

Automated Wrist Pulse diagnosis of Pancreatitis via Autoregressive Discriminant models

S X Lee¹, S Y Leemaqz²

¹*School of Mathematics and Physics, University of Queensland*

²*Adelaide Medical School, University of Adelaide*

Email: shalem.leemaqz@adelaide.edu.au)

Abstract: Wrist pulse signals, commonly used in traditional oriental medicine, reflects important pathological changes in the body which may be utilized to characterize an individual's health status. Being one of the 'four pillars of clinical evaluation', pulse diagnosis plays a critical role in traditional Chinese medicine (TCM), informing the physician about crucial information such as the state of balance of the body and the state of internal organs. The traditional method of examination is via palpation in which the practitioner uses fingertips to feel the radial pulse of the patient. It is thus highly subjective and depends heavily on the practitioner's experience. With the aid of technology, modern measurements of pulse signal can be taken in a more objective manner. However, there is a lack of tools and standards for analysing and interpreting these computerized pulse signals. In particular, diagnosis of inflammation is challenging and current clinical approaches via other methods (such as blood test, urine test, and X-ray) are time-consuming and often inaccurate.

This paper proposes an automated procedure for distinguishing patients with inflammation using their digitized wrist pulse signals, based on a two-stage time-series classification technique involving autoregressive models followed by common classification methods such as Linear Discriminant Analysis (AR-LDA) and Logistic regression (AR-LR). We focus on one of the major inflammation symptoms – pancreatitis, a condition that can potentially lead to fatal complications in severe cases. We work with wrist pulse signals captured from patients using a Doppler ultrasonic blood analyser. After pre-processing, features derived from fitted AR models were fed to train a LDA and a LR classifier. The effectiveness of our approach is demonstrated using a subset of the wrist pulse database from Chen et al. (2009), consisting of 100 healthy persons and 54 Pancreatitis patients. To evaluate the classification performance, the models are trained on 50% of the data while the remaining observations were reserved for testing. An overall accuracy of 83% and 91% was achieved from AR-LR and AR-LDA respectively, both with an area under curve of 0.88. AR-LDA achieved a higher sensitivity (81%) and specificity (96%) with a positive predictive value of 92%. These results showed that AR-LDA is a promising approach for classifying time-series wrist pulse signals. It provides a relatively low-cost, easy-to-implement, and reliable tool for modern computerized wrist pulse diagnosis.

Keywords: *Classification, time series, autoregressive model, linear discriminant analysis, logistic regression*

1. INTRODUCTION

Pulse diagnosis is an essential component of Traditional Chinese Medicine, being one of the ‘four pillars of clinical evaluation’, with history of its use dating back to thousands of years. It is believed that an individual’s wrist pulse reflects important pathological changes in the body which may be utilized to characterize an individual’s health status. The traditional method of examination is via palpation in which the practitioner uses fingertips to feel the radial pulse of the patient, which is highly subjective and depends heavily on the practitioner’s experience. Hence, with the aid of technology, modern measurements of digitized pulse signal can be taken in a more objective manner. However, there is a lack of tools and standards for analysing and interpreting these computerized pulse signals. In particular, diagnosis of inflammation is challenging and current clinical approaches via other methods (such as blood test, urine test, and X-ray) are time-consuming and often inaccurate.

2. DATA

The data from this study is a subset of the wrist pulse database from Chen et al. (2011), consisting of 100 healthy persons and 54 Pancreatitis patients. The wrist pulse signals of patients were captured using a Doppler ultrasound pulser, and their spectrogram over a duration of 1800 seconds were recorded for analysis. Prior to analysis, the maximum velocity envelope was extracted from the spectrograms to reduce signal dimension. High-frequency de-noising and low-frequency drift removal were then performed using wavelet transform techniques (Xu et al., 2005).

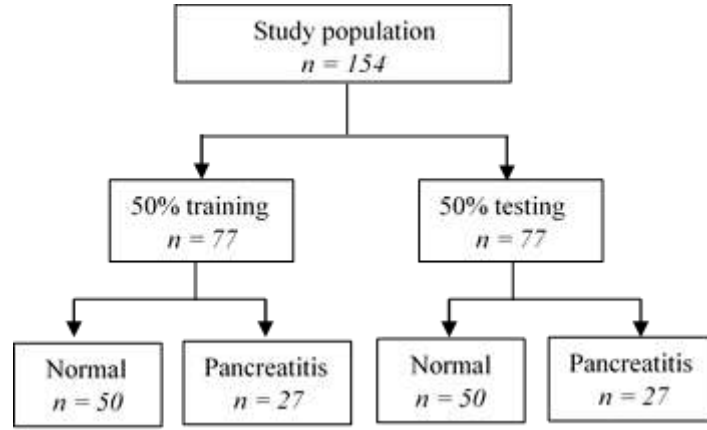


Figure 1. Study population.

