# Application of Machine Learning Algorithms in Presumptive Diagnosis of Urinary System Diseases

Alekhya Pottimuthi, Meghana Nagarahalli Paramesh, Rajashree Dahal and Sneha Vuppala
Department of Computer Science, The University of Alabama in Huntsville

## 1. Introduction

The main idea of this paper is to prepare a machine learning based algorithm of the expert system, which will perform the presumptive diagnosis of two urinary system diseases which are acute Nephritis of renal pelvis origin and acute Inflammation of urinary bladder. Acute nephritis is a bacterial infection of the kidney's tubules and renal pelvis. It is a type of Urinary Tract Infection (UTI) that typically occurs when bacteria from the bladder travel up the ureters to the kidneys. This condition can cause symptoms such as sudden fever with or greater than 40 degrees Celsius which is accompanied by shivers and lumbar pains. Acute inflammation of urinary bladder is a bacterial infection of the bladder. It is a type of UTI that can cause symptoms such as frequent and urgent urination, pain or burning sensation during urination, lower abdominal pain or discomfort, urination in the form of constant urine pushing, and micturition pains. The baseline paper Czerniak & Zarzycki (2003) which we referred used rough set theory in presumptive diagnosis of these urinary system diseases. We are aiming to implement machine learning algorithms for introducing the same expert-based model system that the paper introduced.

## 2. Related Works

The paper that we are following introduces a model of the expert system which performs presumptive diagnosis of two urinary system's diseases acute Nephritis of renal pelvis origin and acute Inflammation of urinary bladder using rough set theory (Czerniak & Zarzycki, 2003). Artificial intelligence technologies are being focused on for implementation in circumstances where complexity occurs. Artificial intelligence has been successfully used in the field of medicine for several purposes, including the categorization and diagnosis of illnesses, the suggestion of treatments, the determination of medication dose, and many more (Baxt, 1995; Saritas, 2012). Likewise, Ozkan et al. (2018) in their paper also carry out classification task to diagnosis UTI with Decision Tree, Support Vector Machines, Random Forest and Artificial Neural Network. Chen et al. (2022) implemented and compared the performance of Neural Network and Logistic Regression algorithms to predict the probability of Urinary Tract Infection which is caused by Cystoscopy and found out that neural network model had superior predictive performance in comparison to Logistic Regression model with sensitivity of just 2% which just predicted that most patients had low risk of infection. Likewise, Goździkiewicz et al. (2022) in their literature review on the use of AI based algorithms in the diagnosis of urinary tract infection and concluded that there is relevance of AI based models in the diagnosis of Urinary Tract Infections but it has not yet been decided which model is preferable in the prediction of infection and figured out XGBoost, Artificial Neural Network (ANN), and Support Vector Machines (SVM) were some of the algorithms found in the review.

So based on these literature review, we have decided to implement ANN, Logistic Regression, Decision Tree Classifier and Support Vector Machines as our model for the diagnosis of urinary system diseases.

## 3. Dataset Description

| | Temperature | Nausea | Lumbar Pain | Urine Pushing | Micturition pains | Burning | Inflammation | Nephritis |
|---|---|---|---|---|---|---|---|---|
| 0 | 35.5 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 1 | 35.9 | 0 | 0 | 1 | 1 | 1 | 1 | 0 |
| 2 | 35.9 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 3 | 36.0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 |
| 4 | 36.0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 115 | 41.4 | 0 | 1 | 1 | 0 | 1 | 0 | 1 |
| 116 | 41.5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 117 | 41.5 | 1 | 1 | 0 | 1 | 0 | 0 | 1 |
| 118 | 41.5 | 0 | 1 | 1 | 0 | 1 | 0 | 1 |
| 119 | 41.5 | 0 | 1 | 1 | 0 | 1 | 0 | 1 |

Figure 1: Sample of Acute Inflammations Data Set after proper preprocessing

The dataset was obtained from UCI Machine Learning Repository which is a center for MachineLearning and Intelligent Systems. It was created by a team of medical experts to perform presumptive diagnosis of urinary system diseases.

