Eric Chiang, Jinrui Liu, Luchao Qi, Jack Wright, Yiyuan Zhang
Group 2

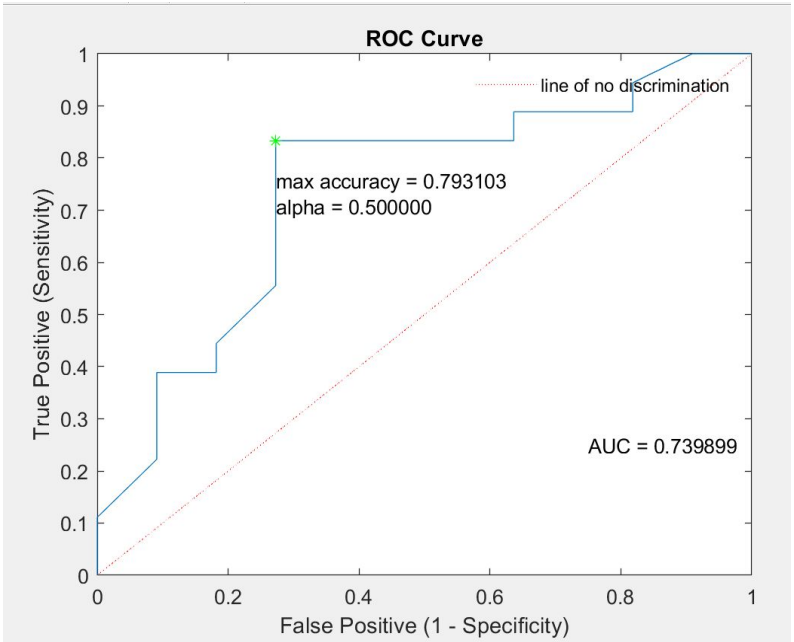
Sepsis Project

PART I: "STATIC MODELS"

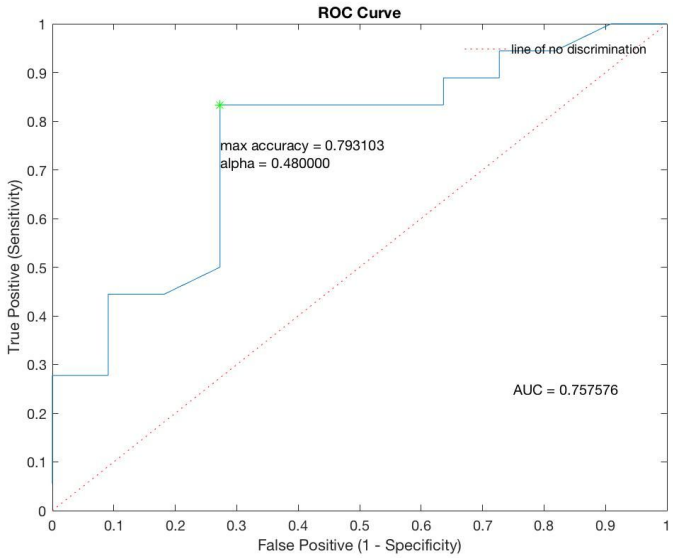
1. Justify the features used in your final model. Describe briefly how you explored the dataset and how you picked your final features (e.g. reference literature, show boxplots or distributions of sepsis vs. non-sepsis values, etc). Explain why you believe they might be informative for predicting sepsis.

For our final model, we selected columns 3-5 as features (i.e. Gender, Age, and Respiratory Comorbidities). Multiple sources indicate increase in sepsis rate in older patients. Similarly, infection rate in males tends to be slightly higher than the rate in females, so we chose to include these two columns as features. Many comorbidities have been investigated as risk factors of sepsis. Typically, incidence of pulmonary infection or other respiratory complications tends to be one of the most common conditions among cohorts of sepsis patients. Therefore, we chose to include this feature. The relationships between sepsis and other issues are quite varied, however. As such, we chose to exclude column 7, since we are unable to effectively distinguish where these additional infections occur in the body without changing the dataset, and some infections have low comorbidity rate with sepsis. Lumping all of these features together is unlikely to be informative. Column 6 was the least clear whether to include or not. Heart measurements are often included in models, but comorbidities of the heart appear to be less definitive than respiratory comorbidities. To alleviate this uncertainty, we simply tested different combinations of features for our model, and made note of the accuracy rate, AUC, and p values. In general, including column 6 didn't improve accuracy, and the p value was often large, so we chose not to include it. Ultimately, we found using columns 3-5 to provide the best balance of accuracy and reasonable p values.

