Understanding Nothing: Interpreting Models Using Sparse Longitudinal Data

Jacob M. Ekstrum B.S.[1†], Noah J. Poczciwinski B.S.[1†], Brett F. Cropp M.S.[1†], John B. Coles Ph.D.[2‡]

[1]CUBRC Inc. Information Fusion Group, Buffalo, NY
[2]Department of Industrial and Systems Engineering, University at Buffalo, Buffalo, NY
[†]Contact: <First Name>.<Last Name>@cubrc.org
[‡]Contact: <First Initial><Middle Initial><Last Name>@buffalo.edu

Summary

- Current models that can explain their decision process, such as linear models and shallow decision trees, lack the accuracy that more structured modeling techniques produce for sparse longitudinal data.
- For those models that can be interpreted, it is difficult to concisely convey important information to care providers and clinicians as the input variables to be ranked increase in both count and sparsity.
- The Heartwood Analytics^{TM} system addresses these problems through its feature importance framework.
- The global approximations embedded in our meta-model are used to generate both group-level and per-patient models of the impact of variables, which are gaining popularity among Heartwood adopters in the complex care and autism care spaces.
- Heartwood Analytics^{TM} uses a custom web interface (Figure 5) to communicate the meta-model results to less-technical audiences, such as clinicians and staff that require more information to reduce autism-related behavioral events.
- Adjustments made to feature values in the interface generate real-time predictions alongside previous predictions in order to communicate feature importance (Figure 2).