The dataset consists of 120 observations for different feature 'Temperature', 'Nausea', 'Lumbar Pain', 'Urine Pushing', 'Micturition Pains', 'Burning', 'Inflammation', and 'Nephritis'. The data ranges from categorical for all the features except 'Temperature' which is a float value. Inflammation and Nephritis are the diseases that needs to be classified based on those 6 features. So, the task that can be done through this dataset is classification. Figure 1 shows how the dataset is like.

## 4. Methods

Figure 2 shows the flowchart of the methods that will be implemented in this project.

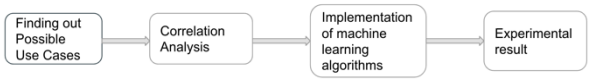


Figure 2: Methodology of the Project

So, basically, we will start figuring out possible use cases and continue with correlation analysis of the dataset to check how all the features affect one another and how they affect the output parameters. The features include temperature, nausea, lumbar pain, urine pushing, micturition pains, burning inflammation and nephritis. The main aim of the paper we are referring out is to build a single expert mode. The expert model that we are trying to introduce will fulfill the following use cases:

- A person can have both Nephritis and Inflammation
- A person can have only Inflammation and not Nephritis
- A person can have Nephritis but not Inflammation
- And finally, a person might not have both.

Since these conditions are mutually exclusive cases, these use cases can be considered as different classes. The given dataset is preprocessed in such a format the problem becomes multi-class classification. In this way, we intend to implement multi-class classification approach to introduce expert based solution. Regarding the implementation of Machine Learning Algorithms, we have surveyed related papers and decided to implement Artificial Neural Network, Decision Tree, Logistic Regression and Support Vector Machines for the multi-class classification problem. After this, the experiments will be carried out based on these models and performance of the models will be compared along with a suitable conclusion.

## 5. Results

### 5.1 Correlation Analysis

The correlation between the target variables (only Nephritis, onlyInflammation, both Nephritis and Inflammation, and none) and the other attributes can help in identifying the most important features in predicting the corresponding use cases. **Only Nephritis:** it is positively correlated with Lumbar Pain, Temperature and Burning. It is negatively correlated with Micturition pains. **Only Inflammation:** It is highly positive correlated with Urine Pushing and weakly positive correlated with Burning. It is negatively correlated with Temperature, Nausea and Lumbar Pain. **Both Nephritis and Inflammation:** It is highly correlated with Nausea, Temperature and Micturition pains. It is negatively correlated with none of the variables. **None:** There is a high positive correlation with Lumbar Pain and all other variables are negatively correlated.

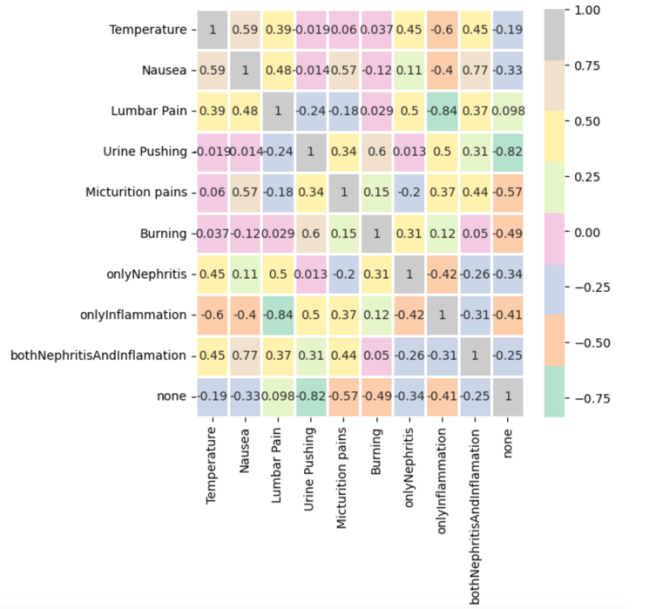


Figure3: Heat map plot showing correlation analysis of all the variables present in Acute Inflammation Dataset

### 5.2 Artificial Neural Network

#### 5.2.1 Baseline settings of the ANN model:

We have the input layer which takes 6 features as the input. We have 2 hidden layers and both of them are batch normalized. LeakyRelu is the default activation function in the hidden layers and the Sigmoid activation function is applied to the output layer. The weights are initialized with a Xavier Uniform initialization. Adaptive learning rate is used (ReduceLROnPlateau). Binary Cross Entropy is used as the loss function. Figure 4 shows the

training and validation loss. Figure 5 shows the training and validation accuracy. The models reach a loss of 0 and an accuracy of 100% after 100 epochs. It is observed that without weight decay as well the model performed the same.

