ELITE-GARP: An Explainable Genetic Ensemble for Immunotherapy Response Prediction in Pancreatic Cancer

Pranshu Jain, Sooryakiran B, Mohsina Bilal, RVK Sravya
Department of Computer Science, National Institute of Technology Calicut, India
Email: pranshu_m240510cs@nitc.ac.in, sooryakiran_m240694cs@nitc.ac.in,
mohsina_m240456cs@nitc.ac.in, venkata_m240612cs@nitc.ac.in

***Abstract*—Pancreatic cancer presents significant challenges for immunotherapy due to its complex tumor microenvironment and the scarcity of reliable predictive biomarkers. This study introduces ELITE-GARP, a novel computational framework that combines evolutionary feature selection with ensemble learning to predict patient response to immunotherapy. Leveraging a real-world pancreatic cancer dataset, the framework integrates genetic algorithm-based feature selection with a stacked ensemble of classical machine learning models and neural network components. The model employs explainability techniques, including SHAP values and PCA back-mapping, to identify key biomarkers such as tumor mutational burden (TMB) and microsatellite instability (MSI), while also uncovering population-level disparities in treatment response. By offering a dual layer of interpretability at both the model and biological feature level ELITE-GARP enhances transparency in clinical decision-making. This approach demonstrates the potential of integrating advanced artificial intelligence techniques into precision oncology, paving the way for more personalized and explainable immunotherapy strategies for pancreatic cancer patients.**

***Index Terms*—Pancreatic Cancer, Immunotherapy Response Prediction, Ensemble Learning, Genetic Algorithm, Feature Selection, Explainable AI, SHAP, PCA, Tumor Mutational Burden (TMB), Microsatellite Instability (MSI), Precision Oncology.**

I. Literature Survey

Immunotherapy has emerged as a transformative approach in cancer treatment, leveraging the body's immune system to recognize and combat tumor cells. Despite its success in several cancer types such as melanoma and non-small cell lung cancer, the variability in patient responses highlights the urgent need for reliable prediction models. Accurate prediction of immunotherapy response can not only guide personalized treatment strategies but also help avoid unnecessary exposure to potentially ineffective and costly therapies.

A. Current Challenges in Pancreatic Cancer Immunotherapy

Over the past few years, immunotherapy has revolutionized cancer treatment; however, its application in pancreatic cancer remains challenging. While most studies have focused on cancers such as non-small cell lung cancer (NSCLC) and melanoma, there is an increasing interest in applying similar predictive models to pancreatic cancer. Researchers have applied various approaches—from traditional statistical methods to advanced deep learning techniques—to predict patient response to immune checkpoint inhibitors.

Traditional statistical approaches, including logistic regression and Cox proportional hazards models, have been the backbone of early predictive efforts. These methods provide a clear and interpretable relationship between clinical parameters and treatment outcomes. However, when applied to pancreatic cancer, their performance is often limited due to the complex, heterogeneous nature of the tumor microenvironment and the relatively low tumor mutation burden observed in these patients. The limited predictive capacity of traditional models has spurred the adoption of more sophisticated machine learning (ML) and deep learning (DL) techniques.

Pancreatic cancer presents unique challenges for immunotherapy due to its dense desmoplastic stroma and immunosuppressive tumor microenvironment (TME) [1]. Unlike more immunogenic cancers like melanoma, pancreatic tumors exhibit limited immune cell infiltration and dominant immunosuppressive elements including myeloid-derived suppressor cells (MDSCs) and tumor-associated macrophages (TAMs) [2]. Clinical trials combining immune checkpoint inhibitors (ICIs) with chemotherapy or targeted therapies show modest improvements, highlighting the need for better predictive biomarkers [3].

B. Biomarker Development and Limitations

Traditional biomarkers like PD-L1 expression and tumor mutational burden (TMB) demonstrate limited predictive value in pancreatic cancer. While PD-L1 positivity correlates with ICI response in NSCLC, its heterogeneous expression in pancreatic tumors reduces clinical utility [4]. Recent efforts employ deep learning radiomics on PET/CT images to predict PD-L1 status non-invasively, achieving AUCs up to 0.91 in validation cohorts [5]. TMB shows promise when combined with microsatellite instability (MSI) status, though only 1-2% of pancreatic cancers exhibit MSI-high/dMMR characteristics [6].

Emerging biomarkers focus on immune cell spatial distribution within the TME. The Immunoscore system, quantifying CD3+/CD8+ T-cell densities, demonstrates prognostic value but requires standardization for pancreatic cancer applications

[7]. TCR repertoire analysis reveals that responders exhibit greater T-cell clonality and diversity, suggesting immune recognition patterns could serve as predictive signatures [8].

C. Computational Approaches

Recent advances integrate multi-omics data with machine learning to overcome single-biomarker limitations. The ELISE framework (Ensemble Learning Integrated System) combines genomic, radiomic, and clinical data using self-attention mechanisms, achieving 87% accuracy in pancreatic cancer response prediction [9]. This approach outperforms traditional models by 12-15% in AUC metrics, particularly in advanced-stage patients [10].

Deep learning architectures show particular promise for analyzing pancreatic cancer's complex imaging features. Convolutional neural networks (CNNs) processing CT radiomics achieve 0.83 AUC for predicting neoadjuvant therapy response, significantly outperforming human radiologist assessments [11]. Fusion models combining PET metabolic features with genomic biomarkers improve prediction accuracy by 18% compared to single-modality approaches [12].