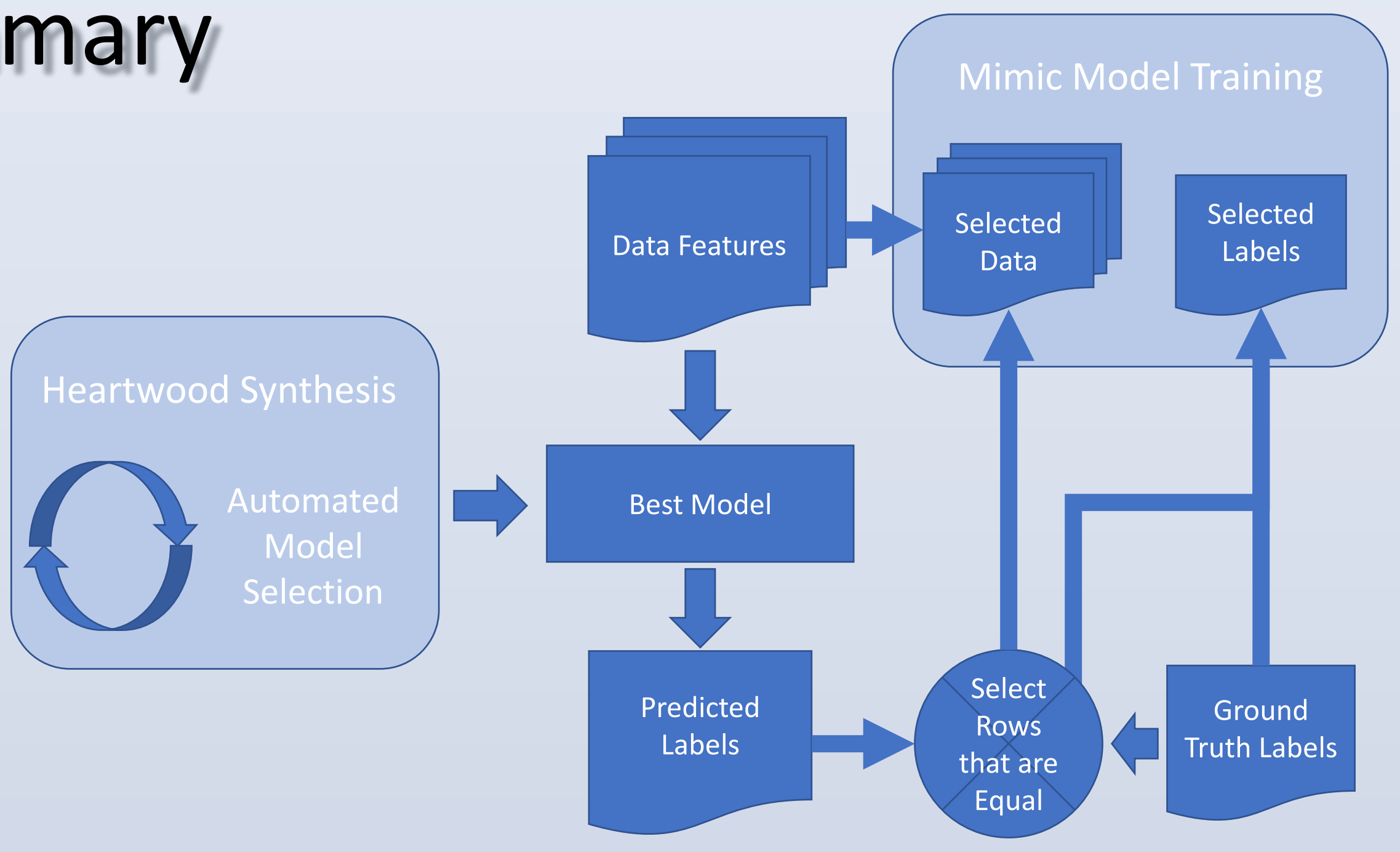
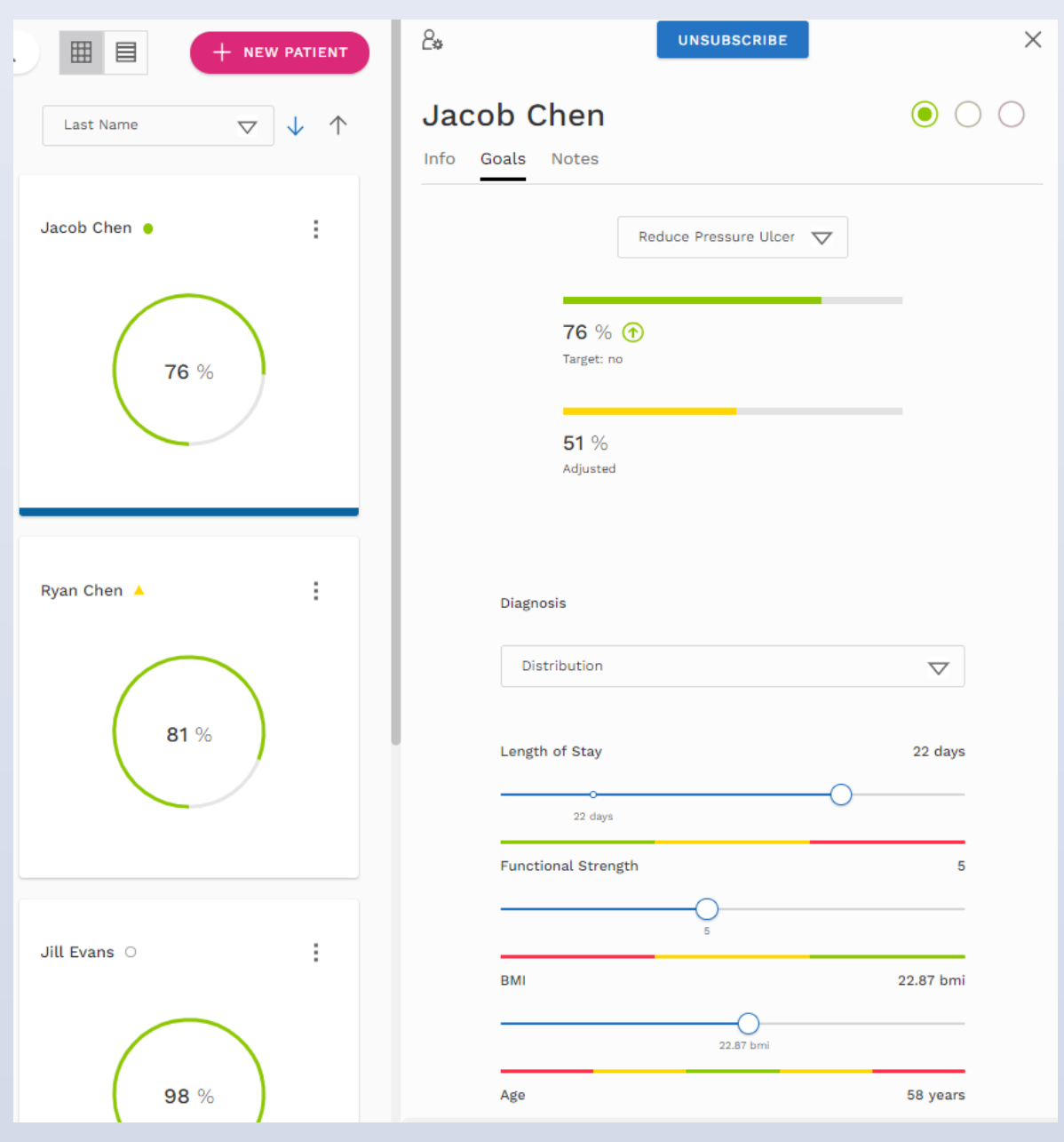