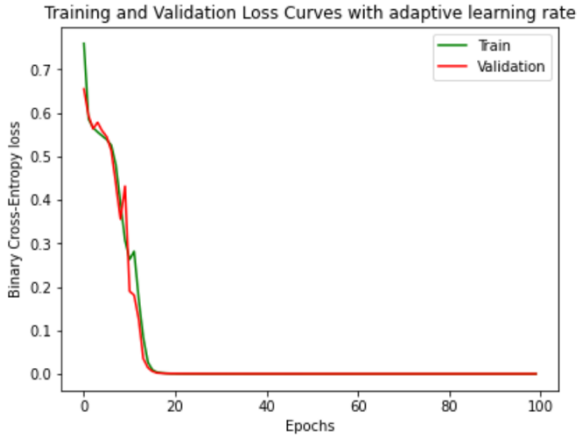


Figure 4: Training and Validation Loss with weight decay

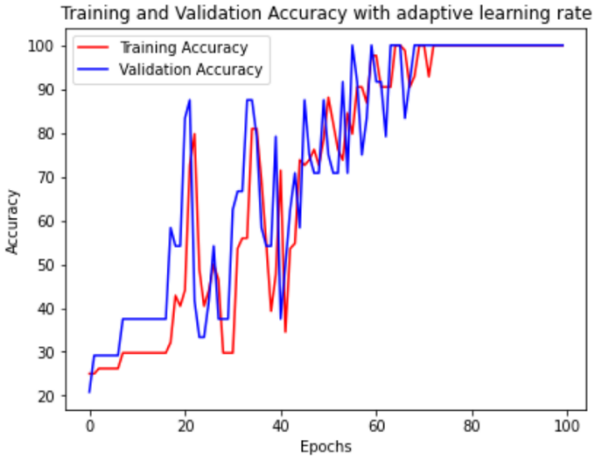


Figure 5: Training and validation accuracy with weight decay

### 5.2.2 Optimizer Used

RAdam (Rectified Adam) is used as the optimizer. Like Adam, RAdam combines the benefits of adaptive learning rates with momentum-based optimization along with gradient descent. However, RAdam differs from Adam in the way it handles the estimation of the adaptive learning rate. Instead of using the biased second moment estimate (moving average) of the gradients (i.e., the variance) to compute the adaptive learning rate, RAdam uses an uncentered second moment estimate that corrects for the bias introduced in the first few iterations. RAdam also introduces a new term to the optimization process that acts as a stabilizer during the early stages of training. This term is based on the variance of the gradients and helps to prevent the optimizer from diverging before it can converge.

### 5.2.3 Partial Dependency Plot Using MLP Classifier from sklearn

MLP Classifier from sklearn is used to plot the partial dependency separately with respect to inflammation and nephritis. Figure 6 shows the partial dependency of the function for Inflammation with respect to all other features. Figure 7 shows the partial dependency of the function for Nephritis with respect to all the other features.

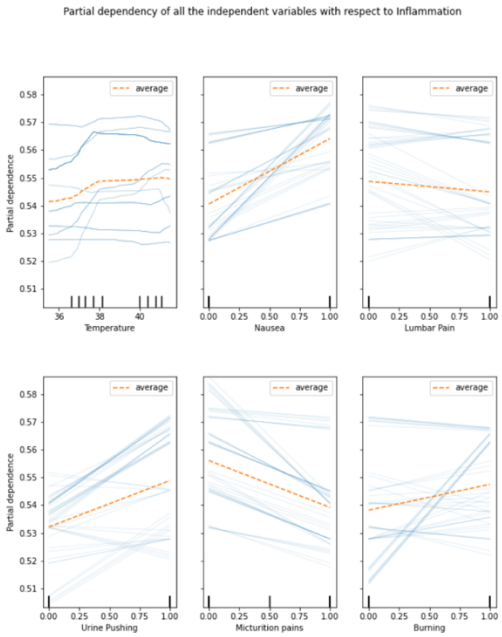


Figure 6: Plot of partial dependency with respect to Inflammation for all the features

