# Predicting radiotherapy toxicity using machine learning and large datasets

**Tim Rattay**, Adam J Webb, Vicki Emms, Bo Wang, Huiyu Zhou, André Panisson, Alan Perotti, Guido Bologna, Sofia Rivera, Marie Bergeaud, Alessio Romita, Karolien Verhoeven, Bram Ramaekers, Willem Witlox, Cheryl Roumen, Hilary Stobart, Gabriella Cortellessa, Francesca Fracasso, Iordanis Koutsopoulos, Ivana Matic, Nina Petrovic, Marina Nikitovic, Tatjana Stanojkovic, Christopher J Talbot

## Background

Radiation therapy remains a cornerstone in cancer treatment, either alone or combined with surgery and systemic treatment. Its primary aim is to maximize the therapeutic ratio - the balance between tumour control probability (TCP) and normal tissue complication probability (NTCP). Despite technological advancements in RT and proton therapy, radiation-induced toxicity remains a concern.

Toxicity can result from cumulative radiation doses across multiple organs and is influenced by patient-specific factors like comorbidities, genetics, and biological microenvironment. Therefore, predicting individual radiosensitivity is key to improving patient outcomes and minimizing side effects. Traditional predictive models have relied on limited statistical methods and fail to integrate complex imaging and genomic data.

Our research aims to address this by applying AI and machine learning (ML) to large existing datasets.

## PRE-ACT

PRE-ACT is a multi-centre European study focused on long-term side effects of radiotherapy in breast cancer patients. Common RT-related complications in breast cancer include skin ulceration, breast atrophy, arm lymphedema, and heart damage. In the first phase, we trained an AI-based risk prediction model based on gradient-boosted decision trees in three European breast cancer cohorts (REQUITE, HypoG-01, CANTO, total n=6,361) to predict arm lymphoedema, defined as ≥5% increase in arm circumference from baseline 15 cm proximal and/or 10 cm distal to the olecranon of the treated side compared to the contralateral side, with an AUC of 0.84 (±0.003) indicating excellent predictive performance. The AI model includes explainability features to ensure transparency and fairness.

The current phase of the project involves a pivotal clinical investigation in France, the Netherlands, and the UK, focusing on patients receiving regional nodal irradiation (Figure 1). In this trial, the XAI-based PRE-ACTOR app communicates individualized risk predictions to physicians and patients in a supportive manner (Figure 2). The study will assess how these personalized insights affect lymphedema rates after RNI, use of supportive interventions (e.g., compression sleeves), and patient quality of life. A control group will receive standard care without AI-driven risk communication.

## RadExIORSBoost

RadExIORSBoost is focused on understanding individual radiosensitivity in prostate cancer patients through transcriptome profiling, cell-based DNA damage assessment, and apoptosis studies. It will also establish a scientific strategy, organize workshops, and strengthen institutional management and research capabilities in this field. The project aims to discover new biomarkers predictive of individual radiosensitivity. The hypothesis is that differences in patients' transcriptome profiles, radiation-induced lymphocyte apoptosis (RILA), and cellular DNA damage levels can distinguish those who will develop adverse effects.

The first phase has involved an analysis of radiation-related transcriptome changes and integrating these data with cell-based assays (RILA, comet assay) and clinical parameters from ~800 patients. Differentially expressed genes from the discovery phase will guide SNP selection in the separate REQUITE prostate cancer patient cohort (n=2,000). These SNPs, together with RILA scores and clinical, pathological, and treatment-related variables, will form the feature set for machine-learning model development. Multiple classification models will be trained and cross-validated on REQUITE and externally validated using patient cohorts from participating institutions.

Insights from these models will guide the design of a clinical study to be completed by the end of the project. Following the study, the accumulated data should support the creation of guidelines for personalized radiation treatment (Figure 3).

