# Classification of Cardiotocographic data using Support Vector Machine

T. Suarez Lissi[1]

***Abstract*— Foetal heart rate signals measurements and classifications obtained were used to train a model based on Support Vector Machine. The model was then used to predict test data to measure its accuracy rate. Here is the presentation of the results.**

**Keywords: *Machine Learning, Support Vector Machine, Cardiotocography*.**

## I. INTRODUCTION

Clinicians rely on their expertise and experience to correctly classify fetal heart health, based on the information obtained in Cardiotocography (CTG). A multi-class Support Vector Machine (SVM) classifier is proposed to aid the clinician in the classification of fetal heart health, using foetal heart rate (FHR) measurements. The classifier will predict 3 different possible fetal heart health results, these are 'Normal', 'Suspect' or 'Pathologic'. The classifier was trained and tested using 2126 FHR measurements and classifications, taking into account 21 different measurements (features). Finally classification accuracy results were obtained using test data as input.

## II. METHODS AND MATERIALS

### A. Support Vector Machine

SVM is a supervised learning tool that is widely used in classification solutions. The simplified concept of SVM is to find the a hyperplane that maximizes the margin between the closest data points. In the case of overlapping data, a slack variable is introduced to represent the deviation of the data point from the hyperplane, and a scalar C that adjusts the weight of this penalization term, making the margin smaller and allowing less training errors for high values, and larger and allowing more training errors for low values. We then need to select the value of this parameter that leads to less classification errors. The formulation of the classifier then, takes the form of a constrained quadratic programming problem [1]:

$$\min_{w,\xi_i} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{L} \xi_i \tag{1}$$

Subject to:

$$y_i(\mathbf{w} \cdot \mathbf{x} + b) \geq 1 - \xi_i \qquad \forall i$$
$$\xi_i \geq 0 \qquad \forall i$$

This is the primal formulation of the problem. Using a function that maps the m dimensional input points to m feature space $\psi(x_i)$ and solving the Lagrangian equation resulting, the problem can be expressed the following way, and is called the dual formulation of the problem [1]:

$$\min_{w,\alpha_i} \sum_{n=l}^{L} \alpha_i - \frac{1}{2} \sum_{i=1}^{L} \sum_{j=1}^{L} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{2}$$

Subject to:

$$\sum_{i=1}^{L} \alpha_i y_i = 0 \qquad \forall i$$
$$0 \leq \alpha_i \leq C$$

We see that now we can make use of kernels to find a better separation hyperplane. In this case the proposed kernel is the Gaussian radial kernel, with the following formulation:

$$K(x, x') = e^{\frac{-\|x-x'\|}{2\sigma^2}} \tag{3}$$

We need to select the hyperparameter $\sigma$ of the gaussian kernel that best fits to the training dataset (same as with hyperparameter C).
Finally, the discriminating function has the following expression [4]:

$$f(x) = sign\Big[\sum_{n=l}^{L} y_i \alpha_i (\psi^T(x_i)\psi(x)) + b\Big] \tag{4}$$

Now we can proceed with the analysis of the data.

### B. Pre-processing

We first start by analyzing the dataset, looking for missing values or features that provide no information, for example, the case where all the observations have the same value. If there are missing values, these are filled with the corresponding feature mean value. In the case of features that do not add information, the proceeding is to remove the entire column of the dataset.
To minimize the impact of the scales of the different features, the standarization of the data is performed. This consist of removing the mean and scaling to unit variance.

### C. PCA

The following step is to analyze the dataset looking for a possible feature selection, thereby reducing the "curse of dimensionality". Taking this into account, Principal Component Analysis (PCA) is performed to reduce the dimensionality of the data set while retaining most variance. We select the number of components that explains approximately 95% of the variance, therefore enabling the possibility of

[1]T. Suarez Lissi is with Universite de Technologie de Troyes, France. And Universidad Tecnologica Nacional, Argentina. The repository with the code implementation can be found on https://github.com/TomasSuarezL/os14-final-practice

reducing the number of dimensions. Details of the results of the application of this technique are explained in Figure 1. The underlying mathematics of this technique would not be specified in this work.

### D. Cross-Validation

From the initial data set, we proceed to split it into training and test sets of 80% and 20% of the total dataset respectively. A stratified approach is considered to balance the difference in observations of the different classes. This will allow us to test the model with this data that was not used for training. Next the before mentioned hyperparameters of the SVM classifier, $C$ and $\sigma$ are found using a grid search cross-validation (CV) with fold split set to 5 ($cv = 5$), that allowes us to find the best values for both hyperparameters, by iterating over the train dataset, trying the different values selected as candidates for parameters that gives the best results. Using this techinque we can narrow down on values that approximate to the optimal for this type of model.

### E. Final Model

Finally the Multi-class SVM classifier with CV method obtained hyperparameters is trained using the training samples that has been scaled and reduced in dimension using PCA, and with the training classifications. If the classes are more than 2, it is necessary to specify an approach to deal with multiple classes [2].

## III. EXPERIMENTS

### A. Background

There are a significant number of previous works that developed different classifiers using different machine learning models [3][4][5]. The accuracy results for SVM classifiers obtained range between 87% and 89% and for other models it could go as high as 99% of accuracy for the 3-class classification, meaning that this is a well studied situation, and there exists very accurate solutions, suggesting that this problem could rely on machine learning solutions effectively. Knowing this accuracy values, we proceed to train and measure the previously proposed 3-Class SVM classifier.

