Prediction of Severe Respiratory Infections in Patients with Diabetes

Nguyen Thi Kim Hien[1], Phan Thanh Phuc[2], Septi Melisa[2],
Muhammad Solihuddin Muhtar[2], Nguyen Phung-Anh[4,5,6], Jason Hsu[3,4,5,6*]

1. Master Program in Global Health and Health Security, College of Public Health, Taipei Medical University, Taipei, Taiwan
2. International Ph.D. Program in Biotech and Healthcare Management, College of Management, Taipei Medical University, Taipei, Taiwan
3. International Ph.D. Program in Biotech and Healthcare Management, College of Management, Taipei Medical University, Taipei, Taiwan
4. Clinical Data Center, Office of Data Science, Taipei Medical University, Taipei, Taiwan
5. Clinical Big Data Research Center, Taipei Medical University Hospital, Taipei Medical University, Taipei, Taiwan
6. Research Center of Health Care Industry Data Science, College of Management, Taipei Medical University, Taipei, Taiwan

* Corresponding author

Background

Patients with type 2 diabetes (T2DM) have a higher susceptibility to infections in comparison to those without type 2 diabetes. Diabetes mellitus is linked to immune dysfunction and alterations in immune system components[1]. Hyperglycemia impairs the innate immune response by adversely affecting the functions of macrophages, dendritic cells, neutrophils, and natural killer cells[2]. Additionally, it disrupts the adaptive immune response, thereby increasing the susceptibility of patients with T2DM to pneumonia and pulmonary tuberculosis[3-6].

Objective

The objective of this study was to develop predictive models to assess the risk of severe respiratory infections over a three-year period in individuals aged 45 and above with T2DM. These models aimed to forecast the occurrence of hospitalized pneumonia and the onset of tuberculosis by utilizing multiple machine learning algorithms, incorporating a comprehensive set of predictive variables.

Method

A retrospective cohort study was conducted using electronic health record (EHR) data obtained from three hospitals affiliated with Taipei Medical University: Taipei Medical University

Hospital, Shuangho Hospital, and Wanfang Hospital. The data were converted to the OHDSI OMOP-CDM for analysis. Individuals over 45 years old with T2DM who had no history of pneumonia within the preceding 30 days or tuberculosis within the preceding 365 days were included in the study as the target cohort.

The primary outcomes of this study were the three-year risks of hospitalized pneumonia and the development of tuberculosis. Participants were administered appropriate medications to ensure accurate diagnoses. This study employed the Patient-Level Prediction (PLP) package, utilizing four machine learning algorithms—logistic regression (LR), LightGBM (LGBM), random forest, and XGBoost—to construct predictive models. These models integrated an extensive array of factors, including patient attributes, coexisting medical conditions, and medication utilization. The performance of the models was evaluated using the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC).

Result

The study's original cohort consisted of 78,322 patients diagnosed with T2DM. After excluding subjects with insufficient follow-up time, the three-year incidence rates of pneumonia and tuberculosis were 4.38% and 0.25%, respectively.

Table 1. The cohort count

Target/ Outcome	Target cohort	Outcome cohort	Incidence (%)
T2DM / Hospitalized pneumonia	42863	1876	4.38
T2DM / Pulmonary tuberculosis	43860	107	0.25

Table 2 presents a comparative analysis of the performance of various machine learning models across the two outcomes, evaluated using AUROC and AUPRC values. This evaluation enables the identification of the most appropriate model for each specific outcome. Notably, in predicting hospitalized pneumonia events within the T2DM cohort, XGBoost demonstrated the highest AUROC of 0.805 and an AUPRC of 0.195. Additionally, logistic regression algorithms

achieved the highest performance in forecasting TB occurrence in T2DM patients, with an AUROC of 0.785 and an AUPRC of 0.007.

Table 2 Performance of prediction models

Outcome	Model type	AUROC (95% CI)	AUPRC
Hospitalized pneumonia	Logistic	0.800 (0.780 - 0.820)	0.175
	XGBoost	0.808 (0.787 - 0.828)	0.195
Pulmonary tuberculosis	Logistic	0.785 (0.721 - 0.849)	0.007
	XGBoost	0.690 (0.599 - 0.781)	0.006

Figure 1 presents the discrimination performance of the best prediction model (XGBoost) for forecasting hospitalized pneumonia in diabetic patients, including ROC curves and precision-recall (PR) plots for the cohort. Figure 2 illustrates the discrimination performance of the optimal prediction model (logistic regression) for predicting the onset of pulmonary tuberculosis.

