

KSN 2017 Abstract

KSN-17-P073

Prediction of hyponatremia using deep learning approach with Ensemble model from EMR

*Sug kyun SHIN¹, Ea-hwa KANG¹, Tae ik JANG¹, Yong gyu LEE¹, Seog young SO², Tak gil SHIM³

¹Kidney center, NHIS Ilsan hospital, Korea, South, ²Management information team, NHIS Ilsan hospital, Korea, South, ³Deep learning team, TMON, Korea, South

Objectives : The reason why hyponatremia is clinically important in elderly patients is that it is an important step in progressing to serious disease. If hyponatremia can be predicted, it will be important for the treatment of elderly patients. Recently, artificial intelligence(AI) is a technology that is attracting attention not only in the IT field but also in all fields. Deep Learning is a key technology for implementing. The purpose of this study is to verify the feasibility of applying this deep learning technique to patient's laboratory records and prescription data from EMR.

Methods : 1. Prerequisite & Exploratory Data Analysis

-Basic and laboratory data): total 182,181(patients) and 853 columns were observed. The columns record the last five screening dates and results data for a total of 85 items for each patient.

-Prescription data: total 182,181 prescriptions and 1,415 drug types were found.

-Classification of sodium level: Label 0: high ≥ 146 , 1: normal: 136 ~ 145, 2: low ≤ 135

2. Preprocessing:

-exclude patients with ambiguities such as sodium grade 136 or 145, approximately 78,315 people were extracted from 182,181 people.

-Variable imputation is performed to fill this empty value. Variable Imputation method used in this paper is calculated by calculating the mean value in column units. Also, if the bin value is more than 90%, the corresponding column is removed regardless of Variable imputation. Through this process, 87 columns were reduced to 37 columns. Distribution of label data after preprocessing to this stage showed Label 0: 2,338, 1: 61,399, and 2: 11,463.

-Oversampling: The preprocessed data is imbalanced classification problem due to relatively insufficient Label 0 data set. Oversampling (SMOTE) method is applied to compensate for this imbalance problem. SMOTE is an Oversampling method that uses K-Nearest Neighbor algorithm to create data with similar similarities to K similar patterns of data. The distribution of label data after

KSN 2017 Abstract

applying SMTE were Label 0: 11,690, Label 1: 12,054, and Label 2: 11,326
3. Deep Learning/Modeling: Learning and modeling have been tried by constructing models made up of two kinds of data sets individually by Ensemble.

-Laboratory dataset: In order to study the test history, 36,070 total data sets were divided into training and validation data sets at a ratio of 7:3, and the training data set was given a random forest algorithm with 500 number of tree.

Results : The validation data set of the remaining 30% of the model was used for the verification, and the accuracy was 91% as in Confusion Matrix below.

-Confusion Matrix and Statistics (Prediction/Real): Overall Statistics are Accuracy: 0.9104, 95% CI: (0.9048, 0.9159) in DATASET 1. Overall Statistics of DATASET 2 was Accuracy: 75.15%

-Dataset 1 & 2 were combined in an ensemble form: Overall Statistics was Accuracy: 92.05%.

Conclusions : In this study, the algorithms developed for predicting images or speech were used to predict hyponatremia using an ensemble model, which is likely to be applicable to future medical situations.

Keywords : Hyponatremia, deep learning technique, ensemble model