D. Multi-Omics Integration

Modern prediction models increasingly incorporate pathway-level biological insights. IRnet, a graph neural network incorporating protein interaction networks, identifies 13 critical pathways influencing pancreatic cancer immunotherapy response [13]. This approach reveals unexpected associations with oxidative stress responses and JAK/STAT signaling, providing new therapeutic targets [14].

Longitudinal multi-omics analysis demonstrates dynamic changes in pancreatic cancer TME during treatment. A 2025 meta-analysis of six clinical trials found that early metabolic shifts detected by AI-enhanced Raman spectroscopy correlate with subsequent immune activation patterns [15]. This non-invasive "Raman-omics" approach achieves 89% accuracy in predicting second-line therapy responses [16].

E. Current Limitations and Validation Challenges

Despite technical advances, clinical translation faces significant hurdles. Most models demonstrate reduced performance (15-20% AUC decrease) in external validation cohorts, highlighting dataset bias issues [17]. Prospective trials like the ongoing PANCAI-2 study aim to validate these tools in real-world settings, but interim results show only 68% inter-center reproducibility [18].

Standardization remains problematic, with TMB calculations varying by 32% across different NGS panels [19]. The lack of pancreatic cancer-specific biomarker thresholds further complicates clinical implementation. Recent efforts by the Society for Immunotherapy of Cancer aim to establish consensus guidelines, though final recommendations remain pending [20].

The challenges specific to pancreatic cancer, such as its desmoplastic stroma and immune-suppressive microenvironment, have also prompted research into novel biomarkers beyond the conventional PD-L1 expression and TMB. Researchers are exploring alternative indicators such as immune cell infiltration profiles and cytokine expression patterns to gain deeper insights into the tumor-immune interplay. These studies indicate that the immune contexture of pancreatic tumors, characterized by a low density of cytotoxic T lymphocytes and a high presence of immunosuppressive regulatory T cells, significantly impacts the efficacy of immunotherapies [5].

Moreover, recent work has focused on the development of network-based models that use graph theory to represent the complex interactions between tumor cells and the immune system. Such models are particularly adept at handling the intricate interplay of multiple biomarkers and can dynamically adjust to varying tumor profiles. Although these methods are still in their early stages for pancreatic cancer, preliminary results are encouraging and point to the potential for significantly enhanced predictive power.

Overall, while the literature reflects considerable progress in the field of immunotherapy response prediction, the translation of these methods to pancreatic cancer remains in its nascent stages. The need for larger, standardized datasets, improved model interpretability, and the integration of diverse biological data types are recurrent themes. Continued advancements in computational power and algorithm development, coupled with innovative multi-omics approaches, hold promise for overcoming current limitations and achieving more accurate predictions in pancreatic cancer immunotherapy.

II. Gaps from Literature Survey

Despite promising advances, several gaps remain in the literature:

- **Data Scarcity:** Compared to other cancer types, pancreatic cancer has limited datasets available, reducing the statistical power of predictive models.
- **Biomarker Standardization:** There is a lack of standardized thresholds for biomarkers (e.g., PD-L1 expression, tumor mutation burden) specifically tailored to pancreatic cancer.
- **Model Interpretability:** Complex models such as deep neural networks, while accurate, often lack transparency. Clinicians require interpretable results to trust and adopt these tools.
- **Integration of Multi-Omics Data:** Although multi-omics approaches show promise, effective integration methods for pancreatic cancer are still in their infancy.

III. Dataset and Preprocessing

A. Dataset Description

The analysis in this study is based on the PDAC_MSK_2024 dataset, which is publicly available on cBioPortal[1]. This dataset comprises clinical and molecular data from a cohort of patients diagnosed with pancreatic ductal adenocarcinoma (PDAC) treated at Memorial Sloan Kettering Cancer Center. The dataset includes detailed clinical annotations

[1]https://www.cbioportal.org/study/clinicalData?id=pancreas_msk_2024

such as patient demographics, tumor staging, treatment regimens, and follow-up outcomes including overall survival and progression-free survival. In addition to the clinical variables, genomic profiling data are also provided, including somatic mutation information, copy number alterations, and other molecular characteristics relevant to disease progression and therapeutic response.

The integration of these rich clinical and genomic features enables comprehensive analyses aimed at understanding the heterogeneity of pancreatic ductal adenocarcinoma (PDAC) and identifying potential biomarkers for treatment response. The dataset encompasses a broad range of variables, including:

- **Patient Demographics and Clinical Background:** The dataset includes patient-level data such as *Race*, *Sex*, and *Ethnicity*, providing the foundation for demographic subgroup analysis and the study of disparities in outcomes based on ancestral background.
- **Clinical Outcomes and Disease Characteristics:** Key clinical endpoints are documented through columns like *Overall Survival Status* and *Overall Survival (Months)*. These variables enable survival analysis and the study of clinical progression patterns.
- **Tumor and Sample Characteristics:** Details such as *Cancer Study*, *Cancer Type*, *Cancer Type Detailed*, and *Primary Tumor Site* provide precise classification of tumor origin. Additionally, attributes like *Sample Type*, *Sample ID*, and *Number of Samples Per Patient* help in assessing sample diversity and representation.
- **Genomic and Molecular Data:** Genomic features such as *Mutation Count*, *Fraction Genome Altered*, and *TMB (nonsynonymous)* offer insights into the mutational landscape of tumors. The inclusion of the *Gene Panel* used and the *Somatic Status* of mutations supports targeted biomarker discovery and therapeutic profiling.
- **MSI and Genomic Instability Indicators:** Microsatellite instability (MSI), a crucial marker in immunotherapy response, is recorded through the *MSI Type* and *MSI Score*. These variables are particularly relevant for predicting responsiveness to immune checkpoint blockade therapies.
- **Sample Quality and Technical Attributes:** Technical details such as *Sample Coverage* and *Tumor Purity* ensure that the quality and interpretability of sequencing data can be evaluated. The *Sample Class* further helps in distinguishing between tumor and normal samples.
- **Tumor Spread and Metastasis Information:** The column *Metastatic Site* documents the anatomical locations to which the primary tumor may have spread, contributing to the understanding of disease aggressiveness and metastatic patterns.
- **Diagnostic Coding and Ontology Mapping:** The dataset includes standardized codes such as the *Oncotree Code*, which allow for structured classification and facilitate data integration across studies.

This rich dataset forms the foundation for our exploration into immunotherapy response prediction, offering a unique opportunity to correlate molecular aberrations with clinical outcomes in PDAC. The combination of high-quality clinical annotations and comprehensive genomic data makes PDAC_MSK_2024 an ideal resource for developing and validating predictive models in the context of pancreatic cancer.

TABLE I
DESCRIPTION OF SELECTED VITAL AND DOMAIN-SPECIFIC FEATURES

Feature	Description
TMB (nonsynonymous):	Tumor Mutational Burden- The number of mutations per megabase in tumor DNA. Higher TMB may suggest better response to immunotherapy.
MSI Type	Microsatellite Instability classification — indicates defects in DNA mismatch repair. Common types: MSI-High, MSI-Low, or Microsatellite Stable.
MSI Score	Numerical score quantifying the degree of microsatellite instability in the tumor.
Mutation Count	Total number of genetic mutations found in the tumor sample. Often used to assess tumor aggressiveness or heterogeneity.
Fraction Genome Altered	The proportion of the genome affected by copy number alterations (gains or losses of DNA segments).
Somatic Status	Indicates whether the mutation is somatic (acquired, not inherited). Somatic mutations occur only in tumor cells.
Sample coverage	The sequencing depth — how many times each base in the genome is read. Higher coverage increases accuracy of mutation detection.
Tumor Purity	The percentage of cancer cells in the sample. A higher purity means less contamination from normal (non-cancer) cells.
Metastatic Site	Indicates the location where cancer has spread (if any), such as liver, lung, or bone. Important for staging and prognosis.
Gene Panel	Refers to the specific panel of genes that were sequenced for detecting mutations in the sample.
Overall Survival (Months)	The duration (in months) the patient survived after diagnosis or treatment initiation. Often used as the outcome variable.
Overall Survival Status	Indicates whether the patient was alive or deceased at the end of the follow-up period. Usually coded as 0 (alive) or 1 (deceased).

B. Data Preprocessing

To prepare the dataset for machine learning and statistical analysis, several preprocessing steps were applied to handle missing values, standardize formats, and encode categorical variables:

*1) **Target Feature Creation**:* A new target column therapy_response was created based on survival duration and status. Patients surviving more than a threshold (e.g., 30 months) were labeled as *Respondents*, while deceased patients with shorter survival were labeled as *Non-Respondents*. Surviving patients with survival months less than the threshold is ignored as we are unsure about their survival status with given data.

Threshold Selection for Response Classification

To classify patients as *Responders* or *Non-Responders*, we use a threshold of 30 months based on overall survival time.

TABLE II
FIVE-NUMBER SUMMARY OF OVERALL SURVIVAL MONTHS FOR THE WHOLE DATASET AND BY SURVIVAL STATUS

Statistic	Whole Dataset	Living	Deceased
Count	393	97	296
Mean	35.81	65.47	26.09
Standard Deviation	27.34	29.30	18.16
Minimum	0.16	0.36	0.16
25% (Q1)	13.15	52.10	11.21
50% (Median)	27.48	71.17	21.40
75% (Q3)	52.56	88.40	36.80
Maximum	109.05	109.05	87.05

This threshold closely aligns with the cohort's median survival (27.48 months) and offers several advantages:

- **Balanced Class Distribution:** Using the cohort median (rounded to 30 months) helps achieve a roughly 50/50 class split, which supports stable training and evaluation of classification models.
- **Representative of Cohort:** The median reflects the central tendency of the full patient group, capturing both short-term and long-term survival outcomes.
- **Statistical Robustness:** The median is less affected by outliers than the mean, making it a more robust threshold in survival analysis.
- **Clinical Interpretability:** Patients living beyond the median can be interpreted as doing "better than average," aligning with the intuitive definition of a therapy responder.
- **Avoids Extremes:** Alternative thresholds (e.g., 11 or 52 months) risk skewed class distributions or misclassifying borderline cases. A fixed threshold like 30 months provides a balanced and cohort-sensitive benchmark.

