# **Using Electronic Health Records to Analyze Aging and Cancer Risks（项目整体标题）**

# **A Passive Multi-Cancer Risk Prediction Framework Using Longitudinal Disease Trajectories for Five High-Incidence Cancers**

# **Project Leaders**

Da Huang

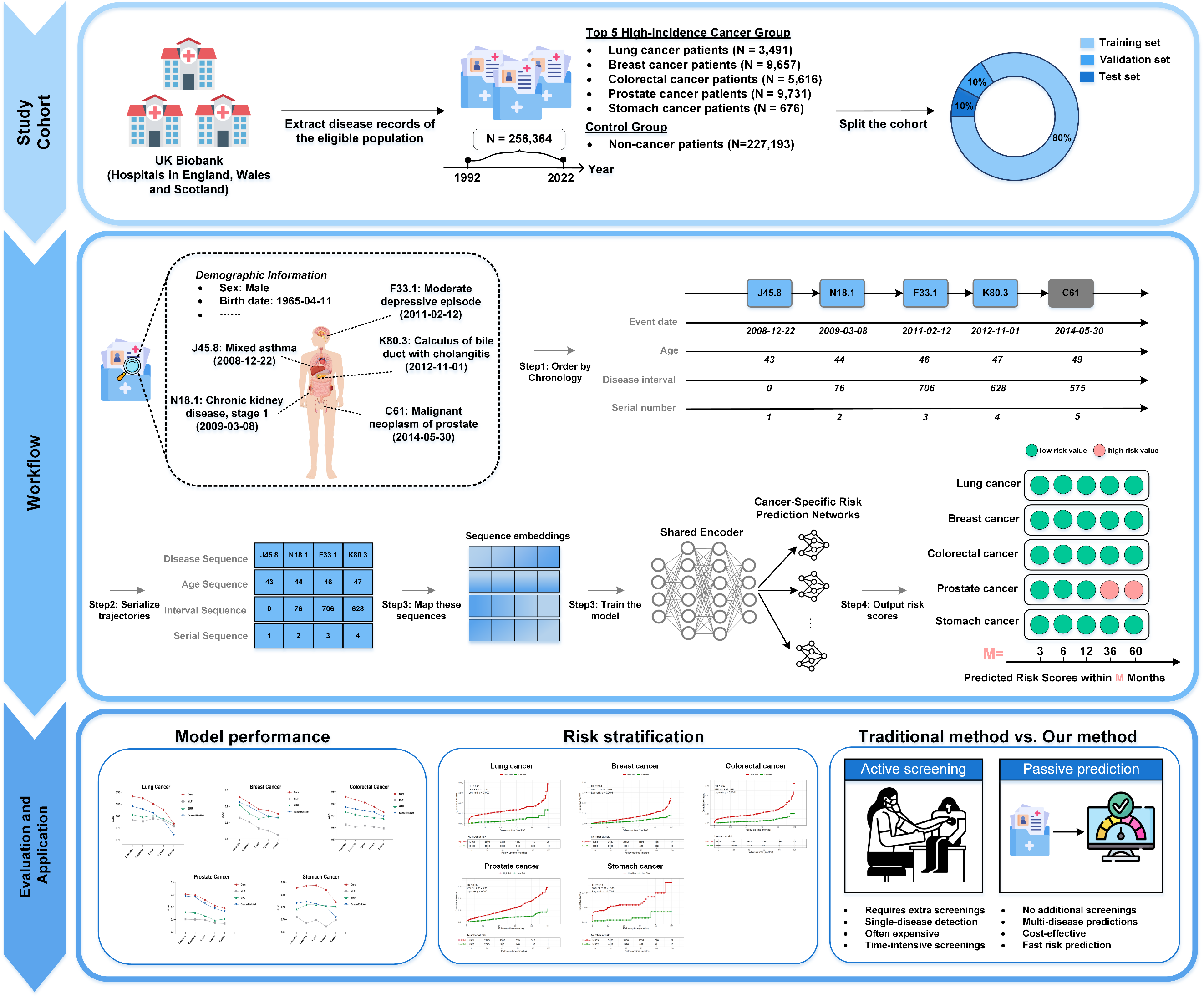
**Partner Organisations**

Cancer remains a leading cause of death globally, with early detection being critical for improving outcomes. However, current screening methods face challenges: imaging and specialized tests (like proteomics) can be resource-intensive and costly, while methods relying on family history or lifestyle are often based on static, potentially inaccurate self-reported data. Furthermore, many prediction models focus only on a single cancer type, missing broader interactions.

