Prediction of Hyponatremia in Cancer Patients Using Machine Learning Based on Oncology CDM

YeJi Lee [1], Hyunwoo Park[1], Yul Hwangbo[1], HyoSoung Cha[2,3]

[1] Healthcare AI Team, National Cancer Center, Goyang, South Korea
[2] National Cancer Control Institute, National Cancer Center, Goyang, South Korea
[3] Graduate School of Cancer Science and Policy, National Cancer Center, Goyang, South Korea

Background

The number of cancer patients worldwide has been steadily increasing. Although the survival rate of cancer patients has improved due to advances in treatment technologies and preventive policies, the management of complications in cancer survivors remains a critical concern. Electrolyte abnormalities are common among cancer patients but are often overlooked. Mild electrolyte imbalances often do not present symptoms and are frequently detected through routine laboratory tests. However, if it severe, it can cause serious complications. In particular, hyponatremia is a common complication in cancer patients and requires early detection, because it is associated with increased long-term hospitalization and mortality. This study developed a previously unexplored model for predicting the occurrence of hyponatremia in cancer patients based on the Oncology Common Data Model (CDM).

Methods

Data source:
The oncology CDM (version 5.4), converted from data of 119,854 cancer patients who visited the National Cancer Center from January 2010 to December 2021, was used in the analysis.

Cohort Selection Criteria:
(1) patients who were not diagnosed with cancer before the onset of hyponatremia were excluded.
(2) Patients diagnosed with liver cancer or liver disease were excluded because hyponatremia could be induced due to decreased liver function.
(3) Patients who developed hyponatremia 365 days after cancer diagnosis were excluded.
(4) To develop a predictive model one day before the onset of hyponatremia, the observation period was defined from the date of cancer diagnosis to the day before the onset of hyponatremia.

Finally, 33,476 patients were included in the analysis.
- Hyponatremia group: 3,102 patients (9.3%)
- Normonatremia group: 30,374 patients (90.7%)

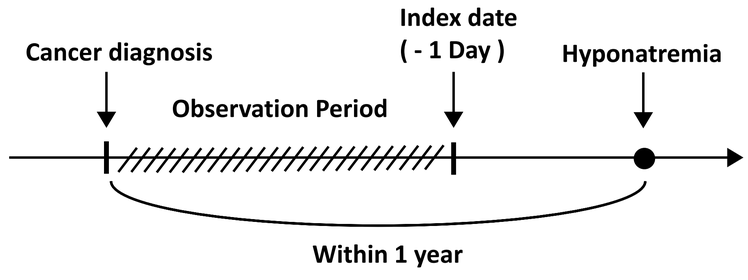

Figure 1. Time window diagram

Outcome:
Occurrence of hyponatremia (serum sodium ≤ 125 mmol/L).

Model Development:
We developed a stacking ensemble model using logistic regression(LR), random forest(RF), support vector machine(SVM), and light gradient boosting(LGB) as a base model, and extreme gradient boosting(XGB) as a meta model.

Model Validation and Evaluation:
The dataset was divided into training(80%) and test(20%) sets, and 5-fold cross-validation was applied to the training set. The performance of the model was evaluated with the base model and stacking ensemble using area under the receiver operating characteristic curve(AUROC), accuracy, sensitivity, specificity, positive predictive value(PPV), negative predictive value(NPV), and F1 scores. The Shapley Additive Explanations(SHAP) algorithm was used to visually express the effect of each variable on the prediction of hyponatremia.

Results

Most models demonstrated excellent performance in AUROC, accuracy, and F1 score, but relatively low sensitivity and Positive Predictive Value(PPV). (Table 1). To address these limitations, we employed a stacking ensemble model, which achieved consistently high performance across all evaluation metrics.

Table 1. Model Performance of Hyponatremia Prediction

Models	AUROC	Accuracy	Sensitivity	Specificity	PPV	NPV	F1
LR	0.899	0.929	0.487	0.976	0.683	0.947	0.765
RF	0.899	0.919	0.184	0.996	0.837	0.920	0.629
SVM	0.899	0.925	0.295	0.992	0.787	0.930	0.695
LGB	0.899	0.937	0.506	0.982	0.752	0.950	0.785
XGB	0.930	0.938	0.511	0.984	0.768	0.950	0.614
Stack Ens	**0.922**	**0.939**	**0.753**	**0.969**	**0.855**	**0.969**	**0.793**

The SHAP summary plot shows the relative importance of each clinical feature in predicting hyponatremia. (Figure 2). The higher the age and the lower the chloride level, the greater the likelihood of developing hyponatremia. This is consistent with previous research findings, as sodium and chloride levels are closely correlated. Cancer-specific information found that SEER stage 7 and M stage 1 were associated with a high incidence of hyponatremia, indicating a high probability of hyponatremia in patients with tumor metastasis.

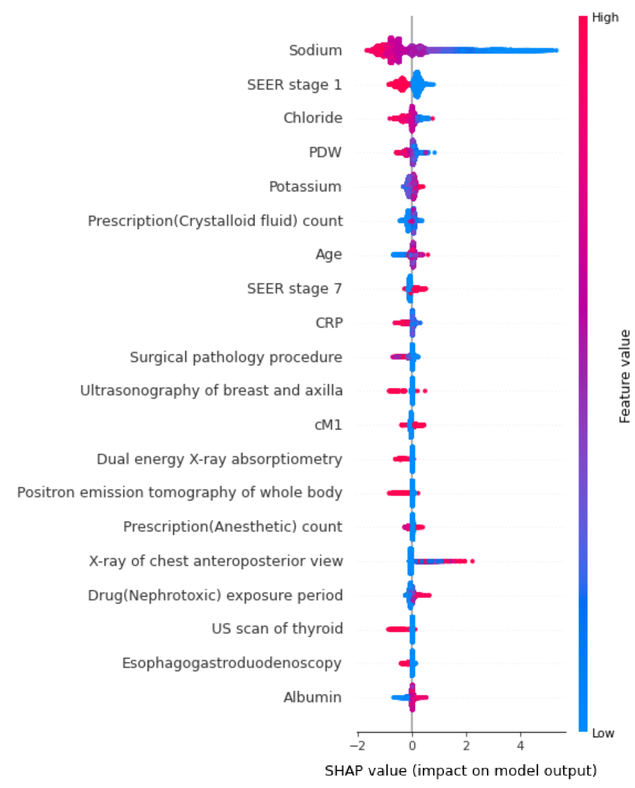

Figure 2. SHAP summary plot of the LightGBM model

Conclusions

In this study, we developed a predictive model to forecast the occurrence of hyponatremia one day in advance in cancer patients using the oncology CDM. Laboratory tests, age, and progression of cancer were identified as major predictors. A review of recent sodium test figures and test dates before hyponatremia confirmed a sharp drop in sodium levels above 15 mmol/L without a sodium test for more than 100 days in some patients. Based on the major predictors identified in this study, providing baseline data for the early detection of hyponatremia enables medical staff to anticipate and prevent its occurrence in advance. Furthermore, patients requiring sodium monitoring can be identified proactively, thereby reducing the risk of overlooking patients for extended periods. However, as this study was based on single-center data, external verification will be performed through multicenter federated learning in the future to reflect the characteristics of various patients. We are currently exploring various federated learning platforms and are considering FeederNet, which is actively utilized in large hospitals.

Contact: leeyeji14@ncc.re.kr

This study was supported by the Bio-Industrial Technology Development Program (20014841) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea).