**Breast Cancer Bone Metastasis Prediction Model: A Retrospective Study Protocol**

**Background**

Bone metastasis in breast cancer is a complex, multi-step biological process. First, primary breast cancer cells proliferate and invade surrounding tissues, escaping into the bloodstream under the regulation of factors such as cadherins, matrix metalloproteinases (MMPs), and integrins. Cancer cells are guided by chemokines to migrate to the skeleton and extravasate from blood vessels to reach bone tissue. Once in the bone, cancer cells can utilize integrins and cadherins to attach and bind to the bone matrix. Ultimately, cancer cells survive, proliferate, and differentiate through interactions with the bone microenvironment, leading to bone metastasis.

The most common site of metastasis in breast cancer is bone, where the balance between osteoclast-mediated bone resorption and osteoblast-mediated bone formation is disrupted. This imbalance leads to osteolytic bone metastases in breast cancer, resulting in bone pain, pathological fractures, spinal cord compression, and other skeletal-related events (SREs). SREs significantly reduce the patient's quality of life and affect the prognosis.

Fabio pointed out that risk factors for breast cancer bone metastasis mainly include the tumor's pathological characteristics and the patient's basic condition. Pathologically, tumor size exceeding 5 centimeters, higher tumor grade, lobular carcinoma subtype, and a greater number of positive lymph nodes all increase bone metastasis risk. Particularly when there are 4 or more axillary lymph node metastases, the 2-year cumulative incidence of bone metastasis can reach 14.9%, and at 10 years it reaches 40.8%. Clinically, age over 60 years and abnormal hemoglobin levels are also important risk factors. Additionally, patients with longer duration of cancer symptoms also have higher bone metastasis risk.

Bone metastasis is the most common metastatic site for solid tumors, with breast cancer accounting for 36% of all bone metastasis cases. According to retrospective analysis of the SEER database, different breast cancer molecular subtypes show significant differences in bone metastasis incidence and metastasis patterns. Luminal-type breast cancer has the highest bone metastasis tendency, with ER+/HER2- bone metastasis incidence at 58.52% and ER+/HER2+ at 47.28%. In contrast, ER-/HER2+ is more prone to liver metastasis, with an incidence of 31.72%, while triple-negative breast cancer is dominated by lung metastasis, with an incidence of 32.09%. And the bone metastasis rates for these two subtypes are 34.49% and 36.39%. Additionally, Molnar et al.'s research found that luminal A (ER+/HER2-) breast cancer not only has high bone metastasis incidence but also manifests as isolated bone metastasis in 59% of cases, and these patients often do not have metastases to other organs when bone metastasis occurs.

The imaging examination system for breast cancer has significant clinical value: ultrasound, mammography, and MRI are the most commonly used imaging modalities for breast cancer diagnosis, holding great importance for its early detection and objective diagnosis, which can effectively reduce the mortality rate of breast cancer patients. Mammography and ultrasound are currently the most common breast examination methods, but they have limitations in examining dense breasts and micro-lesions. MRI has higher sensitivity but relatively insufficient specificity.

Since diagnosis by the human eye mainly relies on subjective judgment of visible morphological changes, it insufficiently utilizes microscopic features and lacks the ability to predict the risk of bone metastasis. Machine learning can extract features from multimodal images (ultrasound, mammography, T1WI, T2WI), extracting high-dimensional information such as deep texture and heterogeneity that are imperceptible to the human eye, and discover hidden correlation patterns between primary lesion imaging features and the risk of bone metastasis through data learning, enabling objective and reproducible risk prediction to predict breast cancer bone metastasis.

**Literature Review**

**Zhong's Study**

Zhong's research indicates that breast cancer (BC) is the most common malignancy in women. According to 2022 U.S. cancer statistics, BC is the most common newly diagnosed malignancy among American women, accounting for 31% of new diagnoses, and is the second leading cause of cancer-related death. With the continuous improvement in breast cancer survival rates, the number of patients developing metastases is also increasing, with bone being the most common site of distant metastasis—nearly 75% of distant metastases are bone metastases (BCBM).

