Project Title
Multivariate time series analysis of physiological and clinical data to predict patent ductus arteriosus (PDA) in neo-natal patients

Student Names and Institutions*Please list the student participants along with their institutions. Please use a separate line for each

*Erica George, Biochemistry and Molecular Biology, University of Maryland, Baltimore County (UMBC)
*Imaobong A. Udofoa, Computer Engineering, University of Maryland, Baltimore County (UMBC)
*Chris Mai, Computer Science, University of Maryland, Baltimore County (UMBC)
*Ofek Lev, Computer Science, University of Maryland, Baltimore County (UMBC)

Advisor Names and Institutions*Please list the advisors along with their institutions. Please use a separate line for each

*Patricia Ordóñez, PhD Candidate, Computer Science, University of Maryland, Baltimore County (UMBC)
*Marie desJardins, PhD, Associate Professor, Computer Science, University of Maryland, Baltimore County (UMBC)
*Kathryn Holmes, M.D., MPH, Assistant Professor of Pediatric Cardiology, Johns Hopkins School of Medicine

Goals and Purpose*Please describe the goals and purpose of the project.

A persistent patent ductus arteriosus results in lack of closure of the ductus arteriosus after birth. Depending on the pulmonary vascular resistance, a patent ductus arteriosus (PDA) may result in a significant left to right shunt in the newborn infant. In premature infants, a patent ductus arteriosus can contribute to considerable morbidity and mortality including, high output heart failure, intracranial hemorrhage, respiratory failure, necrotizing enterocolitis and renal insufficiency. Recognition of a patent ductus arteriosus is made clinically with echocardiographic imaging as gold standard. Therapeutic interventions include prostaglandin inhibitors and surgical ligation. As many patients require transfer to another center for surgical intervention, it is important to recognize these infants early in their course in order to avoid delay of care. Given the ability to track patient hemodynamic and laboratory parameters via computerized data collection, we hypothesize that there is a pattern that is predictive of patients who will be unresponsive to pharmacological closure.

The ultimate goal is to improve quality of patient care with the identification of a clinical pattern predictive of a PDA and determine whether that PDA is amenable to either pharmacological closure versus surgical closure.

Specific questions to be addressed; hypotheses to be investigated

We believe that it may be possible to

- Identify patterns in the data that would signal the development of a significant PDA hours in advance
- Develop a measure of similarity for these multivariate representations allowing physicians to identify patients with similar events
- Create a visualization that will assist providers in examining patient data more thoroughly and efficiently

Process*Please describe the process used in completing the research.

The process for the first two objectives consisted of first converting the data into a representation that was conducive to finding irregular patterns. The students did not participate in this task. Their role was to take the representation and apply machine learning techniques using Weka to determine whether they could cluster or classify the patients into two groups, those who had moderate/severe PDA and those who had trivial/no PDA.

After it was determined that the representation was capable of clustering or classifying the representation, then
we wanted to build a similarity metric for the representation that would improve on the machine learning techniques that performed the best.

Meeting the last objective of creating a visualization consisted of speaking to providers to determine how we could improve on an existing visualization, coding the improved visualization and then performing an evaluation with medical residents at the Johns Hopkins School of Medicine. The latter task consumed most of the semester and summer.

We attempted to convert the visualization project into a NetBeans framework; however, the visualization lost some of its functionality in the process. Thus, we reverted back to coding it in Swing. The CREU students were responsible for the majority of the development of the evaluation process. The students were trained and coached by physicians associated project on how to conduct the evaluation interviews. The students took the required JHU IRB training so that they could acquire the signature of the participant on a consent form approved by the JHU IRB. Unfortunately, the IRB process took so long that in the end only three out of twenty-three interviews were conducted by CREU students, because the school year ended before the IRB was approved.

Each interview consisted of three consecutive stages: an introduction, training, and diagnosis. In the introduction stage, a student met with a participant and followed a predefined interview script describing the interview process and acquired the signature of the participant in a consent form.

The goal of the training stage was to ensure that the participant understood how to use the visualization’s interface. More specifically, the participant needed to become familiar with the difference between the two views of the data; personalized and customized. As part of the training stage, the participant viewed a five minute video that described the visualization and how it could be used. This video was created by one of the CREU students. This video can be viewed online at http://ai.cs.umbc.edu/~ordopa1/creu/videos/.

The participant then received an instruction sheet that contained screen shots of the interface (see Figure 1 & Figure 2). The instruction sheet highlighted the most important points on how to use the visualization. Then they used the interface to visualize the data for two patients- one with no/trivial PDA followed by a one with moderate/severe PDA. The interviewer made certain that the participant could perform basic tasks using the interface such as display univariate plots, use the animation, find the minimum and maximum values for a parameter, and change the target values for a parameter before proceeding. In the final step of training, the participant took an online test on how to use the visualization. This test was developed by a CREU student. The test focused on the participant’s knowledge of the Fixed Dual Personalized and Customized Views, their differences, and the features they have in common. A sample of the online test can be found at http://ai.cs.umbc.edu/~ordopa1/CREUsur/limesurvey/index.php?sid=87761&lang=en. The participant was not allowed to continue with the evaluation until all 12 questions were answered correctly to reduce errors as a result of lack of understanding of the interface.

