

Developing Multimodal Artificial Intelligence-Based Tool for Enhancing Clinical Decision- Making in Personalised Prostate Cancer Treatment

Hein Minn Tun

23H87500@ubd.edu.bn

Universiti Brunei Darussalam

Lin Naing

Universiti Brunei Darussalam

Owais Ahmed Malik

Atlantic Technological University

Muhammad Syafiq Abdullah

The Brunei Cancer Center, (TBCC)

Thu Ta

The Brunei Cancer Center, (TBCC)

Hanif Abdul Rahman

Universiti Brunei Darussalam

Research Article

Keywords: Clinical Decision Support System, Artificial Intelligence in Precision Oncology, Prostate Cancer, Late-stage fusion, Explainable AI, Multimodal AI

Posted Date: July 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10246334/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Prostate cancer is a leading malignancy in men, with rising incidence and suboptimal outcomes due to limitations in conventional treatment approaches based mainly on staging and PSA levels. Recent advances in artificial intelligence (AI) offer opportunities for personalised oncology care. The study aims to develop multimodal artificial intelligence-based clinical decision support system (AI-CDSS) tools by integrating clinical, demographic, genomic, and imaging data to enhance personalised treatment for prostate cancer patients in Brunei.

Method

In this retrospective cohort study (2018–2024), conducted in accordance with the TRIPOD-AI checklist, 82 demographic, clinical, laboratory, genomic, treatment and imaging variables were extracted and integrated into a standardized database. Data preprocessing and analysis were performed using R and Python, including imputation with Extremely Randomized Tree-based iterative imputation method, encoding and normalization. Sequential longitudinal patterns were captured using a recurrent autoencoder, while unsupervised learning (DTWI-HAC) explored latent structures. Supervised models incorporated SMOTE, LSTM-based representation learning, and ensemble and deep learning approaches, evaluated using discrimination metrics and SHAP interpretability. The best-performing tabular and CNN-based imaging models were combined using late-stage fusion methods (weighted averaging, majority voting, and stacking).

Results

The study analysed 212 patients over 6-year (478 observation) demonstrate the unsupervised machine learning model demonstrated a homogeneous structure with a small number of atypical or extreme cases, reflecting the intrinsic clinical similarity of the cohort. The hybrid LSTM–Random Forest architecture demonstrated robust treatment classification performance (overall AUC = 0.75, sensitivity of 32%, and specificity of 83%), effectively capturing temporal PSA dynamics and longitudinal laboratory trends through the LSTM encoder, while leveraging Random Forest for nonlinear decision boundaries. Multimodal late-stage fusion further improved predictive performance; stacking achieved the highest overall AUC with balanced sensitivity, outperforming weighted averaging (0.964).

Conclusion

The study findings and the developed multimodal AI-CDSS models were presented opportunity for prostate cancer treatment prediction. Despite limitations such as small sample size and low sensitivity, these models show promise as clinical decision support tools, requiring further validation in larger prospective studies.

1. Background

Prostate cancer (PCa) was the fourth most common cancer in men, and 1 in 8 men will be diagnosed with it during their lifetime with increasing incidence in Asia [1, 2, 3]. Several risk factors, including advancing age, obesity, genetics, ethnicity, lifestyle and environmental factors, are reported to be associated with prostate cancer risk, but prostate cancer prevention strategies remain challenging and mortality high among Asia, Caribbean and African descents [4–8]. Prostate-specific antigen (PSA) testing is essential for prostate cancer screening and diagnosis but no single PSA level that definitively indicates prostate cancer [8–10]. The diverse treatment modalities ranging from active surveillance to targeted therapy depend on the stages, grades and natural history of disease progression along with the age of the patients along with the benefit and side effect which based on frequency, durability and seriousness of the disease [4, 11–16]. Even though several important guidelines such as NCCN offer recommendations for selecting most appropriate treatment strategies for prostate cancer, clinical decision making in prostate cancer remains highly variable among oncologists [4, 14, 15].

Personalised medicine seeks to address the challenge by tailoring treatment approaches to individual patients' profiles, integrating comprehensive information, including complete clinical history, genomic, and imaging data [17, 18]. Recent advancements in artificial intelligence (AI) offer significant potential in achieving personalised medical care, especially in the field of oncology, as AI algorithms can process and analyse large heterogeneous multi-modal datasets to extract insights that may not be readily apparent to human clinicians [18–25]. In contrast to unimodal AI models that only use a single data source, multimodal AI system, which developed based on multiple data input can integrate the complex interrelationships across varied information sources and enabling a robust output [26, 27]. A systematic review on AI tool performance for prostate cancer care continuum, concerns have been raised about their real-world applicability and need for human-centric design with equity-focused evaluations to ensure robust, ethical, scalable AI deployments in prostate cancer care [28].

The study aims to fill the gaps in the existing clinical practice for oncologist in treatment of prostate cancer by developing and advancing the application of multimodal artificial intelligence based clinical support decision tools (AI-CDSS) to optimize personalized prostate cancer treatment in Brunei Darussalam. Utilizing a developed robust dataset that integrates clinical records, genomic profiles, and imaging data, the study seeks to create sophisticated AI tools capable of accurately predicting disease progression, identifying the most effective treatment strategies, and monitoring therapeutic responses. This research aspires to transform prostate cancer management by delivering precise, individualized treatment recommendations, ultimately improving patient outcomes. Furthermore, it aims to establish a framework for the widespread adoption of AI-CDSS.

2.Method

2.1. Study Design, Population, and Sample

This is a retrospective cohort study designed to analyse prostate cancer patients treated between January 2018 and December 2024 at the Brunei Cancer Centre (TBCC) at Jerudong Park Medical Centre (JPMC). We used TRIPOD-AI (Supplementary Table I) [29]. The inclusion criteria for the study were male patients diagnosed with prostate cancer who had undergone at least one treatment modality, such as hormonal therapy, surgery, radiation therapy, chemotherapy, or surveillance at the TBCC. Patients with secondary metastatic prostate cancer or those with multiple primary tumours were excluded from the study through discussion with the oncologist. 212 eligible prostate cancer patients were selected in the study without sampling.

2.2. Data structure, integration, and preprocessing

Data collection was conducted through systematic extraction and anonymization of participant data. The collected study data were described in the flow chart, which is illustrated in Fig. 1. Out of the 213 patients, 1 patient (ID-004) was excluded from the study due to recurrence prostate cancer with another primary tumour, Ca rectum. Two tabular datasets were developed for the study ; case-structured information, including demographic, survival, clinical and NGS information of 212 observations and cohort structure dataset, expanding initial case structure information with additional time-dependent variables such as treatment, laboratory information, resulting in a final observation of 478. A total of 160 whole-body PSMA PET/CT scans in PNG format, obtained from 121 patients, were included in this study following correspondence with the treatment selection.

The primary outcome (dependent variables) of the study is treatment category. Missing value in both static and longitudinal variables were imputed using Extremely Randomized Trees-based iterative imputation, a non-parametric, iterative imputation technique based on Random Forest algorithm that accommodates mixed type data and non-linear interaction [30]. Initial preprocessing includes data cleaning, harmonization of tumor staging and Gleason Grade Groups, removal of duplicate records and sequence construction of each patient. We performed univariable and multivariable analyses to identify key variable associations, which informed and supported the development of machine learning-based models.

1. Machine Learning Model Development

1. Unsupervised Machine Learning Models

To explore natural groupings among prostate cancer patients and uncover hidden treatment response clusters and latent patterns within the dimensional dataset, we applied multiple unsupervised machine learning techniques with a time-independent approach and a time-dependent approach, as described in (Supplementary Figure I to VI) [31–42].

1. Supervised Machine Learning Models

a) Longitudinal Representation Learning via Autoencoders with Synthetic Minority Oversampling Technique (SMOTE) for supervised machine learning models

1. Data splitting and Missing data handling

The supervised machine learning models were implemented following the analytical workflow illustrated in Fig. 2. Patient-level data were first sorted chronologically using the date of treatment variables, and train-test splitting was performed at the patient level with stratification on treatment outcome to ensure independence between cohorts. The dataset was split into patient-level into training (80%) and testing (20%) cohorts using a group-based split to prevent information leakage [43]. Missing data across both static and longitudinal features were handled using an Extremely Randomized Trees-based iterative imputation approach, which accommodates non-linear relationships and mixed distributions, which categorical variables were encoded using label encoding fitted exclusively on training set [30]. Categorical predictors were encoded using one-hot encoding, while continuous variables were standardized via Z-score normalization to ensure scale comparability. For longitudinal modelling,

patient-specific sequences were constructed by aggregating time-ordered clinical, laboratory and treatment variables with zero-padding applied to achieve uniform sequence length cross patients.

2. Longitudinal Representation learning and class balancing

A recurrent autoencoder based on stacked gated recurrent units (GRUs) was trained on the training cohort to model temporal patterns in laboratory measures and treatment trajectories, with the referencing treatment categories. Laboratory variables (excluding PSA, which was included at both baseline and one year post treatment) were incorporated as baseline inputs, and the resulting latent embeddings were concatenated with static clinical. Model training restricted to the training set to prevent leakage. To address class imbalance in treatment outcome while preserving the longitudinal structure, SMOTE was applied exclusively to the training embedding, thereby avoiding distortion of temporal structure and preventing test-set leakage [44, 45]. The t-SNE (t-Distributed Stochastic Neighbor Embedding) is illustrated in Supplementary Figure II with the distribution of classes in the latent space (embeddings) before and after applying SMOTE.

