University of Groningen

Computing Science

Bachelors Thesis

Classification of Delirium Subtypes and Mortality Rates

Author:
Joseph Jones
S2990652

Supervisor:
Prof. Dr. Michael Biehl
Second Supervisor:
Dr. Kerstin Bunte
External Supervisor:
Dr. Thomas Andrew Jackson
(University of Birmingham)

August 8, 2018

1 Introduction

Delirium is an acute disorder of attention and cognition that affects around an eighth [1] of hospitalised patients. There are 4 different types of delirium; Normal, Hyperactive, Hypoactive and Mixed [2]. In this thesis we will only be concerned with the 3, Hyperactive, Hypoactive and Mixed. A data set from the University of Birmingham containing patient's blood values, cytokine values along with what subtype of delirium they are suffering from and the lifespan of the patient. Generalised Matrix Relevance Learning Vector Quantization will be applied on the data set as a multi-class problem for classifying by subtype and a two-class problem for classifying by mortality rates. A self-organising map will also be used in an attempt to find hidden patterns and clusters as well as a visualisation of the data. The main goal is to achieve a strong, accurate classifier, but relevance learning can also give insight into which features have the most discriminate power, and medical relevance.

2 Problem Definition and Data Set Description

The data set received from the University of Birmingham has separated the patient data into 3 different subtypes of delirium:

1. Hyperactive, which is characterised by restlessness and agitation.
2. Hypoactive, which is characterised by reduced motor activity, unresponsiveness and sluggishness.
3. Mixed, where patients present with symptoms from both the hyperactive and hypoactive subtypes, with patients sometimes switching between the two. [2]

Also most of the patients have information on the overall outcome of the patient: if the patient was still alive 2 years after their values were measured and if not, the amount of days the patient survived after their values were measured. The patients were separated into groups depending on their survival time:

1. Patients who died within 4 months.
2. Patients who died within 1 year.
3. Patients who died within 2 years.
4. Patients who lived longer than 2 years.

The aim of this project is to use machine learning techniques to try to create classifiers to separate the data into the different groups

2.1 Primary Tasks

At the start of the project multiple primary tasks were defined. First was to see if the data clusters around the different subtypes of delirium. This could help in our understanding of the subtypes, it could give some insight of why the patients are displaying different symptoms. Next we wanted to look at the mortality rates of the patients. If the data clusters around the different lifespans of the patients then we might be able to determine the severity of a patients condition. With these aims a number of tasks were laid out:

1. Create Subtype classifier.
2. Options considered for a mortality rate classifier.
    - Patients who died within 4 months vs patients who survived longer.
    - Patients who died within 1 year vs patients who survived longer.
    - Patients who died within 2 years vs patients who survived longer.

3 Machine Learning Methods

In this section I will briefly present the machine learning techniques used throughout the project. The first is generalised matrix relevance learning vector quantisation (GMLVQ), a prototype based supervised method[3] that is part of the family of LVQ techniques. The second is a self-organising map (SOM) which is a prototype based unsupervised machine learning technique[4].

3.1 GMLVQ

As GMLVQ is part of the LVQ family of techniques, it shares a number of properties with some of the other techniques [18][19]. Prototypes are initialised and given a class label. These prototypes are then updated during training based upon their label. For each data the best matching unit (BMU) according to the distance measure is found, if the class of the BMU is of the same class the data point, the prototype is moved closer, else if the class labels are different the prototype is repelled. In LVQ the BMU is found using the euclidean distance between the prototype and the data point. This can be an issue when there are noisy features in the data set and it does not take into account that some features may have more predictive power than others. GMLVQ attempts

to fix this issues by creating a relevance matrix matrix which allows for this, as well as taking pairwise combinations of features into account. The GMLVQ generalised distance is then

$$d^{\Lambda}(\mathbf{w}, \xi) = (\xi - \mathbf{w})^T \Lambda (\xi - \mathbf{w})$$

$$\text{with } \Lambda = \Omega^T \Omega \text{ and } \sum_i \Lambda_{ii} = 1$$

where ξ is a data point, $\mathbf{w}$ is a prototype and Λ is an $n \times n$ positive semi-definite distance matrix. The relevance matrix Λ and the prototypes $\mathbf{w}_i$ are optimised based on a heuristic cost function

$$E = \sum_i \mu_i \text{ where } \quad \mu_i = \frac{d_J^{\Lambda}(\mathbf{w}_J, \xi_i) - d_K^{\Lambda}(\mathbf{w}_K, \xi_i)}{d_J^{\Lambda}(\mathbf{w}_J, \xi_i) + d_K^{\Lambda}(\mathbf{w}_K, \xi_i)}$$

where $d_J^{\Lambda}(\mathbf{w}_J, \xi)$ is the distance between data point ξ and the closest prototype $\mathbf{w}_J$ with the same class label and $d_J^{\Lambda}(\mathbf{w}_K, \xi)$ is the distance between data point ξ and the closest prototype $\mathbf{w}_K$ with a different class label. Correct classification leads to $\mu_i < 0$ whereas incorrect classification leads to $\mu_i > 0$, so the smaller the value of μ_i the more secure the classification. A gradient descent update can be applied leading to the update rule being:

$$\mathbf{w}_{J,K} \leftarrow \mathbf{w}_{J,K} - \eta \frac{\partial \mu_i}{\partial \mathbf{w}_{J,K}}$$

The derivations and complete update rules can be found in [8]. The diagonal elements of the relevance matrix Λ shows the importance of the individual features whereas the off-diagonal show the importance of the pairwise importance. GMLVQ has been proven to be an effective classifier for medical data, [7] [9] therefore GMLVQ is a logical choice for this problem.

3.2 Self-Organising Map

Self-organising maps are an unsupervised machine learning method that creates a low-dimensional topology-preserving view of high-dimensional data[6]. SOMs are composed of a map of nodes is arranged in a two-dimensional grid or lattice. Each node has a set of neighbours associated with it which are affected by the changes/movements of the node. The nodes are identified by their coordinate vector in the grid $r = (r_1, r_2)^T$. Each node is then associated with a prototype with the same dimensions as the input and the SOM is trained. The BMU $\mathbf{w}_c$ is found for each data point ξ using the euclidean distance, then $\mathbf{w}_i$ and its neighbours are updated.

$$\text{Let: } \quad s(\xi) = \text{argmin}_{r \in A} d(\mathbf{w}_c, \xi) \text{ be the winning neuron for } \xi$$

$$\mathbf{w}_c \leftarrow \mathbf{w}_c + \eta \cdot h_p^{SOM}(s(\xi), r) \cdot (\xi - \mathbf{w}_c)$$

where the neighbourhood function is

$$h_p^{SOM}(s, r) = exp(-\frac{||s - r||_A^2}{2p^2})$$

Where the parameter p denotes the neighbourhood range[10]. Unsupervised learning methods can give valuable insights into a dataset and uncover hidden patterns and clusters. For these reasons we will be using a SOM to further analyse the data.

4 Data Analysis and Pre-Processing

This section contains initial data analysis and pre-processing.

4.1 Data Cleansing

	Hyperactive	Hypoactive	Mixed
Number of samples	37	67	21
Total	125		

Table 1: Number of samples per subtype

The full data set contains medical information of 125 different patients suffering from the different subtypes of delirium, as seen in Table 1. From the Blood Values, 7 patients have null values. Patients ID 14 and 119 have null values 3 or more features, therefore these 2 patients will be ignored due to lack of completeness when classifying using the blood values. The remaining 5 patients will have their missing feature values replace with the class specific feature average. This is unfair as the class membership of a new patient is unknown in practice, so the results of the classifiers with the filled missing features may be over-optimistic.

	Hyperactive	Hypoactive	Mixed
Number of samples	37	66	20
Total	123		

Table 2: Data set after removing ID 14 and 119

Of the original 125 data samples, 91 of the patients also had their cytokine values measured. Table 2 shows the spread of these samples across the different subtypes of delirium.

	Hyperactive	Hypoactive	Mixed
Number of samples	27	49	15
Total	91		

Table 3: Number of samples per subtype with cytokine measurements

Finally of the original 125 data samples, the data contains the outcomes of 105 of the patients.

Time Survived Since Samples Measured	Patients who died within	Patients who survived longer
4 Months	26	79
1 Year	40	65
2 Years	63	42

Table 4: Size of the classes for each timespan

Data Set	Number
Blood Values, Subtypes	123
Cytokine Values, Subtypes	91
Blood + Cytokine Values, Subtypes	90
Blood Values, Outcomes	105
Cytokine Values, Outcomes	78
Blood + Cytokine Values, Outcomes	77

Table 5: Sizes of all intersections

4.2 Z-Score Transformation

In order to reduce variability and deal with the large differences in scale in the data we use the z-score values instead of the original values. This is especially important for the self-organising map as the euclidean distance can be heavily influenced by large values, thus a form of normalisation is necessary. The z-score values have the properties of a normal distribution, meaning that the mean is equal to 0 and the variance is 1.

4.3 Outliers

Outliers are values that vary strongly from other observations, their presence can disturb the performance of both GMLVQ and SOM classifiers. Therefore it is important to analyse the data to identify outlying values and process them.

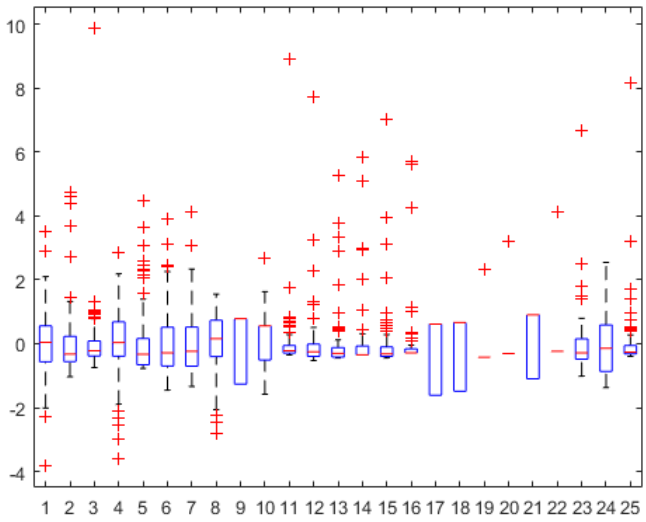

Figure 1: Box plot of all data after z-score transformation. All values below $Q1 - 1.5 \times (Q3 - Q1)$ and above $Q3 + 1.5 \times (Q3 - Q1)$ are marked as outliers. The first 10 features represent the blood values whereas features 11-25 are from the cytokine data.

	Hyperactive	Hypoactive	Mixed	All
Outliers	18	40	11	76
Number of Samples	37	66	20	123
Percentage Outliers	49	60	55	62

Table 6: Summary of Outliers. Each sample that has at least one outlying value is counted as an outlier.

As we can see in Table 6 the outliers represent a large portion of the data. To try and reduce the impact of the outliers on the performance on the classifiers we can apply winsorization on the data set[11]. This limits how extreme the values can be as outlying values below the 5th percentile are set to the 5th percentile and outlying values above the 95th percentile are set to the 95th percentile. This way the effect of outliers on the classifier is more limited while allowing us to use the outlying data sample without disguarding it. As individual nodes in SOMs are heavily influenced by outliers, we will first perform winsorization on the data for the SOM experiments. The outliers were left untouched for the GMLVQ experiments.

5 Experiment Results

In this section we will go over the results of the different classifiers employed during this project

5.1 GMLVQ

All results for the GMLVQ classifier are after a run validation with 50 runs, each run 10% of the data is randomly selected and is used for testing the classifier, the remaining 90% of the data is used to train the classifier. After 50 runs the results are averaged and are then used to evaluate the quality of the model. For 3 class problems the confusion matrix will be shown, for the 2 class problems the receiver operating characteristic (ROC) [5] curve is shown. The feature relevance matrices are also shown, this will allow for the evaluation of features and let us reduce the amount of features. The following GMLVQ experiments were performed using the "no-nonsense LVQ" toolbox for Matlab [12] using the default settings.