To assess the predictive ability of our approaches, data is further split into half, with 77 patients used for training purposes and 77 for validation purposes. The classification results are reported on the 77 testing data (Figure 1).

3. METHODS

Each pulse wave is fitted with an autoregressive (AR) model to form a dimensionally-reduced characteristic matrix of each waveform. The AR order was chosen by averaging the number of lags of all fitted AR models based on AIC (Figure 2). Features derived from fitted AR models were then fed to train a LDA and a LR classifier.

Let $y(t)$ be the velocity (cm/s) of wrist pulse waveform at time t , the AR model is then given by:

$$y(t) = \gamma_0 + \sum_{r=1}^k \gamma_r Y_{t-r} + \epsilon_t \quad (1)$$

where k is the AR order. This gives the matrix X_{nk} for secondary LDA and LR classification:

$$X_{nk} = \begin{bmatrix} \gamma_{10} & \cdots & \gamma_{1k} \\ \vdots & \ddots & \vdots \\ \gamma_{n0} & \cdots & \gamma_{nk} \end{bmatrix} \quad (2)$$

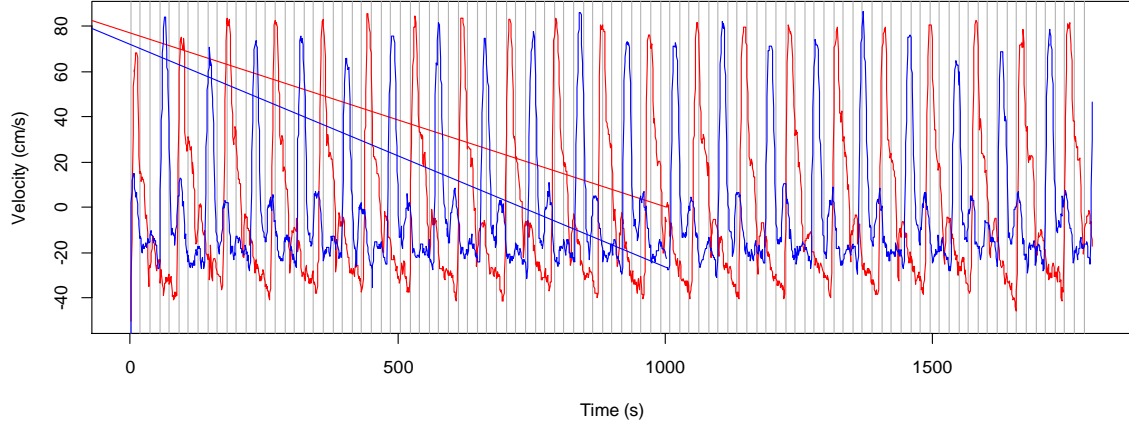


Figure 2. Pulse wave of a normal patient (blue) and patient with Pancreatitis (red). Lag intervals are shown using grey lines.

3.1. AR-LDA Model

Let there be S clusters. Each cluster is modelled by a multivariate normal density and can be characterised by its mean vector μ_s and covariance matrix Σ . Their densities are given by:

$$f_s(x) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu_s)^T \Sigma^{-1} (x-\mu_s)\right), s = 1, 2, \dots, S \quad (3)$$

In linear discriminant analysis (LDA), $f_s(x)$ follows a multivariate normal distribution with a common covariance matrix in all clusters, i.e. $\Sigma_s = \Sigma \forall s$. This results in a linear decision boundary (Hastie et al., 2009).

