Jon Deaton, Diana Gong, Anna Jaffe
BIOE 141B
1/28/17

Milestone 2

2.0: Project Overview

Need statement: People uneducated about and at risk for non-exertional heat stroke in hot regions of India need a method to reduce the likelihood of experiencing heat stroke to decrease mortality.

To address our need statement, we are creating and testing a heat stroke risk prediction wearable device. The overall concept is a system of sensors that take environmental and the user's physiological parameters and connect to a mobile phone to estimate heat stroke risk. A complete project includes sub-projects of development of algorithms for risk assessment, wearable hardware, mobile-software, and experimental testing procedures; however, we have chosen to focus only on the first and last components (depicted in Figure 1) during BIOE 141A/B.

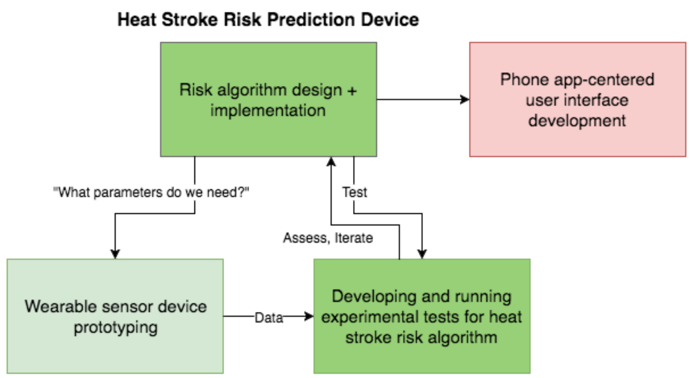


Figure 1: Complete concept subcomponent relationships and intended focus. Red indicates an area that we will not be addressing during BIOE 141A/B. Shades of green indicate areas that we will be addressing with our milestones, with dark green representing more focus and light green representing less focus.

We intend to implement a prototype that senses relevant parameters and need not necessarily be wearable, minimally intrusive, or connect wirelessly to a phone, as we aim to prove concept viability rather than produce a ready-to-use device. Our concept (Figure 2) involves sensors connected to a microprocessor that relays information to a computer, which predicts heat stroke risk using machine learning (ML) algorithms. By the en dof BIOE 141B, we aim to have developed hardware and software that allow us to accurately sense and predict heat stroke, as well as a series of experimental tests verifying that our device functions as expected.

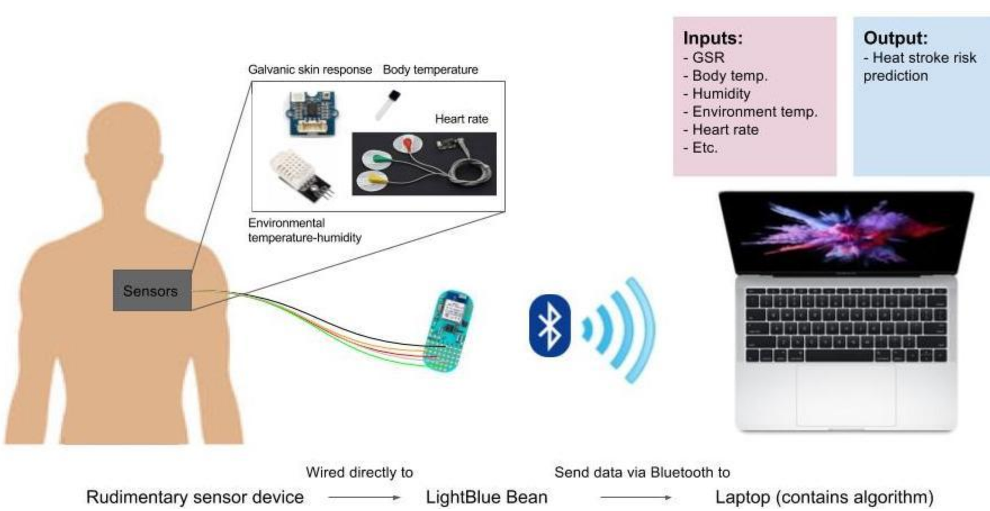


Figure 2: Graphic depicting our concept of a heat stroke risk prediction wearable device prototype.

