A Retrospective Study on a Predictive Model to Probe the Impact of Atmospheric Environment on Breast Cancer Prognosis: A Cohort of 17,438 Patients

Yunqin Yuan[1]; Anqi Yuan[2]; Wanting Wang[3] Zhe Hu[4] Chenhui Zheng[5] Zhibao Zheng[6] Weizhen Liang[7] Yuzhu Zhang[1,*] Yin Pan[2,*] Changjiang Zhang[3,*]
School of artificial intelligence, Taizhou University Taizhou Central Hospital (Taizhou University Hospital)

Introduction

In survival analysis tasks, deep learning–based models such as DeepSurv and DeepHit are designed to handle one-dimensional time-to-event data, with the primary goal of estimating an individual's survival probability as a function of time. However, conventional approaches that process input data in its original one-dimensional form may overlook latent spatial structural information—particularly relevant in certain types of biomedical or clinical data. To address this limitation, we propose an innovative strategy: transforming one-dimensional feature vectors into grayscale images. This conversion encodes the original data into a two-dimensional spatial representation, thereby introducing structural priors that enable deep models to capture more complex and nuanced patterns. Our ultimate objective is to develop a predictive model that analyzes the impact of atmospheric environmental factors on breast cancer prognosis.

Data

All clinical data were retrieved from the Shanghai Jiao Tong University Breast Cancer Database (SJTU-BCDB) and Sun Yat-sen Memorial Hospital, Sun Yat-sen University.This dataset includes data from 17,438 breast cancer patients diagnosed between 2000 and 2022, covering the period from diagnosis to follow-up.

Meteorologica dataset is provided by National Cryosphere Desert Data Center.ChinaMet is a high-resolution (1 km) and long-term (1980–2024) all-sky meteorological forcing dataset for China, developed by integrating multi-source remote sensing data, reanalysis products, and in-situ observations from more than 2,000 meteorological stations. It includes eight meteorological variables: precipitation (pre), 2-meter mean air temperature (tmpmean), maximum temperature (tmpmax), minimum temperature (tmpmin), 10-meter wind speed (wind), relative humidity (rhu), surface pressure (pres), and potential evapotranspiration (pet).

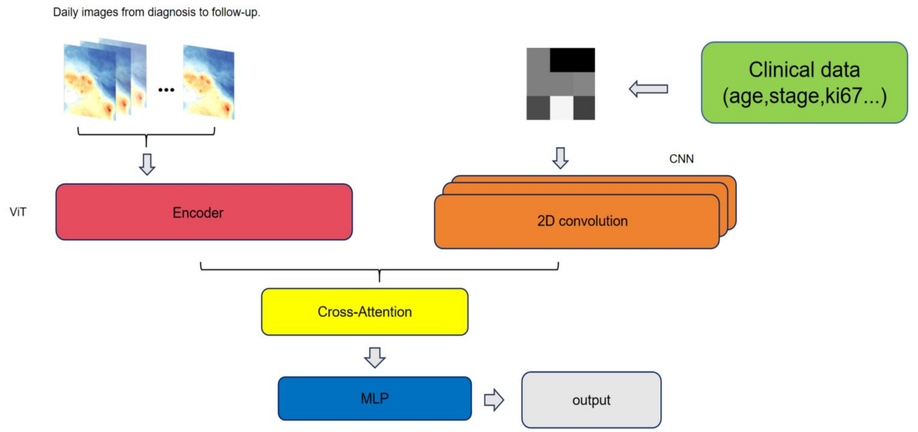


Figure 1. Schematic diagram of model structure

Methods

To more effectively capture the inherent latent structural dependencies in clinical tabular data, we propose a novel preprocessing paradigm that reformulates one-dimensional (1D) patient feature vectors into two-dimensional (2D) grayscale image representations. This spatial embedding enables Convolutional Neural Networks (CNNs), through hierarchical convolutional blocks, to exploit local and global spatial correlations among clinical variables that would otherwise remain obscured in their original vector form.

Concurrently, geolocated meteorological image sequences—such as $PM_{2.5}$ concentrations, temperature, and wind speed—are processed using Vision Transformers (ViTs), whose self-attention–based encoder architecture effectively models the spatiotemporal dynamics of environmental exposures. The resulting clinical and environmental feature embeddings are then integrated via a cross-attention mechanism, facilitating adaptive, context-aware fusion that captures the interplay between individual-level clinical profiles and ambient atmospheric conditions in the context of breast cancer prognosis.

Contact

<Yunqin Yuan>
<School of artificial intelligence, Taizhou University>
Email:1198250378@qq.com
Website:https://www.tzc.edu.cn
Phone:+8613696425538

References

1. Zhang Ling, Li Xin, Zheng Donghai, Zhang Kun, Ma Qimin, Zhao Yanbo, Ge Yingchun. Merging multiple satellite-based precipitation products and gauge observations using a novel double machine learning approach[J]. Journal of Hydrology, 2021, 594:125969.
2. Hu Yingyi, Zhang Ling. Added value of merging techniques in precipitation estimates relative to gauge-interpolation algorithms of varying complexity[J]. Journal of Hydrology, 2024, 645:132214. DOI: https://doi.org/10.1016/j.jhydrol.2024.132214.
3. Zhang J, Li Y, Li Z, et al. Deep-learning-based HER2 status assessment from multimodal breast cancer data predicts neoadjuvant therapy response[J]. Nature Biomedical Engineering, 2025: 1-12.
4. Zhang X M, Zhou H J, Chen Q, et al. Development and validation of an artificial intelligence system for triple-negative breast cancer identification and prognosis prediction: a multicentre retrospective study[J]. EClinicalMedicine, 2025, 89.
5. Fan Y, Sun X, Xiao Y, et al. Deep learning predicts HER2 status in invasive breast cancer from multimodal ultrasound and MRI[J]. Biomolecules and Biomedicine, 2025.
6. Guo J, Chen B, Cao H, et al. Cross-modal deep learning model for predicting pathologic complete response to neoadjuvant chemotherapy in breast cancer[J]. NPJ Precision Oncology, 2024, 8(1): 189.
7. Joo S, Ko E S, Kwon S, et al. Multimodal deep learning models for the prediction of pathologic response to neoadjuvant chemotherapy in breast cancer[J]. Scientific reports, 2021, 11(1): 18800.
8. Comes M C, Fanizzi A, Bove S, et al. Early prediction of neoadjuvant chemotherapy response by exploiting a transfer learning approach on breast DCE-MRIs[J]. Scientific Reports, 2021, 11(1): 14123.
9. Zhang M, Gong L, Cai M, et al. Prediction of Tumor-infiltrating Lymphocytes in Triple-negative and HER2-positive Breast Cancer Using Multi-modality Ultrasound and MRI: A Single-institution Retrospective Study[J]. British Journal of Radiology, 2025: tqaf146.

This presentation is the intellectual property of the author/presenter. Contact them at <<1198250378@qq.com>> for permission to reprint and/or distribute.

© Poster Template by Genigraphics® 1.800.790.4001 www.genigraphics.com