Based on the SEER database, Zhong analyzed 283,373 BC patients between 2010 and 2016. The BCBM patients mainly aged 40-79 years. The incidence was significantly higher in white populations compared to other races, with Luminal A-type BC accounting for 65.7%. Regarding BC prognosis, although the 5-year overall survival rate for non-metastatic BC patients exceeds 80%, patients with distant metastases have only a 25% 5-year survival rate, while BCBM patients have a 5-year survival rate of only 22.8%. Additionally, bone-related events such as fractures, hypercalcemia, or spinal cord compression caused by bone metastasis occur in approximately 45.1% of cases, significantly affecting patient prognosis. Therefore, early identification of high-risk patients who may develop bone metastasis and prediction of their survival rates is crucial for guiding subsequent examinations and treatment of BC patients.

Zhong investigated risk factors for BCBM occurrence through univariate and multivariate analyses. Univariate analysis showed that surgery was the most significant protective factor, while brain metastasis was the most significant risk factor. Other risk factors included grade, T stage, N stage, lung metastasis, and liver metastasis. Multivariate analysis results indicated that risk factors for BCBM include age, race, marital status, grade, surgery, radiotherapy, chemotherapy, brain metastasis, lung metastasis, liver metastasis, ER(Estrogen Receptor), and PR(Progesterone Receptor), with surgery being the most significant protective factor.

Zhong selected 3,492 BCBM patients from the database based on inclusion and exclusion criteria and constructed five machine learning models. Five-fold cross-validation was used to evaluate the performance of the machine learning prediction models, with results averaged after five repetitions and mean ROC curves serving as evaluation indicators. Among the constructed models, the XGBoost model performed best in five-fold cross-validation.

SHAP analysis revealed that surgery, N stage, and T stage are the most significant influencing factors for BCBM. Surgery is a protective factor, while N and T stages are risk factors.

**Li's Study**Li established a prediction model for distant metastasis of BC (Breast Cancer) by utilizing MRI radiomics features combined with a deep learning algorithm. The study included 96 patients in the metastasis group (100 lesions) and 192 randomly selected patients in the non-metastasis group (197 lesions). He manually segmented the DCE-MRI images and extracted morphological and grayscale statistical features from the lesions. A deep learning prediction model was then established through feature dimension reduction and cross-validation. Furthermore, the performance of a deep learning prediction model based only on imaging features was compared with a combined deep learning prediction model incorporating both radiomics and clinical features in terms of accuracy, sensitivity, specificity, and AUC. The results showed that the latter performed better than the former in all three aspects: sensitivity, specificity, and accuracy.

**Wen-hai Zhang's Study**

Wen-hai Zhang adopted a dual-center retrospective study design, enrolling 239 patients and constructing a prediction model using routine clinical blood biomarkers. One center enrolled 176 cases, randomly divided into a training group (123 cases) and test group (53 cases) at an 8:2 ratio. Another center provided 63 cases as an external validation cohort. The external validation cohort came from a geographically proximate but different medical center, thereby validating the model's generalizability across different institutions and patient populations.

Inclusion criteria:

* Confirmed diagnosis of primary breast cancer with newly developed bone metastasis
* Clinical blood biomarker testing completed before treatment (radiotherapy or chemotherapy) or surgical resection
* No history of hypertension, diabetes, or hyperlipidemia
* No history of abnormal blood indicators related to liver, kidney, or cardiovascular function
* No history of other diseases

Exclusion criteria:

* Distant metastasis occurring after treatment (surgical resection or chemotherapy)
* Incomplete clinical blood biomarker data, including tumor markers (AFP, CEA, CA125, CA153, CA199), liver and kidney function tests, blood lipids, or cardiovascular function tests
* Age <18 years
* Metastasis to sites other than bone

Zhang first used Spearman correlation analysis and LASSO regression for feature dimensionality reduction. First, Z-score standardization was applied to normalize each feature, preprocessing the data to fit a standard normal distribution. Then, Spearman rank correlation coefficient (ρ) was used for statistical analysis to measure the correlation between two variables. When ρ approaches 1 or -1, it indicates strong correlation between variables. A threshold of ρ > 0.9 was selected for high correlation. Exceeding the threshold indicates multicollinearity, requiring retention of only one variable with the smallest P-value as the feature variable. Then, L1 regularization of LASSO regression was used for feature dimensionality reduction. The LASSO method penalizes the absolute values of regression coefficients, thereby inducing some coefficients to be zero, facilitating feature selection and generating sparse models. To determine the lambda value for LASSO, the authors used 10-fold cross-validation on the training set to determine the lambda parameter and selected the value that minimized mean squared error, thus preventing overfitting.