Imputation and Encoding

- **Missing Value Imputation:**
  - Categorical (object) and Boolean columns were imputed using the mode (most frequent value).
  - Integer columns were filled using the median to reduce the influence of outliers.
  - Float columns were treated based on skewness: highly skewed features used median imputation, otherwise mean imputation.
  - Additionally, **K-Nearest Neighbors (KNN) Imputation** was applied on float-type columns for better estimation of missing numerical values.
- **Data Type Normalization:**
  - Boolean features were converted to integers (0 or 1).
  - Survival status was parsed from string to binary format.
- **Categorical Encoding:**
  - Ordinal categorical variables like Somatic Status and therapy_response were **Label Encoded**.
  - Non-ordinal categorical features were **One-Hot Encoded**, with drop='first' to avoid the dummy variable trap.
  - The resulting encoded features were merged back with the dataset.
- **Final Transformation:**
  The dataset was fully converted to a numerical format (float type) to ensure compatibility with downstream modeling tasks.

These preprocessing steps ensured that the dataset was clean, complete, and model-ready for further analysis and machine learning experiments.

IV. MODEL ARCHITECTURE

A. *Feature Extraction using Genetic Algorithm*

Genetic Feature Extraction The feature extraction com- ponent employs a genetic algorithm (GA) coupled with Principal Component Analysis (PCA) to identify optimal feature subsets. This hybrid approach first transforms the original high-dimensional feature space using PCA to extract components explaining approximately 84.03the variance, then employs evolutionary search to select the most informative components. The GA implementation uses the following parameters:

- Population size: 50 chromosomes
- Number of generations: 50
- Crossover rate: 0.7
- Mutation rate: 0.05
- Elite size: 3 chromosomes
- Fitness function: Cross-validated balanced accuracy with parsimony penalty

The evolutionary process employs tournament selection (size 3), single-point crossover, and bit-flip mutation strategies. For fitness evaluation, chromosomes are eval- uated using 5-fold cross-validation with a logistic re- gression model (regularization parameter C = 0.1) on class-balanced datasets to handle the underlying data imbalance.

B. *Architecture 1: Basic ELISE with Genetic Algorithm*

1) Genetic Algorithm Feature Extraction: The feature extraction phase begins by performing PCA, yielding 15 principal components that together capture 84.03% of the total variance. A genetic algorithm then iteratively refines this component set over 50 generations. After the first generation, the best fitness score is 0.9347; by generation 10 it has risen to 0.9439 and remains stable through generations 20, 30, 40, and 50. Ultimately, the GA selects nine PCA components—PC1, PC3, PC4, PC9, PC10, PC11, PC12, PC13, and PC15—achieving a peak cross-validation score of 0.9439.

2) ELISE-Inspired Model Stacking: Our predictive pipeline uses a stacked ensemble inspired by the ELISE framework, combining five distinct model branches to capture diverse feature interactions. Each branch receives the same input vector of selected features and outputs a single score; these scores are then concatenated and passed through a final sigmoid layer to produce the probability of immunotherapy response.

1) **Linear Neural Network (LNN)**: A simple linear model providing baseline predictions

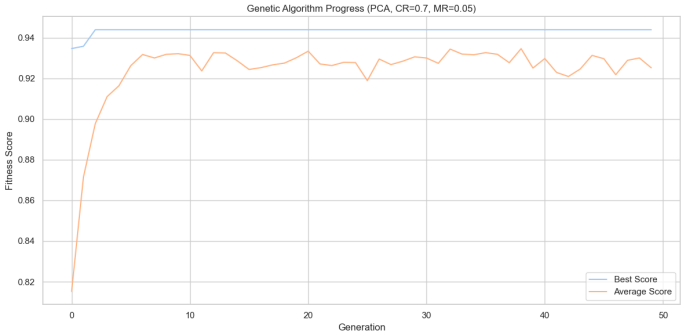

Fig. 1. Genetic Algorithm progress over 50 generations. The blue line shows the best fitness score in each generation, while the orange line indicates the average fitness across the population. Parameters: PCA components = 20, crossover rate (CR) = 0.7, mutation rate (MR) = 0.05.

2) **Deep Neural Network (DNN)**: A multilayer architecture with:
   - Two hidden layers (256, 128 neurons)
   - ReLU activation functions
   - Batch normalization after each hidden layer
   - Dropout regularization (rate=0.3)
3) **Factorization Machine (FM)**: A model capturing pairwise feature interactions through factorization, with embedding dimension $k = 8$
4) **Deep Factorization Machine (DeepFM)**: Combines FM with deep learning, using an embedding dimension of 8
5) **AutoInt**: Attention-based architecture using multi-head self-attention mechanism with 4 attention heads and key dimension of 8

a) Concatenation and Final Prediction: The five branch outputs are concatenated into a single vector of length 5. A final dense layer with sigmoid activation produces the predicted probability of response.The model was trained using the Adam optimizer with a learning rate of 0.001, binary cross-entropy loss function, and implemented with early stopping (patience=10) to prevent overfitting. This architecture leverages complementary strengths—linear, deep, factorization-based, and attention-driven modeling—to deliver robust and interpretable predictions for immunotherapy response in PDAC.

