(1):31–36, 2007.

Alex J. Mitchell and Thomas Selmes. Why don't patients attend their appointments? maintaining engagement with psychiatric services. *Advances in Psychiatric Treatment*, 13(6):423–434, 2007.

Lisa Moran, Manny Noakes, Peter Clifton, Jon Buckley, Grant Brinkworth, Rebecca Thomson, and Robert Norman. Predictors of lifestyle intervention attrition or weight loss success in women with polycystic ovary syndrome who are overweight or obese. *Nutrients*, 11(3):492, 2019.

Elizabeth Murray, Ian R White, Mira Varagunam, Christine Godfrey, Zarnie Khadjesari, and Jim McCambridge. Attrition revisited: Adherence and retention in a web-based alcohol trial. *J Med Internet Res*, 15 (8):e162, Aug 2013.

Cynthia L. Ogden, Margaret D. Carroll, Brian K. Kit, and Katherine M. Flegal. Prevalence of Childhood and Adult Obesity in the United States, 2011-2012. *JAMA*, 311(8):806–814, 02 2014.

ME Pagano, LJ Cassidy, M Little, JM Murphy, and MS Jellinek. Screening 4-and 5-year-old children for psychosocial dysfunction: A preliminary study with the pediatric symptom checklist. *Journal of developmental and behavioral pediatrics*, 15(3): 191–197, 1994.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

Simone Perna, Daniele Spadaccini, Antonella Riva, Pietro Allegrini, Chiara Edera, Milena Anna Faliva, Gabriella Peroni, Maurizio Naso, Mara Nichetti, Carlotta Gozzer, et al. A path model analysis on predictors of dropout (at 6 and 12 months) during the weight loss interventions in endocrinology outpatient division. *Endocrine*, 61(3):447–461, 2018.

Ineke M Pit-ten Cate, Hanen Samouda, Ulrike Schierloh, Julien Jacobs, Jean Francois Vervier, Saverio Stranges, Marie Lise Lair, and Carine de Beaufort. Can health indicators and psychosocial characteristics predict attrition in youths with overweight and obesity seeking ambulatory treatment? data from a retrospective longitudinal study in a paediatric clinic in luxembourg. *BMJ open*, 7(9):e014811, 2017.

Valentina Ponzo, Elena Scumaci, Ilaria Goitre, Guglielmo Beccuti, Andrea Benso, Sara Belcastro, Chiara Crespi, Franco De Michieli, Marianna Pellegrini, Paola Scuntero, Enrica Marzola, Giovanni Abbate-Daga, Ezio Ghigo, Fabio Broglio, and Simona Bo. Predictors of attrition from a weight loss program. a study of adult patients with obesity in a community setting. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity*, 25:1–8, 2020.

Raphael Poulain, Mehak Gupta, Randi Foraker, and Rahmatollah Beheshti. Transformer-based multi-target regression on electronic health records for primordial prevention of cardiovascular disease. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 726–731. IEEE, 2021.

Raphael Poulain, Mehak Gupta, and Rahmatollah Beheshti. Few-shot learning with semi-supervised transformers for electronic health records. In *Proceedings of the 7th Machine Learning for Healthcare Conference*, volume 182 of *Proceedings of Machine Learning Research*. PMLR, 05–06 Aug 2022.

Ramsey Sabit, Timothy L. Griffiths, Alan J. Watkins, Wendy Evans, Charlotte E. Bolton, Dennis J. Shale, and Keir E. Lewis. Predictors of poor attendance at an outpatient pulmonary rehabilitation programme. *Respiratory Medicine*, 102(6): 819–824, 2008.

Maren E Shipe, Stephen A Deppen, Farhood Farjah, and Eric L Grogan. Developing prediction models for clinical use using logistic regression: an overview. *Journal of thoracic disease*, 11(Suppl 4):S574, 2019.

Joseph A Skelton, David C Goff Jr, Edward Ip, and Bettina M Beech. Attrition in a multidisciplinary pediatric weight management clinic. *Childhood Obesity*, 7(3): 185–193, 2011.