Figure 1: Heartwood's Mimic Model Pipeline, which is used for the generation of our longitudinal models for interpretation.



Jacob Chen
76 %
Ryan Chen
81 %
Jill Evans
98 %

Jacob Chen
Reduce Pressure Ulcer
76 %
Target: no
51 %
Adjusted
Diagnosis
Distribution
Length of Stay 22 days
Functional Strength 5
BMI 22.87 bmi
Age 58 years

Figure 2: An example of the Heartwood Analytics dashboard that displays expected patient outcome as a factor of provided "what-if" information with coloring provided by the meta-model.

Methods

Heartwood Synthesis:
Semi-Automated Machine Learning

- Heartwood Analytics^{TM} contains *Heartwood Synthesis*, an automated machine learning platform based off of initial research performed by Cropp et al.[2] (Figure 3).
- *Heartwood Synthesis* has been improved to perform time-series analysis automatically over data that is structured in a sequential manner.
- Once a model is created and selected for the user's time-series data, the feature importance pipeline is automatically executed for that model (Figure 1).

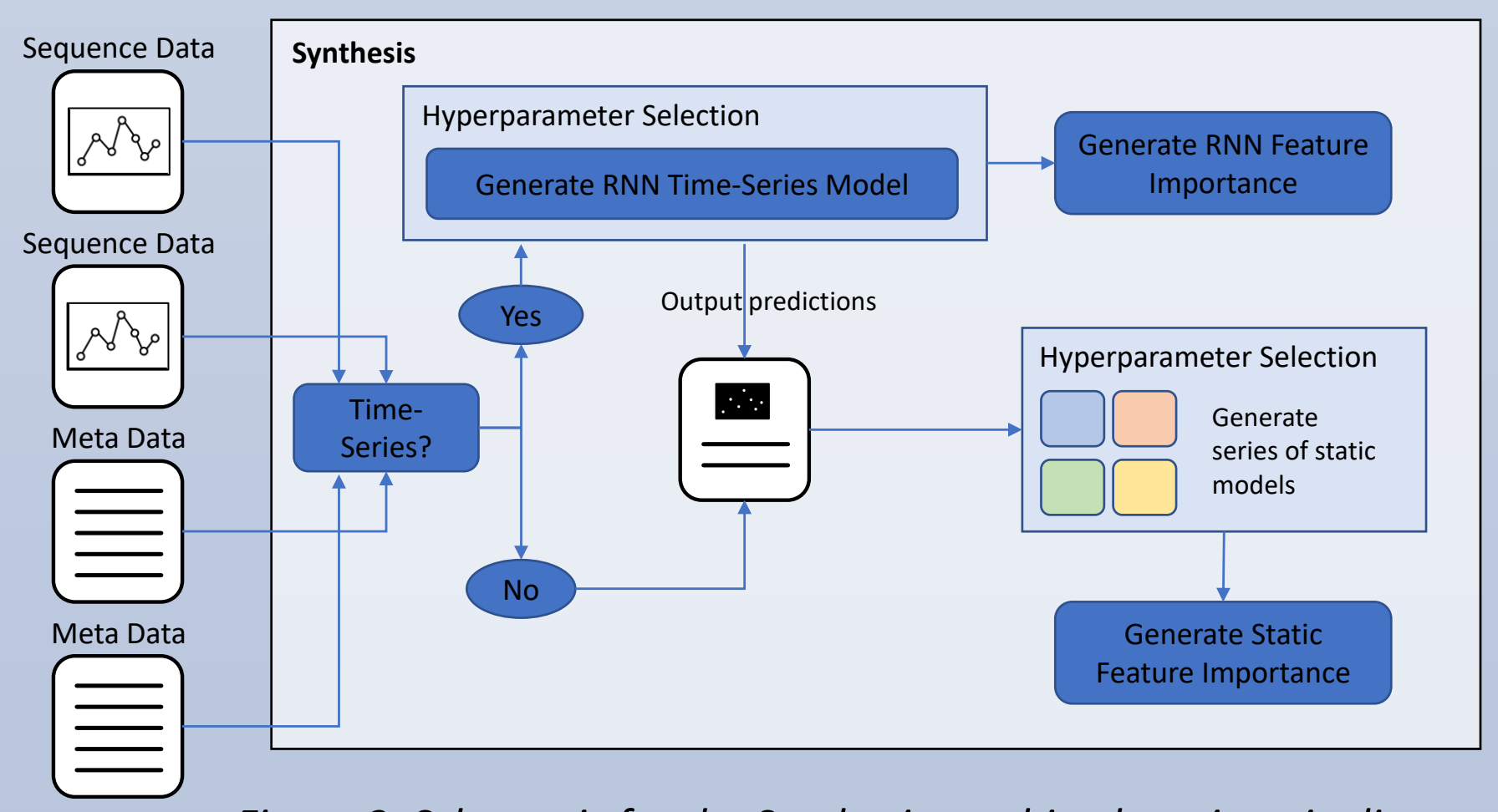
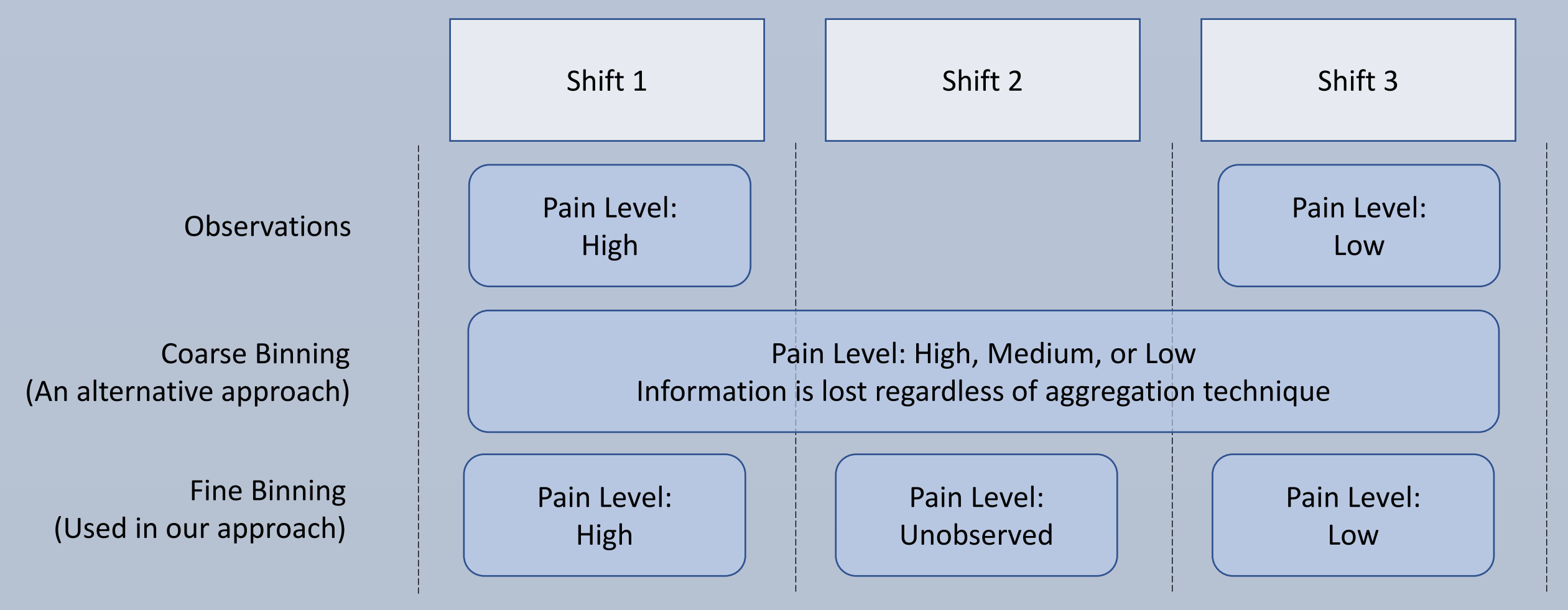