3. Classical Machine Learning Model

Firstly, the logistic regression model was performed with L2(ridge) penalty and regularization parameter $C = 1.0$ with balanced weight on class. Then, K-Nearest Neighbours (KNN) model ($k = 7$, Minkowski distance $p = 2$), Support Vector Machine (SVM) was performed with both RBF kernel for non-linear boundaries ($C = 5$, $\gamma = 0.05$) and Polynomial kernel (degree = 2, $C = 0.5$), Random Forest (RF) was trained using bootstrap aggregation, and grid search was conducted for decision tree, extreme Gradient Boosting (XGBoost), LightGBM, Decision Tree, Random Forest (RF) and CatBoost model to identify best hyperparameter.

b) Other deep neural network-based architectures with SMOTE for supervised machine learning models

1. Data Handling and preprocessing

The analytical workflow for various neural network architectures based on longitudinal representation for supervised machine learning models was illustrated in Fig. 3. Similar pre-processing was conducted except at the phase of longitudinal representation.

2. Longitudinal representation-based supervised machine learning models

Supervised deep learning models were developed to explicitly model longitudinal clinical trajectories alongside static patient features. Firstly, the Bidirectional Long-Short Term Memory (BiLSTM) model was designed to capture both past and future temporal dependencies by processing sequences in an architecture with masking layers to handle padded values, dropout for regularisation, and late fusion with static covariates before classification. Hyperparameter tuning was performed using systematic search strategies, including optimisation of the number of hidden units (64–256), dropout rates (0.2–0.5), batch size, learning rate and the Adam optimiser with an epoch of 100, and class-weighting schemes to address class imbalance [49, 50].

Secondly, a Temporal Convolutional Network (TCN) was implemented using dilated convolutions and residual connections, which enable efficient modelling of long-range temporal dependencies with fewer parameters. Hyperparameters were tuned, including kernel size, dilation depth, number of filters, dropout rate, and learning rate using the Adam optimiser to balance expressiveness and generalisation. Early stopping was applied to mitigate overfitting [51]. Thirdly, a transformer-based architecture was implemented using multi-head self-attention mechanisms, allowing the model to capture global temporal dependencies without the sequential constraints inherent to recurrent networks. The model consisted of an initial dense projection layer, stacked transformer encoder blocks with multi-head attention and a layer normalisation, followed by global average pooling and fusion with static features. Hyperparameters such as the number of attention heads, embedding dimension, feed-forward network size, dropout rate and learning rate were tuned to improve stability and performance [52–54].

Lastly, end-to-end deep learning approaches, a hybrid LSTM autoencoder with a Random Forest framework was developed to decouple temporal representation learning from classification. The LSTM autoencoder served as a temporal "summarizer" with 64 dimensions of latent representation, while a decoder reconstructed the input to ensure meaningful embedding learning. These latent embeddings were then concatenated with static features and used as inputs to a Random Forest classifier. Hyperparameter tuning was conducted using a grid search approach for the Random Forest model, and the best model was configured with 100 trees, a maximum depth of 10, a minimum split of 5 and the Gini impurity criterion [50, 55].

3. Model Evaluation and Explainability

Model performance was evaluated using a unified patient-level framework across all deep learning and hybrid models to ensure that no data leakage occurred between training and testing sets. A comprehensive set of metrics was computed to capture both overall discrimination and class-balanced performance, including Accuracy, Precision, Recall, and F1-score (reported as weighted averages), along with the Area Under the Receiver Operating Characteristic Curve (AUC) calculated using a one-vs-rest strategy with weighted averaging. Confusion matrices were used to derive clinically meaningful measures, Sensitivity, Specificity, and Negative Predictive Value (NPV), which were calculated on a per-class basis and then averaged to account for class imbalance. In addition, multi-class ROC curve analysis was conducted by binarizing true labels in a one-vs-rest manner, generating class-specific ROC curves and macro-average AUC values for each model.

To enhance interpretability, Shapley Additive exPlanations (SHAP) were applied to the best-performing model characterised by the evaluated indicator especially AUC. SHAP summary, dependence, and force plots were used to visualise time-dependent effects of key static and time varying clinical variables. Furthermore, the individual patient was assessed using SHAP to support transparent, clinically interpretable decision-making consistent with prior methodological frameworks [46, 47, 48]. Additionally, the aggregated models on the time features using mean of all numeric variable were conducted in order to explore the time varying features that can describe the treatment pattern for the patient.

2.4. Late-stage Fusion Pipeline from Tabular and Imaging Best Performing Models

The late-stage fusion model, also known as decision-level fusion, was conducted by integrating the independent final prediction from the best performing model from tabular dataset (comprising of sequential laboratory, clinical information) with prediction from Simple Convolutional Neural Network (CNN) Architecture on image data set (PSMA PET/CT scan image), as illustrated in Fig. 4 [56, 57]. Three decision-level ensemble approach was applied: 1) weighting averaging which average the probabilities of outcome from each model 2) majority voting merging 3) Stacking meta classifier that explore best way to combine prediction. Furthermore, to learn optimal combination of output from each modality, the model will be fined tuned on labelled validation when using staking or weight scoring [56, 57]. The fusion model evaluation was conducted by comparing against its individual models and comparing different benchmark fusion technique (e.g., weighting vs average voting). All evaluation metrics such as accuracy, precision, sensitivity, specificity, recall and AUC are calculated as weighted (calculates each class metrics and weights the average by the average by the number of true instances for each class).

3.Results

3.1. Study Participant Characteristics

A total of 212 patients at the case level and 478 observations over the 6 years were included in the study, and the differences between treatment groups were reported in Supplementary Table II. Age groups, survival status, NGS testing, comorbidities (including hepatitis), relapse status, and care setting (inpatient vs outpatient) were all significantly associated with treatment modalities ($p < 0.05$). Tumour-related characteristics showed strong differentiation, with advanced disease more frequently managed using ARPI-based therapy or chemotherapy, while localised disease was treated with local therapy or active surveillance. Key laboratory markers, including sodium, potassium, PSA (baseline and post-treatment), haemoglobin, and neutrophil count, varied across treatment groups ($p < 0.05$). Additionally, multinomial logistic regression further identified age group, NGS status, family history, comorbidities, relapse, tumour characteristics, and laboratory variables (K, Ca, RBC, WBC, eosinophils, and weight) as significant predictors of treatment selection (Supplementary table III). Extreme odds ratios were observed due to the data sparsity.

3.2. Unsupervised Machine Learning

Multiple time-independent unsupervised machine learning models were applied to explore latent structures, patient subgroups, and intervariable relationships within the dataset, without imposing a predefined outcome level, as described in Supplementary Figure III. Despite achieving some degree of separation, all unsupervised learning approach demonstrated with the clustering was dominated by a single large cluster, with most variables and patients concentrated within it, limiting the model's interpretability and reducing its utility for meaningful feature discrimination or clinical stratification.

1. 3.3. Supervised Machine Learning

1. a) Longitudinal Representation Learning via Autoencoders with Synthetic Minority Oversampling Technique (SMOTE) for supervised machine learning models

Table 1

Performance comparison of supervised machine learning models following Longitudinal Representation Learning with SMOTE application for multiclass prostate cancer treatment prediction

Models	Accuracy	Precision	Recall	F1 Score	Sensitivity	Specificity	AUC
Decision Tree	0.328	0.366	0.328	0.320	0.265	0.818	0.553
Random Forest (RF)	0.344	0.446	0.344	0.328	0.238	0.822	0.616
Extreme Gradient Boosting (XGBoost)	0.281	0.385	0.281	0.263	0.233	0.806	0.613
LightGBM	0.297	0.364	0.297	0.285	0.207	0.806	0.600
CatBoost	0.297	0.346	0.297	0.274	0.210	0.808	0.601
Logistic Regression	0.141	0.182	0.141	0.122	0.180	0.803	0.536
K-Nearest Neighbours (KNN)	0.156	0.196	0.156	0.161	0.109	0.782	0.534
Support Vector Machine (SVM) with RBF	0.109	0.193	0.109	0.097	0.318	0.804	0.542
SVM with Polynomial	0.094	0.246	0.094	0.113	0.108	0.795	0.512

The multiple supervised machine learning models following longitudinal representation learning using autoencoders with class balancing using SMOTE were illustrated in Table 1 to evaluate the ability to predict prostate cancer treatment selection using clinical, laboratory and treatment information. The Random Forest (RF) model achieved the highest overall performance, with an accuracy of 0.344, precision of 0.446, F1 score of 0.328, and the highest area under the ROC curve (AUC = 0.616), indicating the superior discriminative ability. Similarly, LightGBM, and CatBoost models demonstrated competitive performance. In contrast, traditional models such as logistic regression, KNN and SVM exhibited lower performance with accuracies below 0.160 and AUC values close to random classification (0.512 to 0.542), suggesting limited capacity to capture nonlinear interaction inherent in the dataset. ROC illustrated in Fig. 5 demonstrate the supervised machine learning model's performance varies across treatment categories, reflecting the differences in how well each algorithm discriminates specific therapeutic pathways.

b) Other deep neural network-based architectures with SMOTE for supervised machine learning models

The hybrid LSTM encoder with a Random Forest Model demonstrated the strongest overall performance, achieving the highest Accuracy (0.418), Precision (0.412), F1-score (0.390), and AUC (0.749), as well as the highest Sensitivity (0.326) and Specificity (0.326), as illustrated in Table 2. In contrast, the purely deep learning models demonstrate modest discriminative ability, with TCN outperforming the Bilateral LSTM and Transformer in terms of Accuracy (0.348 vs. 0.325 and 0.279, respectively) and precision (0.427 vs. 0.334 and 0.326). While the specificity of the model reached more than 80% for most models, sensitivity remained relatively low (the highest LSTM encoder with Random Forest Model with 32%) across all models.