Then, the LightGBM machine learning algorithm was used to construct a prediction model for bone metastasis at initial diagnosis and validated in the training set, internal validation set, and external validation set. Research results showed that the LightGBM model demonstrated good AUC in the training set, internal test set, and external validation set, with high sensitivity, specificity, and accuracy.

Feature importance analysis showed that CEA, creatine kinase, albumin/globulin ratio, apolipoprotein B, and CA153 play key roles in model prediction. Univariate analysis showed that lipoprotein a, CA153, γ-glutamyl transferase, α-hydroxybutyrate dehydrogenase, alkaline phosphatase, and creatine kinase were positively correlated with BC bone metastasis, while leukocyte ratio and total cholesterol were negatively correlated.

**Chao Zhang's Study**

Chao Zhang conducted a retrospective study enrolling 6,703 BC patients. Based on inclusion and exclusion criteria, 186 distant metastasis (DM) and non-DM patients were finally identified, divided into training and validation sets at a 7:3 ratio. By integrating multidimensional data from these patients (including MRI and ultrasound imaging features, clinicopathological features, laboratory examination indicators, etc.), three Cox regression models were constructed and validated: a distant metastasis prediction model, a bone metastasis prediction model, and a visceral metastasis prediction model, presented through nomograms. Results showed that the clinicomics (clinical metadata + radiomics) model's predictive performance was significantly superior to models based solely on clinical or imaging data. This model can be used for individualized metastasis risk stratification of BC patients and to guide precise preventive treatment and follow-up strategies.

Detailed inclusion and exclusion criteria:

* BC confirmed through surgical resection specimens and biopsy
* Diagnostic preoperative MRI and US images
* MRI scans and ultrasound examinations performed before neoadjuvant therapy or surgical resection
* At least five years of follow-up data

Construction of radiomics-based model:

Determination of radiomics features: A total of 2,569 radiomics features were extracted for each patient (855 features from T2WI, 859 features from DCE-MRI, and 855 features from US). Then, reliable radiomics features were determined from all these radiomics features.

1. Screening outcome-related features: Wilson's test was used to identify features highly correlated with biomarkers (P<0.05)
2. Avoiding multicollinearity: Pearson correlation matrix was used to evaluate correlations between features; correlation coefficients greater than 0.8 indicate multicollinearity, representing redundant features, with only one feature with the lowest P-value retained
3. LASSO regression was used to select the optimal prognostic combination of features: By calculating radial scores, the linear combination of selected features for each patient was calculated and weighted by their respective coefficients to obtain RadScore (radiomics-based risk value). Finally, ten-fold cross-validation was performed to evaluate the model's true diagnostic potential

Eight imaging features were finally selected, and a combined radiomics Cox model was established using these features. By comparing the AUC of the combined radiomics model, T2WI radiomics model, US radiomics model, and DCE-MRI radiomics model in the test set, the conclusion was drawn that the combined radiomics model was optimal.

Construction of clinical feature-based model:

Univariate analysis was used to evaluate clinicopathological factors in the training set. Variables with P<0.05 in univariate analysis were included in the Cox regression model.

Screened clinical features related to distant metastasis: lymph node metastasis, elevated CA153, elevated CEA, elevated CA125.

The author constructed a clinicopathological Cox model for distant metastasis through these clinicopathological features and further combined it with the combined radiomics model to construct a clinicomics Cox model for distant metastasis.

Additionally, the author performed stratified analysis of bone metastasis-related factors: reproductive history, lymph node metastasis, elevated CA153, elevated CEA, elevated CA125. Also visceral metastasis-related factors: lymph node metastasis, ER positive, PR positive, elevated CA153, elevated CEA, endocrine therapy. Bone metastasis and visceral metastasis models were constructed separately by combining with the radiomics model.

The author compared the AUC at 1, 2, and 3 years postoperatively for the clinicopathological model, combined radiomics model, and clinicomics model, concluding that the clinicomics model was optimal.

At the end of the article, the author presented the distant metastasis prediction model, bone metastasis prediction model, and visceral metastasis Cox prediction model through nomograms and provided calibration curves and ROC for both training and test sets.