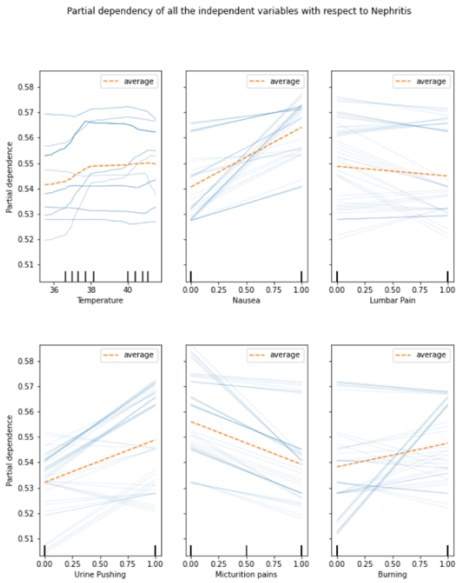


Figure 7: Plot of partial dependency with respect to Nephritis for all the features

## 5.3 Decision Tree Classifier

The dataset was trained setting train size split ranging from 0.1 to 0.9 and corresponding accuracies for Nephritis and Inflammation separately was found out.

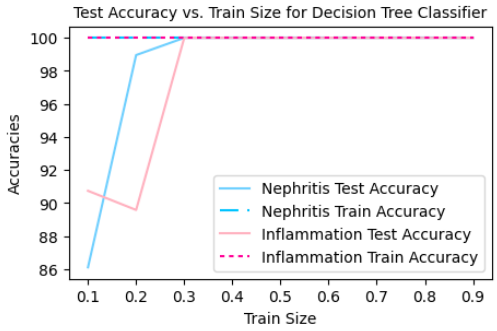


Figure 8: Train and test accuracy for Nephritis and Inflammation classification carried out individually

We can observe in Figure 8 that for both Inflammation and Nephritis the train accuracy was constant and was 100% accurate across all the train split sizes. Likewise, Inflammation test accuracy shows decreasing slope till split of 0.2 and is increasing and maintaining a test accuracy of 100% after split of 0.3. However, in case of Nephritis, test accuracy was found to increase from the beginning and maintains a constant accuracy of 100% after the split of 0.3.

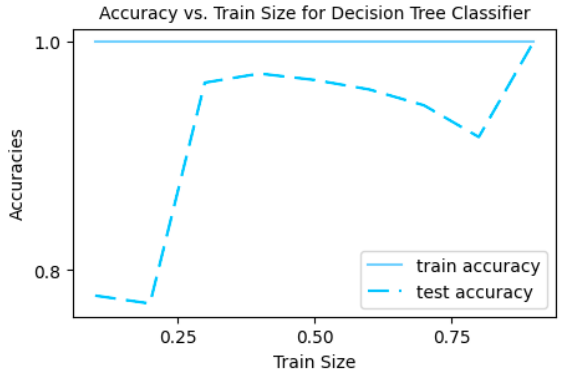


Figure 9: Train test accuracy for Urinary system diseases considering multi-class classification implementing all the use cases

Figure 9 shows the multi-class classification test and train accuracies across different train splits. Here, the performance on train dataset is the same as before. However, the test accuracy shows increasing trend till train size of 0.3 and shows decreasing curve till 0.8 and a sharp increase in train size of 0.9.

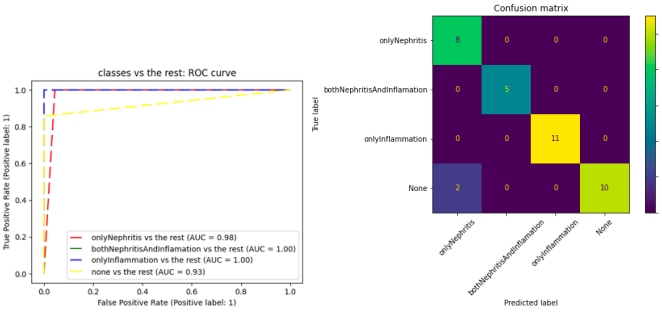


Figure 10: class vs the rest ROC curve (train split=0.7) and confusion matrix

In figure 10, the ROC curve for Decision Tree Classifier for the train size of 0.7 shows that except 'none' cases, all other cases were showing good performance. This fact can be noticed on the confusion matrix where onlyNephritis is predicted twice when its Nephritis actually.