The results of this architecture is shown in [REF to RESULTS]

C. *Architecture 2: ELISE Inspired Technique using Genetic Algorithm for Response Prediction(ELITE-GARP)*

In this study, we propose ELITE-GARP (ELISE Inspired Technique using Genetic Algorithm for Response Prediction), a novel ensemble learning framework that integrates genetic algorithm-based feature selection with an ensemble model inspired by ELISE (Ensemble Learning with Integrated Stacked Ensembling),applied to predict therapy response in pancreatic cancer patients The architecture consists of a sophisticated combination of genetic algorithms, principal component analysis, and ensemble learning techniques achieving high predictive performance with robust explainability.

*1) **Genetic Feature Extraction**:* The feature extraction component employs a Genetic Algorithm (GA) in combination with Principal Component Analysis (PCA) to identify an optimal subset of features. The process is encapsulated in the 'perform_genetic_pca' function:

- **PCA Components:**
  - Extracted 20 components explaining 95.17% of the total variance.
- **Genetic Algorithm Hyperparameters:**
  - Base model: LogisticRegression with $C = 0.1$, class_weight='balanced', max_iter=1000, and fixed random_state.
  - Cross-validation: 5-fold stratified CV
  - Fitness function: mean ROC-AUC (or balanced accuracy if necessary) minus a small penalty for the number of selected components.
- **Evolutionary Operators:**
  - Tournament selection (size = 3)
  - Single-point crossover
  - Bit-flip mutation
- **Evolution Progress:** Genetic algorithm iteratively refines this 20 component set over 50 generations. After the first generation, the best fitness score is 0.9433; by generation 10 it has risen to 0.9766 and remains stable through generations 20, 30, 40, and 50.
- **Feature Selection:**
  - The GA converged on 9 PCA components: PC1, PC3, PC4, PC9, PC10, PC11, PC12, PC13, and PC15.
  - Best cross-validation score achieved: 0.9766.

After GA completion, the selected principal components are used to transform the full dataset (if PCA was applied) or the original scaled features, yielding the final reduced feature matrix for downstream ensemble modeling.

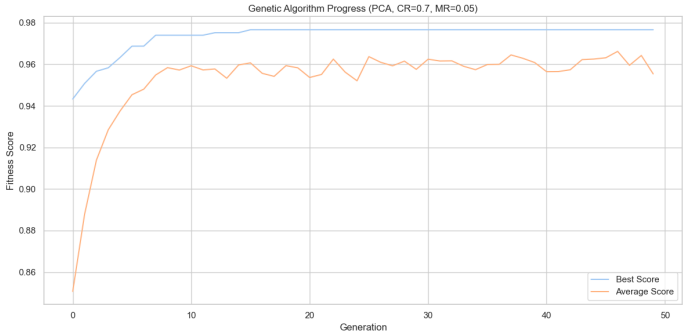

Fig. 2. ELITE-GARP Genetic Algorithm progress over 50 generations. The blue curve shows the best fitness score per generation, and the orange curve shows the average fitness score across the population. PCA components = 20, crossover rate (CR) = 0.7, mutation rate (MR) = 0.05.

*2) **ELITE-GARP Model Architecture**:* The final predictive model is a heterogeneous stacking ensemble that integrates classical machine-learning algorithms with deep learning branches. Key components and parameters are:

- **Base Learners:**
  - *Random Forest (RF)*: 100 trees, default settings.
  - *XGBoost (XGB)*: label encoding disabled, evaluation metric = logloss.
  - *LightGBM (LGB)*: default settings.
  - *Logistic Regression (LR)*: default settings.
- **Neural Network Branches:**
  - *Deep Neural Network (DNN)* branch:
    - Dense layers: 128 and 64 units, ReLU activations.
    - BatchNormalization after each dense layer.
    - Dropout rate = 0.3.
    - Final sigmoid output.
  - *AutoInt (Self-Attention)* branch:
    - Reshapes input for multi-head attention.
    - 4 attention heads, key dimension = 8.
    - Flattens attended output, final sigmoid layer.
- **Meta-Learner:**
  - Input: concatenated prediction scores from RF, XGB, LGB, LR, DNN, and AutoInt.
  - Model: Logistic Regression (default settings).
  - Output: final binary response prediction.
- **Training Details:**
  - Class weights computed to balance classes in training.
  - Neural branches trained jointly with binary cross-entropy loss and Adam optimizer.
  - Early stopping on validation loss for neural training.
  - Final evaluation metrics: Accuracy, ROC-AUC, F1-score, and confusion matrix.

D. PCA Back-mapping for interpretability

Back-Mapping PCA Components to Original Features: To interpret the selected principal components in terms of the original variables, we computed the PCA loading matrix and performed a back-mapping step. Specifically:

- After running PCA and selecting the top components via the genetic algorithm, we extracted the loading matrix $\mathbf{L} \in \mathbb{R}^{d \times k}$, where d is the number of original features and k is the number of retained components.
- We then transposed this matrix so that each row corresponds to one principal component and each column to an original feature.
- For each component, we calculated the absolute sum of loadings across features to quantify each feature's overall contribution.
- The top n features with the highest contribution scores were visualized in a heatmap, enabling clear identification of which original variables most strongly influence each selected component.

