Article

Not peer-reviewed version

Predicting Melanoma Immunotherapy Efficacy: Neural Network Models with Gene Expression and Clinical Data

Praveena Sunkara *

Posted Date: 17 May 2024

doi: 10.20944/preprints202405.1160.v1

Keywords: artificial neural networks; immune checkpoint blockade; overall survival; progression-free survival

 doi:10.20944/preprints202405.1160.v1

Article

Predicting Melanoma Immunotherapy Efficacy: Neural Network Models with Gene Expression and Clinical Data

Praveena Sunkara

Passion Health Primary Care, Dallas, USA; psunkara@passionhealthphysicians.com

Abstract: Immunotherapy, particularly Immune Checkpoint Blockade (ICB), has demonstrated significant efficacy in treating melanoma in recent years. However, accurately predicting treatment success and avoiding ineffective therapies remains an unresolved challenge. Therefore, this study aims to develop statistical models utilizing neural networks to forecast the effectiveness of immune checkpoint therapies for melanoma patients. Our models primarily rely on Artificial Neural Networks (ANN) to anticipate both Overall Survival (OS) and Progression-Free Survival (PFS) among melanoma patients undergoing anti-CTLA4 and anti-PD1/anti-PDL1 therapy. We incorporate gene expression data, measured in Transcripts per Million (TPM), derived from bulk tumor RNA-sequencing datasets. Additionally, clinical variables such as gender, age, and treatment type are factored into our analysis. The ANN underwent optimization to attain the highest feasible precision in anticipating the predetermined survival outcome. Issues stemming from high-dimensional data, such as overfitting, were tackled through regularization and feature selection methods. Consequently, the ANN-based model incorporating feature selection exhibited the capacity to forecast survival (PFS) in response to ICB therapy with a maximal precision of 86%. Conversely, the ANN lacking feature selection but incorporating regularization achieved accuracies of up to 72% for PFS and 71% for OS, correspondingly. In order to confront the challenge posed by limited patient samples and to assess replicability, the model underwent training and validation based on the amalgamation of all five datasets. However, this amalgamation failed to enhance predictive performance, necessitating further investigations.

Keywords: artificial neural networks; immune checkpoint blockade; overall survival; progression-free survival

1. Introduction

Recent research has made significant advancements in treating melanoma through Immune Checkpoint Blockade (ICB). Within cancer cells, checkpoints regulate the strength of the immune response, preventing excessive attack by T cells. Blocking these checkpoints with specific antibodies stimulates the immune system's response against cancer cells. Therapies targeting checkpoint molecules such as Cytotoxic T Lymphocyte Antigen 4 (CTLA4), Programmed Cell Death 1 (PD1), and PD1 ligand 1 (PDL1) have shown early success in clinical trials. Approved treatments now include anti-PD1 drugs like nivolumab or pembrolizumab, and anti-CTLA4 therapy like ipilimumab, either alone or in combination. However, not all patients benefit from these treatments, prompting extensive research into the molecular mechanisms of resistance to ICB. In addition to tumor-intrinsic and tumor-extrinsic biomarkers, patient genomic characteristics are also considered when making treatment decisions. Studies based on Ribonucleic Acid (RNA) analysis have identified gene expression patterns associated with immune cell infiltration within the tumor microenvironment, which in turn correlates with patients' overall survival. Specifically, expressions of Human Leukocyte Antigens (HLA)-A and Cluster of Differentiation (CD)8A show promise as potential targets for CTLA4-directed therapies. Some studies suggest that neoantigens and viral presence can reveal

 doi:10.20944/preprints202405.1160.v1

mutations enabling tumors to resist immune attacks, although their utility as predictive markers remains uncertain. Furthermore, prior failure of Mitogen-Activated Protein Kinase (MAPK)-targeted therapy negatively impacts subsequent responses to ICB in melanoma patients.

