Artificial Neural Networks Based Visual Data Mining for Clinical and Metabolomic Data

Kasun Amarasinghe[1]; Urszula Osinska Warncke[2]; Milos Manic[1]; Dayanjan S Wijesinghe[2]
[1]Virginia Commonwealth University, School of Engineering, Richmond, VA;
[2]Virginia Commonwealth University, School of Pharmacy, Richmond, VA

Modern Heuristics Research Group

Abstract

Introduction

This abstract presents initial steps for mining clinical and metabolomic data of patients with cardiac arrests using machine learning techniques. The presented data mining methodology intends to uncover non-trivial correlations between metabolite concentrations and clinical measurements with the survival rate of the patient. This method uses unsupervised learning techniques for mining relationships that exist in the data and supervised techniques to predict the survivability of the patient from the clinical data and metabolomic data of patients' plasma obtained at time point 1 out of 3.

Methods

Self-Organizing Maps (SOM) were used to perform the unsupervised clustering and visualization of data. SOM comprises of a topological grid of artificial neurons typically arranged in a 1D or 2D lattice. Each neuron has a set of synaptic weights. The knowledge is stored by changing the values of these weights. SOM adjusts itself to the topological properties of the input data set using unsupervised winner-take-all learning algorithm together with cooperative adaptation. Initialization of all neurons are carried out randomly and the learning process is carried out iteratively.

Preliminary Data

During the learning process, for each data record, a best matching unit (BMU) is selected. BMU is selected by calculating the Euclidean distance between the neuron and the input data record. The closest neuron to the data point is considered the BMU. Then, the weights of all neurons are adapted using the cooperative adaptation. This process is iteratively carried out for all the data records until the weights reach an equilibrium. The spatial properties of the SOM neurons visualize the clusters that exist in data. Support Vector Machines (SVMs) are used to perform the supervised classification of the data. SVM is a classification algorithm that calculates the optimal separation plane between classes. SVMs perform this by finding the closest points from each class, rather than using every data point like other algorithms, as they are the most likely to support the equation of the final classification line. The classification resulted in 82.7% accuracy rate for data of 29 patients. Initial classification results are promising. Thus, it shows that mining clinical and metabolite data is a viable path for uncovering non-trivial relationships between the illness and its trends. Therefore, as future work, further analyses and inferences will be conducted using advanced machine learning techniques such as Deep learning.

Novel Aspect

Mining clinical and metabolite data is a viable path for uncovering non-trivial relationships between the illness and its trends.

Introduction

In the United States as of 2016, there has been over 350,000 annual occurrences of out-of-hospital cardiac arrest (CA), with a dismal 12% chance of survival, and an even lower chance of a good neurological outcome measured by Cerebral Performance Category (CPC). Excluding all cardiac diseases but cardiac arrest, it is the second leading cause of death in the U.S. Advancement of care in the emergency medical service setting as well as when the patient is received by the hospital when a cardiac arrest occurs is vital to the survival of these patients, as most of these events happen in out-of hospital settings.[3] Long term treatment plans are also needed for these patients, as neural degradation from oxygen deprivation in survivors greatly decreases quality of life.