5.1.1 Hyperactive vs Hypoactive vs Mixed, Blood and Cytokine Values

First we will use GMLVQ to try and classify the data by subtype. We will start by using both the cytokine and blood values, giving a 25 dimensional feature vector. Features 1-10 are from the

	Hyperactive	Hypoactive	Mixed
Hyperactive	18.7%	66.8%	14.5%
Hypoactive	21.5%	69.9%	8.6%
Mixed	36.7%	56.3%	7%

Table 7: Hyperactive vs Hypoactive vs Mixed confusion matrix

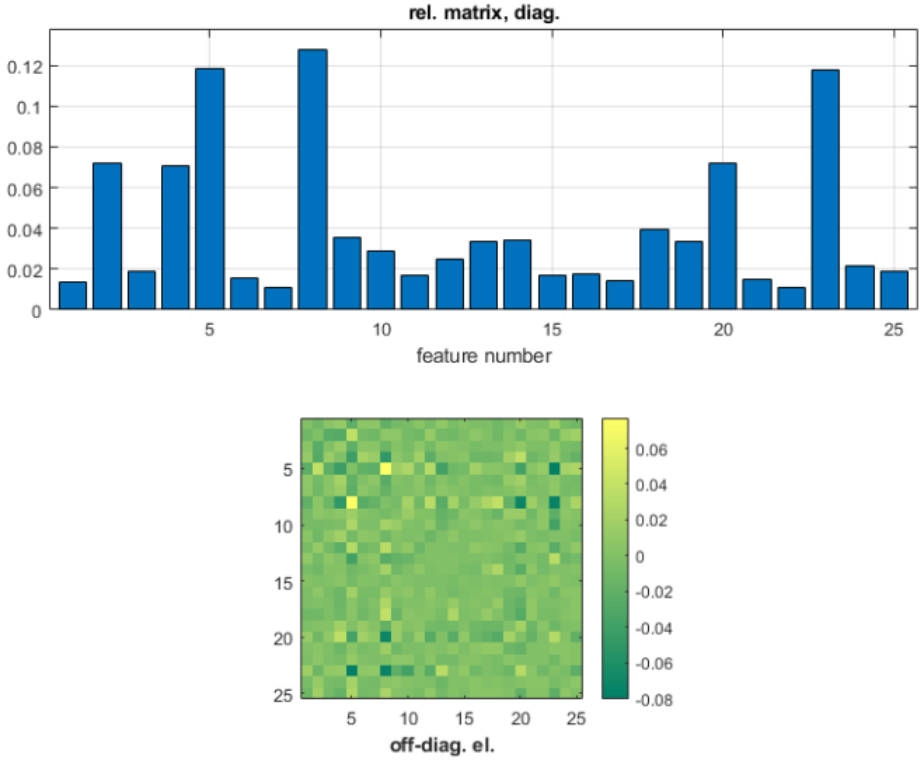

Figure 2: Hyperactive vs Hypoactive vs Mixed. Top - diagonal elements of relevance matrix. Bottom - off-diagonal elements of relevance matrix.

	Hyperactive	Hypoactive	Mixed
Hyperactive	21.7%	68.3%	10%
Hypoactive	19.8%	68.2%	12%
Mixed	24.7%	71.3%	4%

Table 8: Hyperactive vs hypoactive vs mixed with clipped outliers

blood values whereas features 11-25 are from the cytokine.

From the confusion matrices it is clear that the classifier is performing poorly. This is especially apparent with the mixed subtype, with the prototype classifying only 7% of the samples correctly. This could be due to a couple of reasons:
The mixed prototype does not have enough data in comparison to the other subtypes to trains properly. The other reason might be because the mixed subtype itself does not form a cluster in the data. To try and counteract the first issue we can downsize the number of hyperactive and hypoactive data samples while training the classifier to give us even class sizes. To do this a random permutation of the same size as the mixed category is selected from both the hyperactive and hypoactive subtypes for each validation run, all left over data is still used in the testing phase.

	Hyperactive	Hypoactive	Mixed
Hyperactive	19.7%	60.5%	19.8%
Hypoactive	20.1%	62.4%	17.5%
Mixed	44.1%	46.2%	9.7%

Table 9: Hyperactive vs hypoactive vs mixed confusion matrix with equalised class size

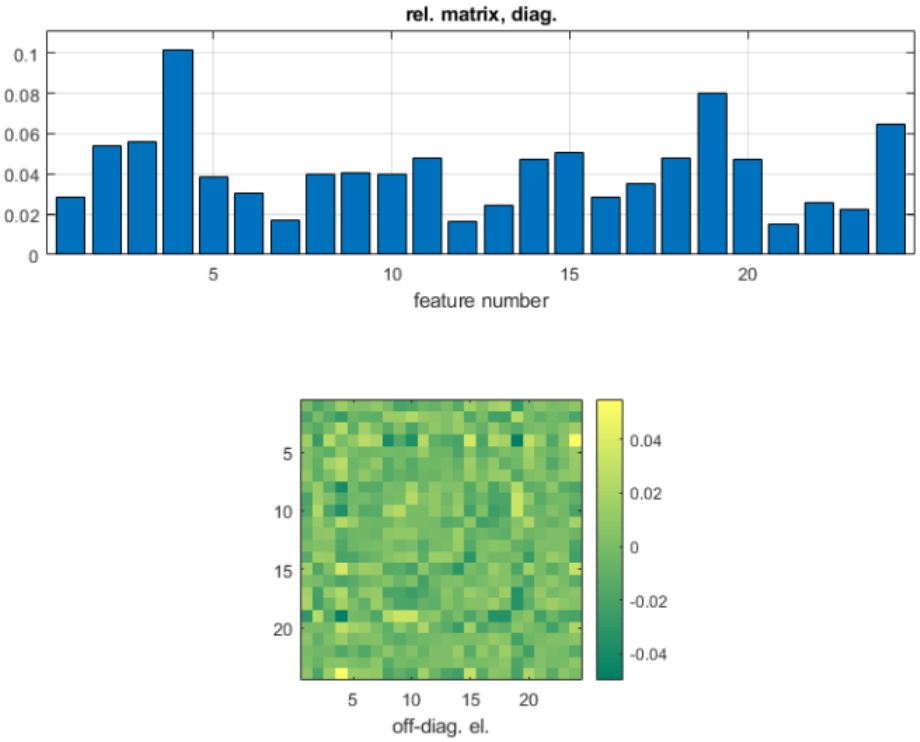

Figure 3: Hyperactive vs. hypoactive vs. mixed, equalised class sizes. Top - diagonal elements of relevance matrix. Bottom - off-diagonal elements of relevance matrix.

The confusion matrix seen in Table 9 shows that equalising the class sizes has had a very minor effect on the mixed prototype. This leads to the conclusion that the mixed subtype does not form a cluster itself. This seems to make sense due to the nature of the mixed subtype: as patients suffering from 'mixed' delirium present symptoms from both of the other subtypes, and their data also represents this. However this makes the subclass unsuitable for GMLVQ, so we may get better results by excluding the class from future tests.

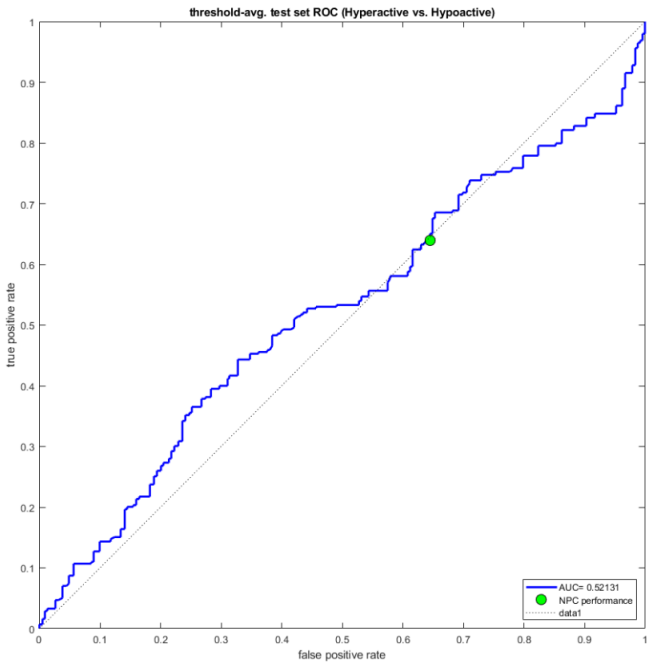

Figure 4: Hyperactive vs. hypoactive ROC Excluding Mixed

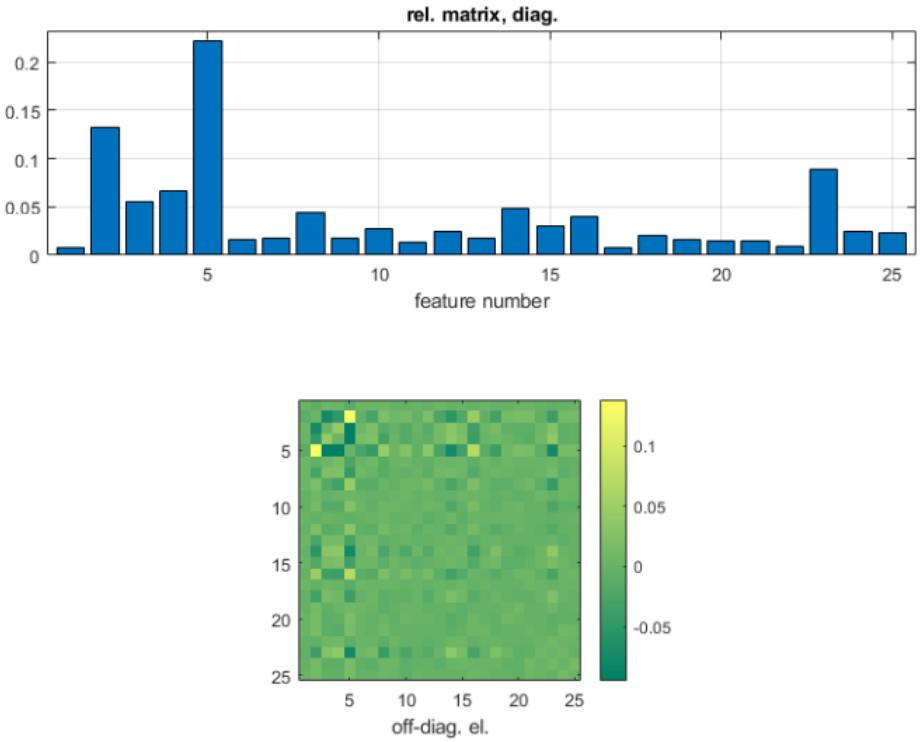

Figure 5: Top - diagonal elements of relevance matrix. Bottom - off-diagonal elements of relevance matrix.

From the ROC curve in figure 4 we observe that the classifier is performing little better than we would expect from a random classifier as the ROC curve is hugging the centre line and the area under the curve (AUC) being only slightly higher than 50%. To attempt to improve the performance of the classifier we will try reducing the amount of features. Looking at the feature relevance matrix 5 we observe that the majority of the relevant features occur in the first 10 features. As the first 10 features are from the blood values, this suggests that perhaps the cytokine values of the patients are not important to the classifier. To test this we shall remove the cytokine features from the data set and then run GMLVQ again. Ignoring the cytokine values increases the size of the data set, allowing the use of the 27 samples that did not have cytokine values.