Recent findings implicate that Deoxyribonucleic Acid (DNA)-and RNA-level genomic information has predictive value for the clinical benefit of both anti-CTLA4 and anti-PD1 therapy. For example, while having many genetic mutations might indicate longer overall survival, it doesn't always mean a better response to anti-PD1 therapy. Therefore, there's a need to explore other genetic and non-genetic factors influencing treatment outcomes. To address this, extensive analysis of large patient groups is necessary. Finding ways to predict treatment response has been challenging due to the complexity of the immune system's response, differences in tumor environments, and the vast amount of data generated. To handle this data, sophisticated and automated techniques are required. Researchers have developed statistical models, like those by [1], which consider associations between genetic and transcriptomic features and treatment response. Because transcriptomic data often contains many interrelated variables, these models must account for nonlinear relationships. Recent studies have shown that Deep Neural Networks (DNNs) [2–4] can effectively model these nonlinear relationships, leading to better predictions of cancer therapy response. Additionally, neural network-based approaches have demonstrated superior performance in classifying gene expression samples such as [5].

Despite the growth of Deep Learning (DL) methods, their application in predicting immunotherapy response remains limited. However, there's a growing need for such methods, especially in melanoma patients undergoing ICB therapy. Artificial Neural Networks (ANNs) serve as the backbone of these predictive models. Using next-generation sequencing data, we aim to predict individual responses to drug therapy. However, the high-dimensional nature of the data and the limited number of samples present challenges. Therefore, we follow established scientific approaches, including quantifying drug response, selecting molecular features, fitting DL models, and evaluating their performance. In our study, we focus on quantifying Progression-Free Survival (PFS) and Overall Survival (OS) as measures of drug response. We also consider other genomic and clinical features, such as tumor mutational burden and patient demographics. Moreover, we account for prior treatments, like MAPK inhibition, if available. By developing and evaluating DL models, we aim to provide clinicians with sophisticated tools for treatment decision-making. However, working with high-dimensional data poses several challenges. Overfitting, where models perform well on training data but poorly on new data, is a common issue [6–12]. To address this, we employ various techniques such as feature selection, regularization, and data augmentation. Additionally, the imbalance in class distribution within our datasets requires special attention, which we address using methods like Synthetic Minority Over-sampling TEchnique (SMOTE) [13–15]. Validating our results is crucial but hindered by limited data availability. Therefore, we emphasize the importance of data sharing and standardization. Despite these challenges, integrating data from multiple sources can enhance the predictive power of our models. While our focus is on PFS/OS, our approach can be adapted for other endpoints or treatment modalities. The objective of this study is to automate the prediction of immune checkpoint therapy success in melanoma patients using gene expression data alongside genomic and clinical features, with patients as the primary outcome variables, addressing high-dimensional data challenges and small sample sizes inherent in DL.

The study is as follows; similar papers are shown in the following section. The materials and methods are provided in Section III. The experimental analysis is carried out in Section IV. The discussion is given in Section V, after which we draw some conclusions and outline our plans for further research in Section VI.

2. Related Works

Recent advances in the treatment of melanoma have demonstrated significant advances in ICB. Laboratories in cancer cells regulate the immune system, which regulates the action of T cells against cancer cells. Monoclonal antibodies that interfere with these checkpoints have shown promise in the immune response against cancer cells. Remarkable success has been observed with antibodies

 doi:10.20944/preprints202405.1160.v1

targeting checkpoint molecules such as [16,17]. Approved therapies include anti-PD1 (nivolumab or pembrolizumab) [18] and anti-CTLA4 (ipilimumab) [16], individually or in combination [19]. However, the efficacy of these treatments varies between patients, prompting extensive research into the molecular mechanisms of ICB resistance [1]. In addition to tumor intrinsic and extrinsic biomarkers, genomic profiling of patients is being investigated for treatment stratification such as [20]. RNA-based studies have identified gene expression signatures of immune penetration within the tumor microenvironment, which correlate with patient survival such as [21,22]. Whereas, this study highlights the expression of their potential CTLA4-targeted antibodies, such as HLA-A and CD8A. However, challenges remain in the search for predictive responses due to the complexity of the immune system and heterogeneous tumor states. Addressing these challenges requires sophisticated research approaches, with predictive models more flexible [1] and deeper tissues [23]. In response prediction modeling complex relationships have shown promise, but their role in ICB treatment of melanoma has been largely unexplored [5]. Given the demand for immunotherapy response prediction methods, especially for melanoma patients undergoing ICB therapy, the integration of DNNs-based models could significantly improve treatment decisions such as [23]. He emphasizes that, such models provide an effective way to capture the complexity of cancer biology compared to traditional linear models. Therefore, the search for effective prediction models in predicting immune responses, especially the use of DL techniques, presents an important frontier in melanoma research, with great potential for personalized treatment approaches.