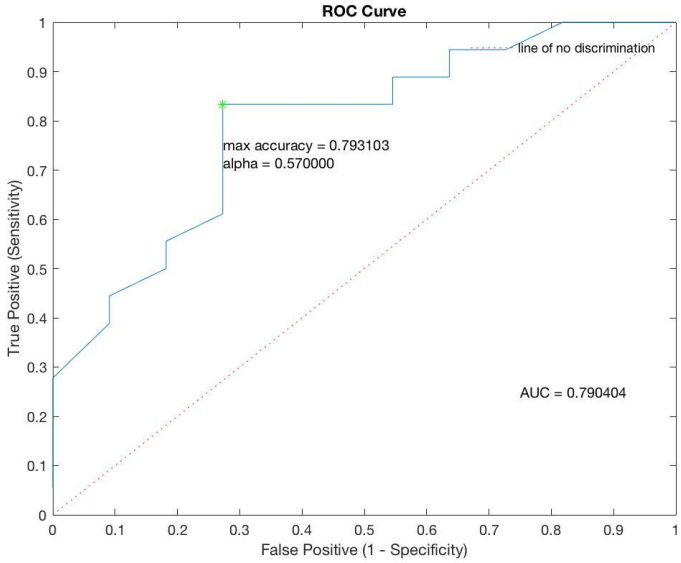
Examples:
3:5 P-values: 0.1695, 0.2329, 0.0564, 0.7516



3:6 P-values: 1.0, 0.2840, 0.0506, 0.9405, 1.0



3:7 P-values: 1.0, 0.2079, 0.0866, 0.9935, 1.0, 1.0



Combinations excluding columns 3 and 4 had worse results than the ones shown above.

Sources:

D. C. Angus, W. T. Linde-Zwirble, J. Lidicker, G. Clermont, J. Carcillo, M. R. Pinsky, Epidemiology of severe sepsis in the United States: Analysis of incidence, outcome, and associated costs of care. *Crit. Care Med.* 29, 1303–1310 (2001).

Katharine E. Henry, David N. Hager, Peter J. Pronovost, Suchi Saria (2015). A targeted real-time early warning score (TREWScore) for septic shock. *Science Translational Medicine*, Vol 7, Issue 299.

Brun-Buisson C, Doyon F, Carlet J, et al. Incidence, Risk Factors, and Outcome of Severe Sepsis and Septic Shock in Adults A Multicenter Prospective Study in Intensive Care Units. *JAMA.* 1995;274(12):968–974. doi:10.1001/jama.1995.03530120060042

Esper AM, Moss M, Lewis CA, Nisbet R, Mannino DM, Martin GS. The role of infection and comorbidity: Factors that influence disparities in sepsis. *Crit Care Med*. 2006;34(10):2576-82.

2. Plot your parameter estimates, αi, or list them in a table, along with their 5% and 95% confidence bounds.

Parameter	Parameter Value	5% Lower CB	95% Upper CB
α_0	-2.7208	-6.1072	0.6656
α_1	1.0391	-0.4503	2.5284

α_2	0.0476	0.0049	0.0903
α_3	-0.3273	-2.0954	1.4407

3. Complete the missing code in test_performance.m and modify classify_test.m to implement and test your GLM on the test data set (for the dynamical model, you can use the training data to make sure your code will run on the test set after you submit).

For this part, we wrote a new script, **classify_test_part1.m**, and made an ROC curve on both the training and testing data. Results are shown in question 6.

4. For the dynamical model, explain your decision rule if you came up with one and justify whichever you think is more useful; time-by-time prediction or patient specific classification.