Youn-Jung Son, Hong-Gee Kim, Eung-Hee Kim, Sangsup Choi, and Soo-Kyoung Lee. Application of support vector machine for prediction of medication adherence in heart failure patients. *Healthcare informatics research*, 16(4):253–259, 2010.

Denise E. Wilfley, Brian E. Saelens, Richard I. Stein, John R. Best, Rachel P. Kolko, Kenneth B. Schechtman, Michael Wallendorf, R. Robinson Welch, Michael G. Perri, and Leonard H. Epstein. Dose, Content, and Mediators of Family-Based Treatment for Childhood Obesity: A Multisite Randomized Clinical Trial. *JAMA Pediatrics*, 171(12):1151–1159, 12 2017.

Meg Zeller, Shelley Kirk, Randal Claytor, Philip Khoury, Jennifer Grieme, Megan Santangelo, and Stephen Daniels. Predictors of attrition from a pediatric weight management program. *The Journal of pediatrics*, 144(4):466–470, 2004.

Yue Zhao, Zhi Qiao, Cao Xiao, Lucas Glass, and Jimeng Sun. Pyhealth: A python library for health predictive models. *arXiv preprint arXiv:2101.04209*, 2021.

Appendix A. Cohort info

This internal WMP dataset specifically covered: (a) parent-reported lifestyle variables (including diet, activity, sleep, and mood) collected in every visit, (b) parent-reported psychosocial variables (two-item hunger vital sign for food insecurity and pediatric symptom checklist for child behavioral concerns (Pagano et al., 1994)), and (c) additional visit data (specifically, the type of providers seen within the weight management program and the days between visits). The EHR dataset was the Nemours portion of the large PEDSnet data repository and included rigorously validated EHR variables including medical conditions, anthropometrics, visits, and demographics (Forrest et al., 2014). PEDSnet is a multi-speciality network that conducts observational research and clinical trials across multiple large children's hospital health systems in the US. The dataset was anonymized, and the study was approved by Nemours Institutional Review Board.

To bucketize the longitudinal EHR data, we combined visits over 15-days time periods. We examined medical diagnosis codes and medication tables, from the available EHR data. Any condition observed at least once during the time window was denoted by 1 in the new sequence, and the measurements were averaged over the time window. If there were no visits for a patient in a time window, the corresponding vector for that period was set to all zeros. We also excluded rare diagnosis codes (i.e., the codes that appeared in less than 2% of patients), which reduced the total number of diagnosis codes from 435 to 24. We used one-hot encoding for the categorical values and normalized the continuous values by performing a min-max scaling on all features.

In our dataset, 28% of the children had only one WMP visit, and 15% had more than ten visits. Figure 4 shows the overall distributions of the number of visits and the duration of WMP attendance in months.

Appendix B. Method details

In the Algorithm 1, each model is trained using the data from a specific observation and prediction window, initialized by the weights from the previous window settings. The procedure in this algorithm receives the input data (X), labels for the attrition and outcome tasks (Y_A and Y_O), and the list of observation and prediction windows. It returns a list of fine-tuned models.

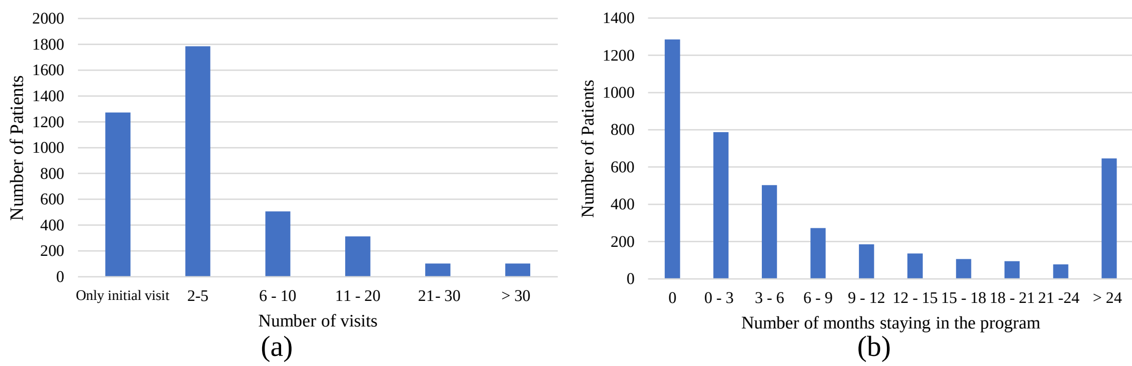


Figure 4: The distribution of (a) the number of visits and (b) the number of months staying in the WMP across the patients in our cohort.