Jon Deaton, Diana Gong, Anna Jaffe
BIOE 141B
1/28/17

2.1: Milestone Overview

Milestone 2: Design, implement, and statistically test prediction algorithms.

Milestone 2 is a crucial component of our project, as it involves developing heat stroke risk assessment algorithms. For our prediction models, we will generate negative data and use the patient data points collected in Milestone 1 that include environmental and physiological parameters. For our output, we aim to have **(1)** a list of algorithms ranked by predictive performance, **(2)** a working implementation of a heat stroke prediction algorithm, which if given appropriate data, will predict the risk of heat stroke, **(3)** statistical results indicating predictive performance of the test, and **(4)** a list of parameters accurately predicting heat stroke and that are collectable with a wearable device. We are measuring milestone success by obtaining estimated sensitivity, specificity, accuracy, and positive predictive value (PPV) for all prediction algorithms, as these metrics will allow us to evaluate our device in a situation where false positives are acceptable, false negatives are not acceptable, and accurate diagnosis is essential to the success of our device. Ideally, we will have one algorithm that has least 75% (chosen as a reasonable goal for this milestone, but leaves room for improvement) in each of these metrics.

2.2: Milestone Methods

Missing Data Points: The case study data for each heat stroke patient that we gathered in Milestone 1 did not have values listed for all 73 parameters that we examined, and we were unable to find similar data for patients who did not have heat stroke. To address the problem of missing data, we created software **(Appendix A.1 > read_data.py)** for (1) automatically filling in missing fields with replacement values in our case study data, and (2) generating data points that represent cases of no heat stroke. We recognized that developing better methods for making distributions from which to draw filler data points is important and have been actively working on this (see **Appendix B** for description).

Supervised Algorithm Implementation: To predict heat stroke risk given demographic, physiological, and environmental data, we developed software **(Appendix A.1 > predictor.py)** that uses data from Milestone 1 as supervised learning training sets. This supervised learning algorithm can use either Support Vector Machines (SVM) or logistic regression to output an indication of risk from zero to one.

Cross-Validation: We also developed software in Python that can perform k-fold cross-validation on these supervised learning algorithms using any subset of features in the training dataset **(Appendix A.1 > cross_validation.py)**. After performing cross-validation, our software reports predictive performance metrics, shows receiver operating characteristic (ROC) and precision-recall curves, and visualizes data points with SVM classification hyperplanes. We used this software to analyze our algorithm's predictive performance on our training datasets.

Unsupervised Model Algorithm Implementation: We implemented model-based prediction algorithms. For heat index (HI) risk estimation we used meteocalc to calculate HI given temperature and humidity, calculated risk based on the guidelines described in National Weather Service Weather Forecast Office. We also implemented an adapted Kalman filter model from Buller et al. (2013)[1] that estimates core body temperature given a series of heart rate measurements. We created a placeholder for implementing a system to predict heat stroke risk given a series of galvanic skin response (GSR) and skin light absorbance measurements. This algorithm checks for rapid decrease in GSR after prior elevation, combined with a decrease in skin light absorbance.

[1] Buller MJ et al. "Estimation of human core temperature from sequential heart rate observations." Physiol Meas. 2013 Jul;34(7):781-98. doi: 10.1088/0967-3334/34/7/781. Epub 2013 Jun 19.

Modular Software Design: Finally, we developed software that encapsulates our prediction algorithms and data handling (**Appendix A.1**). For this portion of the milestone, we used object-oriented design to create classes that modularize patient demographic data, data stream parsing and storage, prediction generation, and execution management (Figure 3).