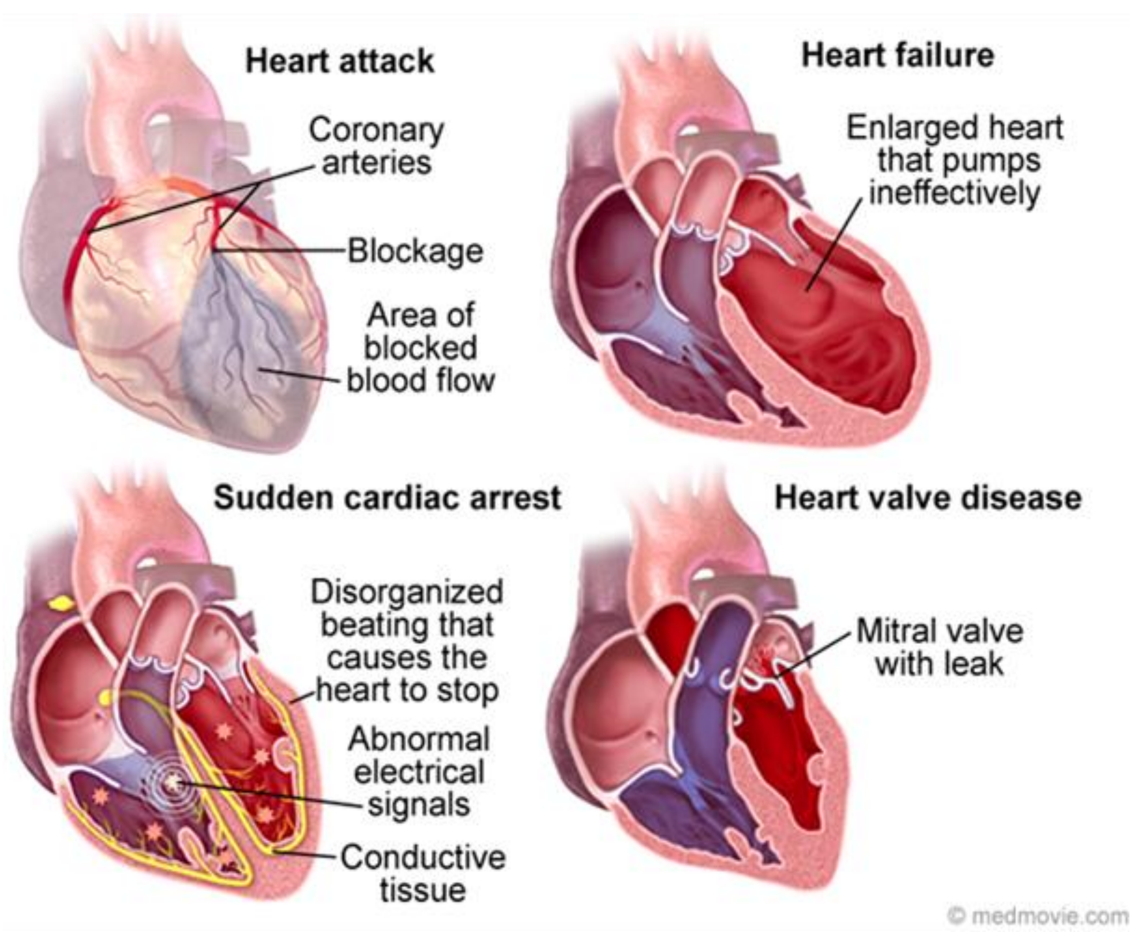

Figure 1: Common cardiac conditions.

Experimental Design, Data Acquisition and Analysis

All samples were kept frozen until ready for analysis. Sample extraction consisted of adding methanol (1.5 ml) to sample aliquot (200ul), vortexing and adding MTBE (5ml). After incubation at RT for 1 hr, phase separation was induced by addition of MS-grade water (1.25ml); followed by 10 min incubation at RT and 10 min centrifugation at 1,000 g. The organic phase was removed, and 2 ml of MTBE/MeOH/water (10:3:2.5) was added for re-extraction. combined organic phase was dried and re-suspended in in 200 µl of $CHCl_3$/methanol/water (60:30:4.5, v/v/v) for future lipid analysis, while the aqueous phase was further processed for HILIC-QTOF MS analysis. Metabolomic and lipidomic data were acquired with TripleTOF 6600 Sciex system as hydrophilic interaction liquid chromatography and liquid chromatography Charged Surface Hybrid, respectively.

Methods

- Combination of supervised learning and unsupervised learning methods used
- Supervised Learning
  - Trained using labeled data
  - Used for predicting survivability of a patient given clinical/metabolomic data
- Unsupervised learning
  - Trained using on-labeled data
  - used for identifying similarities existing data
- Self Organizing Maps (SOM)
  - Developed by Kohonen
  - Comprised of topological artificial neuron grid arranged in a lattice
  - Adjusts itself to the topological properties of the data
  - Provides a visual representation of the data
  - Visual representations can be used to extract behaviors from data