### B. Experiment

The selection of the features resulted in a 21 feature dataset, where the "Repetitive Deceleration" (RD) measure was taken out because it was the same in all cases. The original dataset included a classification field that can take 10 different values that was not used for this experiment, and the 3-class NSP problem was chosen. For this experiment, the dataset used to train the model had 1700 samples (80% of 2126 total samples) and the remaining was used as test set with 426 samples (20% of 2126 total samples) balanced by using a stratified data split, both containing 14 dimensions after applying PCA, as explained in Figure 1. In this case, the dimension of the dataset is 21, this is a relatively low value and should not generate great difficulties, but the results of applying PCA shows that with 14 components we can conserve approximately 95% of the variance, therefore reducing the dimension by 33%, and we see that this is a very effective method.

As was mentioned, the possible outcomes for the predicted values are 'Normal', 'Suspect' and 'Pathologic', so this is a 3-class classification model. A 5-Fold cross-validation grid search was performed using the following combination of parameters:

$$C = [5, 10, 15, 50, 100]$$

and

$$\sigma = [0.01, 0.1, 0.3, 0.5]$$

For this dataset, the found best values are $C = 10$ and $\sigma = 0.1$, resulting in a training score of $\cong 0.92$.

Finally, the Multi-class SVM classifier is trained using both the transformed train samples and labels. The multi-class classification problem was addressed using a one-vs-rest approach. To find the accuracy of the built model, the prepared test set was used to find the corresponding predictions given by the classifier. In this case, 426 classifications were obtained and were compared with the same number of previously known classifications.

## IV. CONCLUSIONS

To compare the results predicted by the classifier with the ones known beforehand (the classifications selected for testing), we proceed to count the errors in classification for each of the 3 classes that the model predicted and compute the ratio of this errors with the total number of predictions done. We can visualize this with a confusion matrix that let us clearly view the correct and wrong predictions for each class. Figure 2 shows the confusion matrix obtained.

With this results, the accuracy of the classifier was calculated yielding a value of 91.31% for the test set, while the score for the training set is 98.35% . Calculating the results for each class, we get the following values:

$$Normal : 94\%$$

$$Suspect : 73\%$$

$$Pathological : 97\%$$

Observing these results as well as the confusion matrix, we note that the model performs significantly worst while trying to predict data from the 'Suspected' class, and this is has a negative impact on the accuracy of the whole model. Additional analysis has to be done to find the causes of this difference, but we suggest that another approach should be implemented to increase the accuracy in the classification of this particular class.

To conclude, we should say that this classifier should not be used for this purpose given the low accuracy of the model and we consider that this type of problem is critical and the solution should be highly effective for the clinician to safely rely on it. Further improvements could be made to increase the accuracy of this SVM classifier, testing different kernel combinations with a wider range of hyperparameters, but most likely the desired accuracy would not be achieved.

## APPENDIX

The following figures complement the work.

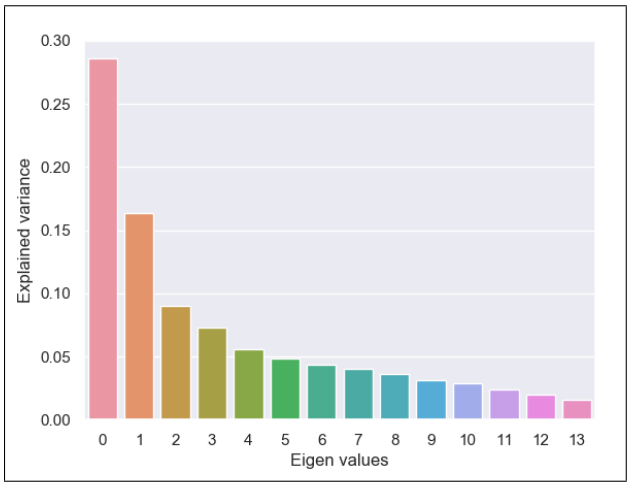


Fig. 1. PCA performed on the dataset. When considering 14 components the total variance is $\cong 0.95$

| | Normal | Suspected | Pathologic |
|---|---|---|---|
| Normal | 316 | 15 | 1 |
| Suspected | 12 | 44 | 3 |
| Pathologic | 3 | 1 | 31 |

Fig. 2. Confusion Matrix obtained as output of test set prediction

## REFERENCES


[1] T. Evgeniou and M. Pontil. Support vector machines: Theory and applications. In *Machine Learning and Its Applications: Advanced Lectures*. IEE, 2001.

[2] C.W. Hsu, and C.J. Lin, A Comparison of Methods for Multi-Class Support Vector Machines. In *IEEE Trans. Neural Networks*, vol. 13, pp. 415-425, 2002.

[3] UZUN, A., KIZILTAS, E.C. and YILMAZ, E. Cardiotocography Data Set Classification with Extreme Learning Machine.

[4] Krupa N., Ali M., Zahedi E., Ahmed S., Hassan F. M. (2011). Antepartum fetal heart rate feature extraction and classification using empirical mode decomposition and support vector machine. Biomed. Eng. Online 10:6. 10.1186/1475-925X-10-6d.

[5] Zhang, Y. and Zhao, Z., 2017, October. Fetal state assessment based on cardiotocography parameters using PCA and AdaBoost. In *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (pp. 1-6). IEEE.