Table 1 shows the performance metrics for Decision Tree Classifier for the train size of 0.1.

Table 1: Performance metrics for Decision Tree Classifier

| | Precision | Recall | F1 score |
|---|---|---|---|
| OnlyNephritis | 0.8 | 1 | 0.89 |
| bothNephritisAnd Inflammation | 1 | 1 | 1 |
| OnlyInflammation | 1 | 1 | 1 |
| none | 1 | 0.83 | 0.91 |
| Macro avg | 0.95 | 0.96 | 0.95 |
| Accuracy | 0.94 | | |

## 5.4 Logistic Regression

Figure 11 shows train and test accuracies of logistic regression model. The x-axis of the graph represents different train size splits and the y-axis represents the training and test accuracy of the model. The training and test accuracy is constant for train split size ranging from 0.2 to 0.9. classification report in Figure 12 shows the performance metrics – Accuracy, Precision, Recall and F1-score for train size of 0.8. In In confusion matrix, the diagonal elements of the matrix represent the correctly classified instances, while the off-diagonal elements represent the incorrectly classified instances. Figure 13 shows ROC curve for this model which is a perfect curve as all the performance metrics are 1.

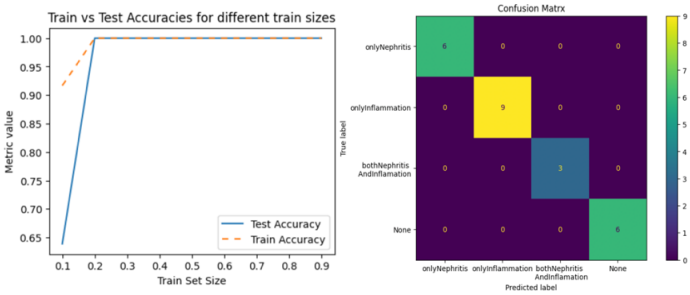


Figure 11: Logistic Regression – Train vs Test accuracies across different train sizes and its corresponding confustion matrix

| | precision | recall | f1-score |
|---|---|---|---|
| onlyNephritis | 1.00 | 1.00 | 1.00 |
| onlyInflammation | 1.00 | 1.00 | 1.00 |
| bothNephritisAndInflamation | 1.00 | 1.00 | 1.00 |
| None | 1.00 | 1.00 | 1.00 |
| accuracy | | | 1.00 |
| macro avg | 1.00 | 1.00 | 1.00 |
| weighted avg | 1.00 | 1.00 | 1.00 |

Figure 12: Performance Metrics for Logistic Regression

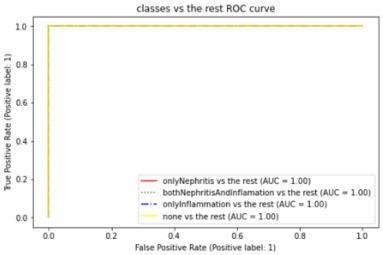


Figure 13: ROC curve – Logistic Regression

### 5.5 Support Vector Machines

Figure 14 shows train and test accuracies of support vector machine model. The x-axis of the graph represents different train size splits and the y-axis represents the training and test accuracy of the model. The test accuracy is constant for train split size ranging from 0.2 to 0.9 whereas train accuracy is constant from the beginning. Classification report in Figure 15 shows the performance metrics – Accuracy, Precision, Recall and F1-score for train size of 0.8. In confusion matrix, the diagonal elements of the matrix represent the correctly classified instances, while the off-diagonal elements represent the incorrectly classified instances.

Since all the performance metrices value are close to 1, we can see that class vs the rest ROC curve overlapped with one another in Figure 16.