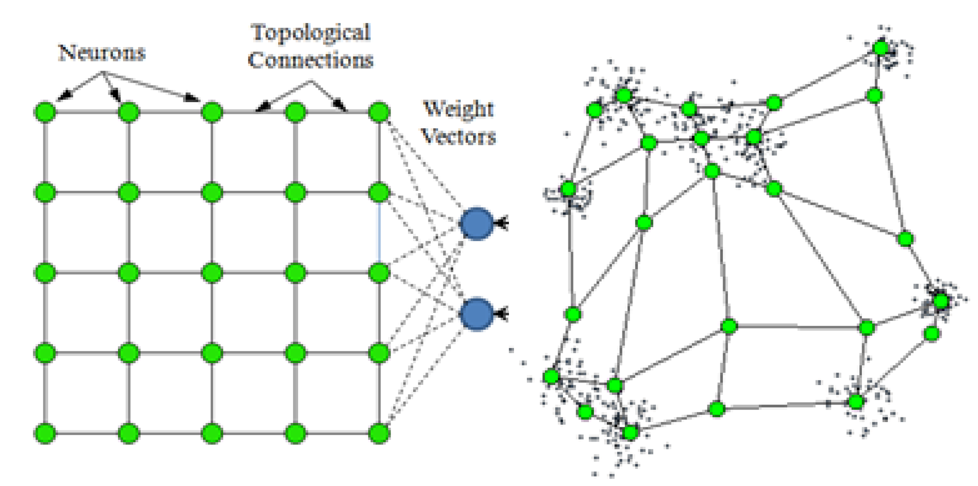

Figure 2: SOM in the output space and in the input space when adjusted to the input points

- Support Vector Machines classification
  - Supervised classification
  - Input data: Clinical/metabolomic data
  - Output classes: patient's' survivability estimation