The diagnosis stage began with obtaining personal background information about the participant (e.g. years in practice) and asking the participant to rank their preferred way to view the patient’s data. The participant then proceeded to use the visualization software to predict PDA in 8 neonatal patients. The participant used the visualization for the first patient and thereafter alternated with the traditional view of the data until all 8 patients were diagnosed. Thus, the participant used the visualization tool and traditional view of the data four times each. The diagnosis stage ended with the participant ranking their preferred way to view a patient’s data one more time and voluntarily submitting any feedback about the software.
Figure 1: Front Page of Instruction Sheet in Evaluation

- **Personalized Fixed Dual View**

  - **Load a patient and change time intervals by selecting a patient ID, time interval or total time to view the data.**

  - **The red dotted circle denotes the patient’s average value for each parameter over the entire time period. Every data point is the standard deviation of a value from the average of the parameter.**

  - **Find the actual value of a data point by clicking on the point in question. Then click on the date in the “Data Window”**.

  - **The colors of the axes represent the category of the parameter. The colors can be changed by clicking on the button.**

  - **Press the “Start” button to start the animation.**

  - **Increase the time to slow down the animation.**
Conclusions and Results Achieved

Please list or describe the conclusions of the work and the results achieved

Because we were not able to create a representation that would cluster or classify patients into the two categories of patients with moderate/severe PDA and those with trivial/no PDA, we did not reach our objective of identifying patterns in the data that would signal the development of a significant PDA hours in advance. As a result, we never attempted developing a measure of similarity for these multivariate representations. We did make significant progress on the third objective of creating a visualization that will assist providers in examining patient data more thoroughly and efficiently.

We created Excel tables to represent the “Traditional” view of the data and allowed the providers to create any type of graph that they desired. However, they were only allowed to view 24 hours of data which was the same amount that was loaded into the visualization. We received complaints from participants that the Excel tables did not adequately represent the “Traditional” visualizations. Thus, a limitation of this evaluation is that we were not able to use a software package that accomplished that.

In determining the resident’s preference between the visualization tool and the traditional view of clinical data, we asked him or her to rank the Traditional, Customized, and Personalized views of the data before and after using the visualization interface to diagnose 8 patients. The Traditional view was defined as the interface that they are accustomed to using in their day to day routine and not the Excel interface that we created to simulate it. Prior to the diagnostic phase, 8 out of 23 residents preferred using the Customized view to the Traditional view, no one preferred the Personalized view, and 15 of the residents preferred the Traditional view. After the diagnostic phase, the same 8 residents still preferred the Customized view except one of them said they liked the Personalize view as much as the Customized view. Only 8 of original residents that preferred the
Traditional view continued preferring it. It follows that 7 of residents switched preferences after using the visualization of which 5 preferred the using Customized view and 2 the Personalized view. None of the residents switched to preferring a Traditional view after the diagnostic stage of the evaluation.

Overall, the providers had better accuracy with the visualization (43.5%) than with the Excel tables (39.1%). We decided to examine the accuracy for each patient to determine whether some patients were much more difficult than others as in the bar graphs in Figure 3. We also examined the patients in the order they were viewed to examine if there were any effects of continued use over time as in the line graphs in Figure 4. Because we lost three participants, each patient was viewed either 10 or 13 times using either the visualization or the Excel interfaces. We also examined the confidence levels and the efficiency as determined by the amount of time the provider took to examine a patient and come to a diagnosis. These graphs indicate that in terms of accuracy the visualization did as well with the visualization as they did with the tables even though the confidence levels indicate that the providers had less confidence in their answers when using the visualization. Considering that the providers had not worked with the visualization for more than 30 minutes before they began to use it, our results indicate that the visualization has a lot of potential. As mentioned before, although the visualization appears to be more efficient than the table, we were not able to simulate a Traditional view of the data. Thus, the time it took residents to select a range in the table and plot it may have been a factor in the lower times to diagnose a patient using the visualization.

Figure 3: Accuracy, Efficiency, and Confidence by Patient
Figure 4: Accuracy, Confidence, and Efficiency by Patient Order

Web links* Please provide links to your project web pages. If there is more than one, please list each on a separate line

http://ai.cs.umbc.edu/~ordopa1/creu/
http://creu2010-umbc.wikispaces.com/
Presentations and Publications

Please list the titles and venues of any presentations or publications resulting from this project. Please list each on a separate line.