Not applicable to this model.

5. Plot the model, $\hat{p}i$, for each patient and overlay the 5% and 95% confidence bounds.

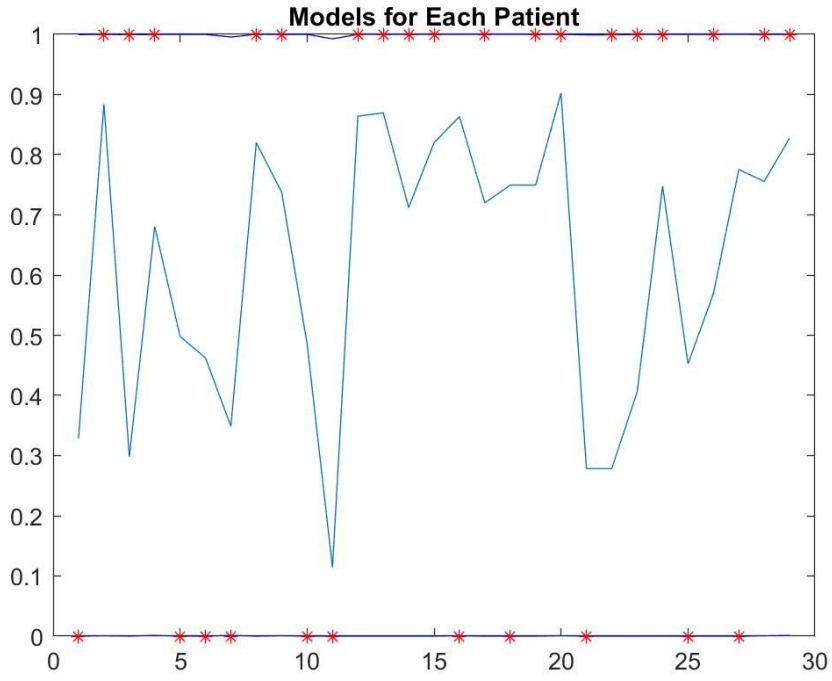


Fig.1. Probability for each patient with training data

6. Report the results of your final GLM model on the training set (and on the test set for the static model). Make sure to include the accuracy of your model, the ROC plot and the AUC value.

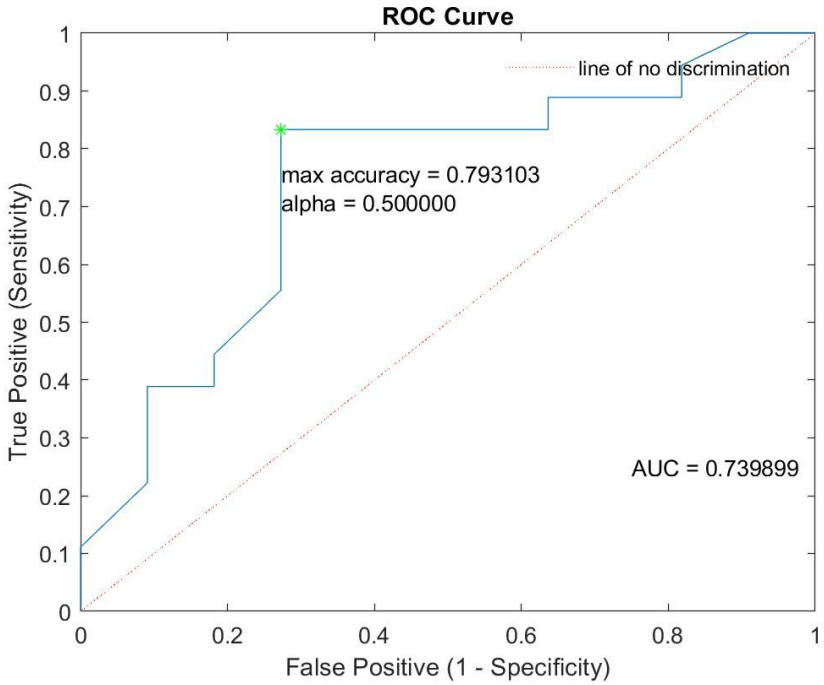


Fig.2. ROC curve with training data

The accuracy is 0.793103, AUC value is 0.739899.