Table 2

Comparative Performance of different deep neural network-based architectures with SMOTE for supervised machine learning models for multiclass prostate cancer treatment prediction

Models	Accuracy	Precision	Recall	F1 Score	Sensitivity	Specificity	AUC
Bilateral LSTM	0.325	0.334	0.325	0.312	0.304	0.828	0.656
Temporal Convolutional Network (TCN)	0.348	0.427	0.348	0.296	0.238	0.815	0.619
Transformer	0.279	0.326	0.279	0.271	0.268	0.816	0.594
LSTM encoder with Random Forest Model	0.418	0.412	0.418	0.390	0.326	0.837	0.749

The best performing model was demonstrated for Chemotherapy combined with ADT/ARPI, captured by an LSTM encoder, with the Random Forest model (AUC = 0.85), illustrated in Fig. 6. Similarly, the hybrid model achieved a moderate discriminative ability for local therapy (AUC = 0.64), while the TCN model provided better performance. Despite better discriminative performance for other treatment categories, the hybrid model performed with lower discriminative ability for ADT monotherapy (AUC = 0.51) compared to transformer-based model (AUC = 0.76). Collectively, the hybrid model with the LSTM encoder, with the Random Forest Classifier, offers the most consistent and highest discriminative ability across all prostate cancer treatment categories in our cohort.

c) Model Explainability using SHAP

SHAP based interpretability results derived from the overall best performing model, namely the hybrid architecture combining an LSTM encoder with a Random Forest Classifier, illustrated in Fig. 7 (A-D). Similarly, the global importance of time varying information encoded as latent time-series features and the best features, aggregates longitudinal laboratory and clinical measures into summary statistics, revealing PSA-related metrics (mean PSA and post-treatment PSA change), hemoglobin, phosphate, and inflammatory markers as dominant drivers (Fig. 7C). Furthermore, Fig. 7D provides the patient-level explanation through a SHAP waterfall plot of patient 13, who was classified as receiving ARPI-based hormonal/targeted combination therapy might offer clinical interpretability.

3.4. Late-stage Fusion Pipeline from Tabular and Imaging Best Performing Models

Table 3
Comparative Performance of best tabular and imaging models after alignment for test set with late-stage fusion models for multiclass prostate cancer treatment prediction

Models	Accuracy	Precision	Recall	F1 Score	Sensitivity	Specificity	AUC
Model tabular LSTMen + RF	0.286	0.243	0.286	0.219	0.286	0.798	0.801
Model Image CNN fusion	0.429	0.619	0.429	0.429	0.429	0.952	0.857
Late-Stage Fusion Model							
Weighted Averaging	0.286	0.179	0.286	0.200	0.286	0.929	0.964
Majority Voting	0.143	0.024	0.143	0.041	0.143	0.881	NA*
Stacking (Meta-Learner)	0.429	0.184	0.429	0.257	0.429	0.571	0.971
*Majority voting model provide hard prediction as a category							
All evaluation metrics are reported as weighted metrics (calculate each class and weights by the number of true instances)							

The best performing model, the LSTM encoder combined with Random Forest classifier (LSTMen + RF) for a sequential tabular dataset machine learning model and CNN-based classifier imaging models from PSMA PET/CT scan image was selected for prediction of the fusion model. Due to the alignment of the tabular test set with available imaging cases, the overall AUC of tabular model and imaging model changed after data harmonization, AUC of 0.801 and 0.857 respectively. Table 3 describe late-stage fusion strategies that demonstrated improved predictive discrimination compared to single-modality approaches. The stacking(meta-learner) model achieved the highest overall AUC (0.971) and maintained an accuracy of 0.429 with a sensitivity of 0.429, highlighting the complementary contribution of sequential clinical embeddings and imaging-derived features. Similarly, weighting averaging improve AUC of 0.964 and achieved high specificity (0.929) with accuracy remaining 0.286. In contrast, majority voting showed inferior performance (accuracy = 0.143) and provided hard class prediction. Despite the strongest discriminative capacity for the stacking model due to the limited dataset weighted averaging provided a more conservative and potentially less overfitted solution.

4.Discussion

The findings of this study highlight the progressive value of increasingly complex machine learning paradigms in prostate cancer treatment prediction. Supervised machine learning models incorporating longitudinal representation learning substantially improved predictive performance, with ensemble-based methods outperforming traditional classifiers [58–64]. The hybrid model with LSTM encoder combined with a Random Forest classifier achieved the strongest results across all evaluation metrics, with an AUC of 75%, sensitivity of 32%, and specificity of 83%. Among treatment categories, the model provides the discriminative performance for Active Surveillance (AUC = 0.92), ARPI-based hormonal/targeted combination therapy (AUC = 0.92) and Chemotherapy combined with ADT/ARPI(AUC = 0.85). Building upon this, the multimodal late-stage fusion

framework further enhanced discriminative ability by integrating tabular and imaging modalities, where stacking achieved the highest AUC (0.971) with balanced sensitivity (0.429), reinforcing the complementary value of multimodal data integration and probabilistic fusion strategies in complex oncological decision-making contexts [56, 57].

The limited sample in the study constrain the formation of stable, well-separated clusters mentioned as the curse of dimensionality by prior studies for all unsupervised learning models including DTWI-HAC [65–72]. Low sensitivity from all supervised learned models indicates higher risk of under-identifying patients who may benefit from specific treatment modalities, reported similarly in previous prostate cancer risk stratification and treatment allocation studies [72, 73]. Despite the class balancing with SMOTE, the model result in high specificity with low sensitivity, reflecting the complexity and heterogeneity of treatment decision-making and existing guidelines in real-world clinical settings [74]. The study model demonstrates that the role of modelling the longitudinal nature of prostate cancer progression and management is essential for identifying the evolving biomarkers such as PSA trajectories, laboratory and biochemical response to therapy, which are poorly represented by cross-sectional models [74–87].

However, caution is warranted when interpreting model outputs, as the hybrid architecture employed in this study, integrating deep representation learning with ensemble-based classification, was specifically designed to mitigate overfitting risks commonly encountered in small to medium-sized cohorts, thereby enhancing generalizability rather than maximizing apparent performance [80–83]. Treatment specific ROC analysis revealed that ARPI-based hormonal/targeted combination therapy and active surveillance were consistently better discriminated than ADT monotherapy. This finding likely reflects clear clinical decision rules and biomarker thresholds guiding these treatments in clinical practice [4, 84, 85]. In contrast ADT monotherapy showed substantial overlap with other treatment categories, reflecting real world clinical ambiguity in deciding treatment selection [4, 84]. Additional analysis with SHAP on best model demonstrated global and patient level interpretability with tumor stage and risk group emerged as dominant static predictors, which consistent with established prostate cancer treatment guidelines [86–88].

From the LSTM encoder, the time varying variables features were demonstrated as time series features rather than original variables, as the currently SHAP model has limited ability to extract the information from latent features [86, 87]. Further analysis using aggregated longitudinal features with mean of original variables demonstrated as mean PSA, PSA post treatment change, hemoglobin, phosphate and inflammatory markers indicate highly influential, supporting prior evidence such as PSA dynamic, anemia and metabolic alteration with disease burden and treatment selection [88–93]. Moreover, the SHAP waterfall plot of individual patients illustrate how individual level characteristics cumulatively contributed to the treatment selection, which enable model interpretability that can enhance clinical applicability [86, 87].

Multimodal integration strategies include early (feature-level), intermediate (representation-level) and late (decision-level) fusion [94–96]. Early fusion concatenates raw features prior to modeling, enabling joint learning but often suffering from dimensional imbalance, scale heterogeneity and missing data issues [95, 98]. Intermediate fusion integrates modality-specific embeddings within a shared network, allowing cross modal interaction while increasing architectural complexity and data requirement [94, 97, 98]. In this study due to the relatively small and partially aligned cohort, late-stage fusion was most appropriate as it minimized overfitting risk, allowed independent optimization of sequential tabular and imaging pipeline, and preserved clinical interpretability of each developed model [96, 99, 100]. The strong AUC observed from stacking fusion model

aligns with the prior studies demonstrating meta learning approaches can enhance multimodal discrimination in oncology prediction tasks [101–108].

The study has several limitations that need to be considered. Firstly, the relatively small sample size limited the ability to generalize the predicting models and performance estimates are likely optimistic due to limited counts. Secondly, the retrospective nature of the study might introduce bias due to the factors that are not measured in the study dataset. Thirdly, the supervised model can learn only from existing practice patterns rather than optimal treatment strategies, patient preference, clinical judgment and healthcare system factors. Fourthly, the reliance on 2D PNG-derived images limits generalizability and may inflate the performance estimates, particularly for stacking late-stage fusion model which are prone to overfitting in small datasets [95, 108]. Lastly, late-stage fusion does not explicitly model cross-modal feature interactions, which intermediate fusion or attention-based architectures could potentially capture [94, 98]. Future studies should provide external validation using multi-institutional cohorts, and evaluation through nested cross-validation or cost-sensitive learning are recommended to enhance robustness and clinical translation.

5. Conclusion

In our study cohort of 212 patients over 478 observations, integrating longitudinal representation learning with multimodal machine learning approaches offers a promising strategy for improving prostate cancer treatment prediction. While unsupervised methods provided limited clinical utility due to constraints related to small sample size, high feature correlation, class imbalance, and irregular follow up intervals, which collectively hindered the formation of stable and interpretable clusters. Supervised models particularly the hybrid LSTM encoder Random Forest effectively captured temporal and nonlinear relationships within complex clinical data, achieving improved predictive performance. The addition of multimodal late-stage fusion further enhanced discriminative ability, highlighting the complementary value of combining sequential clinical information with image-derived features. Despite such as low sensitivity, small sample size and the study highlight the importance of modeling temporal diseases progression and leveraging explainable AI techniques such as SHAP to support clinical interpretability. Overall, these findings support the role of multimodal AI frameworks as clinical decision support tools rather than replacements for clinician judgment, with future work needed to validate and refine these models in larger, prospective, and multi-institutional settings for robust clinical implementation.

Declarations

Ethics approval and consent to participate:

This study was conducted in accordance with the principles of the Declaration of Helsinki. Ethical approval was obtained from the Medical and Health Research and Ethics Committee (MHREC), Ministry of Health, Brunei Darussalam (Approval No. MHREC/MOH/2024/22(1); approved on 25 January 2025). Following ethical approval, access to electronic health records of participants under the supervision of the Brunei Cancer Centre (TBCC) was permitted. All data were fully anonymized prior to analysis, and the requirement for informed consent to participate was waived by the MHREC in accordance with national regulations.

