INPLASY

INPLASY202660043

doi: 10.37766/inplasy2026.6.0043

Received: 8 June 2026

Published: 8 June 2026

High-Recall Abstract Screening with Biomedical Pretrained Language Models: A Meta-Analysis of Prediction Models for Nasopharyngeal Carcinoma Radiotherapy-Induced Complications

Lee, TF; Lin, GZ; Yun, WP; Cheng, CY; Chang, L; Liu, CH; Hsieh, YW; Yeh, SA; Wuu, CS; Chao, PJ.

Corresponding author:
Tsair-Fwu Lee

tflee@nkust.edu.tw

Author Affiliation:
National Kaohsiung University of Science and Technology.

ADMINISTRATIVE INFORMATION

Support - National Science and Technology Council (NSTC).

Review Stage at time of this submission - Completed but not published.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY202660043

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 8 June 2026 and was last updated on 8 June 2026.

INTRODUCTION

Review question / Objective (1) Can biomedical domain-specific pretrained language models (BioBERT, PubMedBERT) achieve ≥97% Work Saved over Sampling (WSS) at 100% recall in title-and-abstract screening? (2) What is the pooled AUC of prediction models for radiation-induced complications in NPC patients receiving modern IMRT/VMAT/TomoTherapy? (3) How do ML, traditional statistical, and NTCP models differ in performance, and what are the main sources of heterogeneity? (4) What is the methodological quality and risk of bias according to PROBAST? (5) What are the common clinical data sources and predictor types?

Rationale Existing reviews are often narrative or limited to single toxicities and rarely integrate advanced NLP screening or PROBAST assessment. This protocol establishes a dual-track design: (a) evaluate BioBERT/PubMedBERT workload reduction at 100% recall, and (b) perform PROBAST-guided random-effects meta-analysis with meta-regression to identify performance determinants and methodological gaps.

Condition being studied Nasopharyngeal carcinoma (NPC) is a malignant tumor arising from the nasopharyngeal epithelium and is relatively common in East and Southeast Asia. Modern radiotherapy techniques, including IMRT, VMAT, and TomoTherapy, have improved tumor control and survival outcomes. However, patients may still develop acute or late radiotherapy-induced complications, such as xerostomia, oral mucositis, temporal lobe injury, hypothyroidism, radiation dermatitis, and hearing loss, which can substantially affect quality of life. This review focuses on prediction models and NLP-assisted screening studies related to these radiotherapy-induced complications in NPC patients.

METHODS

Participant or population Patients with nasopharyngeal carcinoma who received modern radiotherapy, including intensity-modulated

radiotherapy, volumetric-modulated arc therapy, or TomoTherapy, and were evaluated for radiotherapy-induced complications. Eligible populations may include patients with acute or late adverse effects such as xerostomia, oral mucositis, temporal lobe injury, hypothyroidism, radiation dermatitis, or hearing loss.

Intervention The intervention of interest is the use of prediction modeling and NLP-assisted screening approaches in studies of radiotherapy-induced complications among patients with nasopharyngeal carcinoma. These may include machine learning models, deep learning models, traditional statistical prediction models, normal tissue complication probability models, and biomedical pretrained language models such as BioBERT and PubMedBERT. The review will evaluate how these approaches are used to predict or identify radiotherapy-induced complications after modern radiotherapy, including IMRT, VMAT, and TomoTherapy.

Comparator The comparative interventions will include different types of prediction models used for radiotherapy-induced complications in nasopharyngeal carcinoma patients, such as machine learning models, traditional statistical models, and normal tissue complication probability models. For the NLP-assisted screening component, BERT-base, BioBERT, and PubMedBERT will be compared using ROC-AUC, precision-recall AUC, and Work Saved over Sampling.

Study designs to be included Original peer-reviewed studies that develop, validate, or evaluate prediction models for radiotherapy-induced complications in patients with nasopharyngeal carcinoma will be included. Eligible study designs may include retrospective cohort studies, prospective cohort studies, model development studies, external validation studies, and studies reporting diagnostic or prognostic model performance.

Eligibility criteria Additional eligibility criteria will include full-text articles written in English and original peer-reviewed studies reporting quantitative model performance, particularly AUC. Studies must focus on nasopharyngeal carcinoma patients treated with modern photon radiotherapy techniques, including IMRT, VMAT, or TomoTherapy, and must address radiotherapy-induced complications or related prediction outcomes.