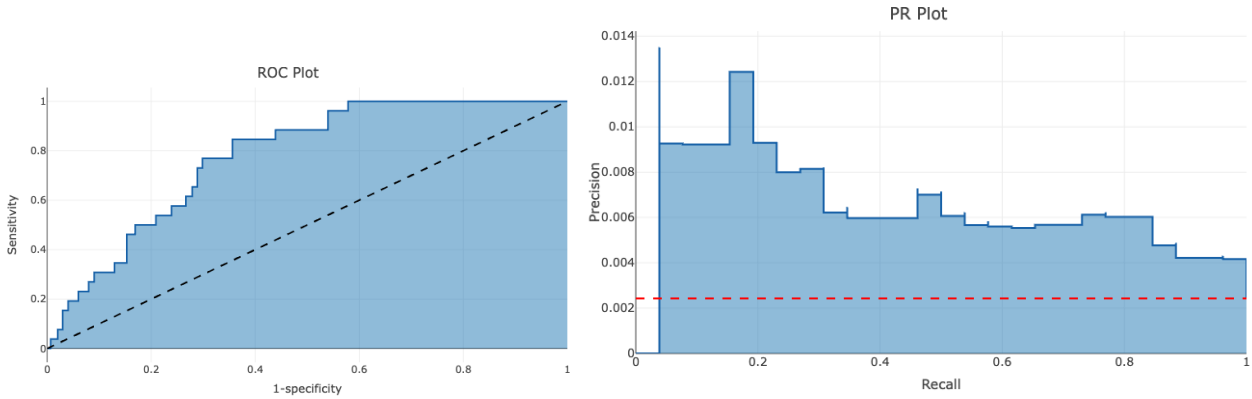

Figure 1 Receiver Operating Characteristic (ROC) Curve and Precision-Recall (PR) Curve: Predicting the Risk of Hospitalized Pneumonia Using XGBoost Algorithm

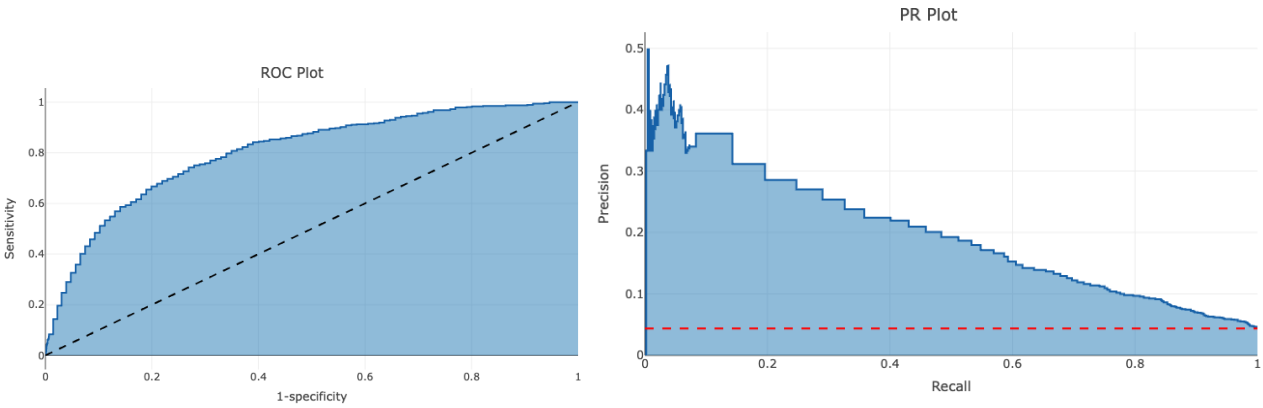

Figure 2 Receiver Operating Characteristic (ROC) Curve and Precision-Recall (PR) Curve: Predicting the Risk of Onset Pulmonary Tuberculosis Using Logistic Regression Algorithm

Conclusion

Diabetes is an independent factor that increases the risk of developing severe pneumonia and is also linked to hospitalization and death related to pneumonia. Moreover, T2DM increases the likelihood of developing active tuberculosis. The prediction models were developed to forecast the risk of pneumonia requiring hospitalization and the onset of tuberculosis using various machine learning algorithms in individuals aged 45 and older with T2DM. The models demonstrated acceptable predictive accuracy and discriminative capability, making them suitable for early estimation of the likelihood of severe respiratory infections. However, further research involving larger sample sizes or prospective cohort studies is necessary to validate and confirm the accuracy of these findings.

References

1. Kumar Nathella P, Babu S. Influence of diabetes mellitus on immunity to human tuberculosis. *Immunology*. Sep 2017;152(1):13-24. doi:10.1111/imm.12762
2. Hirayama D, Iida T, Nakase H. The Phagocytic Function of Macrophage-Enforcing Innate Immunity and Tissue Homeostasis. *Int J Mol Sci*. Dec 29 2017;19(1)doi:10.3390/ijms19010092
3. Kornum JB, Thomsen RW, Riis A, Lervang H-H, Schønheyder HC, Sørensen HT. Diabetes, Glycemic Control, and Risk of Hospitalization With Pneumonia: A population-based case-control study. *Diabetes Care*. 2008;31(8):1541-1545. doi:10.2337/dc08-0138
4. Skowroński M, Zozulińska-Ziółkiewicz D, Barinow-Wojewódzki A. Tuberculosis and diabetes mellitus - an underappreciated association. *Arch Med Sci*. Oct 27 2014;10(5):1019-27. doi:10.5114/aoms.2014.46220
5. Jeon CY, Murray MB. Diabetes mellitus increases the risk of active tuberculosis: a systematic review of 13 observational studies. *PLoS Med*. Jul 15 2008;5(7):e152. doi:10.1371/journal.pmed.0050152
6. Brunetti VC, Ayele HT, Yu OHY, Ernst P, Filion KB. Type 2 diabetes mellitus and risk of community-acquired pneumonia: a systematic review and meta-analysis of observational studies. *CMAJ Open*. Jan-Mar 2021;9(1):E62-e70. doi:10.9778/cmajo.20200013