**Gao's Study**

Based on the SEER database, Gao developed multiple machine learning models to predict the risk of BCBM occurrence in BC patients and identified key predictive features(tumor size, grade, T/N stage, ER/PR/HER2, and brain, liver, and lung metastasis), through the SHAP framework. with the MLPNN model performing best and the RF model being optimal for excluding bone metastasis.

Gao used the Boruta algorithm to screen important predictive factors for BCBM. Results showed that gender, tumor size, multifocal tumors, histological grade, T stage, N stage, ER/PR/HER2, as well as brain metastasis, liver metastasis, and lung metastasis were significant predictive features of BCBM. Among them, lung metastasis was the most significant feature.

**References**

[1] Zhong, X., Lin, Y., Zhang, W., & Bi, Q. (2023). Predicting diagnosis and survival of bone metastasis in breast cancer using machine learning. Scientific reports, 13(1), 18301. https://doi.org/10.1038/s41598-023-45438-z IF: 3.9 Q1

[2] Li, L., Tian, H., Zhang, B., Wang, W., & Li, B. (2022). Prediction for Distant Metastasis of Breast Cancer Using Dynamic Contrast-Enhanced Magnetic Resonance Imaging Images under Deep Learning. Computational intelligence and neuroscience, 2022, 6126061. https://doi.org/10.1155/2022/6126061

[3] Zhang, W. H., Tan, Y., Huang, Z., Tan, Q. X., Zhang, Y. M., & Wei, C. Y. (2024). Development and validation of an artificial intelligence model for predicting de novo distant bone metastasis in breast cancer: a dual-center study. BMC women's health, 24(1), 442. https://doi.org/10.1186/s12905-024-03264-z

[4] Zhang, C., Qi, L., Cai, J. et al. Clinicomics-guided distant metastasis prediction in breast cancer via artificial intelligence. BMC Cancer 23, 239 (2023). https://doi.org/10.1186/s12885-023-10704-w

[5] Gao, Y., Liu, L., Wang, S., Tao, W., Wang, J., Duan, R., Xie, H., Takahashi, H., Hao, J., & Gao, M. (2025). SEER-based machine learning prediction of bone metastasis in breast cancer: model development and validation. Gland surgery, 14(7), 1366--1378. https://doi.org/10.21037/gs-2025-168

[6] Marazzi, F., Orlandi, A., Manfrida, S., Masiello, V., Di Leone, A., Massaccesi, M., Moschella, F., Franceschini, G., Bria, E., Gambacorta, M. A., Masetti, R., Tortora, G., & Valentini, V. (2020). Diagnosis and Treatment of Bone Metastases in Breast Cancer: Radiotherapy, Local Approach and Systemic Therapy in a Guide for Clinicians. Cancers, 12(9), 2390. https://doi.org/10.3390/cancers12092390

**Research Protocol: Retrospective Study on Breast Cancer Bone Metastasis Prediction Model**

**I. Research Background and Objectives**

**1.1 Research Background**

Breast cancer (BC) is one of the most common malignancies in women. The incidence of BC accounts for 7-10% of all malignancies throughout the body and is also the second leading cause of cancer-related death. Although the diagnostic and treatment levels of BC have continuously improved in recent years, with significant increases in overall patient survival rates, the number of patients with distant metastases continues to increase, which has become a key factor affecting the long-term prognosis of BC patients.

Among various types of distant metastases in BC, bone metastasis (BCBM) is most common, accounting for approximately 75% of all distant metastases. BCBM has an extremely significant impact on patient prognosis. Compared to the 5-year overall survival rate of over 80% for non-metastatic BC patients, the 5-year survival rate of patients with BCBM is only 22.8%. Additionally, bone metastasis is often accompanied by skeletal-related events(SREs) such as pathological fractures, hypercalcemia, and spinal cord compression. These complications not only severely reduce patients' quality of life but also significantly affect patient prognosis.

Therefore, early identification of high-risk populations for bone metastasis and accurate prediction of patient survival risks are highly significant. By establishing precise prediction models, clinicians can screen out high-risk patients in the early stages of disease, promptly conduct targeted imaging monitoring and preventive therapeutic interventions, thereby improving the overall prognosis of BC patients. This has important practical significance for optimizing individualized diagnosis and treatment strategies for BC patients and improving clinical decision-making efficiency.

**1.2 Research Objectives**

To construct a breast cancer bone metastasis prediction model based on clinical metadata and imaging data, identify independent risk factors for breast cancer bone metastasis, and provide a basis for clinical decision-making.