3. Materials and Methods

3.1. Data Analysis

In this study, we utilized Boston property price[1] dataset which is encompassed within the Sklearn package. It includes 506 instances with 13 numerical/categorical attributes and the target variable. These attributes comprise details such as locality, surroundings, and crime incidence. The dataset has been utilized to tackle a regression issue by forecasting the median worth of residences. The original data is accessible in the Sklearn manual. All gene expression profiles employed in this investigation were obtained from published melanoma assemblies treated with diverse forms of ICB therapies. Due to differences in size and regimen among datasets, it is imperative to categorize the assemblies based on therapy type and underscore the size of each subset. Within each assembly, a patient may have received either one immune checkpoint antagonist (monotherapy), one antagonist consecutively (sequential therapy), or a blend of antagonists (combination therapy). Furthermore, certain melanoma patients have previously undergone MAPKi therapy. For each dataset, the Transcripts Per Million (TPM) values of gene expression along with data regarding treatment and survival time were extracted from the published collections. Depending on the assembly, supplementary details about patients or tumor milieu were also accessible. Nonetheless, such supplementary information is frequently limited to a single assembly, complicating the amalgamation of assemblies. For instance, age is documented in the [24]—Gide/[18]—Hugo/[25]—Van Allen assembly but not in the [1]—Liu/[26]—Riaz assembly. Given that the Liu dataset, comprising 121 records, is the most extensive and the efficacy of DL techniques aligns with the acquisition of sizable, standardized datasets, this dataset was predominantly employed to assess the ANN methodology. The Liu dataset was procured from the [1]. Patients within this assembly were administered either pembrolizumab or nivolumab as monotherapy or ipilimumab and pembrolizumab/nivolumab as sequential therapy. Some patients also received antecedent MAPKi intervention. Alongside the 20,849 gene expression values in TPM format, six clinical attributes were selected for analysis: gender, heterogeneity, nonsynonymous mutation, cancer stage, elevated Lactate Dehydrogenase (LDH), and treatment type (pembrolizumab or nivolumab). The Gide dataset was sourced from the [24]. The individuals within this dataset have received treatment either through single-agent therapy employing pembrolizumab or nivolumab, or through combined therapy by

[1] https://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html

 doi:10.20944/preprints202405.1160.v1

incorporating ipilimumab as the primary agent, respectively. Patients harboring BRAFV600 mutations were classified as individuals with previous MAPKi exposure. The Riaz dataset was acquired from the [26]. This particular group comprises patients subjected solely to nivolumab therapy. There was no available data regarding prior MAPKi treatment. The Hugo dataset was obtained from the [18]. This set comprises patients treated with pembrolizumab monotherapy. Patients with previous MAPKi exposure were included in this cohort. The Van Allen dataset was sourced from the [25]. This dataset involves patients undergoing ipilimumab monotherapy. Patients who received a BRAF inhibitor subsequent to ipilimumab treatment were categorized as individuals with prior MAPKi exposure.