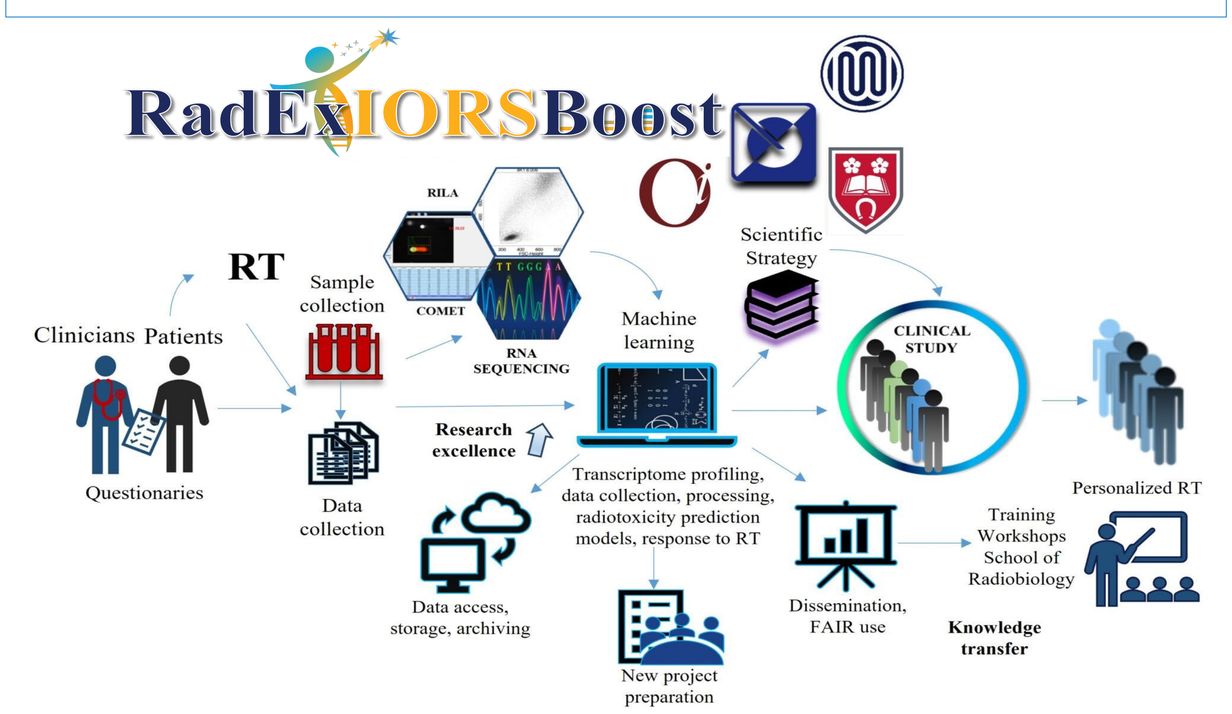


***Figure 3.*** *RadExIORSBoost schema showing integration of different project phases.*

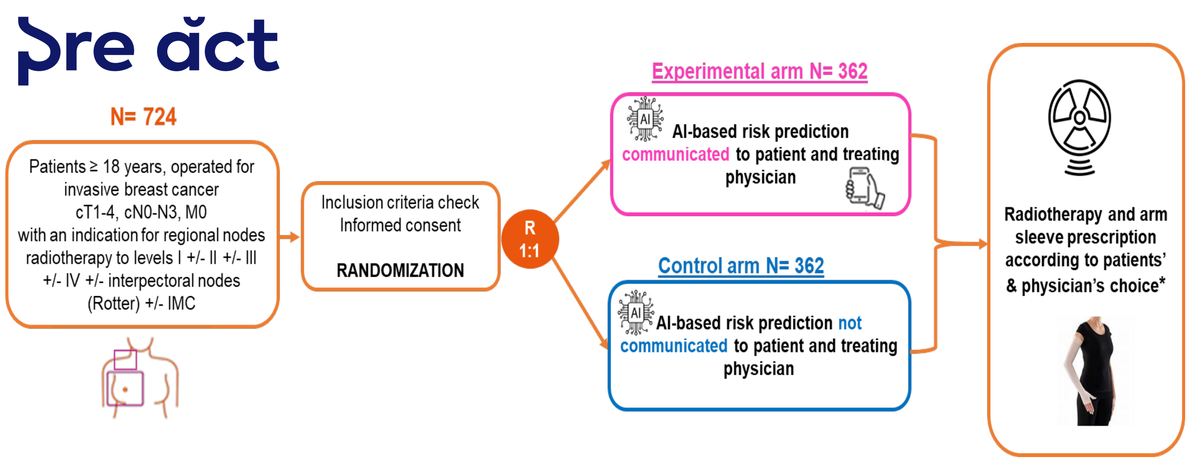


***Figure 1.*** *Clinical investigation schema. Radiotherapy treatment options will include volume-based radiotherapy techniques including conformal 3D RT or IMRT (fixed or rotational) or protons if available and indicated according to institutional and local guidelines, free breathing or deep inspiration breath-hold, irradiation of any lymph node level (level 1 (lower axilla) +/- 2 (upper axilla) +/- 3 (infra clavicular) +/- 4 (supra clavicular) +/- Interpectoral +/- Internal mammary chain), various fractionation regimens (normofractionated, hypofractionated, ultra-hypofractionated) as per local practice and guidelines.*

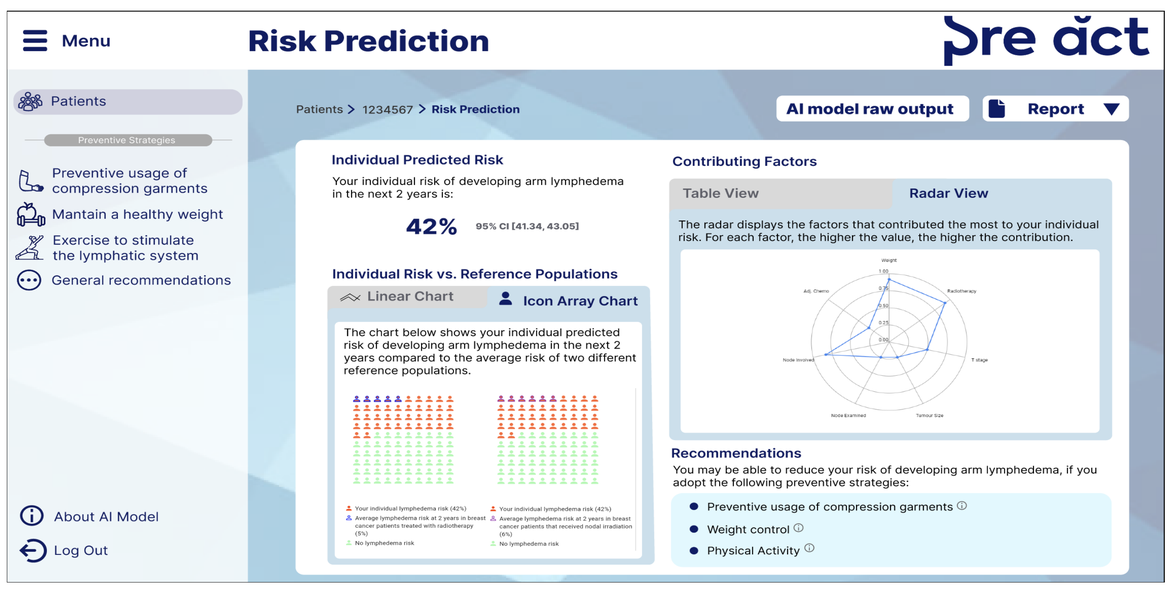


***Figure 2****. PRE-ACTOR app interface showing visualisation of predicted risk and contributing patient and treatment features, using simulated patient data. The risk prediction is shown in relation to the overall reference population and patients within the reference population who had nodal irradiation.*

Collaborators

This work uses data provided by patients as part of their participation in medical research studies. It is conducted in the context of the Horizon Europe project PRE-ACT (Prediction of Radiotherapy side effects using explainable AI for patient communication and treatment modification), supported by the European Commission through the Horizon Europe Program (Grant Agreement number 101057746), by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 22 00058, and by the UK government (Innovate UK application number 10061955). RadExIORSBoost is funded by the European Union under Horizon Europe programme Grant agreement number 101158832.