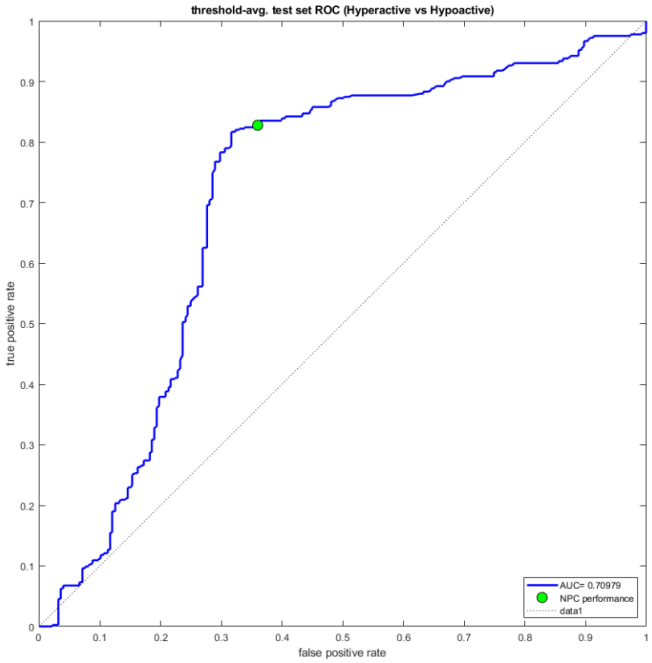

Figure 6: Subtypes ROC Excluding Mixed and Cytokine

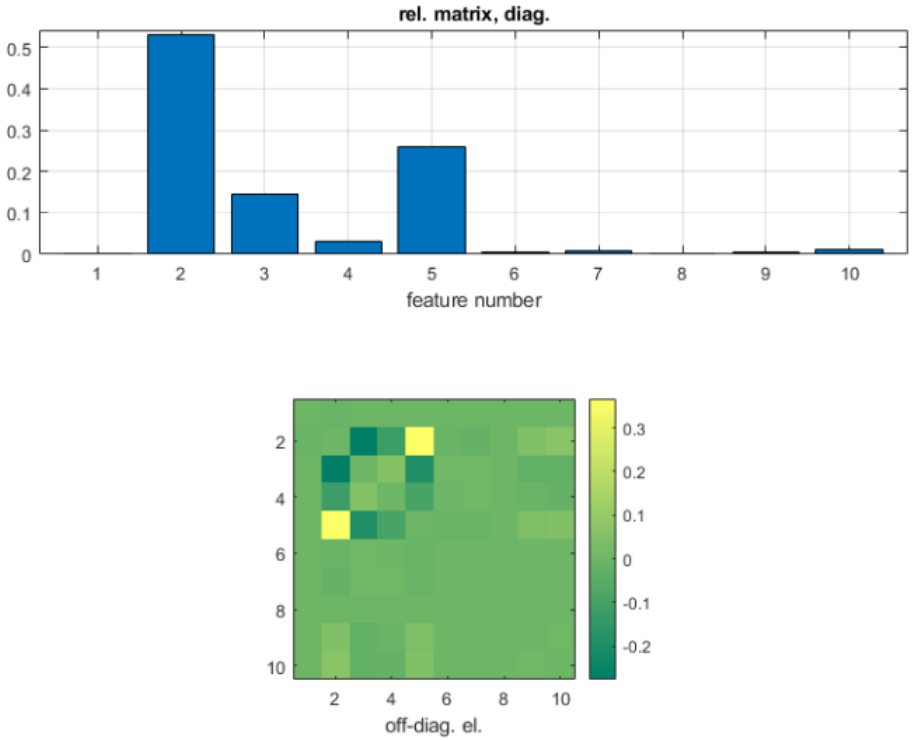

Figure 7: Feature Relevance Matrix Excluding Mixed and Cytokine. Top - diagonal elements of relevance matrix. Bottom - off-diagonal elements of relevance matrix.

Now with the reduced amount of features and the larger data set we observe a large improvement in the performance of the classifier. The classifier is now performing much better than random, although it is not accurate enough to be relied upon as the sole diagnosis tool, as seen in the ROC curve presented in figure 6. The kink in the graph shows that there is a lot of overlap between the 2 classes near the NPC, meaning that the decision boundary needed to be finely tuned to get a reasonable performance. The feature relevance matrix in figure 7 shows that features 2 and 5 display a much larger predictive power than the other features. Feature 2 is urea, a marker of dehydration and feature 5 is C-reactive protein (CRP), a marker of infection. From a medical perspective this could give some valuable insight to the differences between the hyperactive and hypoactive subtypes of delirium.

5.1.2 Mortality Rates

In this section the patient's blood and cytokine values will again be considered together, along with the patient's age and sex. Rather than building a classifier to group by the different delirium subtypes we focus on the mortality and survival times of the patients. Again all results are the averaged results over 50 runs with 10% of the data left out of each run for testing. The first groups considered will be patients who survived longer than 4 months vs. patients who died within 4 months followed by patients who survived 1 year vs. died within 1 year and survived 2 years vs. died within 2 years.

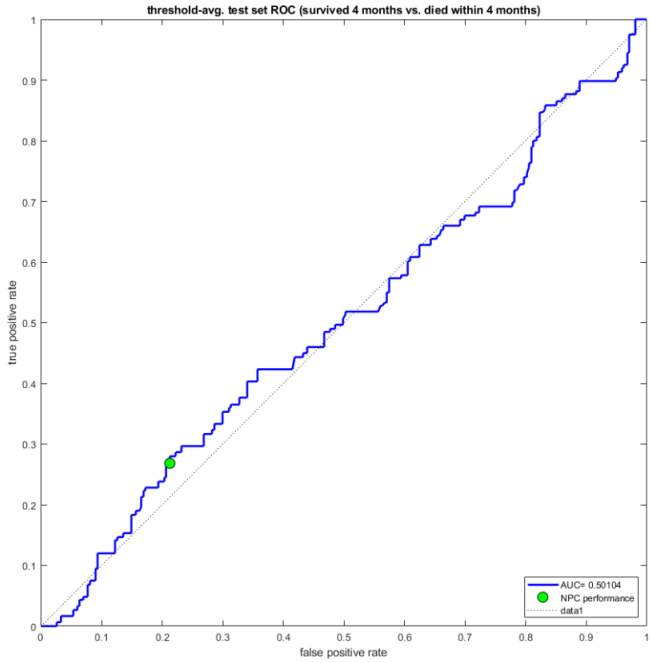

Figure 8: Survived 4 months vs. died within 4 months ROC curve.

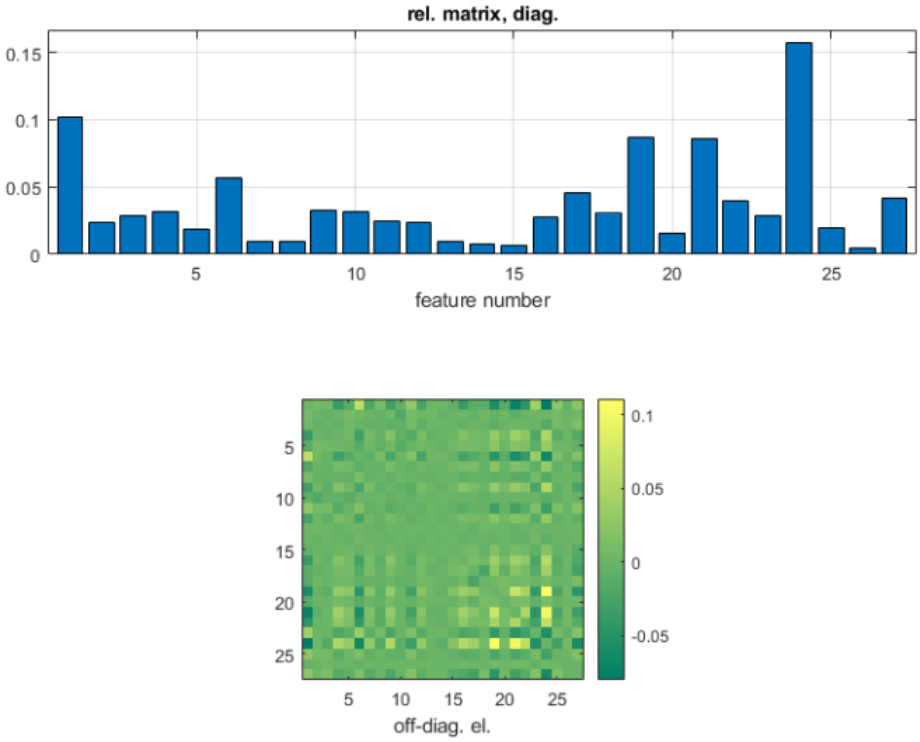

Figure 9: Survived 4 months vs. died within 4 months. Top - diagonal elements of relevance matrix. Bottom - off-diagonal elements of relevance matrix.

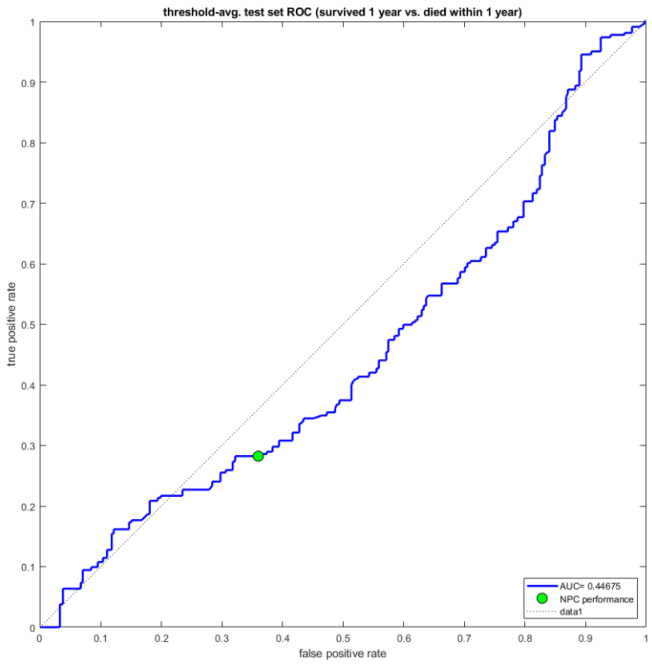

Figure 10: Survived 1 year vs. died within 1 year ROC curve.

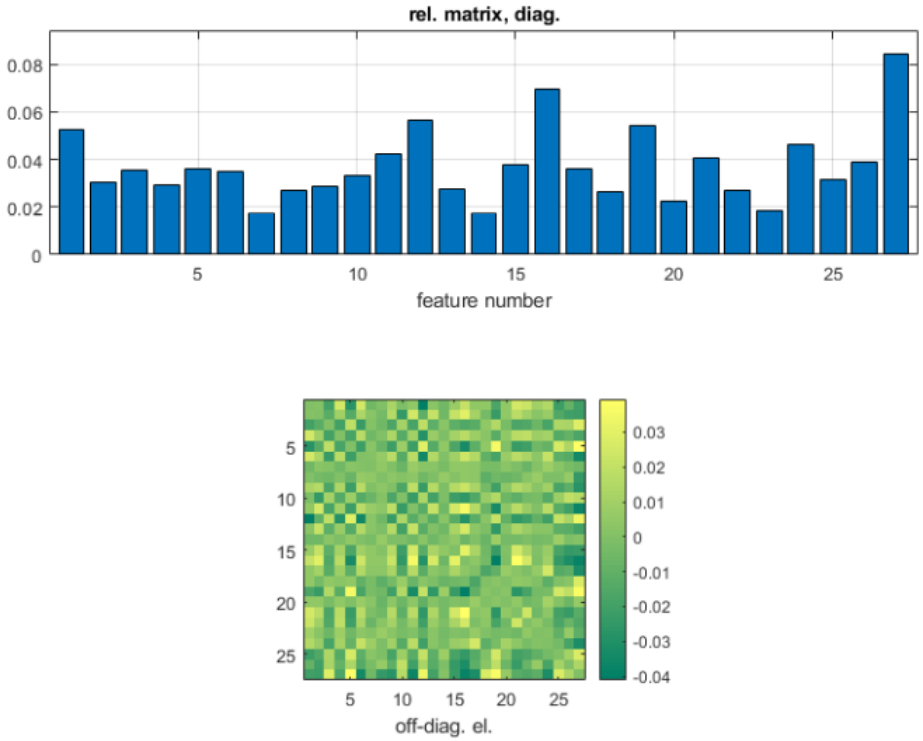

Figure 11: Survived 1 year vs. died within 1 year. Top - diagonal elements of relevance matrix. Bottom - off-diagonal elements of relevance matrix.