Competing interest:

All authors do not have a conflict of interest to declare.

Clinical Trial Number

Not applicable.

Consent for publication:

Permission to publish the findings of this study was obtained from the relevant institutional authorities. All data were anonymized prior to publication

Funding:

This work was supported by Universiti Brunei Darussalam through research (UBD/RSCH/1.6/FICBF/2025/008), recipient: Hanif Abdul Rahman.

Author Contribution

All authors were involved in conceptualizations of this report. HMT was responsible for data curation, investigation, methodology, resources, visualization, and writing the original draft. HAR was involved in methodology development, supervision, methodology, resources, and writing, reviewing, and editing the manuscript. OAM provided supervision and reviews. LN contributed through supervision, writing, reviewing, and editing the manuscript. SAR and TT contributed through conceptualization, methodology and resources. All authors approve the content of the manuscript.

Acknowledgement

The authors would like to thank the Ministry of Health, Brunei Darussalam, and the Brunei Cancer Centre (TBCC) for granting permission to access the required data for this study. We also acknowledge the support of the Medical and Health Research and Ethics Committee (MHREC) for ethical approval and Universiti Brunei Darussalam for the guidance and support. The authors are grateful to all individuals and institutions who contributed to the successful completion of this research.

Data Availability

All data generated or analyzed during this study are included in this published article. The datasets generated and/or analyzed during the current study are not publicly available to preserve the anonymity of study participants.

Code Availability

The multimodal AI models were developed in Python using PyTorch, with image processing, tabular analysis, survival modeling, ensemble learning, and interpretability performed using standard libraries including MONAI, scikit-learn, pycox, XGBoost, CatBoost, and SHAP. All experiments were conducted in a controlled environment to ensure reproducibility, and the full source code is available at:

<https://github.com/heinminntun/Developing-Multimodal-AI-CDSS-for-prostate-cancer.git>

References

1. World Health Organization (WHO). Cancer. Accessed August 12, 2024. <https://www.who.int/news-room/fact-sheets/detail/cancer>
2. Siegel RL, Kimberly D, Miller NS, Wagle, Jemal A. Cancer Statistics, 2023. *CA: A Cancer Journal for Clinicians* 73, no. 1 (January 2023): 17–48. <https://doi.org/10.3322/caac.21763>
3. Cancer (IARC), The International Agency for Research on. Global Cancer Observatory. Accessed May 9, 2025. <https://gco.iarc.fr/>
4. DeVita V, Lawrence T, Rosenberg S. DeVita, Hellman, and Rosenberg's Cancer: Principles and Practice of Oncology, 12th Edition. eTextbooks for Students, January 1, 2023. <https://stars.library.ucf.edu/etextbooks/760>
5. SEER, Cancer Statistics Review. 1975–2015 - Previous Version - SEER Cancer Statistics Review. Accessed May 9, 2025. https://seer.cancer.gov/archive/csr/1975_2015/index.html
6. Perdana N, Riza CA, Mochtar R, Umbas. The Risk Factors of Prostate Cancer and Its Prevention: A Literature Review. *Acta Med Indones*. July 2016;48(3):228–38.
7. Hassanipour S, Delam H, Arab-Zozani M, Abdzadeh E, Hosseini SA, Nikbakht HA, Malakoutikhah M, Ashoobi MT, Fathalipour M, Salehiniya H, Riahi S. Survival Rate of Prostate Cancer in Asian Countries: A Systematic Review and Meta-Analysis. *Ann Glob Health*. 2020;86(1):2. 10.5334/aogh.2607. . PMID: 31976303; PMCID: PMC6966336.
8. National Cancer Institute, Prostate-Specific Antigen (PSA). Test - NCI. Article, December 12, 2024. Nci global. <https://www.cancer.gov/types/prostate/psa-fact-sheet>
9. Gulati R, Gore JL, Etzioni R. Comparative effectiveness of alternative prostate-specific antigen–based prostate cancer screening strategies: Model estimates of potential benefits and harms. *Ann Intern Med*. 2013;158(3):145–53.
10. RI G, Pf P, Rt G, Izmirlian G, Miller Ab H, Tp R, TI et al. December. Prostate Cancer Screening in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial: Update on Findings from the Initial Four Rounds of Screening in a Randomized Trial. *BJU International* 102, 11 (2008). <https://doi.org/10.1111/j.1464-410X.2008.08214.x>
11. Sandhu S, Moore CM, Chiong E, Beltran H, Bristow RG, Williams SG. Prostate Cancer. *The Lancet* 398, 10305 (September 18, 2021): 1075–90. [https://doi.org/10.1016/S0140-6736\(21\)00950-8](https://doi.org/10.1016/S0140-6736(21)00950-8)
12. Wasim S, Lee SY, Kim J. Complexities of Prostate Cancer. *Int J Mol Sci*. 2022;23(22):14257. 10.3390/ijms232214257. . PMID: 36430730; PMCID: PMC9696501.
13. National Cancer Institute. Prostate Cancer Treatment (PDQ®) - NCI. Cancer Info Summary, August 9, 2024. NCI global enterprise. <https://www.cancer.gov/types/prostate/hp/prostate-treatment-pdq>

14. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®). Prostate Cancer, Version 2.2025 [cited 2025 May 7]. Available from: https://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf
15. Wasim S, Lee SY, Kim J. Complexities of Prostate Cancer. *Int J Mol Sci.* 2022;23(22):14257. 10.3390/ijms232214257. . PMID: 36430730; PMCID: PMC9696501.
16. Pound CR, Partin AW, Eisenberger MA, Chan DW, Pearson JD, Walsh PC. Natural History of Progression after PSA Elevation Following Radical Prostatectomy. *JAMA.* 1999;281(5):1591–97. <https://doi.org/10.1001/jama.281.17.1591>.
17. Gallagher K. Not Your Father's Prostate Cancer. Prostate Cancer Foundation, October 3, 2016. <https://www.pcf.org/c/not-your-fathers-prostate-cancer/>
18. Topol EJ. High-Performance Medicine: The Convergence of Human and Artificial Intelligence. *Nature Medicine* 25, 1 (January 2019): 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
19. Fountzilas E, Pearce T, Baysal MA, Chakraborty A, Tsimberidou AM. Convergence of Evolving Artificial Intelligence and Machine Learning Techniques in Precision Oncology. *Npj Digit Med.* January 31, 2025;8(1):1–19. <https://doi.org/10.1038/s41746-025-01471-y>.
20. McGale J, Liao MJ, Lopci E, Marabelle Aurélien, Dercle L. Artificial Intelligence: A Transformative Tool. *Precision Oncol Oncotarget.* August 26, 2024;15:588–89. <https://doi.org/10.18632/oncotarget.28639>.
21. Liao J, Li X, Gan Y, Han S, Rong P, Wang W, Li W, Li, Zhou. Artificial Intelligence Assists Precision Medicine in Cancer Treatment. *Front Oncol.* 2023;12(4):998222. <https://doi.org/10.3389/fonc.2022.998222>.
22. Kantoff PW, Celestia S, Higano ND, Shore E, Roy Berger EJ, Small DF, Penson CH, Redfern, et al. Sipuleucel-T Immunotherapy for Castration-Resistant Prostate Cancer. *N Engl J Med.* July 29, 2010;363(5):411–22. <https://doi.org/10.1056/NEJMoa1001294>.
23. ArteraAI. ArteraAI Prostate Test. Accessed May 12, 2025. <https://artera.ai/arteraai-prostate-cancer-test>
24. The Paige Prostate Suite. Assistive Artificial Intelligence for Prostate Cancer Diagnosis: Emerging Health Technologies. CADTH Horizon Scans. Ottawa (ON): Canadian Agency for Drugs and Technologies in Health, 2024. <http://www.ncbi.nlm.nih.gov/books/NBK608438/>
25. Food, Drug Administration USA, FDA Grants. 510(k) Clearance to AI-Powered Prostate Cancer Diagnostic. Accessed May 12, 2025. <https://www.urologytimes.com/view/fda-grants-510-k-clearance-to-ai-powered-prostate-cancer-diagnostic>
26. Lipkova J, Chen RJ, Chen B, Ming Y, Lu M, Barbieri D, Shao AJ, Vaidya, et al. *Oncol Cancer Cell.* October 10, 2022;40:1095–110. <https://doi.org/10.1016/j.ccell.2022.09.012>. Artificial Intelligence for Multimodal Data Integration in.
27. Yang H, Yang M, Chen J, Yao G, Zou Q, Jia L. Multimodal Deep Learning Approaches for Precision Oncology: A Comprehensive Review. *Briefings in Bioinformatics* 26, 1 (January 1, 2025): bbae699. <https://doi.org/10.1093/bib/bbae699>
28. Tun H, Minn L, Naing OA, Malik. and Hanif Abdul Rahman. Artificial Intelligence (AI)-Based Tools in the Diagnosis and Management of Prostate Cancer: A Systematic Review and Meta-Analysis. *Prostate Cancer and Prostatic Diseases*, December 12, 2025, 1–15. <https://doi.org/10.1038/s41391-025-01060-w>
29. Collins GS, Karel GM, Moons P, Dhiman et al. TRIPOD + AI Statement: Updated Guidance for Reporting Clinical Prediction Models That Use Regression or Machine Learning Methods. *Res Methods Report BMJ* 385 (April 2024): e078378. <https://doi.org/10.1136/bmj-2023-078378>