The discriminant boundary between two clusters i (Normal) and j (Pancreatitis) occurs where a new observation has an equal chance of falling into one of the clusters, which is determined by the discriminant function $\delta_s(x) = \log(\pi_s f_s(x))$, where π_s is the probability of falling into cluster s . Cluster membership of an observation is determined by the largest posterior probability of membership, i.e. a new data x is classified into class j when $\pi_j f_j(x) > \pi_i f_i(x)$. In other words, the clusters are selected by

$$s^*(x) = \arg \max_s \delta_s(x) \quad (4)$$

where $\delta_s(x) = \log(\pi_s) - \frac{1}{2} \log |\Sigma_s| - \frac{1}{2} (x - \mu_s)^T \Sigma_s^{-1} (x - \mu_s)$. In practice, the estimate of π_s can be calculated from the training data, $\hat{\pi}_s = \frac{N_s}{N}$, where N_s is the number of observations in the training data classified as cluster s , and N is the total number of observations in the dataset. The estimate of π_s is defined by $\hat{\pi}_s = 1 - \hat{\pi}_i = \frac{N - N_s}{N}$.

3.2. AR-LR Model

The Autoregressive Logistic regression (AR-LR) model classifies variables into two classes ‘Normal’ and ‘Pancreatitis’ by modeling the posterior probability of class membership via a linear function of the explanatory variables (Hastie et al., 2009). Instead of directly modeling the posterior probability of ‘Pancreatitis’ membership, $p = E[Y|X]$, the *logit transformation* is used (Wichern and Johnson, 2007). This ensures the predicted response from the logistic regression is bounded between 0 and 1. Let X_1, X_2, \dots, X_k denotes the AR coefficients in the k th order. Then the logit transformation of p , which is defined as the logarithm of the odds ratio, is modeled by the linear function:

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (5)$$

It is convenient to express the linear regression function in matrix form $\beta^T X$, where $\beta = [\beta_0, \beta_1, \dots, \beta_k]^T$ and $X = [1, X_1, X_2, \dots, X_k]^T$. The probability of ‘diseased’ can also be written as $p = \frac{1}{1 + e^{-\beta^T X}}$ in matrix form. Note that the odds of a variable X is simply given by $\exp(\hat{\beta})$ (Hosmer and Lemeshow, 2005).

4. RESULTS

Results from both the AR-LR and AR-LDA models showed clear separation between Normal and Pancreatitis patients, indicating that features from AR coefficients contains potential markers for classifying Pancreatitis wrist pulse.

An optimal AR order of 18 were chosen based on AIC of AR models from the testing data. The 18 AR coefficients for each patient is then used to train the AR-LR and the AR-LDA model. The sensitivity, specificity, and area under ROC curves (AUC) are reported on the 50% testing data. The overall accuracy is the proportion of true positives and true negatives in the testing data.

Amongst the 77 patients from the testing data unseen by the models, the AR-LR model achieved a sensitivity of 74% and a specificity of 88%, with an overall accuracy of 83.1% (Table 1). The positive predictive value (PPV) is 76.9% while maintaining a negative predictive value (NPV) of 86.3%.

Table 1. Classification of Normal and Pancreatitis patients from AR-LR model.

	Normal	Pancreatitis	Total
Predicted Normal	44	7	51
Predicted Pancreatitis	6	20	26
Total	50	27	77

The AR-LDR model obtained a slightly better result compared to the AR-LR model. It achieved a sensitivity of 81% and a specificity of 96%, with an overall accuracy of 90.9%. The positive predictive value (PPV) is 91.7% while maintaining a negative predictive value (NPV) of 90.6%. When comparing the area under ROC curves (AUC), the AR-LR and AR-LDA models yield identical results AUC=0.887 (Figure 3).

Table 2. Classification of Normal and Pancreatitis patients from AR-LDA model.

	Normal	Pancreatitis	Total
Predicted Normal	48	5	53
Predicted Pancreatitis	2	22	24
Total	50	27	77

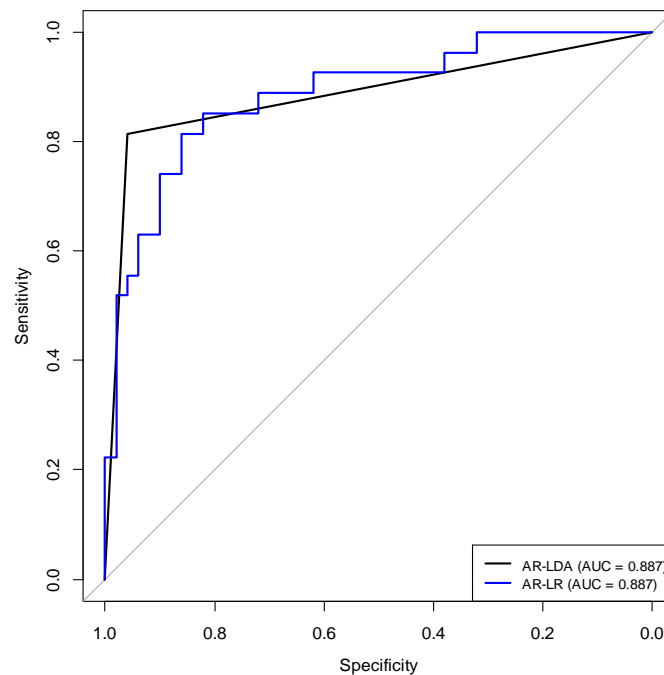


Figure 3. ROC curves of AR-LDA (black) and AR-LR (blue) models.