Figure 3. Schematic for the Synthesis machine learning pipeline, which automatically performs analysis based on the structure of the data present.

Addressing Sparsity

- We have found that a large portion of longitudinal sensor data is collected at drastically varying frequencies.
- We cannot assume that these frequencies will be multiples of each other, so we must choose to either lose information (via coarse binning) or introduce data sparsity to our models (via fine binning).
- Model accuracy is dependent on the information present, so we leverage fine binning as a preprocessing step (Figure 4).
- Our sparse data fields are not interpolated due to the fact that interpolation could constrain the learning process and induce bias.
- Missing values in sparse categorical fields are replaced by special missing value tokens which are dependent on whether the missing field was discovered by the algorithm or created by the algorithm.
- Fields are then embedded into a high-dimensional feature space by neural network layers that contribute to the end-to-end training pipeline.



	Shift 1	Shift 2	Shift 3
Observations	Pain Level: High		Pain Level: Low
Coarse Binning (An alternative approach)	Pain Level: High, Medium, or Low Information is lost regardless of aggregation technique		
Fine Binning (Used in our approach)	Pain Level: High	Pain Level: Unobserved	Pain Level: Low

Figure 4: Two different binning techniques. Our approach leverages a fine binning approach. Note how instead of imputing an interpolated value for shift 2, our approach explicitly marks that no data was observed, allowing the meta model to generate its own conclusions about the importance of missing data.

Time-Series Global Interpretability

- Standard interpretability measures, such as linear mimic models and gradient boosted tree mimic models[1], operate with strong independence assumptions over the data[4].
- Longitudinal data, by definition, implies that feature-level independence over time cannot be assumed[5].
- We contribute to the model interpretability community by leveraging a vanilla recurrent neural network to model global time-series interpretability.
- This neural network models a first-order Markov decision process, simplifying interpretability greatly due to the fact that *the final classification/regression score is independent of all prior information given the model's most recent prediction.*[3]
- Due to the fact that each recurrent step uses the same weights, we can fold high-dimensional spaces with $f * t$ features into a much lower space with $f + 1$ features.

Providing Real-Time Clinical Insights

- Heartwood Analytics^{TM} contains a web interface designed to assist healthcare staff in their daily decision-making processes.
- The output of the Heartwood Synthesis feature importance process is provided to this interface, allowing users to view the most impactful longitudinal features that were learned by the model and the mimic model.
- This interface provides both quick "what-if" information based on the linear mimic model (Figure 2) and detailed feature analysis for general informational purposes (Figure 5).
- Due to the flexibility of our approach, Heartwood can successfully provide both individual and population-level insights and is deployed in multiple care centers and contexts.