30. Geurts P, Ernst D, Wehenkel L. Extremely Randomized Trees Mach Learn. 2006;63(1):3–42.
<https://doi.org/10.1007/s10994-006-6226-1>.
31. Jolliffe IT, Cadima J. Principal Component Analysis: A Review and Recent Developments. Philosophical Transactions., Series A, Mathematical. Physical, and Engineering Sciences 374, no. 2065 (2016): 20150202.
<https://doi.org/10.1098/rsta.2015.0202>
32. Rousseeuw PJ. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. J Comput Appl Math. November 1987;20:53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
33. Ester M, Kriegel H-P, Jörg Sander, and, Xu X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. January 1, 1996, 226–331.
<https://ui.adsabs.harvard.edu/abs/1996kddm.conf.226E>
34. Agrawal R, Imieliński T, Swami A. Mining Association Rules between Sets of Items in Large Databases. SIGMOD Rec. 1993;22(2):207–16. <https://doi.org/10.1145/170036.170072>.
35. Goodfellow I, Bengio Y, Courville A. (2016). 14. Autoencoders. Deep learning. Adaptive computation and machine learning. Cambridge, Mass: The MIT press. ISBN 978-0-262-03561-3.
36. Kumar U, Legendre CédricP, Lee J-C, Zhao L. On Analyzing GNSS Displacement Field Variability of Taiwan: Hierarchical Agglomerative Clustering Based on Dynamic Time Warping Technique. Comput Geosci. December 2022;169:105243. <https://doi.org/10.1016/j.cageo.2022.105243>.
37. Tiozzo A. Dynamic Time Warping for Time Series Classification,(November 2022). <https://medium.com/eni-digitaltalks/dynamic-time-warping-for-time-series-classification-6482cf9d7c3b>
38. Shokoohi-Yekta M, Jin BHH, Wang J, Keogh E. Generalizing DTW to the Multi-Dimensional Case Requires an Adaptive Approach. Data Mining and Knowledge Discovery 31, no. 1 (2017): 1–31.
<https://doi.org/10.1007/s10618-016-0455-0>
39. Leeuw Jde. Modern Multidimensional Scaling: Theory and Applications (Second Edition). Journal of Statistical Software 14September (2005): 1–2. <https://doi.org/10.18637/jss.v014.b04>
40. Müllner D. Modern Hierarchical, Agglomerative Clustering Algorithms. arXiv:1109.2378. Preprint, arXiv, September 12, 2011. <https://doi.org/10.48550/arXiv.1109.2378>
41. Rousseeuw PJ. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. J Comput Appl Math. November 1987;20:53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
42. Steyaert S, Pizurica M, Nagaraj D, et al. Multimodal Data Fusion for Cancer Biomarker Discovery with Deep Learning. Nat Mach Intell. 2023;5(4):351–62. <https://doi.org/10.1038/s42256-023-00633-5>.
43. Scikit-Learn. GroupShuffleSplit. Accessed November 28, 2025. https://scikit-learn/stable/modules/generated/sklearn.model_selection.GroupShuffleSplit.html
44. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-Sampling Technique. J Artif Intell Res. June 2002;16:321–57. <https://doi.org/10.1613/jair.953>.
45. Li Y, Yang Y, Song P, Duan L, Ren R. Space Sci Rep. 2025;15(1):23521. <https://doi.org/10.1038/s41598-025-09506-w>. An Improved SMOTE Algorithm for Enhanced Imbalanced Data Classification by Expanding Sample Generation.
46. Krzyżiński M. Mikołaj Spytek, Hubert Baniecki, and Przemysław Biecek. SurvSHAP(t): Time-Dependent Explanations of Machine Learning Survival Models. Knowl Based Syst. February 2023;262:110234.
<https://doi.org/10.1016/j.knosys.2022.110234>.

47. Sato K, Sakamoto S, Saito S, et al. Time-Dependent Personalized Prognostic Analysis by Machine Learning in Biochemical Recurrence after Radical Prostatectomy: A Retrospective Cohort Study. *BMC Cancer*. 2024;24(1):1446. <https://doi.org/10.1186/s12885-024-13203-8>.
48. Ponce-Bobadilla A, Victoria V, Schmitt CS, Maier S, Mensing, Stodtmann S. Practical Guide to SHAP Analysis: Explaining Supervised Machine Learning Model Predictions in Drug Development. *Clin Transl Sci*. 2024;17(11):e70056. <https://doi.org/10.1111/cts.70056>.
49. Abduljabbar RL, Dia H, Tsai P-W. Development and Evaluation of Bidirectional LSTM Freeway Traffic Forecasting Models Using Simulation Data. *Scientific Reports* 11, no. 1 (2021): 23899. <https://doi.org/10.1038/s41598-021-03282-z>
50. Lipton ZC, Kale DC, Charles Elkan, and, Wetzel R. Learning to Diagnose with LSTM Recurrent Neural Networks. arXiv:1511.03677. Preprint, arXiv, March 21, 2017. <https://doi.org/10.48550/arXiv.1511.03677>
51. Bai, Shaojie JZ, Kolter, Koltun V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. arXiv:1803.01271. Preprint, arXiv, April 19, 2018. <https://doi.org/10.48550/arXiv.1803.01271>
52. Li S, Jin X, Xuan Y et al. Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting. arXiv:1907.00235. Preprint, arXiv, January 3, 2020. <https://doi.org/10.48550/arXiv.1907.00235>
53. Song H, Rajan D, Thiagarajan JJ, Spanias A. Attend and Diagnose: Clinical Time Series Analysis Using Attention Models. arXiv:1711.03905. Preprint, arXiv, November 19, 2017. <https://doi.org/10.48550/arXiv.1711.03905>
54. Vaswani A, Shazeer N, Parmar N et al. Attention Is All You Need. arXiv:1706.03762. Preprint, arXiv, August 2, 2023. <https://doi.org/10.48550/arXiv.1706.03762>
55. Miotto R, Li L, Kidd BA, Dudley JT. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. *Sci Rep*. May 2016;6:26094. <https://doi.org/10.1038/srep2609>.
56. Carrillo-Perez F, Morales JC, Castillo-Secilla D, Gevaert O, Rojas I. J Personalized Med. 2022;12(4):601. <https://doi.org/10.3390/jpm12040601>. and Luis Javier Herrera. Machine-Learning-Based Late Fusion on Multi-Omics and Multi-Scale Data for Non-Small-Cell Lung Cancer Diagnosis.
57. Baltrušaitis T, Ahuja C. and Louis-Philippe Morency. Multimodal Machine Learning: A Survey and Taxonomy. arXiv:1705.09406. Preprint, arXiv, August 1, 2017. <https://doi.org/10.48550/arXiv.1705.09406>
58. Breiman L. Random Forests Mach Learn. 2001;45(1):5–32. <https://doi.org/10.1023/A:1010933404324>.
59. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 13, 2016, 785–94. <https://doi.org/10.1145/2939672.2939785>
60. Ke G, Meng Q, Finley T et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Adv Neural Inf Process Syst* 30 (2017). https://papers.nips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html
61. Hastie T, Tibshirani R, Friedman J, The Elements of Statistical Learning: Data Mining, Inference, and, Prediction. Second Edition | Springer Nature Link. Accessed February 6, 2026. <https://link.springer.com/book/10.1007/978-0-387-84858-7>
62. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine Learning Applications in Cancer Prognosis and Prediction. *Comput Struct Biotechnol J*. January 2015;13:8–17.

- <https://doi.org/10.1016/j.csbj.2014.11.005>.
63. Hochreiter S. and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation* 9, 8 (1997): 1735–80. <https://doi.org/10.1162/neco.1997.9.8.1735>
64. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. *JMLR Workshop and Conference Proceedings* 56August (2016): 301–18. Epub 2016 Dec 10. PMID: 28286600; PMCID: PMC5341604. <https://pmc.ncbi.nlm.nih.gov/articles/PMC5341604/>
65. Esteva A, Robicquet A, Ramsundar B, et al. A Guide to Deep Learning in Healthcare. *Nat Med*. 2019;25(1):24–9. <https://doi.org/10.1038/s41591-018-0316-z>.
66. Miotto R, Wang F, Wang S, Jiang X, Joel T, Dudley. Deep Learning for Healthcare: Review, Opportunities and Challenges. *Brief Bioinform*. 2018;19(6):1236–46. <https://doi.org/10.1093/bib/bbx044>.
67. Lee J, Warner E, Shaikhouni S, et al. Disease Sci Rep. 2022;12(1):4832. <https://doi.org/10.1038/s41598-022-08974-8>. Unsupervised Machine Learning for Identifying Important Visual Features through Bag-of-Words Using Histopathology Data from Chronic Kidney.
68. Gardner-Hoag J, Novack M, Parlett-Pelleriti C, Stevens E, Dixon D, Linstead E. *JMIR Med Inf*. 2021;9(6):e27793. <https://doi.org/10.2196/27793>. Unsupervised Machine Learning for Identifying Challenging Behavior Profiles to Explore Cluster-Based Treatment Efficacy in Children With Autism Spectrum Disorder: Retrospective Data Analysis Study.
69. Bellman R. *Adaptive Control Processes: A Guided Tour*. Princeton: Princeton University Press; 1961. <https://doi.org/10.1515/9781400874668>.
70. Chen L. Curse of Dimensionality. *Encyclopedia of Database Systems*. Boston, MA: Springer; 2009. https://doi.org/10.1007/978-0-387-39940-9_133.
71. Kopland MCG, Giltay EJ. Dynamic Time Warp (DTW) as a Scalable, Data-Efficient, and Clinically Relevant Analysis of Dynamic Processes in Patients with Psychiatric Disorders: A Tutorial. *J Eat Disorders*. October 2025;13:230. <https://doi.org/10.1186/s40337-025-01414-8>.
72. He H, Garcia EA. Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering* 21, no. 9 (2009): 1263–84. <https://doi.org/10.1109/TKDE.2008.239>
73. Roach M, Weinberg III, Sandler V, H. and, Thompson I. Staging for prostate cancer. *Cancer*. 2007;109:213–20. <https://doi.org/10.1002/cncr.22403>.
74. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-Sampling Technique. *J Artif Intell Res*. June 2002;16:321–57. <https://doi.org/10.1613/jair.953>.
75. Goldstein BA, Ann Marie Navar MJ, Pencina, John PA, Ioannidis. Opportunities and Challenges in Developing Risk Prediction Models with Electronic Health Records Data: A Systematic Review. *J Am Med Inform Assoc*. 2017;24(1):198–208. <https://doi.org/10.1093/jamia/ocw042>.
76. Rajkomar A, Oren E, Chen K, et al. Scalable and Accurate Deep Learning with Electronic Health Records. *Npj Digit Med*. 2018;1(1):18. <https://doi.org/10.1038/s41746-018-0029-1>.
77. Li S, Jin X, Xuan Y et al. Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting. arXiv:1907.00235. Preprint, arXiv, January 3, 2020. <https://doi.org/10.48550/arXiv.1907.00235>
78. Song H, Rajan D, Thiagarajan JJ, Spanias A. Attend and Diagnose: Clinical Time Series Analysis Using Attention Models. arXiv:1711.03905. Preprint, arXiv, November 19, 2017. <https://doi.org/10.48550/arXiv.1711.03905>