This project introduces an innovative, passive multi-cancer risk prediction framework. We leverage existing longitudinal data already present in patient electronic health records – their "disease trajectories" over time. Using advanced deep learning (AI), our framework analyzes these trajectories across multiple high-incidence cancers simultaneously (such as lung, breast, and colorectal).

This unique approach allows us to identify both shared risk patterns common across different cancers and cancer-specific progression signals. By incorporating factors like age progression and the timing between diagnoses, and utilizing the hierarchical structure of disease codes (ICD-10), we enhance predictive accuracy.

Crucially, our method requires no active screening or additional invasive tests. It offers a cost-effective, scalable solution for large-scale population screening by passively utilizing existing healthcare data. This facilitates earlier risk identification, paving the way for timely interventions and ultimately aiming to reduce the global cancer burden through data-driven insights.



# **EHR-AGE: Deep Learning-Powered Organ Aging Prediction from Electronic Health Records for Precision Medicine**

# **Project Leaders**

Xiangmin Luo

**Partner Organisations**

Accelerated aging of the body's organ systems is associated with an increased risk of disease. Although deep learning models using healthy population training and diseased population testing are currently proposed, the potential of mixed-population based biological age prediction for multi-organ systems remains unexplored.

This study proposes a new paradigm for systematically quantifying multi-organ aging trajecues to address key limitations in current biological age studies: (a) Existing models generally ignore the dynamic effects of disease status on organ-specific aging rates; (b) Relying primarily on imaging models ignores the potential utility of electronic health record (EHR) data. The model was developed based on multidimensional phenotypic data (including anthropometric indicators, physiological function parameters), structured clinical records (International Classification of Diseases 10th Edition [ICD-10] code and clinical diagnostic text) and prescription medication records of 457,044 participants aged 40-85 years in the UK Biobank. Our model shows accurate multi-organ system age estimates with a mean absolute error of 3.59 to 3.65 years, and a strong correlation (0.80-0.81) between the predicted ages of each organ and the actual ages.

The predicted organ ages also showed a risk stratification ability comparable to actual age. The model was explained by the conclusion analysis, and the age-specific patterns of different organ systems were revealed. After system bias, the organ system age gap (the difference between the predicted age of the organ system and the actual age) can reflect the health of the organ. In order to better analyze the relationship between different organ systems, we classify the groups with different organ systems. The average Age Gap of each organ system is 0.92±4.07 years, the average Age Gap of hepatic Group is 0.92±4.07 years: 0.08±3.92 years old, immune Group mean Age Gap: 0.12±3.89 years old, metabolic Group mean Age Gap: 0.59±3.92 years old, musculoskeletal Group mean Age Gap: 0.69±4.07 years, the mean Age Gap of the pulmonary Group: 0.48±4.38 years, and the mean Age Gaps of the renal Group: -0.001±4.10 years) and the healthy group (mean Age Gaps: Compared to 0.004±3.58 years), the organ age difference was higher, averaging -0.005±0.88 years. Using Cox proportional hazard regression models, we assessed the potential effect of different organ age gaps on the risk of developing disease over a 5-year period in the participants used. Our research suggests that organ aging can serve as a valuable biomarker of organ disease risk (cardiovascular Group HR=1.17, 95%CI:1.15-1.18, p<0.001; hepatic Group HR=1.03, 95%CI:1.02-1.04, p<0.001; immune Group HR=1.06,95%CI:1.04-1.08, p<0.001; metabolic Group HR=1.17,95%CI:1.15-1.19, p<0.001; musculoskeletal Group HR=1.15,95%CI:1.13-1.16, p<0.001; pulmonary Group HR=1.11,95%CI:1.09-1.05, p<0.001; renal Group HR=1.15,95%CI:1.13-1.17, p<0.001), which means that it is not likely to play a role in future disease risk screening for those who already have the disease, not only for early detection online, but also for risk stratification and personalized disease screening.