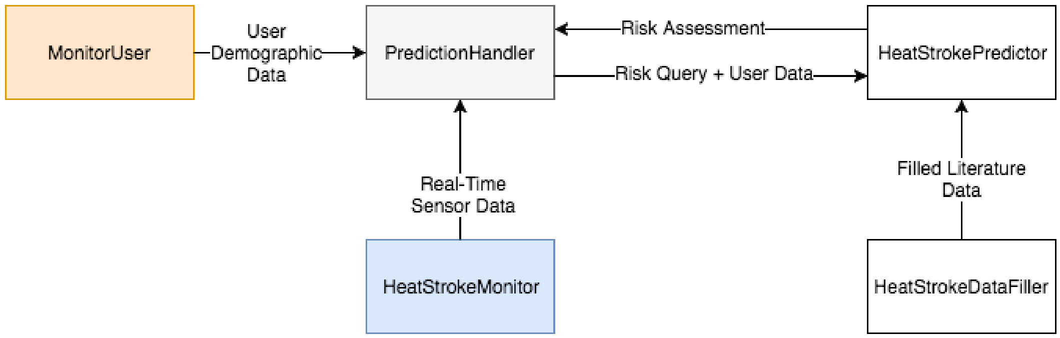


Figure 3: Block diagram depicting class structure. Relationships between classes used in the software (implemented in Python) for Milestone 2. The MonitorUser class is a wrapper for patient demographic data. Static user data is stored in an XML file between runs, and the specific user can be specified on startup. HeatStrokeMonitor is a class that interfaces with the bluetooth Serial port through which data is transmitted from the physical monitor (sensor system), and also stores data retrieved from the data stream in time-associated tables. The PredictionHandler class (currently incomplete) periodically requests data from the HeatStrokeMonitor class, combines this data with the demographic data from the MonitorUser class, and sends the combined data to the HeatStrokePredictor class. The HeatStrokePredictor class is designed to take user data streams and produce a risk assessment by retrieving case study data from the HeatStrokeDataFiller class.

2.3: Milestone Results

Both our SVM and logistic regression classifiers classify our data with ~97% accuracy, ~92% F-Score (output from **Appendix A.1 > cross_validation.py**), and >95% specificity, sensitivity, and PPV (Figure 4). It is important to note that these values are high predominantly due to our method of filling in missing data, because some features are filled using non-overlapping distributions for positive and negative cases, making the data easily classifiable. Developing the software that can visualize the distributions of values that are chosen in our dataset (Figure 5) helped us to choose more reasonable distributions from which to sample data. Visualizing these scatterplots with the SVM classifier line also shows us how our prediction will be affected by the data that we use in the system (e.g., Figure 5B vs. Appendix C shows how changing the negative dataset to cover a wider range of possible cases makes the classification line vertical, indicating that heart rate is not a distinguishing feature).