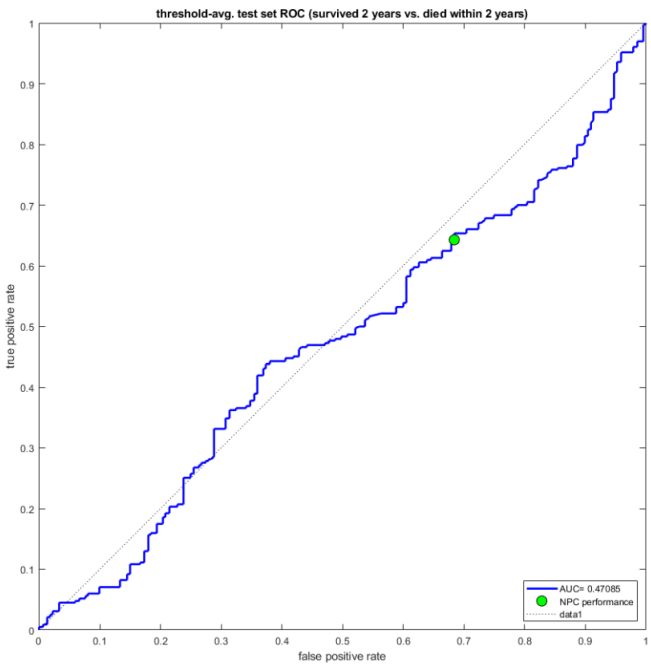

Figure 12: Survived 2 years vs. died within 2 years ROC curve.

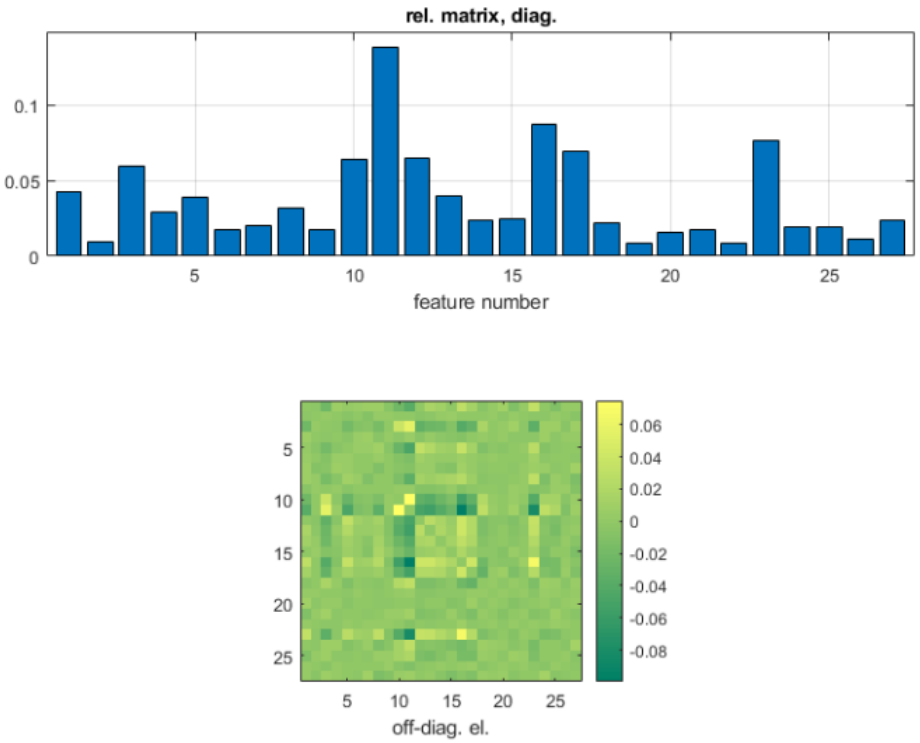

Figure 13: Survived 2 years vs. died within 2 years. Top - diagonal elements of relevance matrix. Bottom - off-diagonal elements of relevance matrix.

As seen in Figures 8, 10, 12, the performance of all of the classifiers on the different survival rates are very similar, with the ROC curve hugging the centre line. This tells us that the classifiers have a performance akin to a random classifier. To attempt to improve the performance of the classifiers we can try reducing the number of features. To do this we remove all features except top 10 most important features from each classifier.

Rank	Feature	Value
1	HCT	0.15668
2	Alb	0.110885
3	IL6	0.101641
4	WBC	0.0993705
5	Neut	0.0623199
6	IL1b	0.0605727
7	Na	0.0555619
8	IL10LD	0.0523885
9	TNFa	0.0490014
10	Urea	0.0460849

Table 10: Survived 4 months vs. died within 4 months classifier feature rankings.

Rank	Feature	Value
1	Sex	0.0841948
2	Na	0.0694477
3	IL1bLD	0.0563250
4	Alb	0.0542117
5	IL6	0.052523
6	HCT	0.0463932
7	IL1raLD	0.0423249
8	WBC	0.0405513
9	IL1Ra	0.03948
10	Age	0.0385949

Table 11: Survived 1 year vs. died within 1 year classifier feature rankings.

Rank	Feature	Value
1	IL1raLD	0.138603
2	Na	0.0872272
3	NueutWBCratio	0.0762888
4	TNFaLD	0.0750265
5	Urea	0.0693507
6	IL1bLD	0.0649643
7	IL10	0.0599627
8	IL6	0.042815
9	Cortisolnmoll	0.0409021
10	IL1Ra	0.0386647

Table 12: Survived 2 years vs. died within 2 years classifier feature rankings.

The GMLVQ classifier was then run again on the data sets with the reduced amount of features.

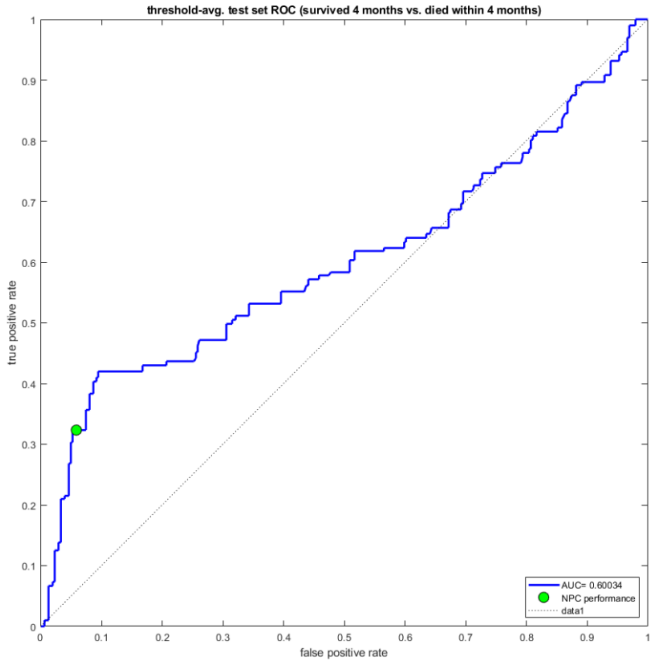

Figure 14: Survived 4 months vs. died within 4 months with reduced number of features ROC curve.

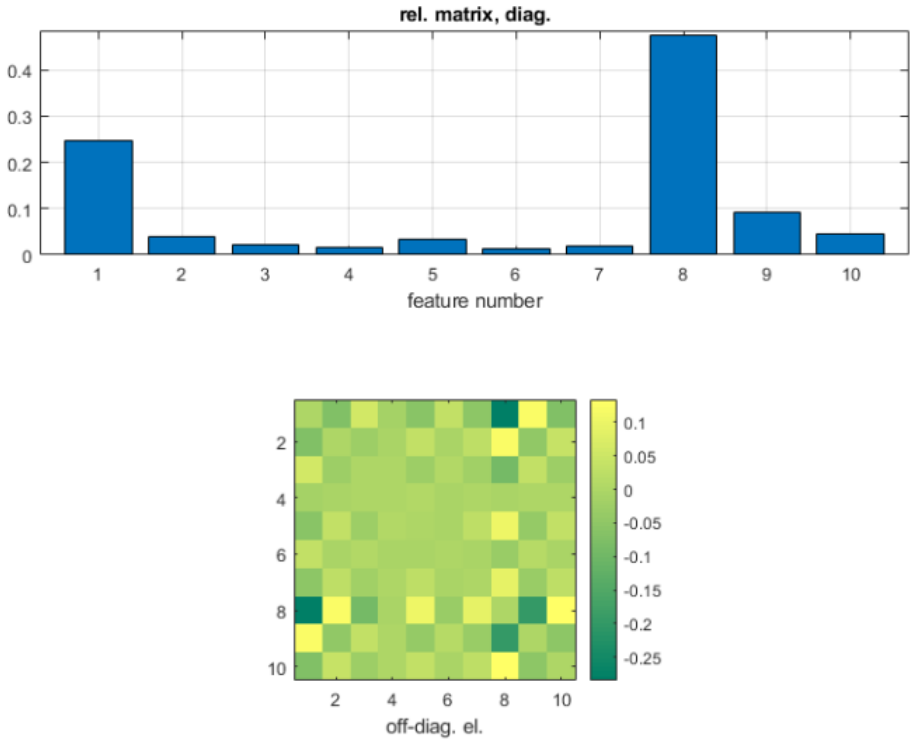

Figure 15: Survived 4 months vs. died within 4 months with reduced number of features. Top - diagonal elements of relevance matrix. Bottom - off-diagonal elements of relevance matrix.

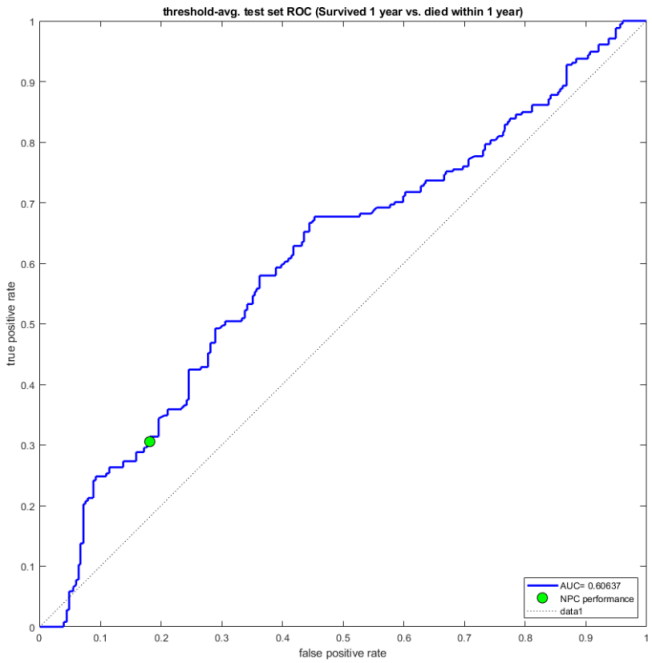

Figure 16: Survived 1 year vs. died within 1 year with reduced number of features ROC curve.