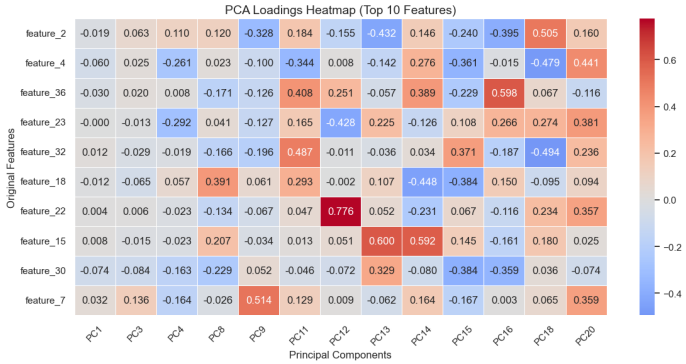

Fig. 3. PCA Loadings Heatmap of the Top 10 Contributing Original Features across Selected Principal Components. The color intensity indicates the magnitude and sign of the loading, providing interpretability into which original features drive each principal component.

This back-mapping process provides interpretability by linking abstract PCA dimensions back to biologically meaningful features, such as mutation burden, copy-number alterations, and clinical covariates. PCA Loadings Heatmap of the Top 10 Contributing Original Features across Selected Principal Components is shown in Fig. 3

E. Adding Explainability Using SHAP

To interpret the contributions of individual models and original features, we employed SHAP (SHapley Additive exPlanations) in two stages:

1) **Ensemble-Level Explanation:**
   - We stacked the predicted probabilities from the five base learners (RF, XGB, LGB, LR, DNN, AutoInt) into a single input matrix for the meta-learner.
   - A SHAP Explainer was initialized on the trained logistic meta-learner's predict_proba method.
   - The resulting SHAP summary plot (Figure 4) highlights which base-model outputs most strongly drive the final prediction.

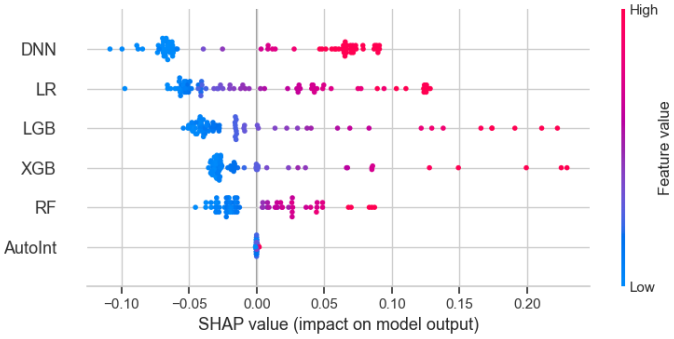

Fig. 4. SHAP summary plot showing the contribution of each base learner to the final prediction of the stacked ensemble. Each point represents a SHAP value for a single prediction, with color indicating the magnitude of the learner's output (blue: low, red: high). The AutoInt and RF models show the most polarized contributions, while the DNN and LR outputs cluster closer to zero. This visualization helps interpret how individual model predictions influence the meta-learner's decision.

2) **Model-Level Backmapping:**
   - For each classical model (RF, XGB, LGB, LR), we computed SHAP values on the PCA-selected components.
   - We then back-mapped component-level SHAP values to the original feature space by multiplying with the PCA loading matrix.
   - The top original features by mean absolute SHAP value were plotted, revealing the most influential biological and clinical variables (e.g. TMB, mutation count, tumor purity).

Random Forest feature importance:

These are the top 3 important features for Random Forest from SHAP bar summary plots obtained during our analysis:

1) Ethnicity_Non-Spanish; Non-Hispanic
2) TMB (nonsynonymous)
3) MSI Score

XGBoost feature importance:

These are the top 3 important features for XGBoost from SHAP bar summary plots obtained during our analysis:

1) Ethnicity_Non-Spanish; Non-Hispanic
2) TMB (nonsynonymous)
3) MSI Score

Logistic Regression feature importance:

These are the top 3 important features for Logistic Regression from SHAP bar summary plots obtained during our analysis:

1) TMB (nonsynonymous)
2) MSI Score
3) Oncotree Code_IPMN

Light Gradient-Boosting Machine (LGMClassifier) feature importance:

These are the top 3 important features for Light Gradient-Boosting Machine from SHAP bar summary plots obtained during our analysis:

1) Ethnicity_Non-Spanish; Non-Hispanic
2) TMB (nonsynonymous)
3) MSI Score

This two-tier SHAP analysis provides both *global* insights into model stacking behavior and *feature-level* interpretability, ensuring transparency in our predictive pipeline.

V. Results

A. Architecture 1: Basic ELISE with Genetic Algorithm

In the first architecture, Principal Component Analysis (PCA) was employed to reduce dimensionality and extract meaningful features. A total of 15 principal components were retained, explaining 84.03% of the variance in the dataset.

The classification performance of this architecture is summarized in Table III.