Results

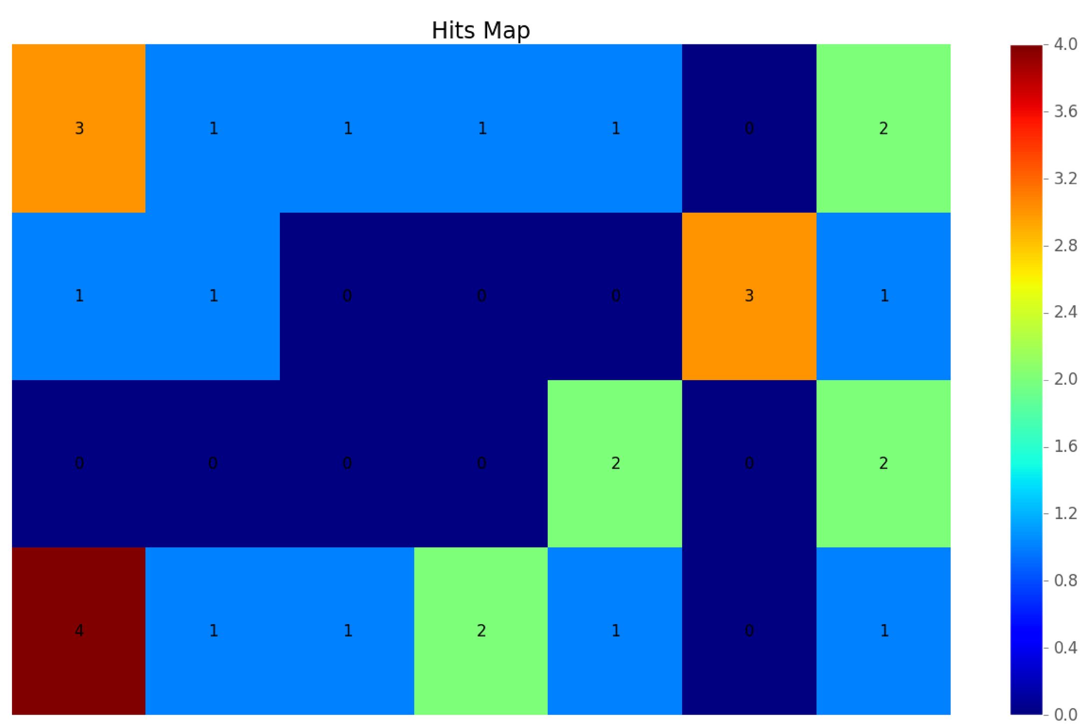

Figure 3: Self Organizing Map: Clustering

Results

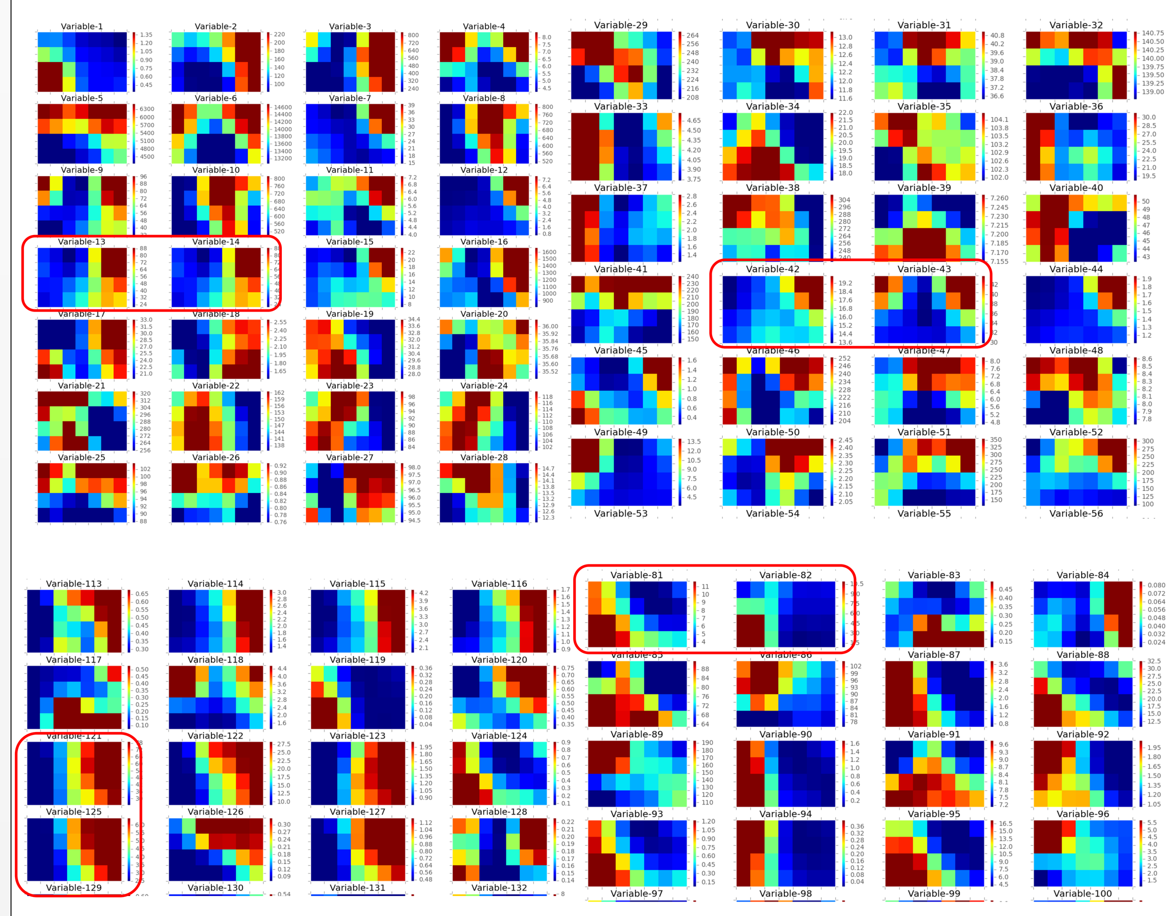

Figure 4: behavior of features in the SOM clusters

- Support Vector Machines
  - Data set of 29 patients
  - Two classes: survival and non-survival
  - Entire dataset used for training
  - Classification accuracy: 82.7%
- Therefore, initial classification accuracy is promising!
- Can use the data driven methodologies to predict the patient's chance of survival and risk

Conclusions and Future Directions

- Significant and similarly behaving metabolites were identified through an unsupervised visual analysis
- By finding these metabolites and their associated pathways, we can develop new hypotheses
- Enables to get a glimpse of the underlying causes of neurological outcomes following cardiac arrests
- Allows us to predict which patients are at risk.

References

1. "Out-of-Hospital Cardiac Arrest." Cardiac Arrest Statistics. American Heart Association, n.d. Web. 14 Apr. 2017.
2. "Deaths and Mortality." Centers for Disease Control and Prevention. National Center for Health Statistics, 17 Mar. 2017. Web. 14 Apr. 2017.
3. Girotra, Saket, Paul S. Chan, and Steven M. Bradley. "Post-resuscitation Care following Out-of-hospital and In-hospital Cardiac Arrest." Heart 101.24 (2015): 1943-949. Web.
4. T. Kohonen, "Automatic Formation of Topological Maps of Patterns in a Self-Organizing System, " in Proc. SCIA, E. Oja, O. Simula, Eds., pp. 214-220, 1981.