Mathematically Predicting Treatment Response of Hepatocellular Carcinoma Patients

Team Live and Let Liver
Lucas Cadalzo, Lorenz Gahn, Ken Groszman
Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005

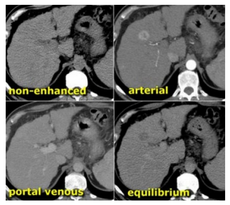
Motivation

- Hepatocellular Carcinoma (HCC) is one of the most lethal forms of cancer, with most patients not surviving more than 20 months [1]
- Treatment options, such as transcatheter arterial chemoembolization (TACE) are invasive, expensive, and often ineffective in extending the lives of patients [2]
- Oncologist methodology for deciding which treatment to use are subjective and based on experience
- We seek a **mathematical model that predicts the responsiveness of HCC patients to TACE**

Background: Hepatocellular Carcinoma (HCC)

Hepatocellular Carcinoma (HCC)
- Third most lethal variety of cancer, often associated with cirrhosis of the liver and hepatitis [3]
- Life expectancy is 6-20 months [1]
- HCC causes more than 660,000 deaths per year
- Common treatment is TACE, which is quite invasive, but not all patients respond to it [2]

Patient Imaging
- Patients are imaged using computerized tomography (CT)
- Four images are taken at different times
- Contrast agent is injected into patient's arm and flows through circulatory system, illuminating the tumor first



Description of Data

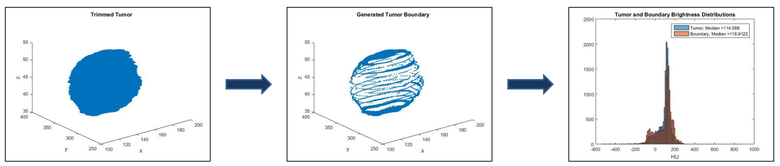
Our patient feature data comes from three sources:

1. CT Scan Feature Data
   - Each patient's 4 CT scans were processed with software called Atropos
   - This generates ~450 features per patient

2. Patient Medical Data
   - Mentor provided anonymized patient medical records
   - Includes ~ 15 features including age, hepatitis, cirrhosis, etc.

3. Self-Generated Data
   - Some features were expected to be predictive, but were not in dataset
   - We generated tumor encapsulation and vascularity features in MATLAB

Tumor Encapsulation Feature:
Measures whether the tumor contains a dense boundary of connective tissue



Regression-Based Feature Selection

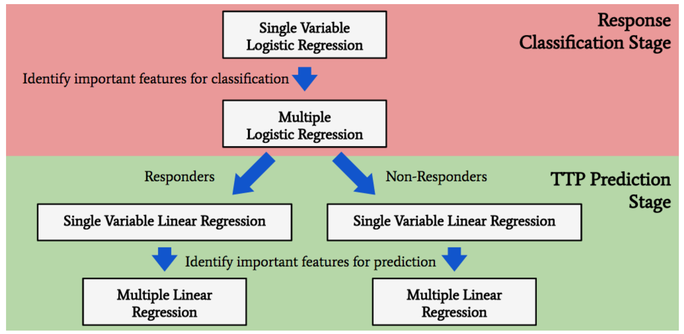
- With a dataset that contains 500 features but only 20 patients, it was important to select only important features to include in the model
- We used a single-variable regression-based method for selecting important features to include in our model

Method:
- For each feature, perform a regression of that feature compared to a response variable:
  - Categorical variable – logistic regression – binary response variable
  - Continuous variable – linear regression – TTP is response variable
- Compare R^2 or deviance values to determine predictive features

Model Overview

- The model was split into two stages:

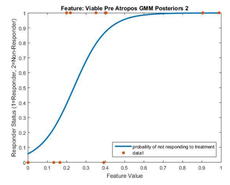
1. Response-Classification Stage:
   - Select features for R/NR classification with logistic feature selection
   - Use multiple logistic regression to classify patients
2. TTP Prediction Stage
   - Select features for TTP prediction with linear feature selection
   - Use multiple linear regression to predict TTP for each group



Response Classification Stage

- Generate a multiple logistic regression model using three most important features identified by single logistic regression:
  1. Del_Atropos_GMM_Posteriors3
  2. viable_Pre_Atropos_GMM_Posteriors2
  3. viable_Pre_Atropos_GMM_Posteriors3

$$Response = 2 - \frac{1}{1 + e^{-(16.7 - 9.0x_1 - 23.4x_2 - 287.9x_3)}}$$



Average classification accuracy of 92% using cross-validation

- Next step: improve classification accuracy with multilayer neural network

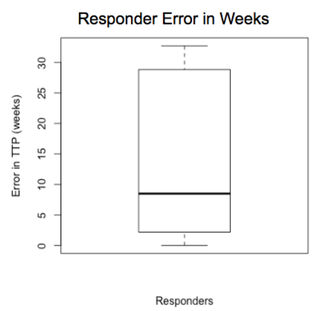
TTP Prediction Stage

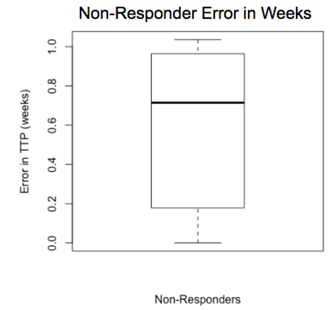
- After patients are classified as responders or non-responders, a multiple linear regression model is used to fit to each group
- Different features are significant for each case:
- Important responders features:
  1. Mean.Del_ATROPOS_GMM_POSTERIORS2
  2. Mean.Del_ATROPOS_GMM_POSTERIORS1
  3. Mean.Del_SKEWNESS_RADIUS_3
  4. hep_b
- Important non-responders features:
  1. Mean.Pre_SKEWNESS_RADIUS_5
  2. Metastasis
- Responders: Mult. R-squared: 0.87, Adj. R-squared: 0.77

$$TTP = 1.76 * 10^2 - 1.011 * 10^2 x_1 + 4.906 x_2 + 9.6 * 10 x_3 - 1.12 * 10^7 x_4$$

- Non-Responders: Mult. R-squared: 0.80, Adj. R-squared: 0.75

$$TTP = 5.04 - 7.21 * 10^{-5} x_5 + 0.96 x_6$$





Conclusion and Future Plans

Results:
- ***92% classification accuracy for Responders vs. Non-Responders***
- Median TTP prediction errors for split models:
  - Responders: 8 weeks
  - Non-responders: 0.7 weeks

Next steps:
1. Improve classification error using multilayer perceptron neural network
2. Obtain more data and further test model
3. Hand over to MD Anderson to continue work

Acknowledgements

We would like to acknowledge the Rice University Computational and Applied Mathematics Department, and specifically Dr. Beatrice Riviere, for their support and advice. We also want to thank Dr. David Fuentes and Dr. Ahmed Khalfe from the MD Anderson Cancer Center for their help. Finally, Dr. Erszebet Merenyi from Rice University ECE provided excellent advice with regard to neural networks.

References

[1] Forner et al. "Hepatocellular Carcinoma." *Lancet (London, England)* 379, no. 9822 (March 31, 2012): 1245–55.
[2] Altekruse er al. "Hepatocellular Carcinoma Incidence, Mortality, and Survival Trends in the United States From 1975 to 2005." *Journal of Clinical Oncology* 27, no. 9 (March 20, 2009): 1485–91.
[3] Chen, et al. "Risk of Hepatocellular Carcinoma Across a Biological Gradient of Serum Hepatitis B Virus DNA Level." *JAMA* 295, no. 1 (January 4, 2006): 65–73.