79. Vaswani A, Shazeer N, Parmar N et al. Attention Is All You Need. arXiv:1706.03762. Preprint, arXiv, August 2, 2023. <https://doi.org/10.48550/arXiv.1706.03762>
80. Breiman L. Random Forests Mach Learn. 2001;45(1):5–32. <https://doi.org/10.1023/A:1010933404324>.
81. Hastie T, Tibshirani R, Friedman J, The Elements of Statistical Learning: Data Mining, Inference, and, Prediction. Second Edition | Springer Nature Link. Accessed February 6, 2026. <https://link.springer.com/book/10.1007/978-0-387-84858-7>
82. Bengio Y, Courville A, Vincent P. Representation Learning: A Review and New Perspectives. IEEE Trans Pattern Anal Mach Intell. 2013;35(8):1798–828. <https://doi.org/10.1109/TPAMI.2013.50>.
83. Jin Lin-peng, Dong J. Ensemble Deep Learning for Biomedical Time Series Classification. Computational Intelligence and Neuroscience 2016 (2016): 6212684. <https://doi.org/10.1155/2016/6212684>
84. Mottet N, van den Bergh Rcn, Briers E, et al. EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer-2020 Update. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. Eur Urol. 2021;79(2). <https://doi.org/10.1016/j.eururo.2020.09.042>.
85. Cooperberg MR, Peter R, Carroll. Trends in Management for Patients With Localized Prostate Cancer, 1990–2013. JAMA. 2015;314(1):80–2. <https://doi.org/10.1001/jama.2015.6036>.
86. Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems (Red Hook, NY, USA), NIPS'17, December 4, 2017, 4768–77. <https://dl.acm.org/doi/10.5555/3295222.3295230>
87. Ribeiro M, Tulio S, Singh, Guestrin C. 'Why Should I Trust You?': Explaining the Predictions of Any Classifier. arXiv:1602.04938. Preprint, arXiv, August 9, 2016. <https://doi.org/10.48550/arXiv.1602.04938>
88. Molnar C. (2022) Interpretable Machine Learning. A Guide for Making Black Box Models Explainable, Self-Published. - References - Scientific Research Publishing. Accessed February 6, 2026. <https://www.scirp.org/reference/referencespapers?referenceid=4062912>
89. Bergh RCN, van den S, Roemeling MJ, Roobol T, Wolters FH, Schröder, Bangma CH. Prostate-Specific Antigen Kinetics in Clinical Decision-Making During Active Surveillance for Early Prostate Cancer—A. Rev Eur Urol. 2008;54(3):505–16. <https://doi.org/10.1016/j.eururo.2008.06.040>.
90. Lilja H, Ulmert D, Vickers A. Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. Nat Rev Cancer. 2008;8:268–78. <https://doi.org/10.1038/nrc2351>.
91. Beer TM, Catherine M, Tangen LB, Bland IM, Thompson, Crawford ED. Prognostic Value of Anemia in Newly Diagnosed Metastatic Prostate Cancer: A Multivariate Analysis of Southwest Oncology Group Study 8894. The Journal of Urology 172, 6 Pt 1 (2004): 2213–17. <https://doi.org/10.1097/01.ju.0000147771.92104.83>
92. Lundquist K, Antti H. Metabolomic Insights into Prostate Cancer Treatment and Relapse. Cancers. 2025;17:24. 3993. <https://doi.org/10.3390/cancers17243993>.
93. Stark T, Livas L, Kyprianou N. Inflammation in Prostate Cancer Progression and Therapeutic Targeting. Translational Androl Urol. 2015;4(4):455–63. <https://doi.org/10.3978/j.issn.2223-4683.2015.04.12>.
94. Huang S-C, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of Medical Imaging and Electronic Health Records Using Deep Learning: A Systematic Review and Implementation Guidelines. NPJ Digit Med. 2020;3:136. <https://doi.org/10.1038/s41746-020-00341-z>.
95. Ramachandram D, Taylor GW. Deep Multimodal Learning: A Survey on Recent Advances and Trends. IEEE Signal Process Mag. 2017;34(6):96–108. <https://doi.org/10.1109/MSP.2017.2738401>.

96. Gadzicki K, Khamsehashari R, Zetsche C. Early vs Late Fusion in Multimodal Convolutional Neural Networks. 2020 *IEEE 23rd International Conference on Information Fusion (FUSION)*, July 2020, 1–6. <https://doi.org/10.23919/FUSION45008.2020.9190246>
97. Stahlschmidt SörenR, Benjamin Ulfenborg, and, Synnergren J. Multimodal Deep Learning for Biomedical Data Fusion: A Review. *Brief Bioinform.* 2022;23(2):bbab569. <https://doi.org/10.1093/bib/bbab569>.
98. Guarrasi V, Aksu F, Caruso CM, et al. A Systematic Review of Intermediate Fusion in Multimodal Deep Learning for Biomedical Applications. *Image Vis Comput.* May 2025;158:105509. <https://doi.org/10.1016/j.imavis.2025.105509>.
99. Kline A, Wang H, Li Y, et al. Multimodal Machine Learning in Precision Health: A Scoping Review. *NPJ Digit Med.* 2022;5(1):171. <https://doi.org/10.1038/s41746-022-00712-8>.
100. Sun Z, Lin M, Zhu Q, et al. A Scoping Review on Multimodal Deep Learning in Biomedical Images and Texts. *J Biomed Inform.* October 2023;146:104482. <https://doi.org/10.1016/j.jbi.2023.104482>.
101. Ting KM, Witten IH. Issues in Stacked Generalization. *J Artif Intell Res.* May 1999;10:271–89. <https://doi.org/10.1613/jair.594>.
102. Hosni M, Abnane I, Idri A, Carrillo de Gea JMC, Alemán JLF. Reviewing ensemble classification methods in breast cancer. *Comput Methods Programs Biomed.* 2019;177:89–112.
103. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine Learning Applications in Cancer Prognosis and Prediction. *Comput Struct Biotechnol J.* 2015;13:8–17. <https://doi.org/10.1016/j.csbj.2014.11.005>.
104. Simaiya S, Kaur R, Sandhu JK, et al. A Novel Multistage Ensemble Approach for Prediction and Classification of Diabetes. *Front Physiol.* December 2022;13:1085240. <https://doi.org/10.3389/fphys.2022.1085240>.
105. Sistaninejhad B, Rasi H, Nayeri P. A Review Paper about Deep Learning for Medical Image Analysis. *Computational and Mathematical Methods in Medicine* 2023 (May 2023): 7091301. <https://doi.org/10.1155/2023/7091301>
106. Lambin P, Leijenaar RTH, Deist TM et al. Radiomics: The Bridge between Medical Imaging and Personalized Medicine | *Nature Reviews Clinical Oncology*. Accessed February 24, 2026. <https://www.nature.com/articles/nrclinonc.2017.141>
107. Vallières M, Kay-Rivest E, Léo, Perrin J et al. Radiomics Strategies for Risk Assessment of Tumour Failure in Head-and-Neck Cancer. *Scientific Reports* 7, no. 1 (2017): 10117. <https://doi.org/10.1038/s41598-017-10371-5>
108. Aerts HJWL, Rios E, Velazquez, Ralph TH, Leijenaar, et al. Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nat Commun.* 2014;5(1):4006. <https://doi.org/10.1038/ncomms5006>.