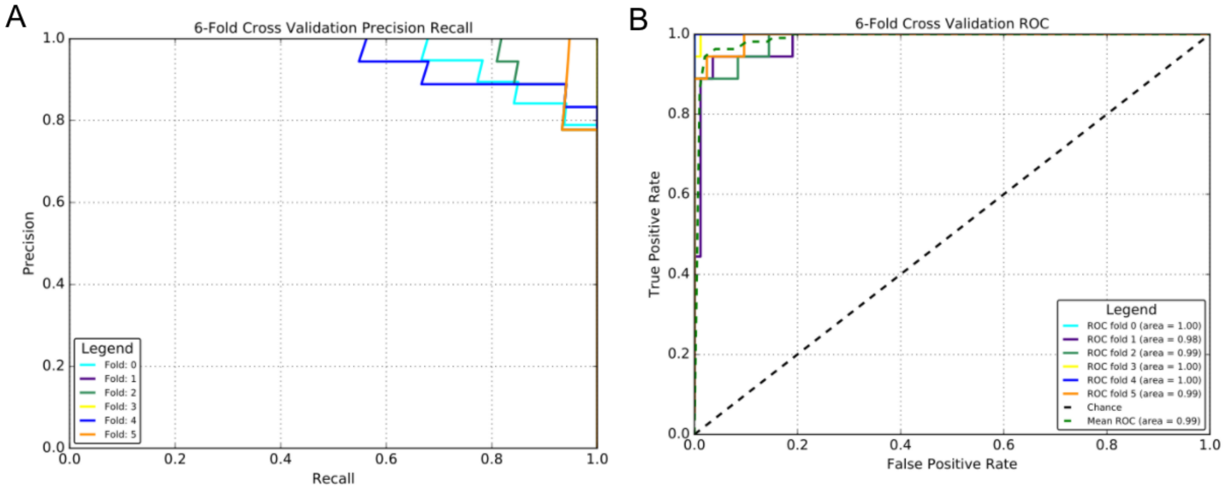


Figure 4: Precision-recall curve using K-fold cross-validation (A) and ROC curve (B). A) Plot of precision vs. recall for an SVM classifier on the features of patient temperature (1), HI (2), relative humidity (3), environmental temperature (4), and heart rate (5). A perfect classifier would have recall and precision equal to one. Each line in the plot represents the performance for the SVM classifier for one of the six folds used in cross-validation. Precision-recall curves in general depict the tradeoff between false positives and false negatives. For the purposes of this project, we are willing to trade off precision (PPV) for sensitivity. This is because it would be worse for someone to get a heat stroke without being warned than for someone to be warned who is not at risk for heat stroke. **B)** ROCs generated by a SVM classifier using patient temperature (1), HI (2), relative humidity (3), environmental temperature (4), and heart rate (5).