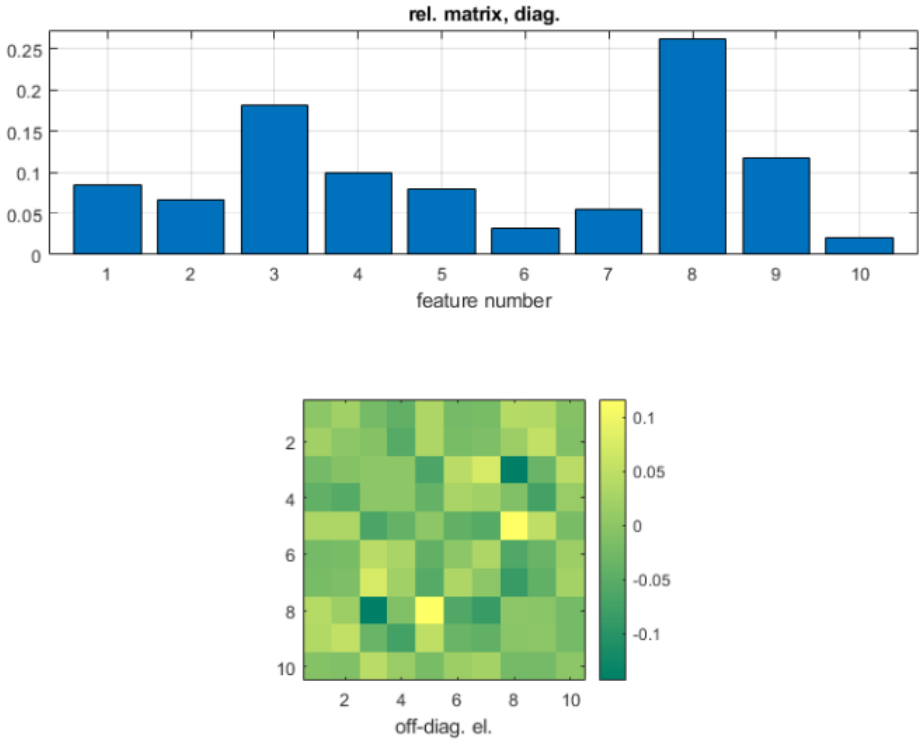

Figure 17: Survived 1 year vs. died within 1 year with reduced number of features. Top - diagonal elements of relevance matrix. Bottom - off-diagonal elements of relevance matrix.

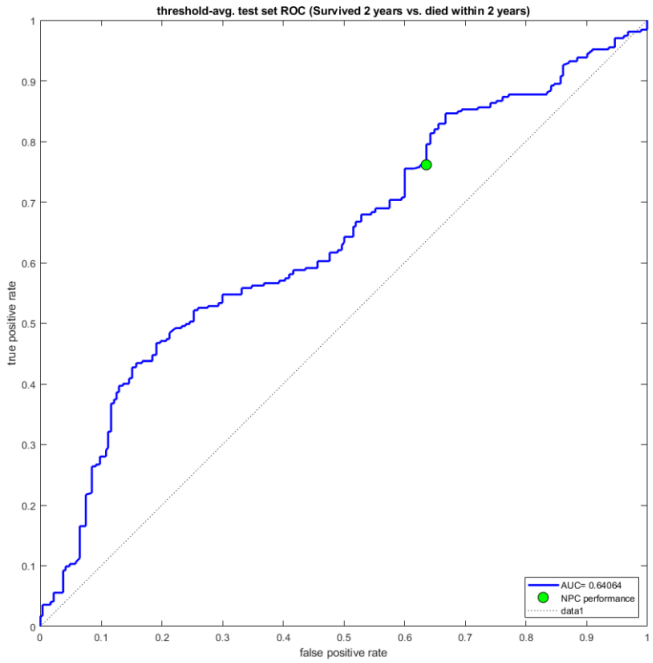

Figure 18: Survived 2 years vs. died within 2 years with reduced number of features ROC curve.

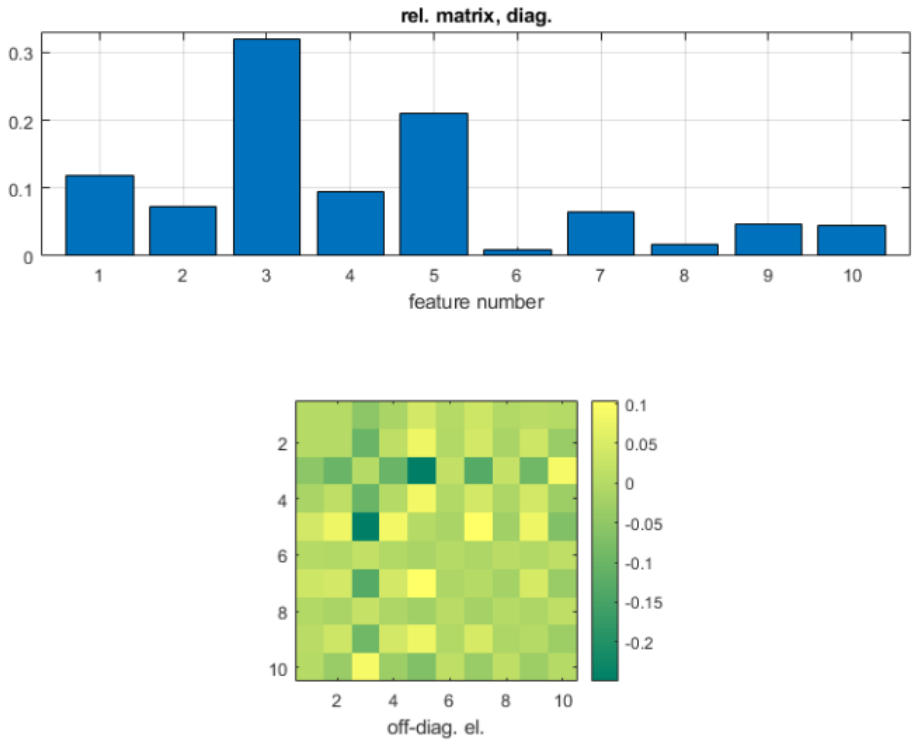

Figure 19: Survived 2 years vs. died within 2 years with reduced number of features. Top - diagonal elements of relevance matrix. Bottom - off-diagonal elements of relevance matrix.

In figures 14, 16, 18 we can observe that reducing the amount of features has slightly improved the performance of each of the classifiers to now be slightly better than we would expect a random classifier to perform. However none of these classifiers are accurate enough to give us significantly relevant medical information about the severity of a patient's condition.

5.2 Self-Organising Map

In this section we will be applying a self-organising map to the data to try and highlight hidden patterns or clusters in the data set. For the initial SOM experiment we will use a map of size 10×10 with a hexagonal neighbour function. The following SOM experiments were performed using the "Neural Network" toolbox for Matlab [13] using the default settings.

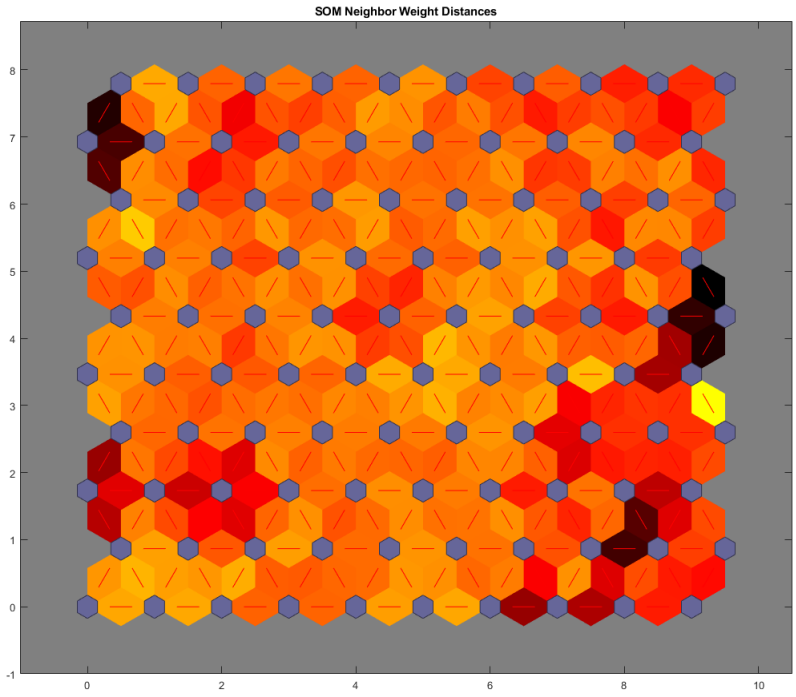

Figure 20: 10×10 SOM U-Matrix

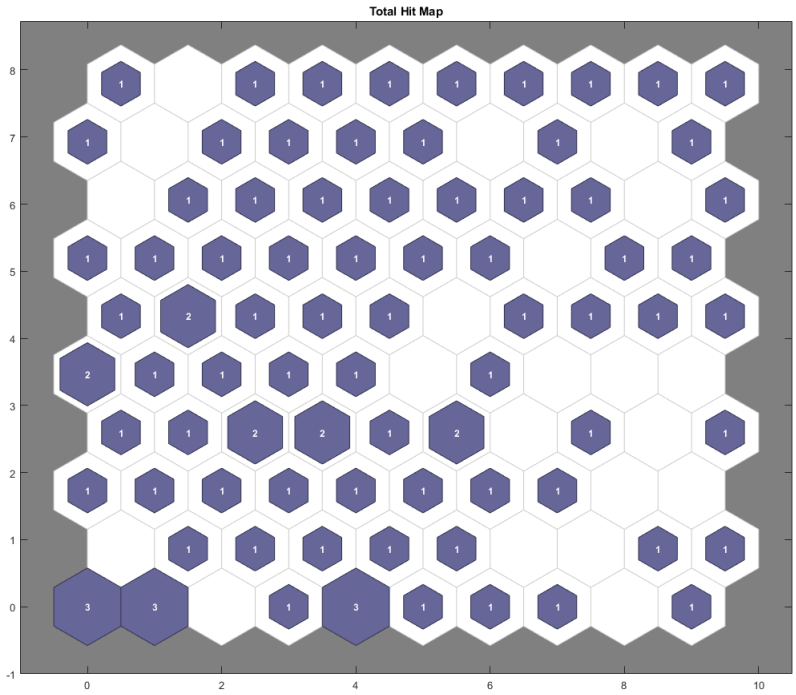

Figure 21: 10×10 SOM Hit Map

In figure 20 we see the U-matrix for the 10×10, the blue hexagons represent the different nodes with the lines between the hexagons representing the distance between the nodes, darker colours meaning the nodes are further apart. The nodes are almost entirely evenly spread out across the 10×10 grid, with no clear decision boundary separating potentially hidden clusters. The U-matrix gives very little insight into any patterns in the data. To try and glean some information out of the SOM we look at 21, the hit map for the 10×10 SOM. The blue hexagons represent the number of data points each of the nodes has captured. Again we see that the data is very spread out between the different nodes with no clear clusters. We can also look at the hit maps of the individual classes that were defined for the supervised learning.