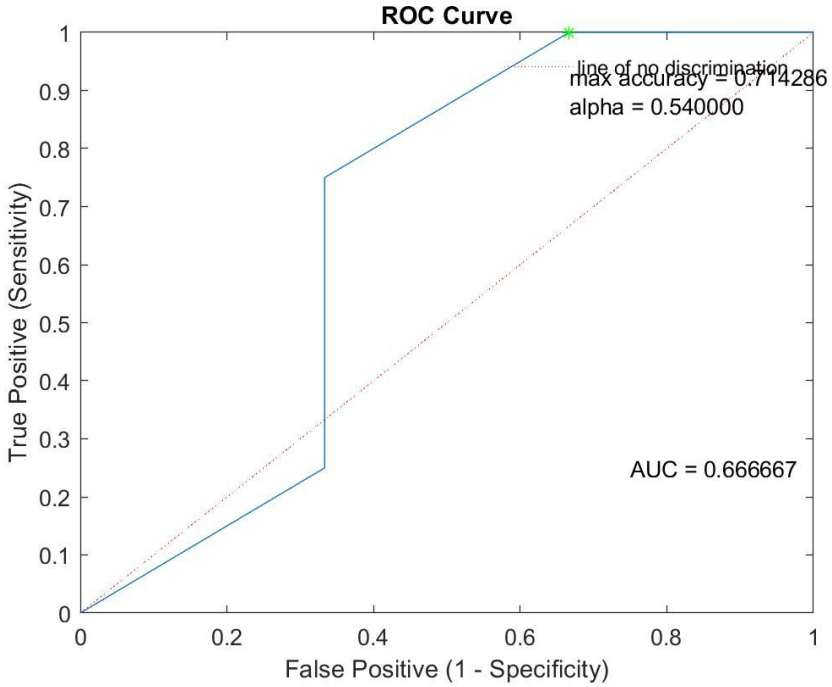


Fig.3.. ROC curve with testing data

The accuracy is 0.714286, AUC value is 0.666667.

7. Discuss performance differences on the training vs. the test set. Is this what you expected?

Results using test data set are below:

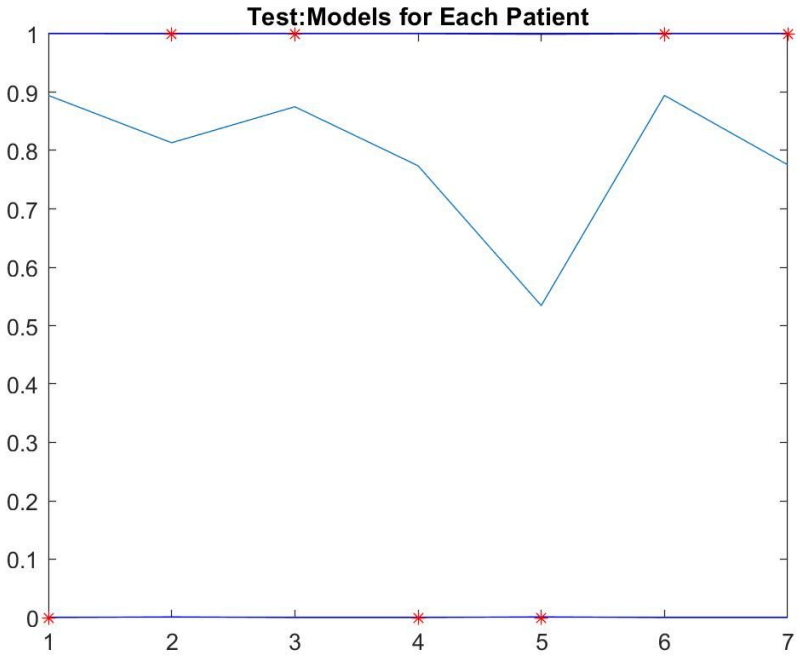


Fig.4. Static models for each patient with testing data

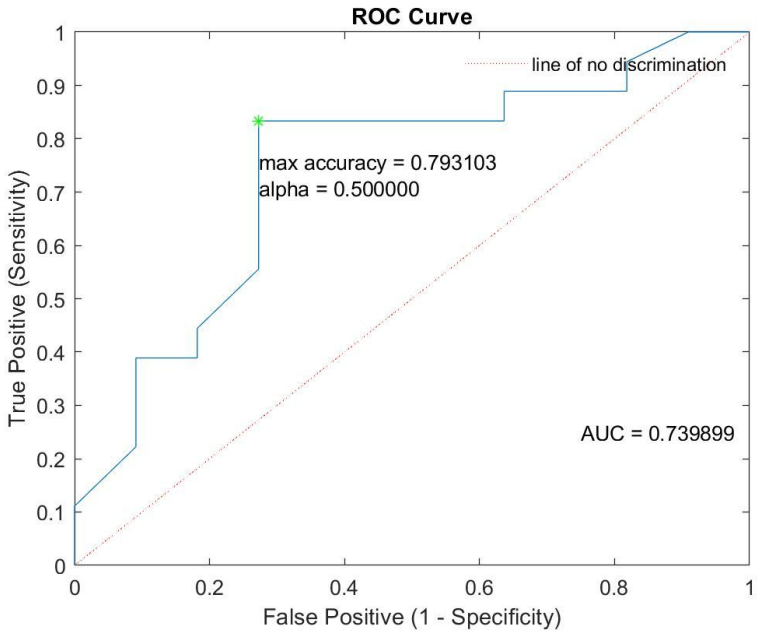


Fig.5. ROC curve with training data

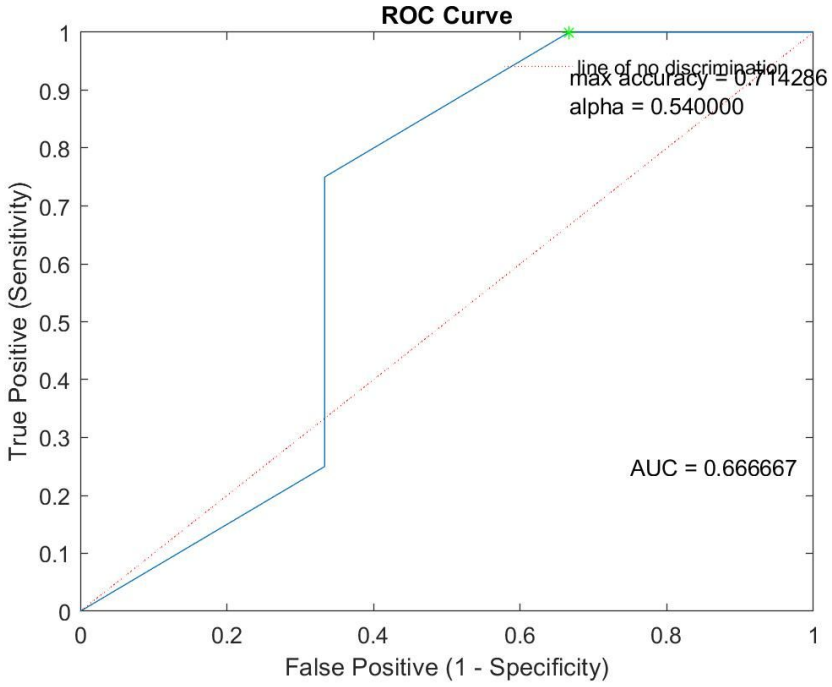


Fig.6. ROC curve with testing data

The above ROC curves illustrate the performance of our model on both the training and testing data. By using gender, age, and respiratory comorbidities as features, we achieved our best results on the training data. Our accuracy was 0.793103 with an alpha of 0.5 and an AUC of 0.739899. By applying this model to the testing data set, it performed similarly, but slightly worse, which is not unexpected. The accuracy on this set was 0.714286, with an alpha of 0.54 and an AUC of 0.666667.