Figures

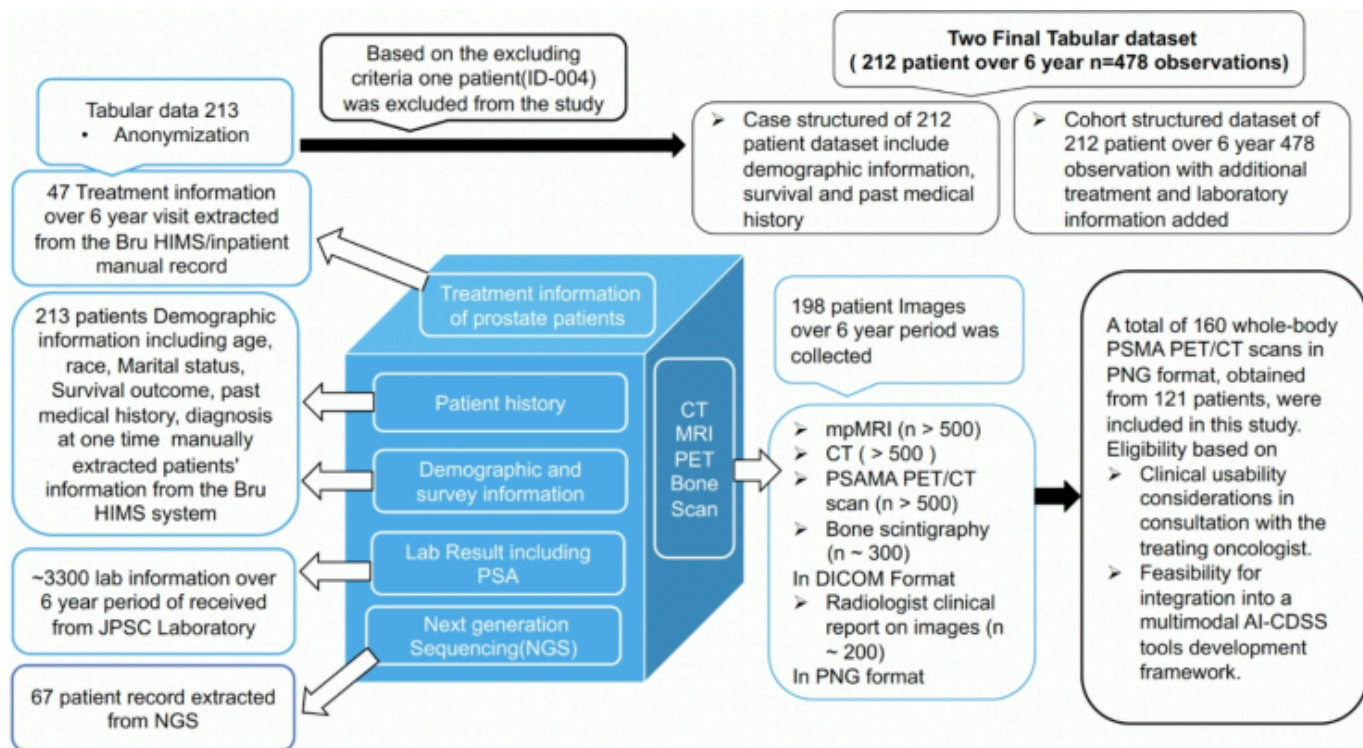


Figure 1

Data Flow Diagram of the Study Cohort with selected sources

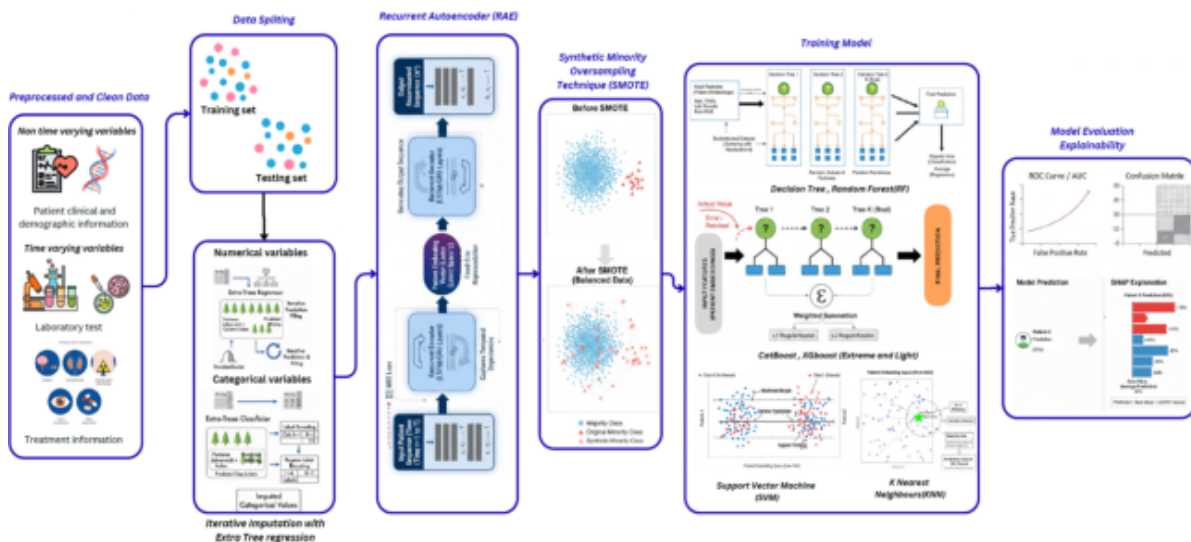


Figure 2

Graphical Demonstration for Longitudinal Representation Learning via Autoencoders with Synthetic Minority Oversampling Technique (SMOTE) for a Supervised Machine Learning Pipeline for Prostate Cancer Treatment Selection

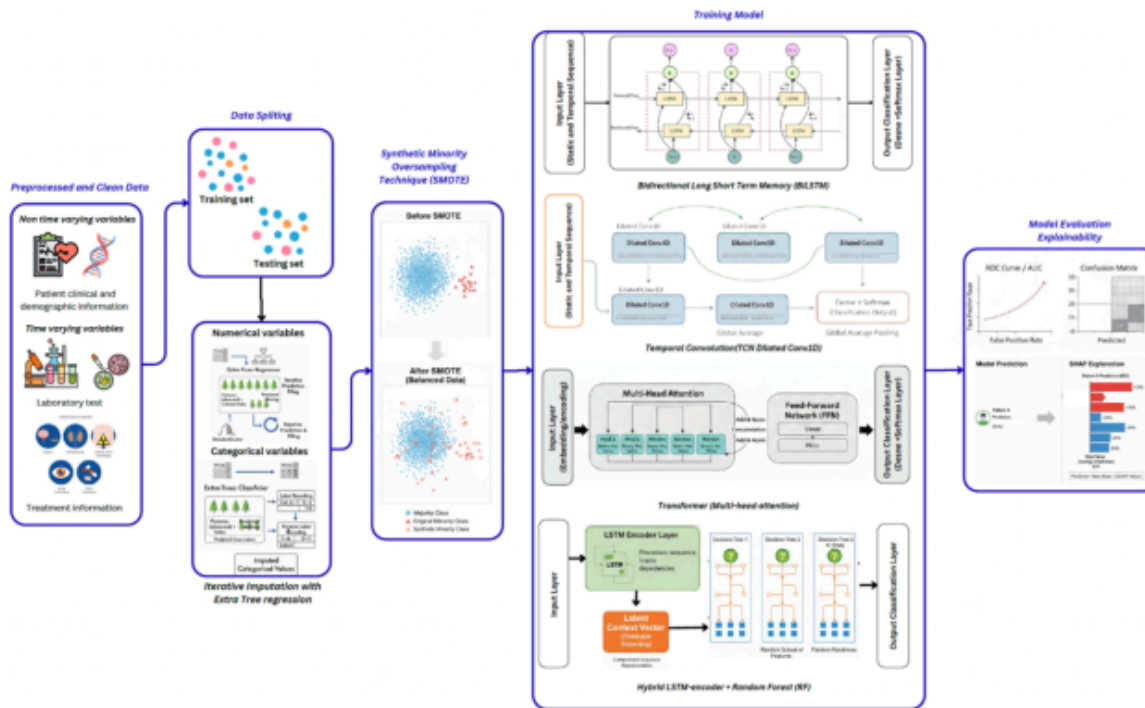


Figure 3

Graphical Representation for Time-Series-Aware Supervised Machine Learning Framework for Prostate Cancer Treatment Selection

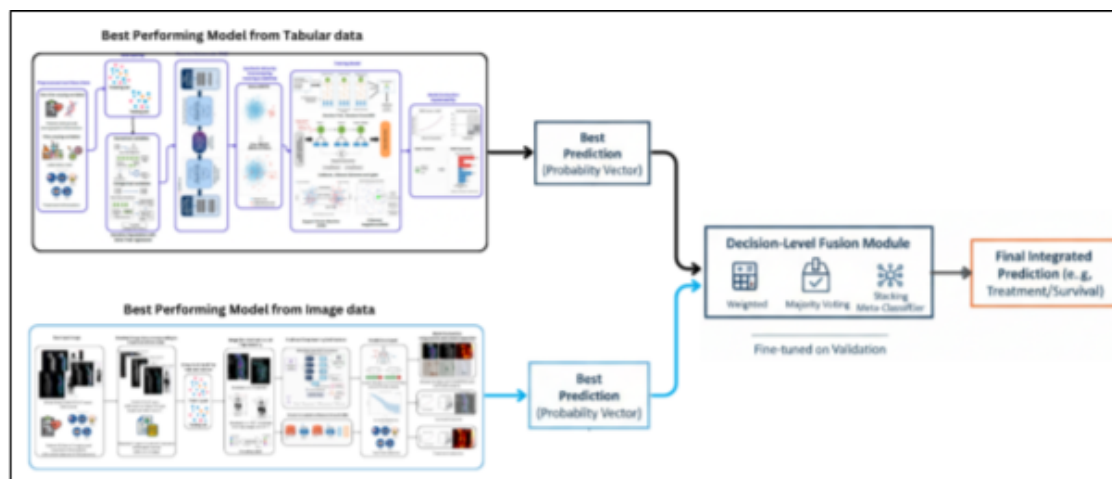


Figure 4

Schematic of the late-stage fusion pipeline utilised for the best-performing multimodal data integration, combining optimised tabular and imaging data