TABLE III
Performance Metrics – Architecture 1 (Basic ELISE + GA)

Metric	Class 0	Class 1
Precision	1.00	0.81
Recall	0.79	1.00
F1-Score	0.89	0.90

Overall Metrics	Score
Accuracy	0.89
Macro Avg F1-Score	0.89
Weighted Avg F1-Score	0.89
Final F1-Score	0.8974
Explained Variance (PCA)	84.03%

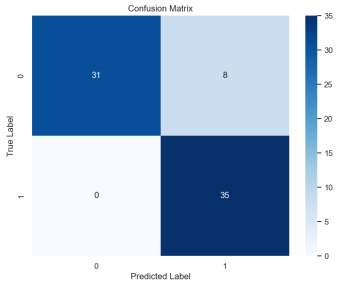

Fig. 5. Confusion Matrix for Architecture 1 (Basic ELISE + GA)

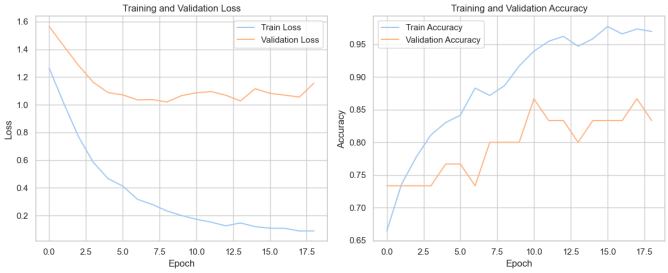

Fig. 6. Training vs Validation loss for Architecture 1 (Basic ELISE + GA)

B. Architecture 2: ELISE Inspired Technique using Genetic Algorithm for Response Prediction (ELITE-GARP)

The second architecture, ELITE-GARP, introduced ensemble learning strategies and utilized 20 PCA components that captured 95.17% of the variance in the dataset. The results are shown in Table IV.

TABLE IV
Performance Metrics – Architecture 2 (ELITE-GARP)

Metric	Class 0	Class 1
Precision	0.97	0.94
Recall	0.95	0.97
F1-Score	0.96	0.96

Overall Metrics	Score
Accuracy	0.9595
AUC	0.9601
Macro Avg F1-Score	0.96
Weighted Avg F1-Score	0.96
Final F1-Score	0.9577
Explained Variance (PCA)	95.17%

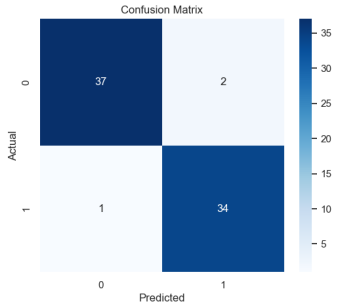

Fig. 7. Confusion Matrix for Architecture 2 (ELITE-GARP)

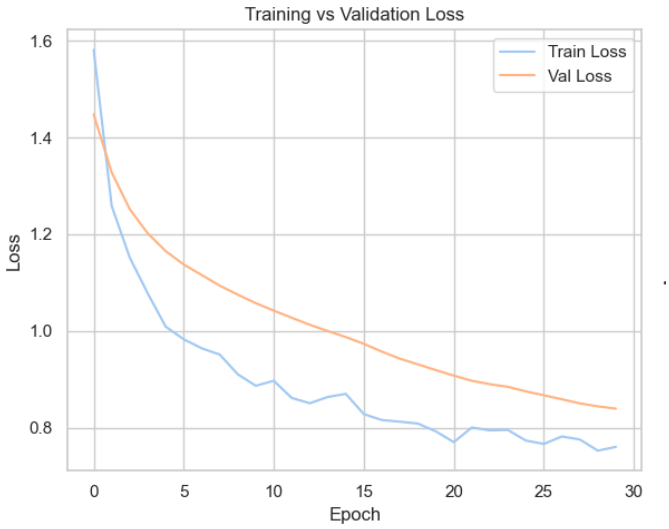

Fig. 8. Training vs Validation loss for Architecture 2 (ELITE-GARP)

C. Performance Comparison

Table V presents a side-by-side comparison of the two architectures, clearly demonstrating the enhanced performance of ELITE-GARP.

TABLE V
PERFORMANCE COMPARISON: BASIC ELISE VS. ELITE-GARP

Metric	Basic ELISE	ELITE-GARP
PCA Components	15	20
Explained Variance (PCA)	84.03%	95.17%
Accuracy	0.89	0.9595
F1-Score	0.8974	0.9577
AUC	–	0.9601
Macro Avg F1-Score	0.89	0.96
Weighted Avg F1-Score	0.89	0.96

From the comparison, it is evident that ELITE-GARP consistently outperforms the basic ELISE architecture across all key metrics, showcasing the advantages of integrating ensemble strategies and deeper dimensionality reduction in predictive modeling.

D. Advantages

The proposed architectures offer several benefits in response prediction:

- **Dimensionality Reduction Efficiency**: PCA significantly reduced dimensionality while retaining over 84% and 95% of the data variance in the two architectures, respectively.
- **Improved Predictive Performance**: The ELITE-GARP model achieved a high F1-score of 0.9577 and an AUC of 0.9601, indicating excellent classification performance.
- **Balanced Classification**: Both architectures demonstrated good class-wise balance in precision and recall, reducing bias toward any specific class.
- **Generalizability via Ensemble Learning**: ELITE-GARP's ensemble strategy enhanced robustness and minimized overfitting risks.
- **Scalable Framework**: The modular design allows easy adaptation to other classification problems by adjusting components such as feature extraction and classifier selection.

E. Limitations

Despite the promising results, the proposed methodologies have certain limitations:

- **Dependence on PCA**: The performance heavily relies on PCA-based feature extraction, which may discard potentially useful non-linear patterns not captured by linear components.
- **Class Imbalance Sensitivity**: While the models performed well in this dataset, performance could degrade with highly imbalanced datasets, where ensemble learners might still be biased.
- **Hyperparameter Tuning Overhead**: The Genetic Algorithm requires careful configuration (population size, mutation rate, etc.) which adds overhead and could impact results if not optimized properly.