PART II: "DYNAMIC MODELS"

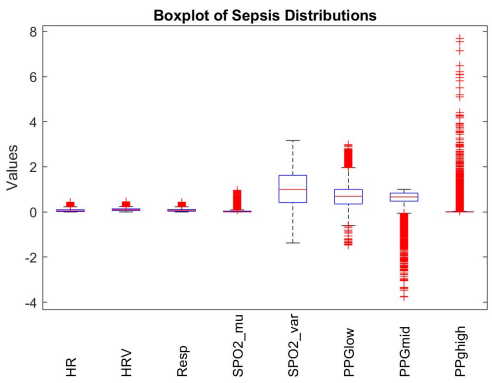
1. Justify the features used in your final model. Describe briefly how you explored the dataset and how you picked your final features (e.g. reference literature, show boxplots or distributions of sepsis vs. non-sepsis values, etc). Explain why you believe they might be informative for predicting sepsis.

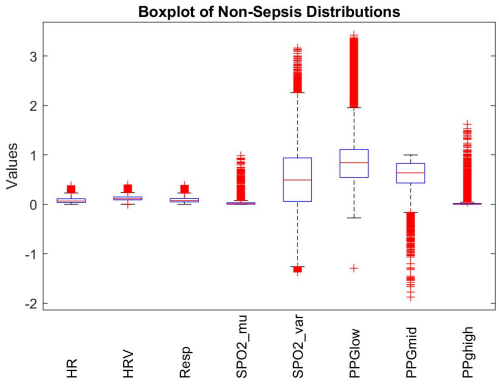
We added the following features in our dynamic model

- Heart rate
- Respiratory rate
- Mid-PPG
- Low-PPG

Firstly, we chose to retain the features selected for the static model for our dynamic model. Thus, we needed only to decide which of the dynamic models we should use. We started this by

using k-fold stratified validation to compare the “goodness” of a model with a combination of specific features using the data provided. Unfortunately, it wasn’t distinctly clear which variables we should use from this. We then turned to a combination of looking at the distribution of variables between sepsis and non-sepsis patients and literature searches to verify which variables to use.





According to *Fábio M. de Castilho et al.* and *Senthil K. Nachimuthu et al.* in early stage sepsis, patients demonstrate reduced heart rate, heart rate variability (HRV) as well as increased respiratory rate, comparing to healthy patients. Thus, we adopted those three parameters into our model and for sepsis detection. Despite no study has been conducted showing that HRV can be used as an independent predictor of sepsis, for which further study need to be conducted, HRV is statistically proven to be useful for sepsis detection when it corresponds with other parameters.

According to Hummler et al. [1], the accuracy of pulse oximetry is decreased during low perfusion caused by sepsis, which indicates that mean and variance of S_pO_2 data measurement may be unreliable under sepsis condition and they may not be good sepsis predictors. Thus, we choose to exclude S_pO_2 data in our model.

For PPG data, we choose to include low frequency power of PPG because, according to the box plot, the outliers of low frequency power have quite different patterns of distribution. In addition, Middleton et al.[2] found that 0.1Hz ear blood flow oscillation (corresponding to mid frequency power of PPG) is linked to tissue metabolic changes in sepsis. So, we decided to include mid frequency power of PPG in our model.

From both the box plots and literature search, we found heart rate, respiratory rate, low frequency power of PP, and mid frequency power of PPG to be our predictor variables from the waveform data. We appended these to the static variables: gender, age, and respiratory comorbidities.

We found the following P values using these parameters. All P values were less than 0.05 suggesting they all contributed to the model.