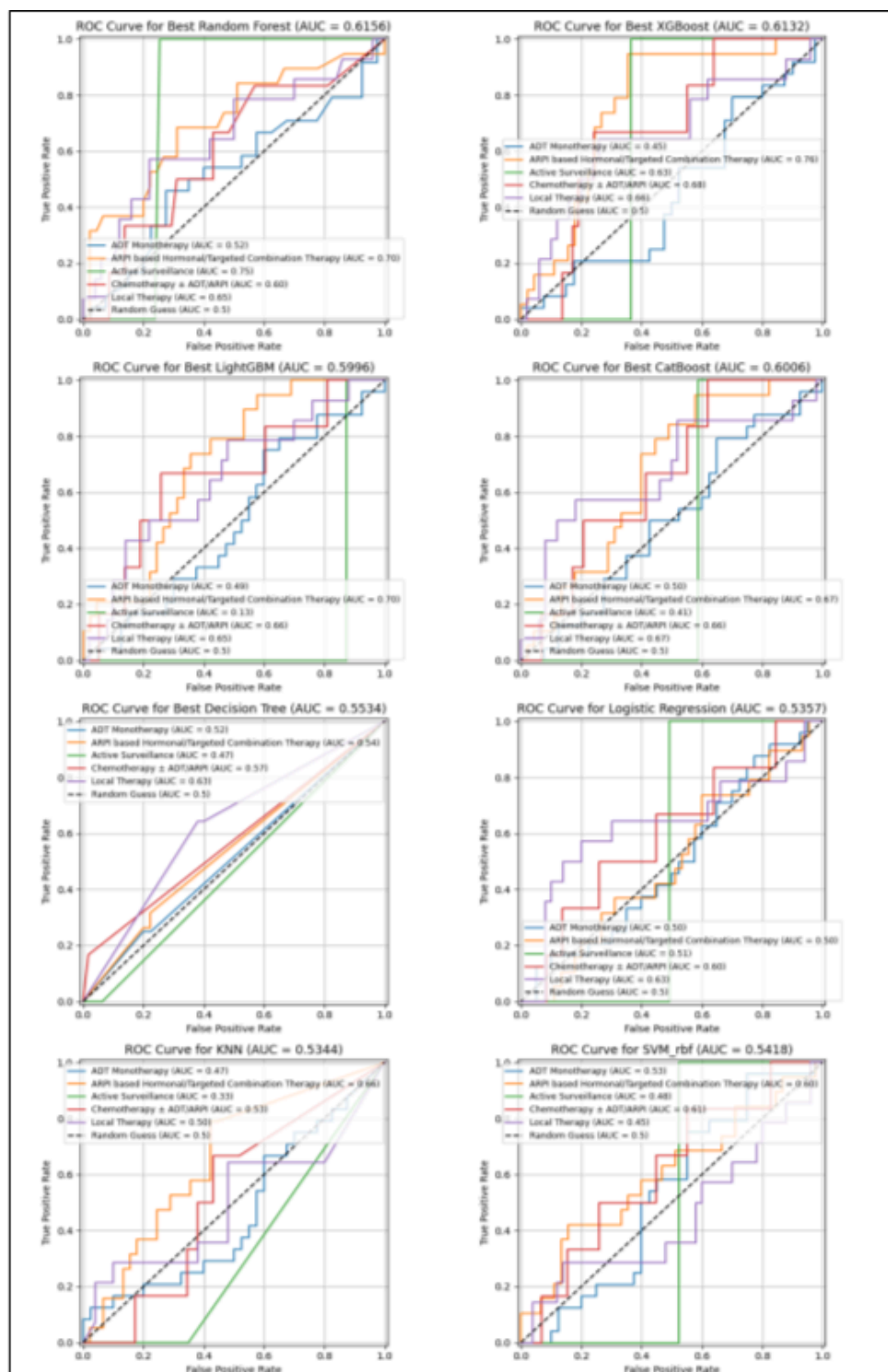


Figure 5

Receiver operating characteristic (ROC) curves for supervised machine learning models following Longitudinal Representation Learning with SMOTE application for predicting prostate cancer treatment categories for Brunei Cancer Centre (TBCC) from Jan 18 to Dec 24

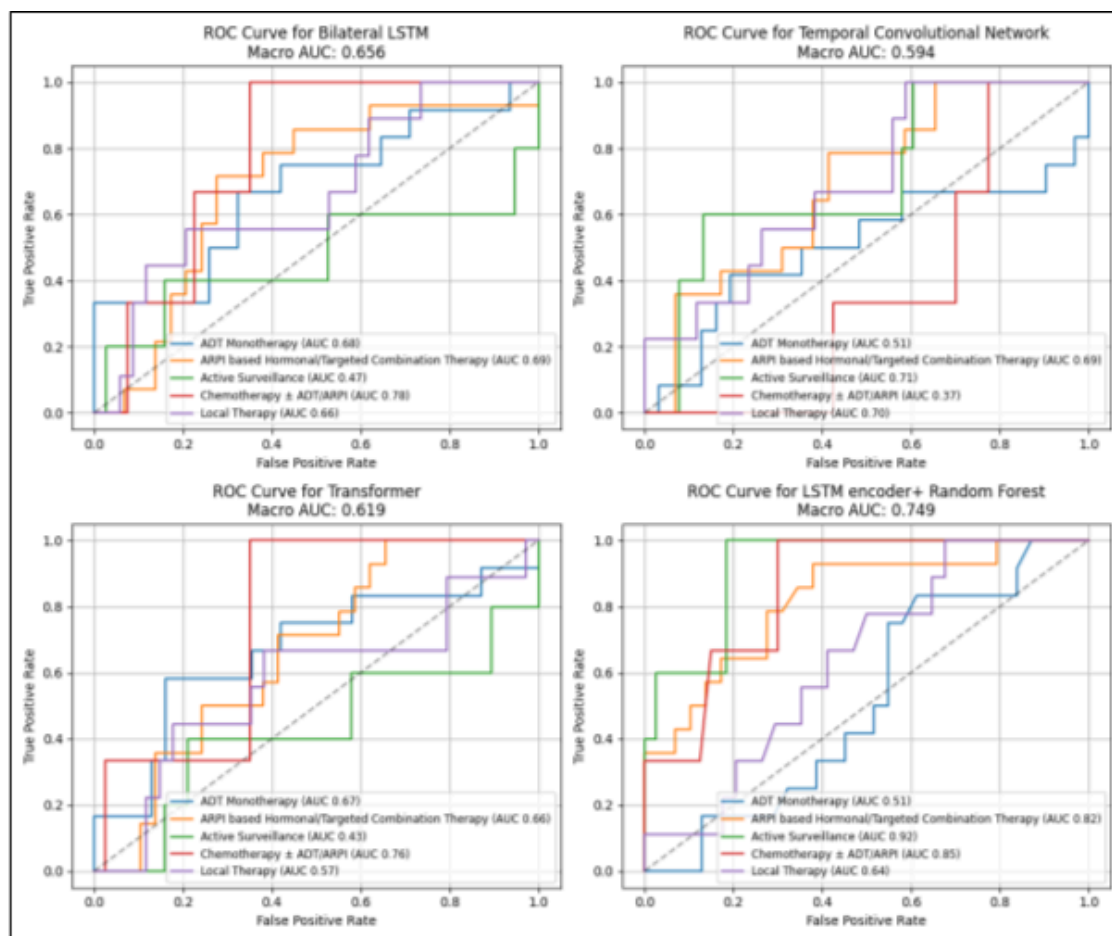


Figure 6

Class-specific Receiver operating characteristic (ROC) curves across prostate cancer treatment categories for different deep neural network-based architectures with SMOTE for supervised machine learning models

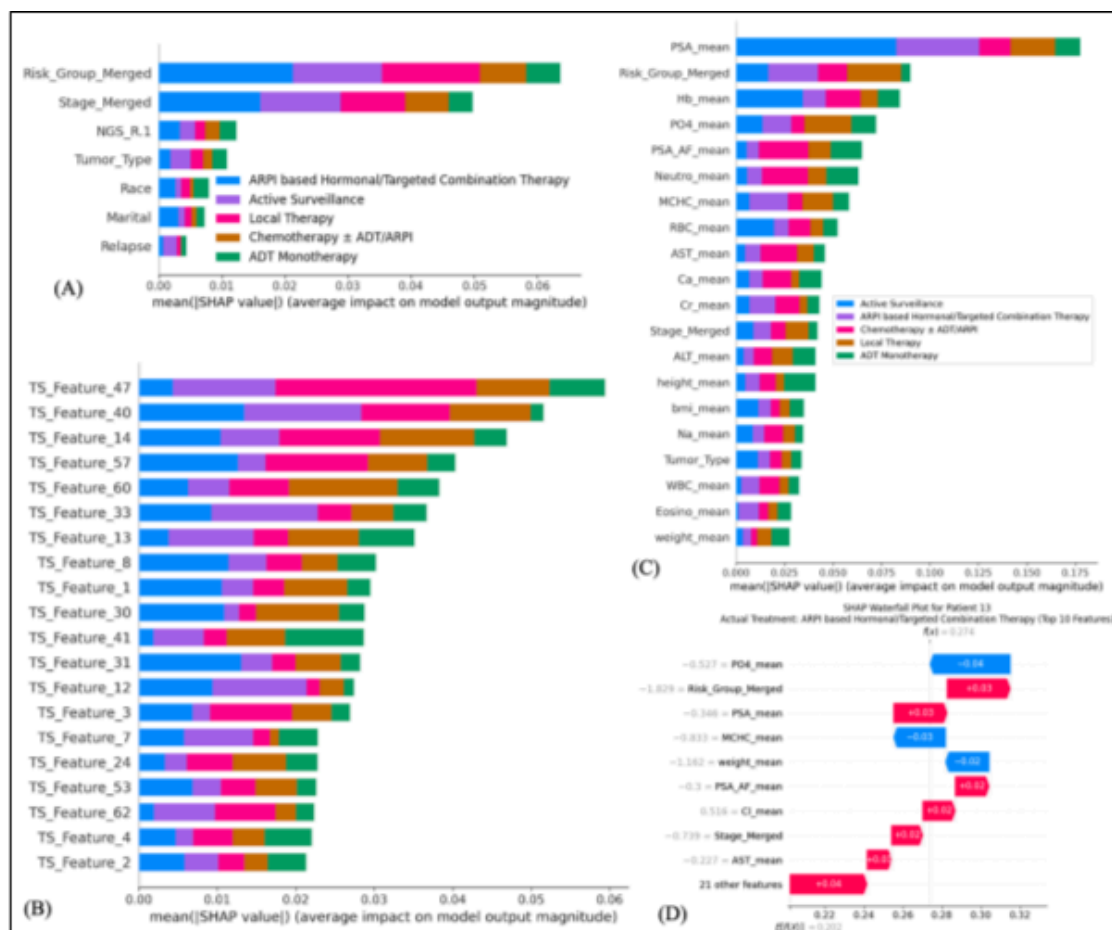


Figure 7

SHAP-Based Interpretation of the Hybrid LSTM Encoder and Random Forest Model: A) Bar plot on Static Variables, B) Bar plot on Time varying variables as Time series features, C) Bar plot on aggregated features using mean for demonstrating time varying variable in model, D) SHAP Waterfall plot for random patient (index 13) top 10 features Prediction Explanation.