**II. Research Design**

**2.1 Study Type**

Retrospective cohort study

**2.2 Study Population**

Inclusion criteria:

* BC patients diagnosed through surgical resection specimens or biopsy
* Complete clinical data
* Complete imaging data, including mammography, MRI, and ultrasound examinations
* No abnormal blood indicators related to liver, kidney, or cardiovascular function

Exclusion criteria:

* BCBM already present at diagnosis
* Incomplete clinical or imaging data
* Combined with other malignancies
* Missing follow-up data

**2.3 Sample Size Estimation**

Target sample size requires 150-225 patients (positive model accounting for approximately 20-40%).

**III. Data Collection**

**3.1 Clinical Metadata**

3.1.1 Demographic Characteristics  
Age, gender, height, weight, BMI, education level, smoking history, drinking history, race, marital history, reproductive history, menopausal status, family history

3.1.2 Tumor Characteristics

* Pathological type
* Histological grade
* Clinical stage (TNM staging)
* Molecular subtype: ER, PR, HER-2, Ki-67

3.1.3 Laboratory Indicators

* White blood cells, hemoglobin, platelets, serum total protein, serum albumin, alanine aminotransferase, aspartate aminotransferase, serum creatinine, blood urea nitrogen, serum sodium, serum chloride, serum potassium, blood glucose, fibrinogen, D-dimer and blood oxygen saturation, arterial blood gas pH
* Acid phosphatase, alkaline phosphatase, lactate dehydrogenase, blood calcium, globulin and albumin, creatine kinase, apolipoproteins
* CEA, CA153, CA125

3.1.4 Treatment Information

* Surgical method
* Chemotherapy regimen
* Radiotherapy regimen

**3.2 Imaging Data**

* Mammography
* Breast ultrasound
* Breast MRI

**3.3 Outcome Indicators**

Primary outcome: Occurrence of bone metastasis  
Secondary outcomes: Bone metastasis-free survival time, overall survival time, metastasis to other sites

**IV. Statistical Analysis Methods**

**4.1 Data Preprocessing**

Imaging data: All images undergo unified preprocessing, including anonymization, resampling, gray-scale normalization, noise reduction, and artifact removal to ensure data consistency. Use 3D Slicer to extract radiomics features.

Clinical metadata:  
First, conduct data import and quality checks, delete duplicate records, and assess missing values for each variable. Variables with missing rates exceeding 20% are deleted; variables with missing rates between 5%-20% are processed using multiple imputation methods to generate imputed datasets; for variables with missing rates below 5%, continuous variables are filled with the median and categorical variables with the mode. After handling missing values, detect outliers for continuous variables and determine whether they are input errors based on clinical reality, then correct or delete them.

Categorical variables use one-hot encoding, and binary variables use 0/1 encoding. All continuous variables undergo Z-score standardization to eliminate dimensional effects and improve model convergence speed and predictive performance.

Conduct multicollinearity tests on variables and handle variables with severe multicollinearity. Calculate the Pearson correlation coefficient matrix between continuous variables. For multifactor models containing categorical variables, calculate the variance inflation factor (VIF). Generally, VIF>10 indicates severe multicollinearity, and only the variable with the smallest P-value is retained.

Determine variables with clinical significance for BCBM occurrence based on univariate and multivariate analyses. Determine variables with greater impact on BCBM based on the Boruta algorithm. Take the intersection of both to determine the final feature variables.

**4.2 Machine Learning Models**

Logistic Regression, RF, XGBoost, SVM, LightGBM

**4.3 Model Validation**

Dataset division: Training set (70%) and validation set (30%)  
Cross-validation: 10-fold cross-validation

**4.4 Model Performance Evaluation**

* Discrimination: Plot ROC curves with sensitivity as the ordinate and 1-specificity as the abscissa. Area under the curve (AUC) is used to evaluate model discrimination. Compare AUC-ROC of training and validation sets.
* Calibration: Plot calibration curves for both training and validation sets to verify model robustness and reliability. Use Hosmer-Lemeshow goodness-of-fit test to determine model fit.
* Clinical utility: Plot decision curves for both training and validation sets and compare with all predictions and no predictions to determine the model's clinical application value across most threshold probability ranges.

Compare models' AUC, sensitivity, specificity, accuracy, precision, and F1-score.