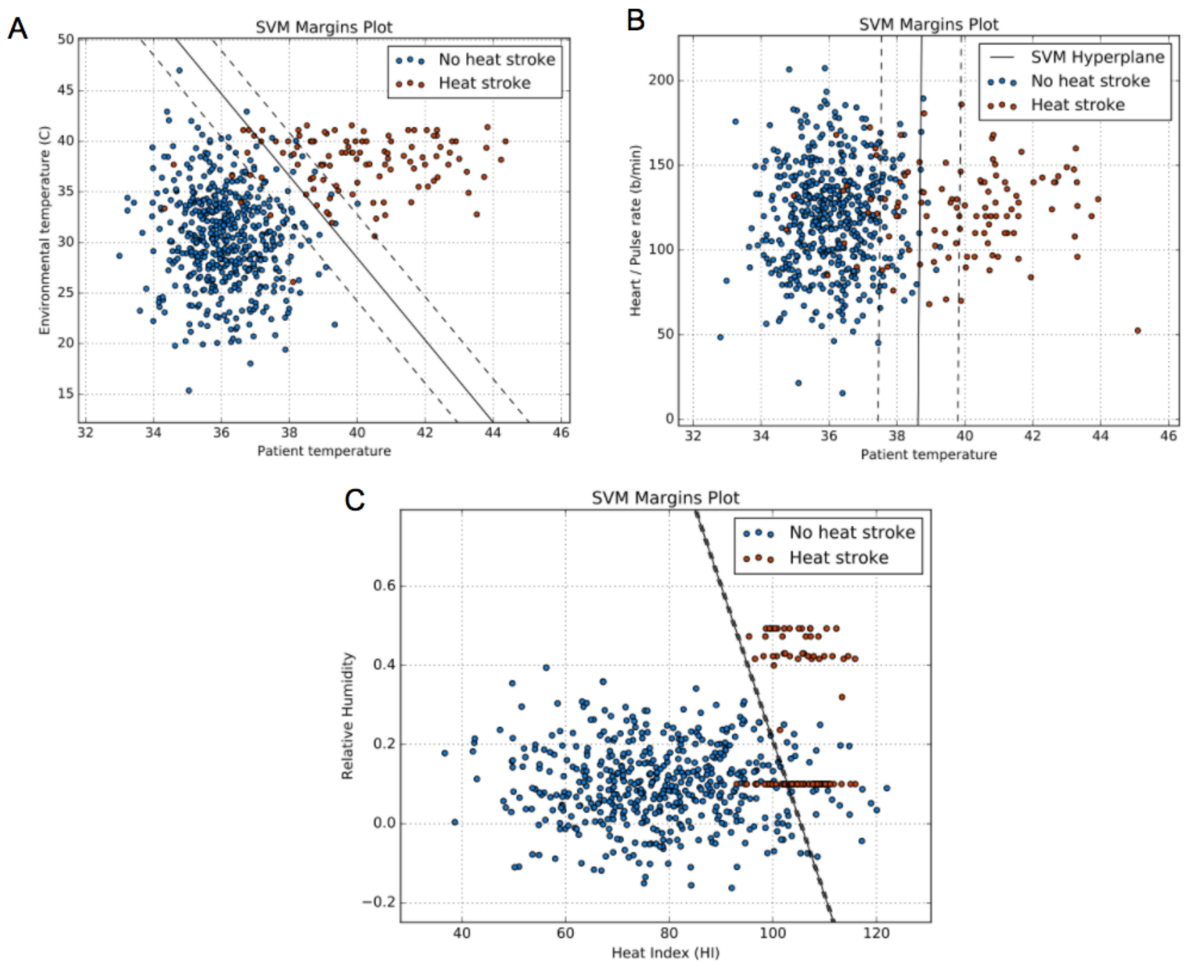


Figure 5: Examples of SVM margins plots. The solid line in each plot shows an SVM classification of the data with margins (dotted lines). Many of the heat stroke data points are from true patient populations; however, all of the blue data is drawn from distributions of physiologically normal ranges. **A)** Environmental temperature vs. patient core temperature. **B)** Heart rate vs. patient temperature. Given that many points from both classifications lie on both sides of the line indicates that these two features alone may be insufficient in predicting heat stroke. **C)** Relative humidity vs. HI.

2.4: Milestone Interpretation/Discussion

- (1) We initially planned to implement different prediction algorithms and create a ranked ordering by performance, but decided that this was not useful. In working on this milestone, we found that developing one solid framework and continuing to tune the information given to the algorithm to optimize performance is the best use of our time. Furthermore, because both of the ML algorithms that we have created (logistic regression and SVM) can be made to perfectly classify our data depending on the distributions used to fill missing parameter values, predictive performance is based on the provided data rather than the algorithm itself.
    - We instead focused on generating data points for non-heat stroke cases. We are creating more data than originally anticipated, so we will continue to find better “physiologically normal” data as we continue with the other milestones. Ideally, some of this information will come from Stanford labs with controlled environments that measure physiological parameters.
    - We have also developed methods to visualize our data (Figure 5).

- (2)/(3) We have created a working heat stroke risk assessment algorithm, evidenced by the statistical results (>75% accuracy, F-Score, specificity, sensitivity, PPV) indicating the predictive performance of each test.
- (4) Because we are generating so much data based on our assumptions, we can essentially control which parameters in our algorithm are predictive of heat stroke. Therefore, determining the most important parameters is not relevant at the moment. If we are able to later find empirical physiologically normal data, we can discover which parameters are truly predictive.
- We completed Milestone 2 by the due date (1/29/17).
    - We are continuing to work on better data-filling methods with more justified data (Appendix B).

2.5: Supporting Material/Appendices

Appendix A: Main Github page for heat stroke prediction code.
Appendix A.1: "src" folder.

Appendix B: Matrix containing relevant parameters, with mean and standard deviation for each. Links to spreadsheet with calculations and histogram for each parameter are included.

Because the patient data we obtained from literature in Milestone 1 did not include measurements for all 76 parameters that we examined (e.g., systolic blood pressure, serum albumin), we have begun an improved method (that has not yet been implemented) of generating data to fill in the missing variables. With the method that we are currently using, we estimated a value for each parameter and used the same value across all patients. For example, a patient missing an entry for respiratory rate would have been assigned 16 breaths/min. Our new method involves taking all measurements for each category across all our collected patient data points, determining if the distribution looks roughly normal, and calculating mean and standard deviation **(Appendix B, Figure B.1, Figure B.2)**. Then, for each patient and each missing parameter, we randomly generate a number from the appropriate distribution. With this new method, each patient missing an entry for hemoglobin, for example, will be assigned a value randomly drawn from a normal distribution with mean 32.3778 and standard deviation 9.8125.