3.2. Model Analysis

The described ANN process is documented using the publicly accessible web application Jupyter Notebook in conjunction with the Python programming language. Numerous Python packages were utilized for constructing the ANN workflow. These packages were installed via pip (20.3.3) and deployed on Debian GNU/Linux 10 (buster). The workflow can be segmented into four primary phases. Initially, the data necessitates preprocessing and division into training and testing datasets. An optional stage follows, concerning dimensionality reduction through feature selection. Subsequently, the ANN is trained and tested, encompassing parameter tuning with GridSearch, ANN training with optimized parameters, and final model validation using the held-out test dataset. Data preprocessing emerges as a pivotal aspect of the ANN modeling workflow. Initially, all samples featuring missing values in any attribute were eliminated, resulting in 506 samples for the Boston Housing dataset and 118 for the Liu dataset. Given that the dataset comprises both numerical and categorical variables such as treatment, all input variables necessitate conversion into numerical representations. In the Boston Housing dataset, all 13 attributes are inherently numeric, while categorical attributes in Liu were transformed utilizing the LabelEncoder function from Sklearn. For binary predictions, outcomes were stratified using an appropriate threshold value. In the Boston housing dataset, a MEDV (Median value of owner-occupied homes in $1000) less than 22 is categorized as low-price houses, and greater than 22 as high-price houses. In Liu, patients with PFS of less than 365 days or OS of less than 730 days are classified as short survivors, while those with PFS exceeding 365 or OS surpassing 730 are considered long survivors.

Addressing imbalanced data, the SMOTE-NC oversampling technique is applied to divide the data into two groups, generating synthetic data using a *k*-nearest neighbor algorithm [27–32]. Feature scaling is conducted to standardize feature scales and enhance the gradient descent algorithm's efficiency. Standardization, facilitated by the StandardScaler method from Sklearn, is preferred over Min-Max-Scaling due to its resilience to outliers. Data scaling is solely computed on the training set to prevent data leakage. Recursive Feature Selection (RFE) from Sklearn is employed for feature selection, utilizing Support Vector Regression (SVR) for continuous prediction or Support Vector Classification (SVC) for binary classification. ANN architecture comprises an input layer accommodating gene expression TPM values and clinical variables, followed by one or more hidden layers. Weights and biases initialization employs 'glorot_uniform' and 'zeros' respectively, while training employs the Adam optimizer with variable learning rates. The hidden layer incorporates dense layers followed by dropout layers, utilizing ReLU activation functions for input layers, and variable L_1 and L_2 regularization for kernel regularization. The number of neurons and hidden layers are adjustable parameters. The output layer of the neural network comprises a single node, with activation and loss functions tailored based on the nature of predictions—sigmoid activation and binary_crossentropy loss for binary outcomes (e.g., predicting PFS over or under 365 days), and linear activation with Mean Squared Error (MSE) loss for continuous outcomes (e.g., predicting exact PFS days). The optimizer employed is Adam, with additional variants such as Stochastic Gradient Descent (SGD) with momentum and Nesterov momentum. Several parameters are fine-tuned to optimize the neural network's performance. These include the number of neurons per layer, the number of hidden layers, the learning rate, the L_1 regularization rate, the L_2 regularization rate, the type of optimizer, and the dropout rate. The data is initially split into 80% for training the model and

 doi:10.20944/preprints202405.1160.v1

20% for testing the model's performance on unseen data. Only the 80% training data (n = 94) are used for tuning and training the model. Validation is performed with the remaining 20% (n = 24), as described earlier. The neural network's variable parameters undergo fine-tuning via GridSearchCV with 5-fold cross-validation and a 20% validation split. Once optimal parameters are identified, training commences in shuffle mode to ensure data independence. Two Keras callbacks, EarlyStopping and ReduceLROnPlateau, expedite training and optimize the learning rate. EarlyStopping ceases training if validation loss stagnates for 20 epochs, while ReduceLROnPlateau adjusts the learning rate by a factor of 0.1 if validation loss stalls for five epochs until reaching a minimum of $1e^{-6}$. To evaluate the trained ANN, multiple 80/20 splits are executed using a ShuffleSplit cross-validator, employing tailored metrics. For binary classification, these metrics encompass sensitivity, specificity, positive and negative predictive values, and the confusion matrix. Conversely, for continuous outcomes, evaluation metrics include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2.