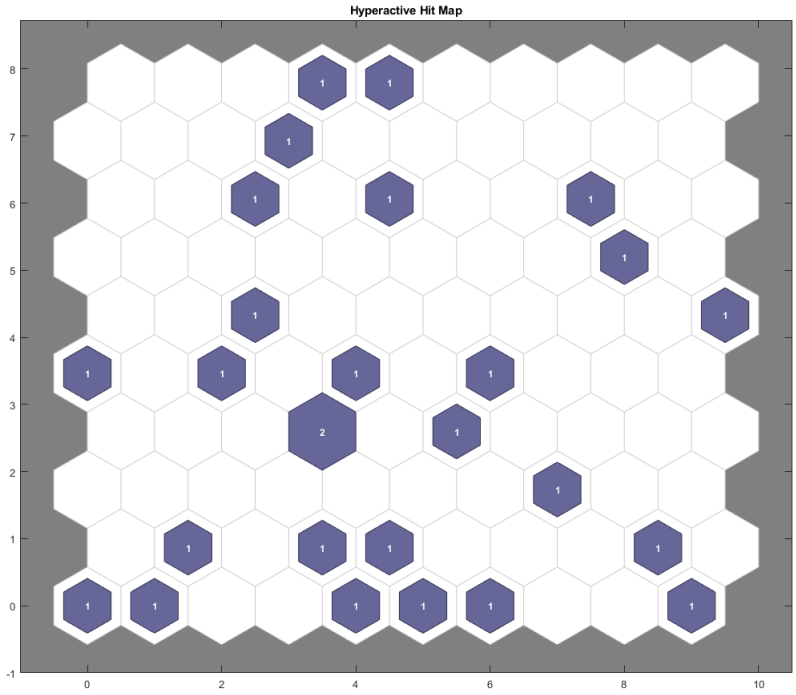

Figure 22: 10×10 Hyperactive Hit Map

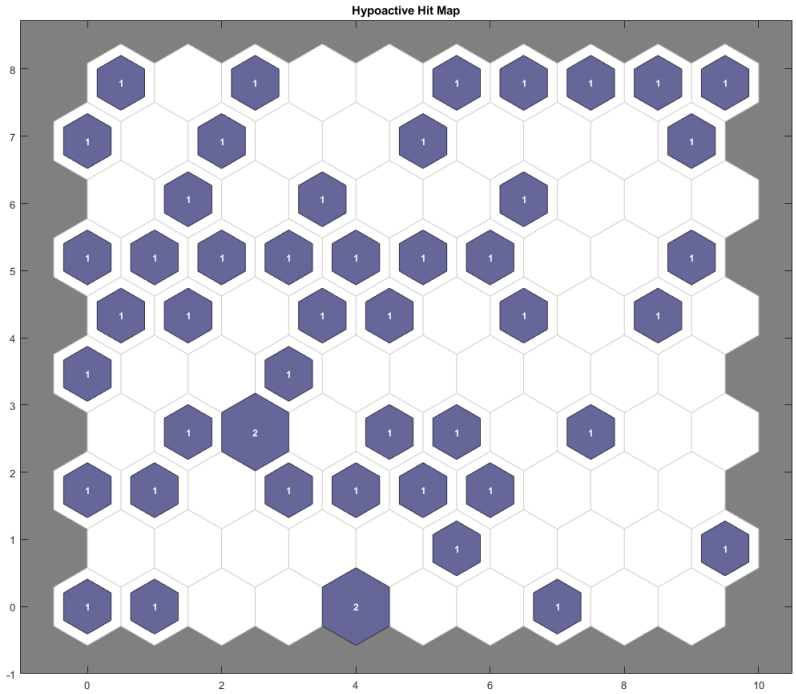

Figure 23: 10×10 Hypoactive Hit Map

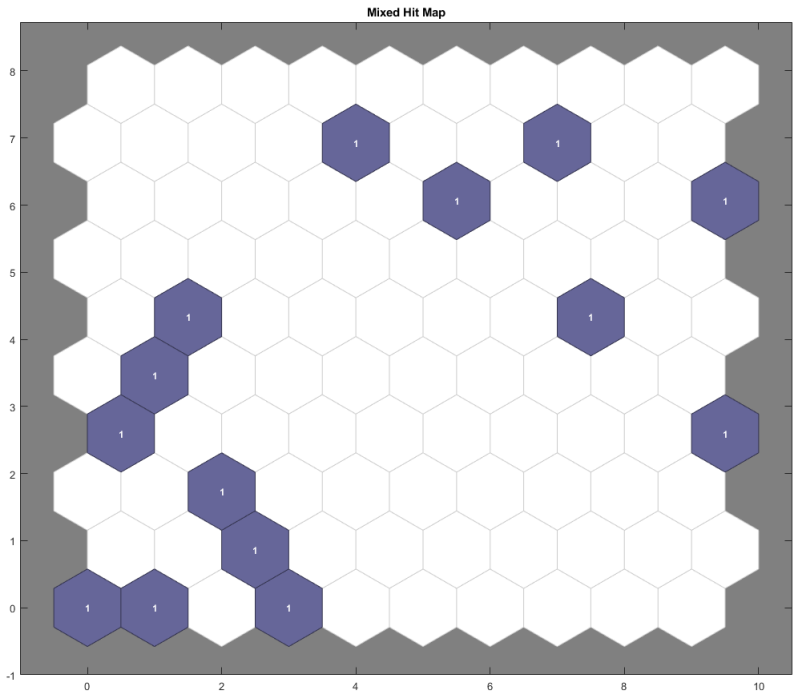

Figure 24: 10×10 Mixed Hit Map

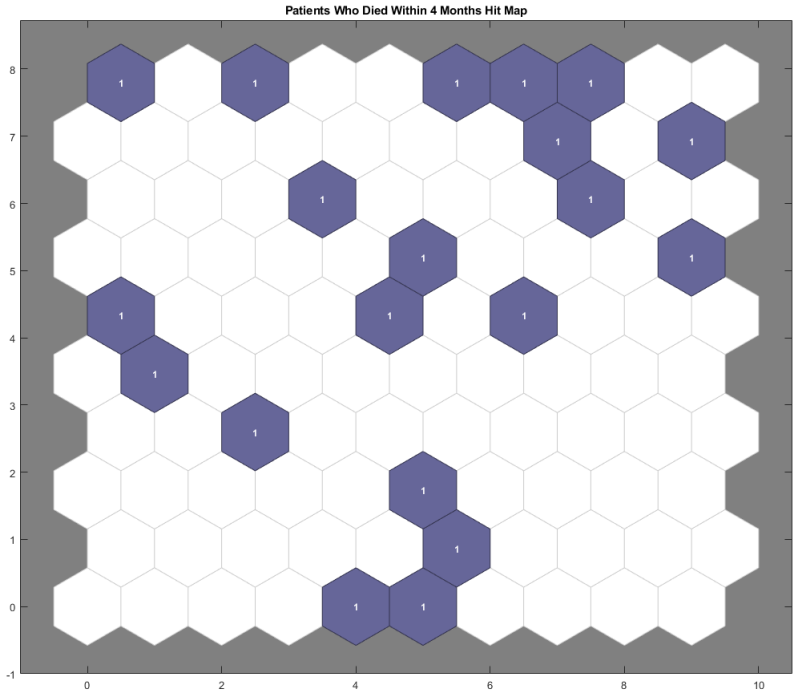

Figure 25: 10×10 Patients who Died Within 4 Months Hit Map

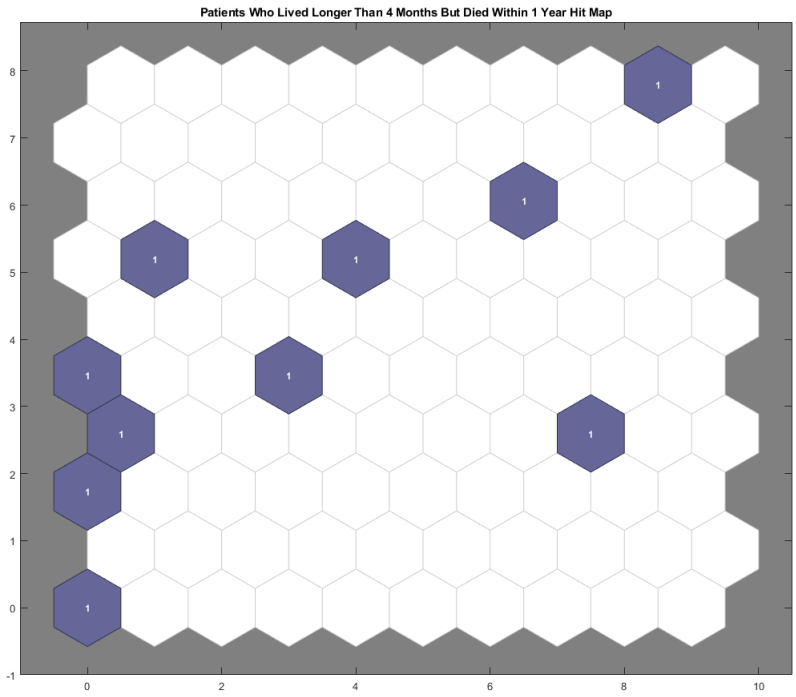

Figure 26: 10×10 Patients who Survived Longer Than 4 Months Died Within 1 Year Hit Map

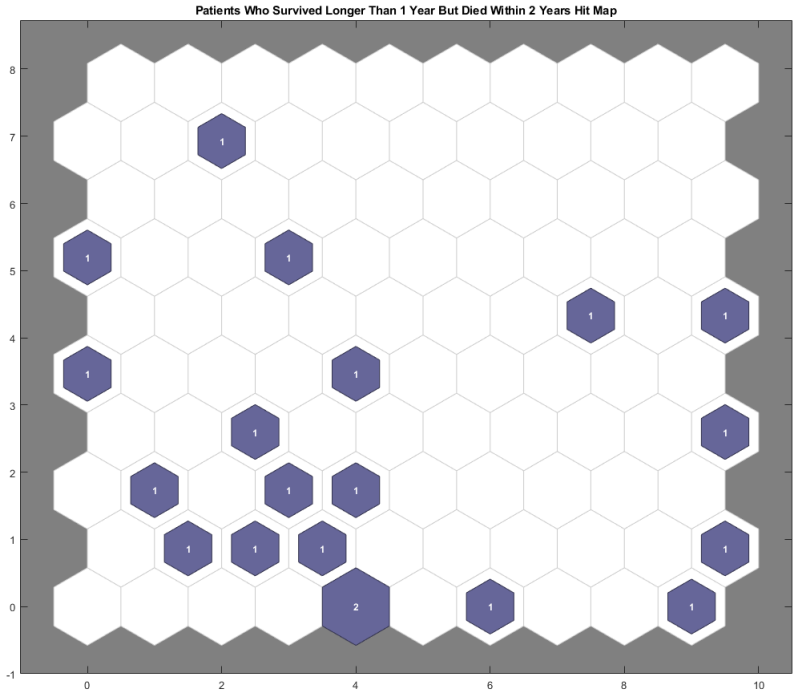

Figure 27: 10×10 Patients who Survived Longer Than 1 Year Died Within 2 Years Hit Map

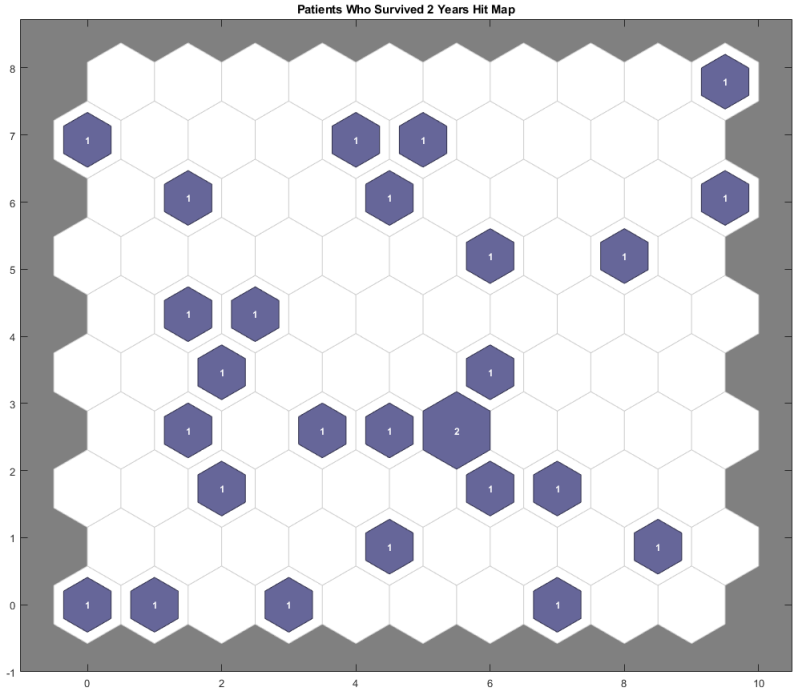

Figure 28: 10×10 Patients who Survived Longer than 2 Years Hit Map

In figures 22, 23 and 24 we see the hit maps for the different subtypes of delirium. Again, these hit maps show the data very spread out with very few patterns with none of the data clustering strongly by any of the similar nodes. The same is true for figures 25, 26, 27 and 28, which show the hit maps for the patients with different lifespans. This poor performance of the SOM may be due to having too large a grid size, so we will try reducing the grid size from 10×10 to 6×6 and see if it has a noticable impact on the performance of the SOM.