Parameter	Old Value	New Value(s)	Notes
Patient temperature	37	normal dist with mean 40.55871429 and stdev 1.575845415	https://docs.google.com/
Sex	F	50/50 M/F	http://onlinelibrary.wiley.c
Rectal temperature (C)	37	normal dist with mean 42.16334364 and stdev 1.10101409	https://docs.google.com/
Respiratory rate	16	normal dist with mean 32.37777778 and stdev 9.812535788	https://docs.google.com/
Daily ingested water (L)	3.7		
Sweating	0.5		all literature data is a 0-
Skin color	1	uniform dist from 0 to 1 where 0.17 is 0, 0.1 is 0.5, and 0.73 is 1	https://docs.google.com/
Heart/Pulse rate	80	normal dist with mean 122.9210526 and stdev 23.22858191	https://docs.google.com/

Figure B.1: Representative screenshot of missing parameter matrix. Parameter, previous values, new values, and notes and/or links to spreadsheet with calculations are shown.

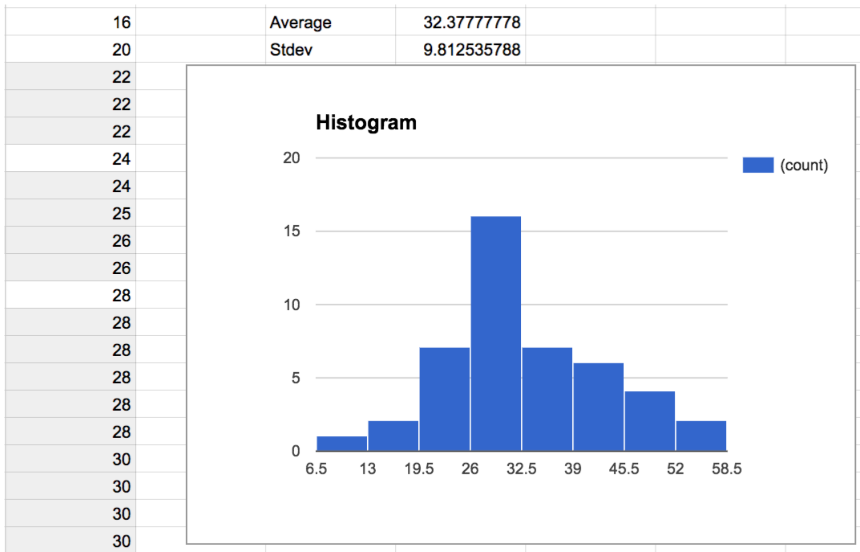
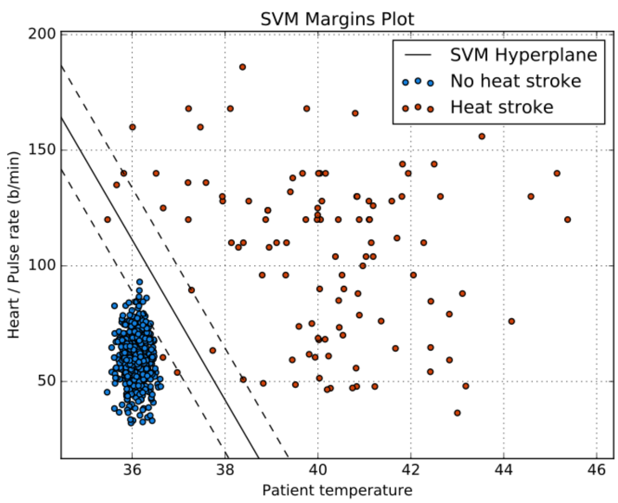


Figure B.2: Representative screenshot of literature data used to calculate average, standard deviation, and histogram for each parameter. Respiratory rate is shown as an example.

Appendix C: Poorly generated negative data.



These data do not represent a suitable range of situations (empirically determined) and thus are not useful for classifying heat stroke risk. Making SVM plots was useful for visualizing how reasonable our negative data (no heat stroke) was. Following the generation of this graph, we edited our negative data to generate the SVM plots in Figure 5.