

신경근육재활 및 전기진단

발표일시 및 장소 : 10 월 19 일(토) 14:40-14:50 Room C(5F)

OP3-3-5

Machine Learning-Based Prediction of Diabetic Polyneuropathy

Dae Youp Shin^{1*}, Bora Lee², Tae Uk Kim¹, Seong Jae Lee¹, Jung Keun Hyun^{1,3}, Seo Young Kim^{1†}

Dankook University Hospital, Department of Rehabilitation Medicine¹, Deargen, Inc., Division of A.I. research², Dankook University, Institute of Tissue Regeneration Engineering (ITREN)³

Objective

To find the most relevant predictor for the detection of diabetic sensorimotor neuropathy (DSPN) among patients with type 2 diabetes mellitus (DM) using machine learning (ML) algorithms, and whether ML-based method is better than traditional statistics for the prediction of DSPN.

Method

Five hundred twenty seven DM patients were analyzed, and patients who had polyneuropathies other than DSPN, mononeuropathies, or radiculopathies were excluded. Subjects were divided into two groups according to the electrophysiological results based on the guidelines of the American Diabetes Association; DSPN group (n=129) and control group without DSPN (n=398). Clinical features, medical history, method of treatment, possible risk factors based on previous studies, electrophysiological and clinical pathology results were used for analysis. ML was performed with XGBoost. Average values of each code of the individual patients was used as the input value. Patients with missing values for more than half of the feature codes were excluded, and the final cohort was set with 104 test samples and 349 control samples. To find the best parameters for the model, it was trained by changing the following parameters; max_depth, subsample, and colsample_bytree. To compare ML methods with traditional statistical methods, all parameters were compared between the control and DSPN groups using independent t-test, chi-square test, factor analysis, and regression analysis to derive predictable factors for DSPN.

Results

A total of 56 variables were extracted and clustered into 3 groups using k-means clustering. For clustering, differences in observed values between test/control group for each month were calculated and normalized by each test code. 3 clusters were created according to distribution of difference of observed value. Cluster 1 showed relatively high difference, cluster 2 showed relatively low difference and cluster 3 was in-between. The model was evaluated 10 times using 5-fold cross validation. Finally, 79.4% accuracy and 0.745 area under the curve (AUC) was achieved in cluster 2, and Hemoglobin A1c (HbA1c), C-reactive protein, hemoglobin and total protein were indicated as the strong predictors for DSPN. Through the statistical method, we found that high levels of HbA1c, insulin users, subjects with retinopathy and lower levels of C-peptide were the most important predictors of

DSPN, but results were lower than those obtained using the ML method when the predictions were confirmed using equations incorporating individual parameters. Both methods had the same problem in that it took a lot of time because the data had to be inputted and integrated by human hands.

Conclusion

We found a ML system can predict DSPN, and showed better generalized predictive patterns than traditional statistics. However, in order to become a more efficient analysis method in clinical practice, more efforts are needed to shorten the duration and increase the prediction rate.

Table 1. The AUC and accuracy for the different clusters when predicting DSPN by machine learning

	AUC (Area under the curve) (0~1)	Accuracy(%)
56 (selected)	0.7393	80.20%
Cluster1 (17)	0.7325	80.00%
Cluster2 (28)	0.7450	79.41%
Cluster3 (11)	0.6267	75.43%

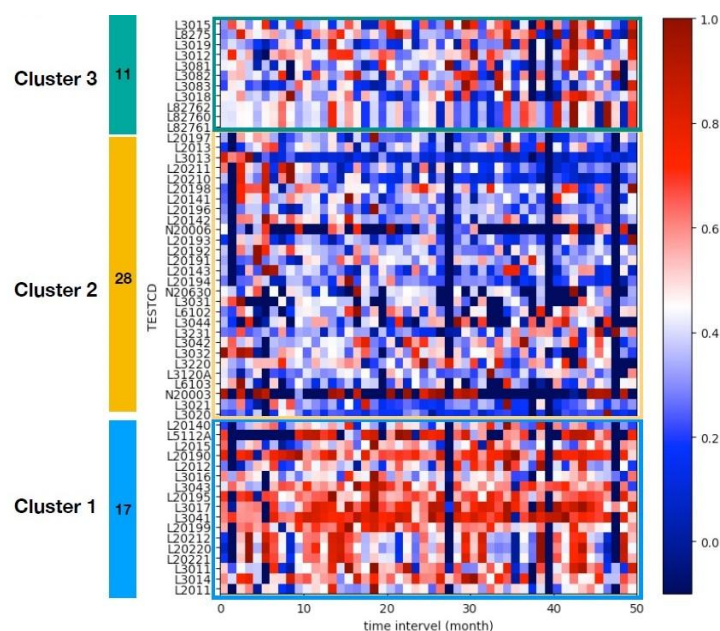


Figure 1. Heatmap drawn as the difference in observed values between the test group and the control group for each month normalized by each test code (e.g. 0-1 months, 1-2 months, etc.)