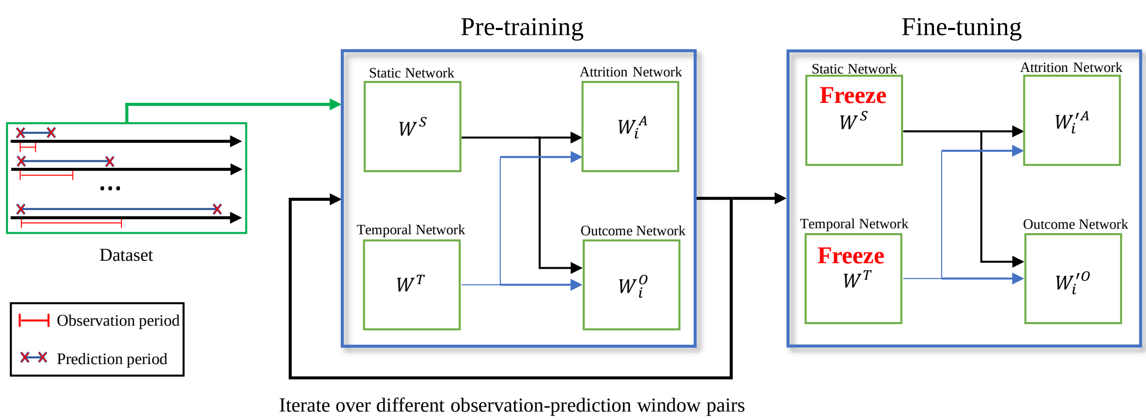


Figure 5: Proposed training process using the multi-task and transfer learning themes. W^S, W^A, W^T, and W^O are the weights of the static, attrition, temporal, and outcome sub-networks, respectively. We extract the features and the corresponding labels for each patient based on the rolling observation and prediction windows and feed them to the network. After the pretraining of a general model, we then initialize the weights of the specialized models with the weights from the general model. For each specific observation and prediction window setting, the model is then fine-tuned using only the relevant samples. W_i shows the ith fine-tuned configuration.

Appendix C. Top predictors

Table 4: The top five features predict the weight outcome in various observation and prediction window settings, as determined by the scaled SHAP values (raw SHAP Value $\times 10^{-3}$) shown in parentheses. BMI% shows the BMI% trajectory recorded during the observation window. Food ins:Food insecurity. Visits int: Visits intervals.

Observation/Prediction window (in months)			
1/1.5	2/3	3/4.5	6/9
Age (10)	Age(58)	Insurance(20)	Age(21)
Race(5)	Visits int.(14)	Age(18)	Visits int.(8)
Insurance(4)	Insurance(14)	Ethnicity(12)	Food ins.(7)
Lifestyle Score(3)	Sex(8)	Race(11)	BMI %(5)
Food ins.(3)	BMI %(6)	visits int.(10)	Sex(5)

Algorithm 1: Training pseudocode

Input: X, Y_A, Y_O, Observation-Prediction window list **Output:** Trained models list
Preprocess X
PM = random (PM) ▷ `Random initialization of the pretrained model`
for *Every observation-prediction window pair* **do**
  X' = Select cohort based on the windows
  M = PM ▷ `Load weights from the previous PM to model M`
  Pretrain M with X'
  PM = M ▷ `Store M as the current pretrained model PM`
  Freeze the first two components in M
  Fine-tune the rest of M
  Add M to the trained models list
end
return Trained models list