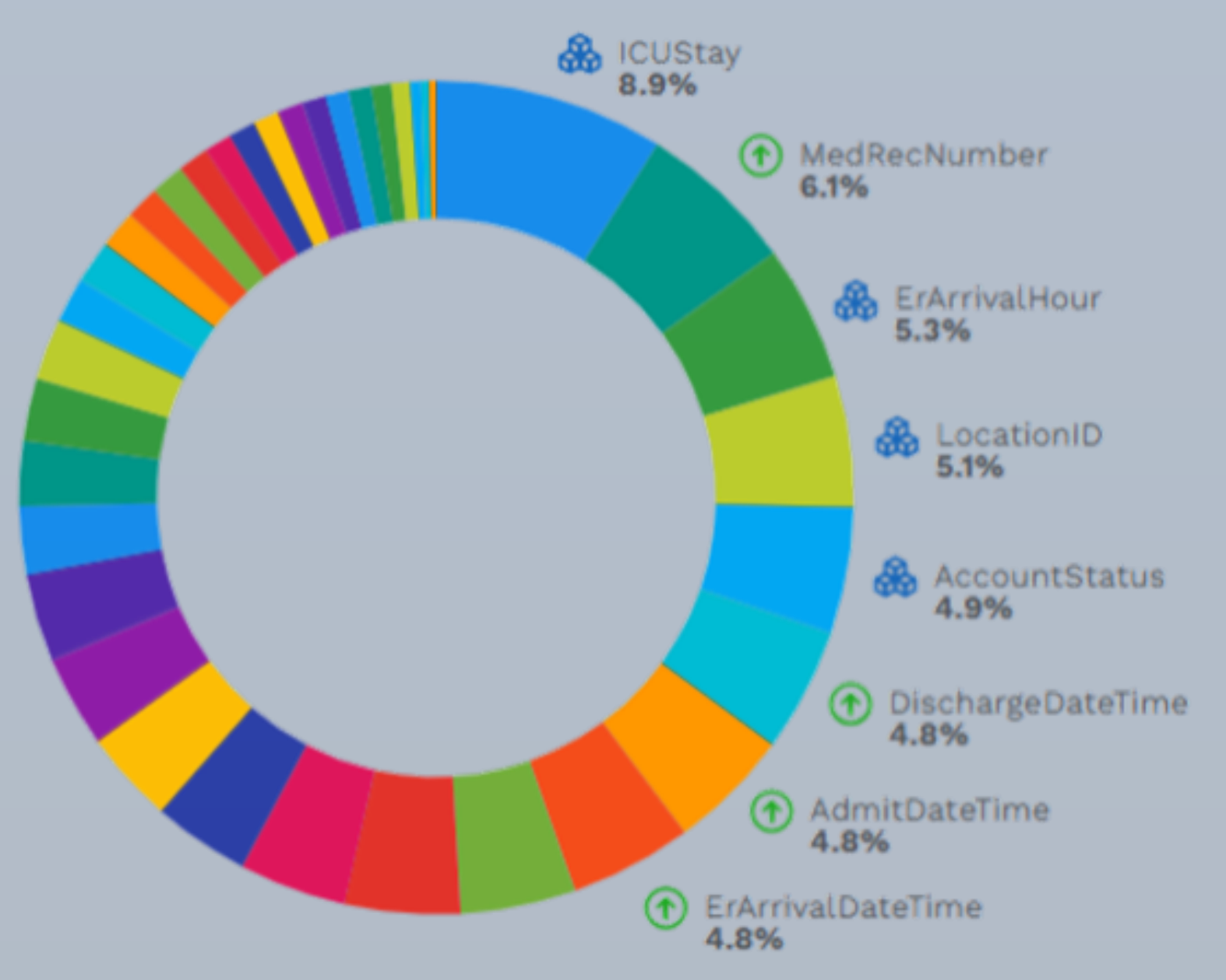


Figure 5: An example of the Heartwood Analytics feature importance user interface that provides users with detailed information about which captured data elements provide the greatest impact on an individual's disposition, allowing for more effective decision-making.