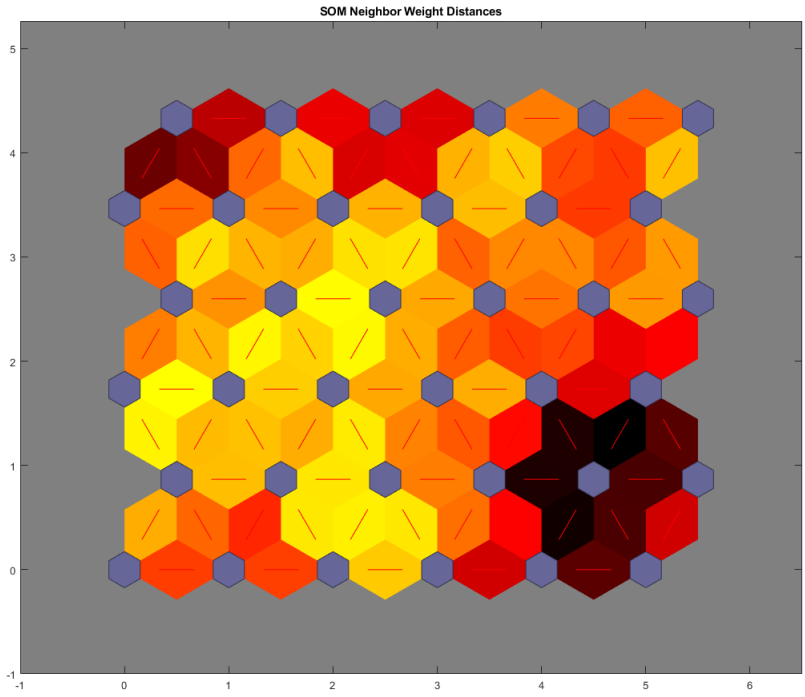

Figure 29: 6×6 SOM U-Matrix

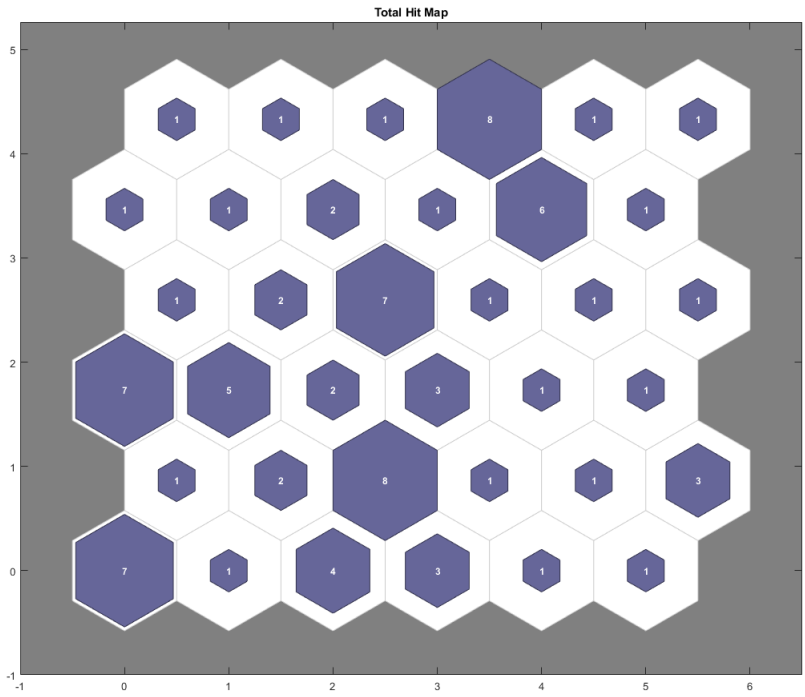

Figure 30: 6×6 SOM Hit Map

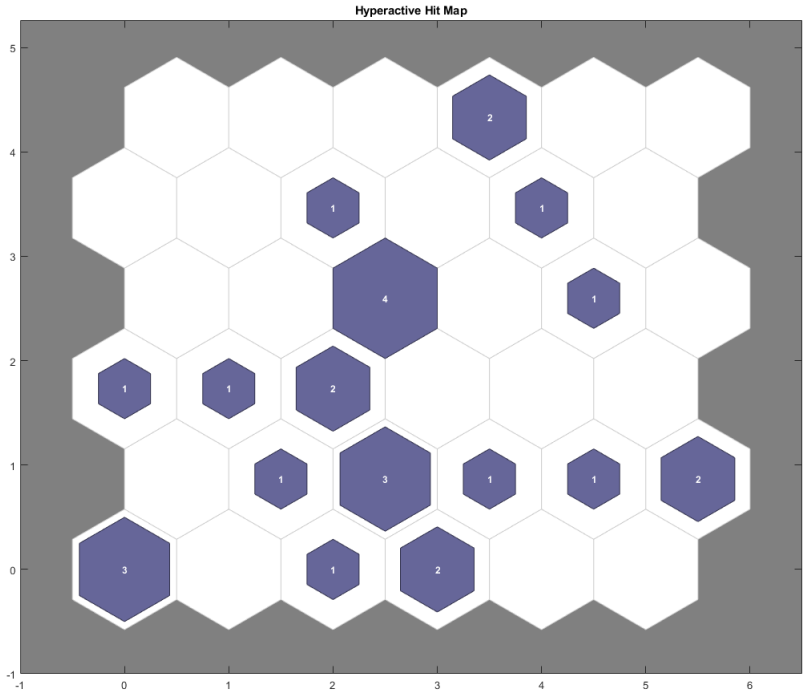

Figure 31: 6×6 Hyperactive Hit Map

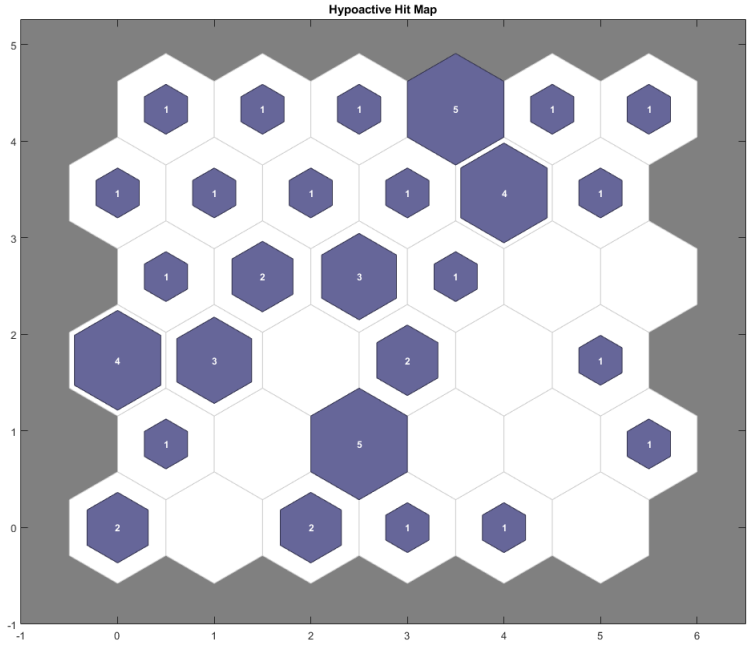

Figure 32: 6×6 Hypoactive Hit Map

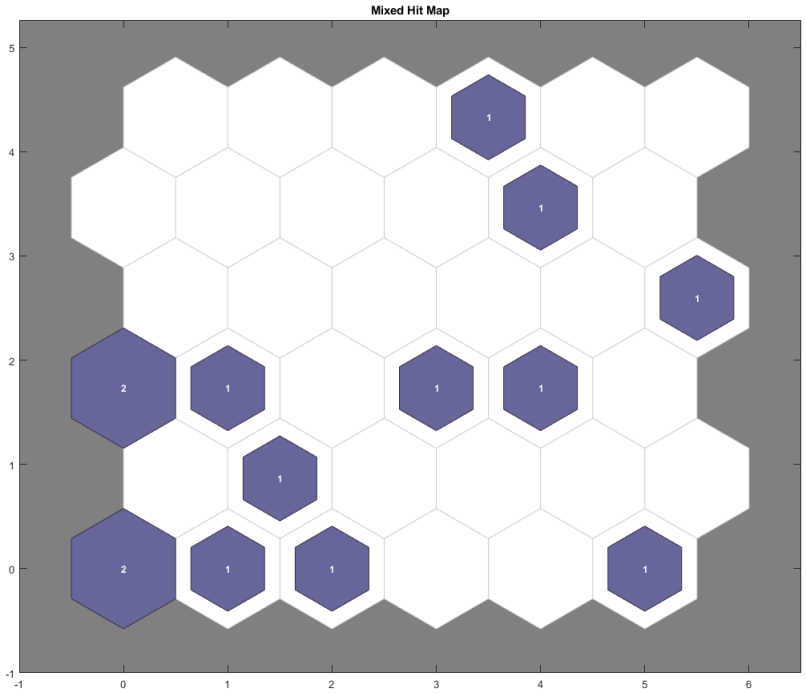

Figure 33: 6×6 Mixed Hit Map

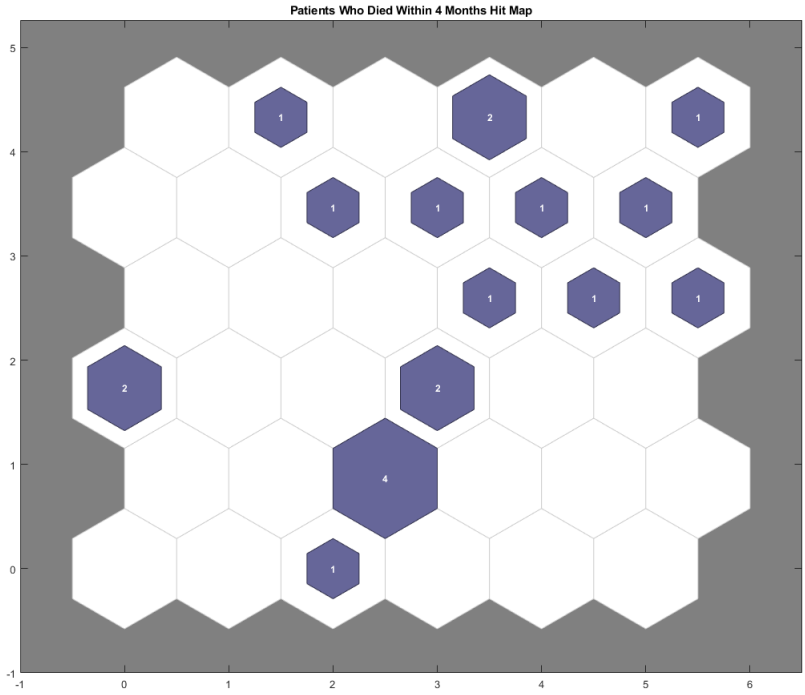

Figure 34: 6×6 Patients who Died Within 4 Months Hit Map

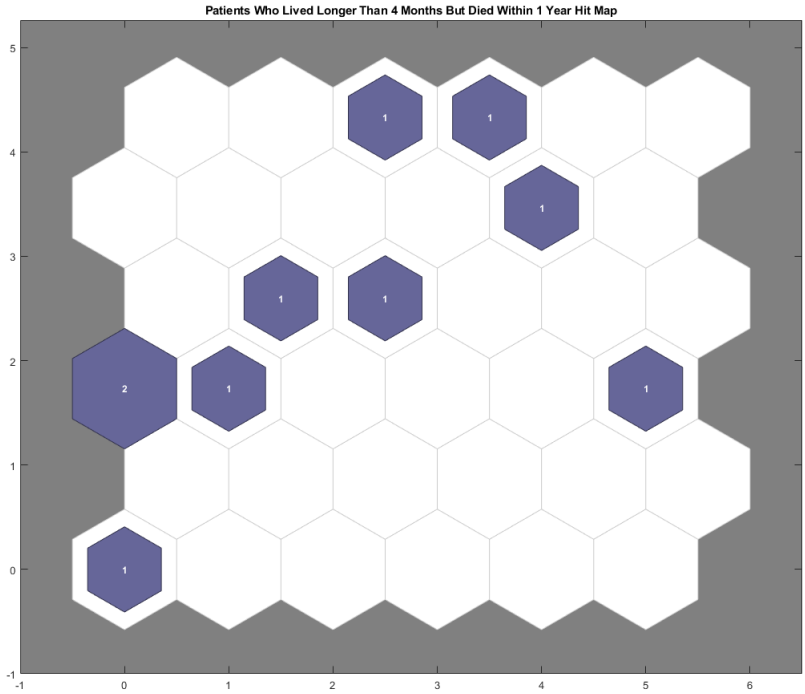

Figure 35: 6×6 Patients who Survived Longer Than 4 Months Died Within 1 Year Hit Map