**4.5 Model Interpretation**

Apply SHAP framework to calculate each feature's contribution value to each sample's prediction result.  
Display global importance ranking of features through SHAP summary plots.

**V. Research Timeline**

Months 1-4: Protocol design, ethics application, data collection and organization  
Months 5-8: Data cleaning, statistical analysis, model construction and validation  
Months 9-12: Paper writing and submission

**VI. Expected Results and Significance**

**6.1 Expected Results**

* Construct an accurate and stable BCBM prediction model
* Organize research methods, model construction process, and validation results and publish papers

**6.2 Research Significance**

* Provide tools for clinicians to identify high-risk bone metastasis patients early
* Guide individualized follow-up
* Provide decision-making basis for preventive interventions for bone metastasis
* Lay the foundation for subsequent prospective studies

**Supplementary and Alternative Approaches**

Establishing machine learning models for bone metastasis sites in breast cancer. According to existing research, breast cancer bone metastases have the highest incidence in the spine, with thoracic and lumbar vertebrae being the most common metastatic sites. Therefore, radiomics feature extraction primarily focuses on a total of 17 vertebral bodies in the thoracic spine (T1-T12) and lumbar spine (L1-L5). MRI sequences are collected for each patient, including T1WI and T2WI. T1WI sequences are sensitive to bone marrow signal changes and can clearly display early manifestations of bone marrow infiltration, with metastatic lesions typically appearing as low signal intensity. T2WI sequences help evaluate the extent of lesion edema and activity status. Through multi-sequence combined analysis, abnormal vertebral changes can be captured more accurately, laying the foundation for subsequent radiomics feature extraction.

A multi-level comparative study is designed, including both cross-sectional comparisons (vertebral comparisons between different patients, comparative analysis of patients with and without bone metastases) and longitudinal comparisons (comparisons of different vertebrae within the same patient, paired analysis of vertebrae with and without metastases). First, raw imaging data in DICOM format is exported. Subsequently, standardized image preprocessing is performed, including gray-level normalization, denoising filtering, image resampling to unify voxel spacing, and bias field correction to eliminate signal inhomogeneity in MRI scans. Regions of interest (ROI) are delineated for each vertebral body, and radiomics features are extracted from T1WI and T2WI sequences respectively using the Radiomics plugin, including first-order statistical features, shape features, texture features, and wavelet transform features. Intraclass correlation coefficients (ICC) are calculated, retaining only highly reproducible features with ICC≥0.75. VIF is then calculated, removing features with VIF>10. To avoid feature redundancy and overfitting, LASSO regression is employed for feature selection, with the optimal regularization parameter lambda determined through cross-validation, ultimately retaining key features with non-zero coefficients for subsequent model construction.

The filtered radiomics features are integrated with the filtered clinical features, and various machine learning algorithms such as Random Forest (RF), Support Vector Machine (SVM), XGBoost, and LightGBM are employed to construct predictive models. The dataset is randomly divided into training and validation sets at a 7:3 ratio. Ten-fold cross-validation is used in the training set to optimize model parameters, and model performance is evaluated on an independent validation set. Model classification performance is evaluated by plotting ROC curves to calculate AUC values, accuracy, sensitivity, specificity, and other metrics, and SHAP values are utilized to interpret feature importance and decision logic of the models.

Additionally, imaging features can also be extracted using the pooling layers of deep transfer learning CNN (DTL-CNN): In the image preprocessing stage, standardization is first performed, then vertebral ROI regions are cropped to a fixed size (such as 224×224 pixels) to meet the input requirements of pre-trained networks, and data augmentation (rotation, flipping, scaling, etc.) is performed to expand training samples. Classic CNN architectures pre-trained on ImageNet (such as ResNet50, VGG16, DenseNet121, or EfficientNet) are then selected, with convolutional layer weights frozen to serve as feature extractors. T1WI and T2WI images of each vertebral body are input into the network separately or after fusion, and deep feature vectors are extracted at the Global Average Pooling layer before the fully connected layers or at the pooling layer of the last convolutional block. These high-dimensional features can automatically capture abstract imaging representations of vertebrae. Dimensionality reduction is similarly applied to the extracted deep features, with key features selected based on feature importance. Finally, the deep imaging features extracted by CNN are integrated with clinical features to construct hybrid predictive models.