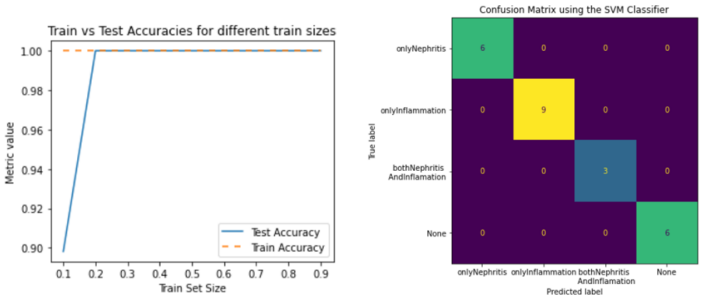


Figure 14: Support Vector Machines – Train vs Test accuracies across different train sizes and its corresponding confustion matrix

| | precision | recall | f1-score |
|---|---|---|---|
| onlyNephritis | 1.00 | 1.00 | 1.00 |
| onlyInflammation | 1.00 | 1.00 | 1.00 |
| bothNephritisAndInflamation | 1.00 | 1.00 | 1.00 |
| None | 1.00 | 1.00 | 1.00 |
| accuracy | | | 1.00 |
| macro avg | 1.00 | 1.00 | 1.00 |
| weighted avg | 1.00 | 1.00 | 1.00 |

Figure 15: Performance Metrics for Support Vector Machines

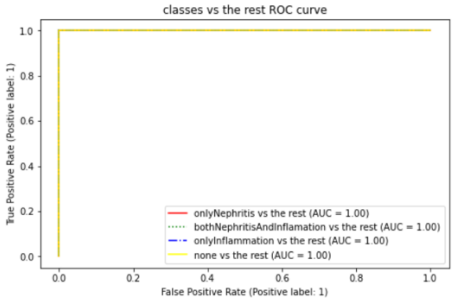


Figure 16: ROC curve –Support Vector Machines

## 6. Discussion and Conclusion

Since the dataset was very small but the performance metrics were close to 1, it indicates high accuracy. The medical datasets are very sensitive and taking decision based on this output might impact lives at different levels. Training Logistic Regression at different cross folds also did not do anything to make model generalizable. Obtaining high performance metrics at even lower train split size shows that the model is not able to learn the dataset properly. Despite these facts, one thing that can be noted is that since some of the features for both the disease classification had common symptoms, it might be because the model had to learn on the remaining few features to learn the features and learning from the remaining three features are just a combination of 8 distinct conditions (2*2*2). Therefore all the models were able to learn in the fewer split sizes. To gain more insights implementing larger data, GAN model can be implemented to generate synthetic dataset on this case so that the model can learn the dataset. However, for this further study is still required. In addition, after this, further research can also be carried out by integrating weighted average of each model to give the final prediction.

Also, the approach introduced in this dataset is somehow good as we can obtain performance metrics which can be interpreted with different figures easily unlike the baseline paper that we have been following.

Likewise, since all the features were implemented at once in all the models, selected features after proper correlation analysis might also give different results where the model is not overfitted.

## 7. References:


Baxt, W. G. (1995). Application of artificial neural networks to clinical medicine. *The Lancet*, *346*(8983), 1135–1138. https://doi.org/10.1016/s0140-6736(95)91804-3

Chen, T., Hsu, Y. F., & Chen, C. H. (2022). Comparison of Neural Network and Logistic Regression Analysis to Predict the Probability of Urinary Tract Infection Caused by Cystoscopy. BioMed Research International, 2022, 1–6. https://doi.org/10.1155/2022/5775447

Czerniak, J. M., & Zarzycki, H. (2003). Application of rough sets in the presumptive diagnosis of urinary system diseases. In *Springer eBooks* (pp. 41–51). Springer Nature. https://doi.org/10.1007/978-1-4419-9226-0_5

Goździkiewicz, N., Zwolińska, D., & Polak-Jonkisz, D. (2022). The Use of Artificial Intelligence Algorithms in the Diagnosis of Urinary Tract Infections—A Literature Review. Journal of Clinical Medicine, 11(10), 2734. https://doi.org/10.3390/jcm11102734

Ozkan, I. A., Koklu, M., & Sert, I. U. (2018). Diagnosis of urinary tract infection based on artificial intelligence methods. *Computer Methods and Programs in Biomedicine*, *166*, 51–59. https://doi.org/10.1016/j.cmpb.2018.10.007

Saritas, I. (2012). Prediction of Breast Cancer Using Artificial Neural Networks. *Journal of Medical Systems*, *36*(5), 2901–2907. https://doi.org/10.1007/s10916-011-9768-0