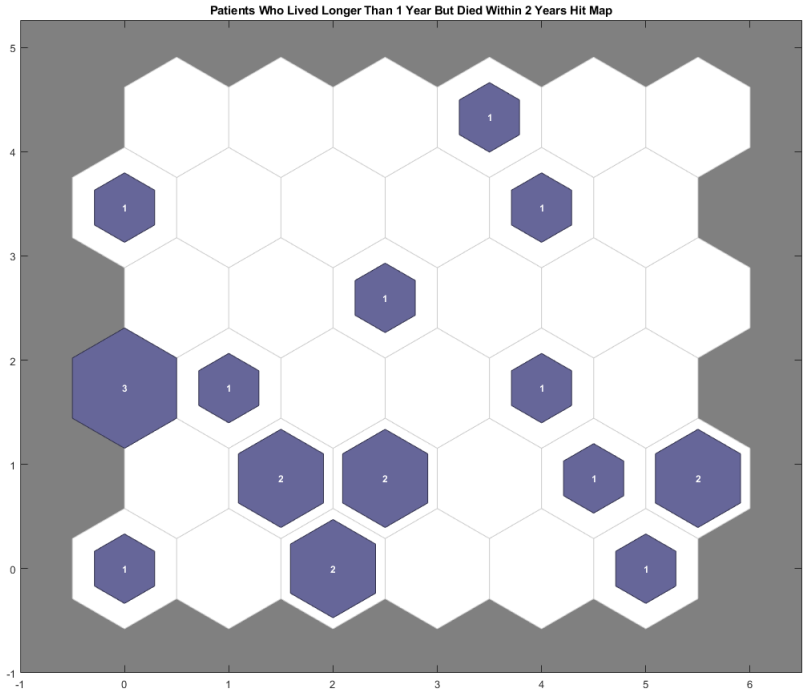

Figure 36: 6×6 Patients who Survived Longer Than 1 Year Died Within 2 Years Hit Map

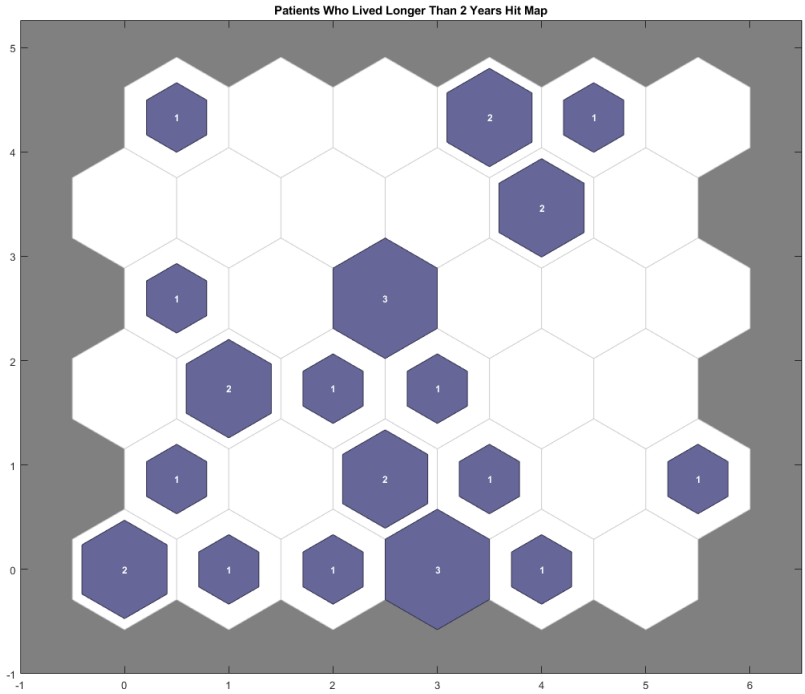

Figure 37: 6×6 Patients who Survived Longer than 2 Years Hit Map

The resluts of the 6×6 SOM have not improved much on the 10×10 one, with no clear decision boundaries and no strong clusters.

6 Conclusion

In conclusion we were able to create a classifier that could differentiate between patients suffering from hyperactive and hypoactive subtypes of delirium using GMLVQ, but this classifier was not hugely accurate and would not be usable as the sole diagnostic tool. It did however uncover a relationship between a patient's Urea and CRP values which served as the main decisive power of the classifier, which may be medically interesting.
We were also able to create a classifier that could predict a patient's lifespan better than a random classifier would have, but none of them were accurate enough to give much indication of what the causes of the differing lifespans might be. Perhaps this indicates that a patient's blood and cytokine values only play a small role in a patient's prospective lifespan.
The unsupervised machine learning technique SOM gave very little insight to the nature of the data set, with no clear decision boundaries between potential classes.
The data set had a few limitations, there were relatively few patients, with a number of them needing to be dropped due to incompleteness. Different and new patterns may be uncovered with a larger and more complete data set, but the results shown give a good preliminary look at what relationships exist between the data set, the different subtypes of delirium and patient lifespans.
Future work could see the use of LGMLVQ [20], GMLVQ using local matrices for each class which could improve the performance of the classifier. The data set could also be extended with other features from other medical test, urine samples for instance. Other machine learning techniques could be applied, Random Forests [14] and Support Vector Machines [15] have both been proven as effective classifying techniques[16][17]. With the presence of control patients who do not have delirium, we could apply GMLVQ and other machine learning techniques to predict the risk of patients developing delirium as well as determining the different subtypes, as seen in other studies [21].

References

[1] Maclullich AM, Anand A, Davis DH, Jackson T, Barugh AJ, Hall RJ, Ferguson KJ, Meagher DJ and Cunningham C (2013) "New Horizons in the Pathogenesis, Assessment and Management of Delirium," Age and ageing, 42(6), pp. 667–74. doi: 10.1093/ageing/aft148.

[2] Morandi, A., Di Santo, S. G., Cherubini, A., Mossello, E., Meagher, D., Mazzone, A., Bianchetti, A., Ferrara, N., Ferrari, A., Musicco, M., Trabucchi, M. and Bellelli, G. (2017) "Clinical Features Associated with Delirium Motor Subtypes in Older Inpatients: Results of a Multicenter Study," The American journal of geriatric psychiatry : official journal of the American Association for Geriatric Psychiatry, 25(10), pp. 1064–1071.

[3] Alpaydin, E. (2014) Introduction to machine learning. Third edition. edn. Cambridge, Massachusetts: MIT Press (Adaptive computation and machine learning). Available at: ProQuest Ebook Central.

[4] Kyan, M., Muneesawang, P., Jarrah, K. and Guan, L. (2014) Unsupervised learning : a dynamic approach. Hoboken, New Jersey: John Wiley & Sons (IEEE Press series on computational intelligence). doi:10.1002/9781118875568

[5] Fawcett, T. (2006) "An Introduction to Roc Analysis," Pattern Recognition Letters, 27(8), pp. 861–874. doi: 10.1016/j.patrec.2005.10.010.

[6] Kohonen, T. (1998) "The Self-Organizing Map," Neurocomputing, 21(1), pp. 1–6. doi: 10.1016/S0925-2312(98)00030-7.

[7] Biehl, M., Schneider, P., Smith, D., Stiekema, H., Taylor, A., Hughes, B., ... Arlt, W. (2012). Matrix relevance LVQ in steroid metabolomics based classification of adrenal tumors. In M. Verleysen (Ed.), Proc. 20th European Symposium on Artificial Neural Networks: ESANN 2012 (pp. 423-428). d-side publishing.

[8] Schneider, P., Biehl, M. and Hammer, B. (2009) "Adaptive Relevance Matrices in Learning Vector Quantization," Neural computation, 21(12), pp. 3532–3561.

[9] Mudali, D., Biehl, M., Meles, S., Renken, R., Garcia Garcia, D., Clavero, P., Arbizu, J., Obeso, J., Oroz, M., Leenders, K., & Roerdink, J. (2016). "Differentiating Early and Late Stage Parkinson's Disease Patients from Healthy Controls." Journal of Biomedical Engineering and Medical Imaging, 3(6), 33. doi:http://dx.doi.org/10.14738/jbemi.36.2280

[10] Biehl, M., Hammer, B. and Villmann, T. (2016) "Prototype-Based Models in Machine Learning," Wiley Interdisciplinary Reviews. Cognitive Science, 7(2), pp. 92–111.

[11] Hastings, C. Mosteller, F. Tukey, J. and Winsor, C. (1947) "Low moments for small samples: a comparative study of order statistics." The Annals of Mathematical Statistics, pages 413–426,.

[12] Biehl M. A no-nonsense beginners tool for GMLVQ (version 2.3). 1 2017. Available online, http://www.cs.rug.nl/ biehl/gmlvq

[13] Mathworks Neural Network toolbox. (2018) Available online, https://www.mathworks.com/products/neural-network.html

[14] Breiman, L. (2001) "Random Forests," Machine Learning, 45(1), pp. 5–32. doi: 10.1023/A:1010933404324.

[15] Mammone A, Turchi M and Cristianini N (2009) "Support Vector Machines," Wiley Interdisciplinary Reviews: Computational Statistics, 1(3), pp. 283–289. doi: 10.1002/wics.49.

[16] Alickovic, E. and Subasi, A. (2016) "Medical Decision Support System for Diagnosis of Heart Arrhythmia Using Dwt and Random Forests Classifier," Journal of Medical Systems, 40(4), pp. 1–12. doi: 10.1007/s10916-016-0467-8.

[17] Guyon, I., Weston, J., Barnhill, S. and Vapnik, V. (2002) "Gene Selection for Cancer Classification Using Support Vector Machines," Machine Learning, 46(1-3), pp. 389–422. doi: 10.1023/A:1012487302797.

[18] Sato, A. and Yamada, K. (1996) "Generalized Learning Vector Quantization," Advances in neural information processing systems, 8(8), pp. 423–423.

[19] Hammer, B., Strickert, M. and Villmann, T. (2005) "On the Generalization Ability of Grlvq Networks," Neural Processing Letters, 21(2), pp. 109–120.

[20] Schneider, P., Biehl, M. and Hammer, B. (2009) "Adaptive Relevance Matrices in Learning Vector Quantization," Neural computation, 21(12), pp. 3532–3561.

[21] Wong A., Young A.T., Liang A..S, Gonzales R., Douglas V.C., Hadley D. (2018) "Development and Validation of an Electronic Health Record–Based Machine Learning Model to Estimate Delirium Risk in Newly Hospitalized Patients Without Known Cognitive Impairment." JAMA Network Open. 2018;1(4):e181018. doi:10.1001/jamanetworkopen.2018.1018