*P = 1.0e-16 * 0.3302,*
*1.0e-16 * 0.0000,*
*1.0e-16 * 0,*
*1.0e-16 * 0.0000,*
*1.0e-16 * 0,*
*1.0e-16 * 0,*
*1.0e-16 * 0,*
*1.0e-16 * 0.0000*

Sources:

Hummler, H.D., Engelmann, A., Pohlandt, F. et al. Intensive Care Med (2006) 32: 1428. https://doi.org/10.1007/s00134-006-0254-y

Middleton, P.M., Tang, C.H.H., Chan, G.S.H. et al. Med Biol Eng Comput (2011) 49: 337. https://doi.org/10.1007/s11517-010-0713-z

Nachimuthu, Senthil K and Peter J Haug. “Early detection of sepsis in the emergency department using Dynamic Bayesian Networks” *AMI... Annual Symposium proceedings. AMISymposium* vol. 2012 (2012): 653-62.

de Castilho, Fábio M et al. “Heart rate variability as predictor of mortality in sepsis: A prospective cohort study” *PloS one* vol. 12,6 e0180060. 27 Jun. 2017, doi:10.1371/journal.pone.0180060

2. Plot your parameter estimates, αi, or list them in a table, along with their 5% and 95% confidence bounds.

Parameter	Parameter Value	5% Lower CB	95% Upper CB
α_0	0.425717	0.342703	0.50873
α_1	0.27347	0.244405	0.302536
α_2	0.032466	0.031489	0.033442
α_3	-0.61104	-0.65051	-0.57158
α_4	-50.6053	-52.3721	-48.8385
α_5	47.97423	46.15085	49.79762
α_6	-1.04825	-1.0718	-1.0247

α_7	0.263555	0.222509	0.304601

3. Complete the missing code in test_performance.m and modify classify_test.m to implement and test your GLM on the test data set (for the dynamical model, you can use the training data to make sure your code will run on the test set after you submit).

This was done with a new file, *classify_test_part2.m*. This script has a commented section that can be uncommented to run with training data. In addition, ROC curves are generated for both the training and testing data. In this case, since testing data wasn't provided, duplicate ROC curves were generated for the training data. Once testing data is provided and the appropriate section is commented out, ROC curve for the testing data will show up. The ROC curve is shown in question 6.

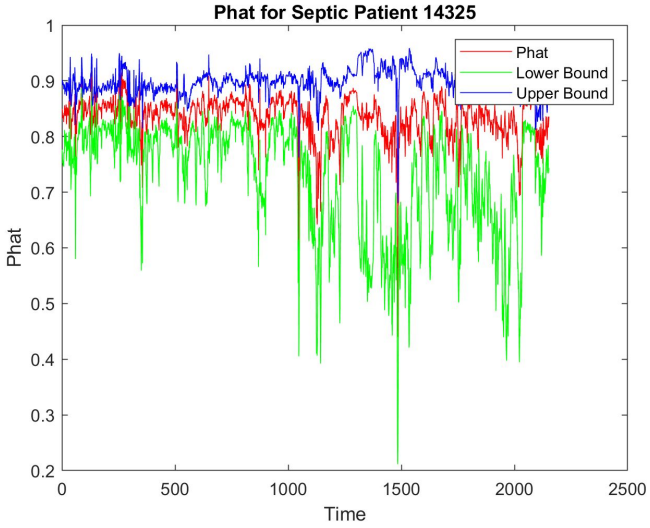
4. For the dynamical model, explain your decision rule if you came up with one and justify whichever you think is more useful; time-by-time prediction or patient specific classification.

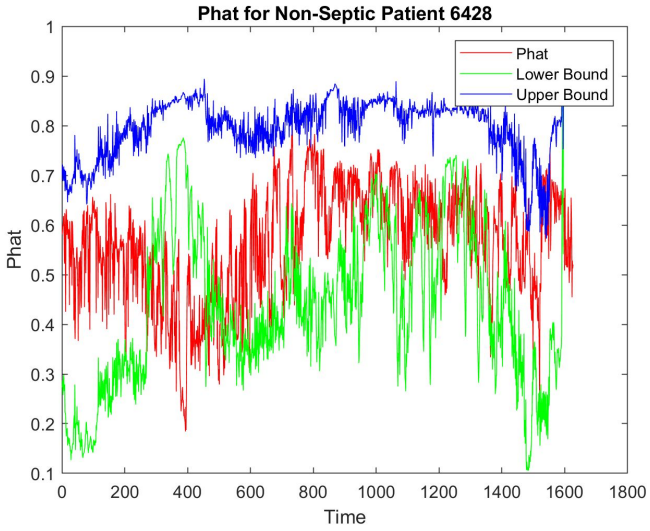
The decision rule we used was patient-specific classification. This was because a patient-specific classification rule allowed us to utilize all the data across all possible time instead of trying to determine sepsis from one individual snapshot of time.