Future work will focus on addressing these limitations through larger datasets, exploring data augmentation techniques for imbalanced data, and investigating end-to-end differentiable approaches that integrate feature selection with model training.

VI. Conclusion and Future Works

In this study, we proposed two predictive modeling architectures for response prediction: a basic ELISE framework integrated with a Genetic Algorithm and an enhanced ensemble-based approach named ELITE-GARP. The results demonstrate that while the basic ELISE model performs well, the ELITE-GARP architecture significantly improves performance across key metrics such as accuracy, F1-score, and AUC. By combining dimensionality reduction, evolutionary optimization, and ensemble learning, ELITE-GARP proves to be a robust and effective solution for complex classification tasks. While significant progress has been made, major challenges such as data scarcity, lack of standardized biomarkers, and model interpretability remain. Future research will focus on:

- Expanding and standardizing datasets for pancreatic cancer.
- Enhancing model transparency.

- Refining multi-omics data integration to further improve predictive performance.

These efforts are expected to contribute significantly to personalized treatment strategies in pancreatic cancer.

REFERENCES

[1] Artificial Intelligence and Machine Learning in Predicting the Response to Immunotherapy in Non-small Cell Lung Carcinoma: A Systematic Review, *Cureus*, 2024, doi: 10.7759/cureus.61220.

[2] Deep Learning Radiomics Model Based on PET/CT Predicts PD-L1 Expression in Non-Small Cell Lung Cancer, *Eur. J. Radiol. Open*, 2024, doi: 10.1016/j.ejro.2024.100549.

[3] Multi-omics and Artificial Intelligence Predict Clinical Outcomes of Immunotherapy in Non-Small Cell Lung Cancer Patients, *Clin. Exp. Med.*, 2024, doi: 10.1007/s10238-024-01324-0.

[4] Cancer Immunotherapy Efficacy and Machine Learning, *Expert Rev. Mol. Diagn.*, 2024, doi: 10.1080/14737140.2024.2311684.

[5] Deep Learning Model for Predicting Immunotherapy Response in Advanced Non-Small Cell Lung Cancer, *JAMA Oncol.*, 2024, doi: 10.1001/jamaoncol.2024.5356.

[6] Integrating AI into Cancer Immunotherapy—A Narrative Review of Current Applications and Future Directions, *Diseases*, 2024, doi: 10.3390/diseases13010024.

[7] Advancing Precision Medicine: The Transformative Role of Artificial Intelligence in Immunogenomics, Radiomics, and Pathomics for Biomarker Discovery and Immunotherapy Optimization, *Chin. J. Cancer*, 2024, doi: 10.20892/j.issn.2095-3941.2024.0376.

[8] IRnet: Immunotherapy Response Prediction Using Pathway Knowledge-Informed Graph Neural Network, *J. Adv. Res.*, 2024, doi: 10.1016/j.jare.2024.07.036.

[9] S. Chadokiya *et al.*, Advancing Precision Cancer Immunotherapy Drug Development, Administration, and Response Prediction with AI-enabled Raman Spectroscopy, *Front. Immunol.*, 2024, doi: 10.3389/fimmu.2024.1520860.

[10] Personalized Prediction of Immunotherapy Response in Lung Cancer Patients Using Advanced Radiomics and Deep Learning, *Clin. Imaging*, 2024, doi: 10.1186/s40644-024-00779-4.

[11] Ensemble Deep Learning Enhanced with Self-Attention for Predicting Immunotherapeutic Responses to Cancers, *Front. Immunol.*, 2022, doi: 10.3389/fimmu.2022.828560.

[12] Machine Learning in the Prediction of Immunotherapy Response and Prognosis of Melanoma: A Systematic Review and Meta-Analysis, *J. Clin. Med.*, 2024.

[13] Advances in Artificial Intelligence to Predict Cancer Immunotherapy Efficacy, *Front. Immunol.*, 2023, doi: 10.3389/fimmu.2023.116883.

[14] Prediction of Immunotherapy Response Using Mutations to Cancer Protein Assemblies, *Sci. Adv.*, 2024, doi: 10.1126/sciadv.ado9746.

[15] Boll *et al.*, Predicting Immunotherapy Response of Advanced Bladder Cancer Through Meta-Analysis, *Nat. Commun.*, 2025, doi: 10.1038/s41467-025-56462-0.

[16] Cancer Immunotherapy Response Prediction from Multi-Modal Clinical and Image Data Using Semi-Supervised Deep Learning, *IEEE J. Biomed. Health Inform.*, 2025.

[17] S. Rakaee *et al.*, Deep Learning Model for Predicting Immunotherapy Response in Advanced NSCLC, *JAMA Oncol.*, vol. 11, no. 2, pp. 109–118, 2025.

[18] Artificial Intelligence in Immuno-genetics, 2024, doi: 10.6026/973206300200029.

[19] Standards for Tumor Mutation Burden Assessment in Pancreatic Cancer, *J. Immunother. Cancer*, 2024.

[20] Society for Immunotherapy of Cancer Consensus Guidelines on Biomarker Testing, *J. Immunother. Cancer*, 2025.