Compared to previous studies (Table 3), the AR-LDA model is one of the overall top-ranked models along with the AR and SW method (Chen et al., 2009a) and KNN-LDA (Shen and Lu, 2010). Although the AR-LDA model had a lower sensitivity compared to some models, it had the highest specificity ($s = 96\%$) and positive predictive value ($PPV = 92\%$).

Table 3. Accuracies compared to previous studies (best values shown in bold italics).

Method	Sensitivity	Specificity	PPV	NPV	Overall
WT method (Zhang et al., 2008)	0.61	0.82	0.61	0.82	0.75
AR model (Chen et al., 2009a)	0.74	0.86	0.71	0.88	0.82
AR and SW (Chen et al., 2009a)	0.83	0.94	0.87	0.92	0.91
Gaussian FCM (Chen et al., 2009b)	0.96	0.84	0.73	0.98	0.88
KNN-LDA (single-period) (Shen and Lu, 2010)	0.96	0.90	0.82	0.98	0.92
KNN-LDA (multi-period) (Shen and Lu, 2010)	0.44	0.70	0.41	0.73	0.62
AR-LR	0.74	0.88	0.77	0.86	0.83
AR-LDA	0.81	0.96	0.92	0.91	0.91

5. CONCLUSION

These results showed that AR-LDA is a promising approach for classifying time-series wrist pulse signals. It provides a relatively low-cost, easy-to-implement, and reliable tool for modern computerized wrist pulse diagnosis.

Features extracted by Autoregressive models provided distinctive features for classification models, including Logistic regression and LDA, which allows accurate identification of Pancreatitis patients from their wrist pulse. The high positive predictive values from the AR-LDA model will be more relevant in clinical settings, where practitioners are interested in the risk of patients having Pancreatitis given a positive identification of wrist pulse characteristics.

The results from this study not only showed that scientific identification of digitized wrist pulse signal is practical and accurate, but also suggested some scientific evidence of the utilization of wrist pulse in traditional oriental medicine for characterisation of health status.

REFERENCES

- Chen, Y., Zhang, L., Zhang, D. & Zhang, D. (2011). Computerized Wrist Pulse Signal Diagnosis Using Modified Auto-Regressive Models. *Journal of Medical Systems*, 35, 321-328.
- Chen, Y., Zhang, L., Zhang, D., & Zhang, D. (2009a). *Pattern classification for Doppler ultrasonic wrist pulse signals*. Paper presented at the 5th ICBBE conference.
- Chen, Y., Zhang, L., Zhang, D., & Zhang, D. (2009b). Wrist pulse signal diagnosis using modified Gaussian models and Fuzzy C-Means classification. *Medical Engineering & Physics*, 31, 1283-1289.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). New York: Springer-Verlag.
- Hosmer, W. D., & Lemeshow, S. (2005). *Applied Logistic Regression* (2nd Ed. ed.). Hoboken, NJ, USA: John Wiley & Sons.
- Shen, B., & Lu, G. (2010). Wrist Pulse Diagnosis Using LDA. In D. Zhang & M. Sonka (Eds.), *Medical Biometrics: Second International Conference, ICMB 2010, Hong Kong, China, June 28-30, 2010. Proceedings* (pp. 325-333). Berlin, Springer-Verlag.
- Wichern, D. W., & Johnson, R. A. (2007). *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.
- Xu, L., Zhang, D., & Wang, K. (2005). Wavelet-based cascaded adaptive filter for removing baseline drift in pulse waveforms. *IEEE Transactions on Biomedical Engineering*, 52, 1973-1975.
- Zhang, D., Zhang, L., Zhang, D., & Zheng, Y. (2008). *Wavelet based analysis of Doppler ultrasonic wrist-pulse signals*. Paper presented at the ICBBE 2008.