Datasets Used for Development

- Heartwood contains multiple examples of distributions for modeling that are used for internal research, benchmarking, and development.
- One of these examples, the pain dataset, is a univariate time-series dataset that can be used with ordinal regression or classification. The independent variable is measured pain over time, while the dependent variable is a ranking based on the severity of the pain. Severity is a mildly noisy aggregation of the most recent features, weighted according to the amount of time that has passed.
- Heartwood's multivariate nutrition dataset models the impact of eating and exercise habits on overall fitness. Variables such as amount of exercise performed (categorical), amount of macronutrients consumed (ordinal), and time of interaction have quadratic interactions over time to produce a classification measure of "fit" or "unfit."

Results

Ongoing Validation in Clinical Settings

As a result of the initial successes seen by Cropp et al.[2], Heartwood analytics has seen steady growth and adoption in the healthcare industry. CUBRC's primary partners operate in the area of complex care and continue to field, test, and iteratively improve Heartwood alongside our team each day.

Currently, models that were selected, trained, and analyzed by Heartwood have been deployed for predicting maladaptive behaviors in complex care environments at varying time intervals. As one example, partners use Heartwood models and the Heartwood feature importance framework to understand the driving factors correlated with aggressive behaviors in inpatient settings. With this knowledge, they are able to supplement their decision-making and focus on more impactful preventative care measures.

Pain Level Prediction

- Pain level modeling in Heartwood (see Methods) selected a convolutional neural network with over 99% accuracy.
- Mimic model accuracy retained the majority of information, maintaining 98% accuracy.
- A person's pain was determined by the mimic model to be highly correlated over time (previous timestep had a weight of 0.98) and to have a powerful inverse correlation with pain at the current timestep (weight of -1.16).

Nourishment Prediction

- Nourishment modeling (see Methods) selected an LSTM network with poor accuracy for both the nutrition and exercise data.
- Mimic model accuracy, however, retained the majority of learned information, maintaining nearly perfect accuracy in mimicking successful predictions.
- Previous predictions had 58% of the weight, followed by date with 27% and quantity of food consumption at 15%.
- This emphasizes that our approach is highly dependent on the ability of the underlying model.
- Poorly curated models (due to data filtering, signal-to-noise ratio, etc.) may emphasize the wrong features, which can be identified in our visualizations.