Studies will be excluded if they are reviews, meta-analyses, conference abstracts, theses, case reports, editorials, letters, non-human studies, non-clinical studies, or studies without sufficient model performance data. Studies involving 3D conformal radiotherapy, particle therapy, non-NPC populations, or outcomes unrelated to radiotherapy-induced complications will also be excluded.

Information sources Information sources will include electronic searches of PubMed, Web of Science, Scopus, and the Cochrane Library. The reference lists of eligible studies and relevant review articles will also be manually screened to identify additional records.

Grey literature, conference abstracts, theses, editorials, letters, and case reports will not be included. Trial registers will not be searched because this review focuses on published prediction model studies and NLP-assisted screening for radiotherapy-induced complications in nasopharyngeal carcinoma. If necessary, study authors may be contacted to clarify missing or unclear model performance data.

Main outcome(s) The main outcomes will be model performance measures for predicting radiotherapy-induced complications in patients with nasopharyngeal carcinoma. The primary effect measure will be the area under the receiver operating characteristic curve (AUC), extracted from eligible prediction model studies. When available, AUC values from internal validation, external validation, or test datasets will be recorded separately.

Additional outcome(s) For the NLP-assisted screening component, additional outcomes will include ROC-AUC, precision-recall AUC, Work Saved over Sampling at 100% recall, and Work Saved over Sampling at 85% recall. These outcomes will be used to evaluate the ability of biomedical pretrained language models to reduce screening workload while maintaining high recall.

Radiotherapy-induced complications of interest may include xerostomia, oral mucositis, temporal lobe injury, hypothyroidism, radiation dermatitis, and hearing loss. Outcomes may be assessed at acute or late post-radiotherapy time points, depending on the definitions used in the included studies.

Quality assessment / Risk of bias analysis The methodological quality and risk of bias of the included prediction model studies will be assessed

using PROBAST. PROBAST evaluates four domains: participants, predictors, outcome, and analysis. Each domain will be judged for risk of bias, and the first three domains will also be assessed for applicability concerns. The overall risk of bias will be classified as low, high, or unclear according to PROBAST guidance.

Strategy of data synthesis For the NLP-assisted screening component, model performance will be evaluated using stratified 5-fold cross-validation. The performance metrics will include ROC-AUC, precision-recall AUC, Work Saved over Sampling at 100% recall, and Work Saved over Sampling at 85% recall.

For the meta-analysis of prediction models, AUC values will be extracted from eligible studies. AUCs will be transformed using a logit transformation and pooled using a random-effects model. Heterogeneity will be assessed using I^2 statistics and Cochran's Q test. Subgroup analysis will be conducted according to model type, including machine learning models, traditional statistical models, and normal tissue complication probability models. Meta-regression will be performed to explore potential sources of heterogeneity, including model type, sample size, validation dataset type, and complication type.

Sensitivity analysis will be performed using a leave-one-out approach. Publication bias will be assessed using funnel plots, Egger's test, and trim-and-fill analysis.

Subgroup analysis Subgroup analysis will be conducted according to model type, including machine learning models, traditional statistical models, and normal tissue complication probability models. The pooled AUC will be compared across these model categories when sufficient data are available.

Sensitivity analysis Sensitivity analysis will be performed using a leave-one-out approach. Each included study will be sequentially removed from the meta-analysis to examine whether the pooled AUC is strongly influenced by any single study. The results will be used to assess the robustness of the pooled model performance estimates.

Country(ies) involved The study is being conducted by authors affiliated with institutions in Taiwan, with potential inclusion of published studies from multiple countries as part of the systematic review.

Keywords NLP; pretrained language models; nasopharyngeal carcinoma; radiotherapy complications; prediction models; machine learning; deep learning; normal tissue complication probability; meta-analysis; PROBAST.

Contributions of each author

Author 1 - Tsair-Fwu Lee.
Author 2 - Guang-Zhi Lin.
Author 3 - Wei-Ping Yun.
Author 4 - Chih-Yu Cheng.
Author 5 - Liyun Chang.
Author 6 - Chao-Hong Liu.
Author 7 - Yang-Wei Hsieh.
Author 8 - Shyh-An Yeh.
Author 9 - Cheng-Shie Wuu.
Author 10 - Pei-Ju Chao.