The way we did this was by taking finding the average Phat value for each patient across the measurement time. If this average Phat value was above 0.7, we determined that patient had sepsis. If the average Phat value was below this, the patient did not have sepsis.

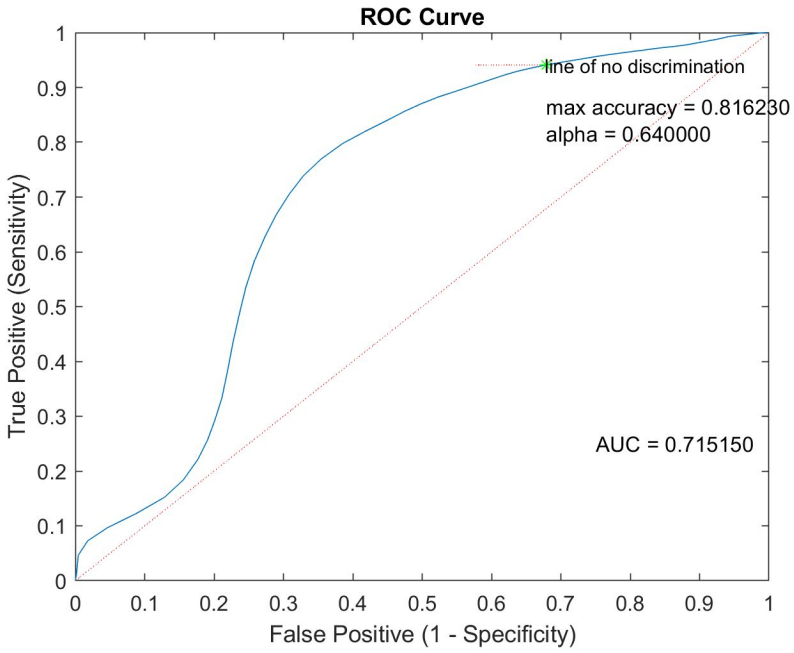
5. Plot the model, $\hat{p}i$, for each patient and overlay the 5% and 95% confidence bounds.

Below are typical examples of septic and non-septic patients. Some things to note are that non-septic patients generally have Phat values that are less than septic patients on average. This was the basis of our decision rule. In addition, the confidence intervals sometimes cross the Phat value. This was due to the way we were instructed to create the confidence intervals. **Classify_test_part2.m** plots all patient's Phat values.





6. Report the results of your final GLM model on the training set (and on the test set for the static model). Make sure to include the accuracy of your model, the ROC plot and the AUC value.



Using this model, we found an accuracy of 0.816230, with alpha equal to 0.64 and an AUC of 0.71515, as indicated in the above feature. Our dynamic model performed better than the static model on the training data, which was to be expected.

Using the decision rule, we were able to achieve an percentage correct of 93.4879%. Although this seems high, we understand that the Septic/Non-Septic split is around 80/20, but having a percentage correct higher than that split suggests the model did learn something from the training data.

7. Discuss performance differences on the training vs. the test set. Is this what you expected?

Because testing data for the waveform data wasn't provided, we weren't able to do this comparison. However, it'll be interesting to see how our model performs with the test data graders have. In order to use the test data, uncomment the relevant loading of the testing code and comment the loading of the training code as shown below.

```
%bring in testing data (UNCOMMENT THIS SECTION IF waveform_data_testing.mat
%IS PROVIDED)
% clinical_data_t = load('clinical_data_testing.mat');
% waveform_data_t = load('waveform_data_testing.mat');

%COMMENT THIS SECTION IF waveform_data_testing.mat
%IS PROVIDED
clinical_data_t = load('clinical_data_training.mat');
waveform_data_t = load('waveform_data_training.mat');
```