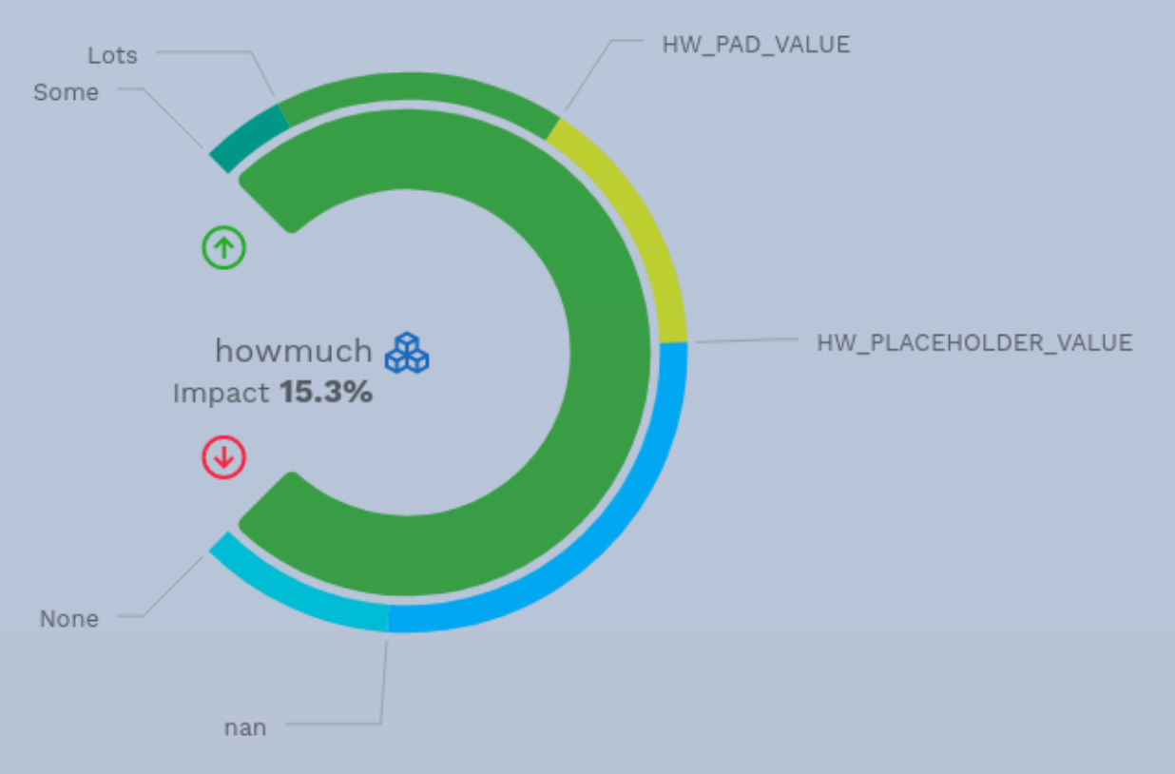


Figure 6: The feature importance chart produced for the quantity of consumption (how much) feature group in the exercise-based nourishment model. Eating an appropriate amount supports the model's prediction of being healthy, while failed observations or failure to eat is correlated with model predictions trending towards unhealthy dispositions.

Conclusions

Our approach to time series modeling is a sufficiently generic approach that can satisfy needs across a variety of longitudinal datasets. For those datasets with sparse data, we propose a unique way to analyze these fields without removing information from the dataset. Overall, we believe that our modeling and visualization approaches combine to produce actionable intelligence for healthcare staff. Through current and future fielding of our technology, we anticipate that our methodology will hold with small modifications depending on the complexity of the feature space being studied.

Limitations

- Many real-world scenarios are complex processes that require deep architectures to fully understand which features serve as indicators. As the complexity of an architecture increases, an interpretable mimic model will trade off larger amounts of performance in order to maintain sufficient communication ability.
- Our mimic model is trained via gradient descent, which does not guarantee that a global minimum will be found when the model reaches convergence. Therefore, model performance must be considered when using our method to supplement decision-making.
- Our tool must be used as supplementary knowledge instead of providing an explicit automated intervention due to the lack of theoretical bounds for deep network model error and bias.

Future Work

Future work will revolve primarily around modifying the global interpretability pipeline to focus on indicator features. For example, instead of indicating the relative impact of a feature, our model will be able to provide insight into the areas in which a feature is a sufficient indicator (the risk of entering a diabetic coma is affected by blood sugar when blood sugar is above a specific threshold). By using this technique, our model's accuracy could improve significantly with minimal cost to interpretability.

Funded by CUBRC, Inc. under the Health Informatics Internal Initiative

References

1. Che, Zhengping et al. "Interpretable Deep Models for ICU Outcome Prediction." AMIA Annual Symposium proceedings. AMIA Symposium vol. 2016 371-380. 10 Feb. 2017.
2. Cropp, Brett et al. "An Automatic Data Mining Approach Using Linked Data Technologies." AMIA Annual Symposium proceedings. AMIA Symposium vol. 2016. November 12-16, 2016.
3. Givan, Bob and Parr, Ron. "An Introduction to Markov Decision Processes." Rice University, https://www.cs.rice.edu/~vardi/dag01/givan1.pdf. Accessed 29 October 2019.
4. Shalizi, Cosma. "Logistic Regression." Carnegie Mellon University, 2012, https://www.stat.cmu.edu/~cshalizi/uADA/12/lectures/ch12.pdf. Accessed 29 October 2019.
5. "What are Longitudinal Data?" National Longitudinal Surveys, https://www.nlsinfo.org/content/getting-started/what-are-longitudinal-data. Accessed 29 October 2019.