4. Experimental Analysis

4.1. Sanity Check with Boston Housing Dataset

To assess the predictive ability of the neurons under ideal conditions, we performed experiments using the well-known Boston Housing dataset as shown in Table I. This list is typically used to predict housing prices based on 13 distinct characteristics. To simulate the binary prediction task, we divided the target variables into two groups: "affordable" and "affordable" building prices, using a threshold value of 22 (equivalent to $22,000). Parameters chosen for GridSearch optimization include: neuron count = 8,16,32,64 respectively; SecretPopulation = 1,2,3; Best performer = SGD, Adam; study rate = 0.01; Release = 0.0,0.2,0.4,0.6; L1 Constants = 0.0,0.2,0.4; L2 constant processing rate = 0.0,0.2,0.4. The constructed workflow achieved up to 95% accuracy for five train/test splits typically 80/20. However, the prediction performance of the workflow varied depending on the choice of fragmentation, resulting in false predictions and false negatives, and in addition, the specific train/test fractionation had the best coefficient affected by the grid search method. Notably, the number of hidden neurons ranged from 1 to 3 and the number of neurons per layer ranged from 32 to 64 neurons. For a more detailed analysis and considering that the Boston housing dataset is designed for regression functions rather than binary classification, we evaluated the predictive performance of the project using continuous effects. For five different train/trial separations, the root network produced a RMSE of 3.41, MAE of 2.94, R-class, with the same GridSearch parameters used. Furthermore, the optimal parameters differed depending on the specific train/test split, although no constants were used for any of the five splits.

Table I. BOSTON HOUSING DATASET PREDICTION: BINARY AND CONTINUOUS.

Binary Outcome		**Continuous Outcome**	
Best Parameter	**Metrics**	**Best Parameter**	**Metrics**
Dropout rate:0.2	Accuracy 0.85 +/-(0.01)	Dropout rate:0.0	MAE 2.29 +/-(0.07)
L1 reg:0.0	Sensitivity 0.81 +/-(0.01)	L1 reg:0.0	RMSE 2.90 +/-(0.09)
L2 reg:0.0	Specificity 0.88 +/-(0.02)	L2 reg:0.0	R^2 0.90 +/-(0.01)
Learning rate:0.01	PPV 0.85 +/-(0.02)	Learning rate:0.01	
Number neurons:64	NPV 0.85 +/-(0.01)	Number neurons:64	
Number hidden: 2		Number hidden: 3	
Optimizer: Adam		Optimizer: Adam	

 doi:10.20944/preprints202405.1160.v1

Dropout rate:0.2 L1 reg:0.0 L2 reg:0.0 Learning rate:0.01 Number neurons:64 Number hidden: 3 Optimizer: Adam	Accuracy 0.89 +/-(0.01) Sensitivity 0.87 +/-(0.02) Specificity 0.90 +/-(0.02) PPV 0.88 +/-(0.02) NPV 0.89 +/-(0.01)	Dropout rate:0.0 L1 reg:0.0 L2 reg:0.0 Learning rate:0.01 Number neurons:64 Number hidden: 2 Optimizer: Adam	MAE 1.97 +/-(0.05) RMSE 2.55 +/-(0.08) R^2 0.92 +/-(0.00)

4.2. Predictions on Liu Dataset

The current section present the application of the developed workflow to address the research query regarding predicting individual responses to therapy based on gene expression, in conjunction with described genomic and clinical features. The overarching objective of guiding therapy decisions through a patient's genetic makeup necessitates rigorous validation of predictive models concerning the outcome of ICB therapy based on individual genomic profiles. Given the pivotal role of extensive and standardized datasets in DL success, initial preference is given to the Liu dataset due to its extensive RNA sequencing data as shown in Table II. This dataset comprises 20849 gene expression values along with six clinical variables, including gender, heterogeneity, nonsynonymous mutation, cancer stage, LDH elevation, and treatment (pembrolizumab or nivolumab). Binary and continuous predictions for PFS and OS were conducted through various splits using identical grid search parameters. The selection of optimal parameters involved a meticulous process, combining testing and knowledge from neural network prediction methods in cancer modeling. Nonetheless, comprehensive details on network configurations, overfitting mitigation strategies, and specific learning parameters were often lacking in available literature. For continuous outcome prediction, addressing the challenge of "exploding gradients" in ANNs proved imperative. This phenomenon arises when significant weight updates during training lead to numerical overflow or underflow. While implementing the gradient norm clipping function in Keras resolved the issue, prediction metrics remained unsatisfactory. Notably, the R^2 values across all splits were consistently negative, indicating arbitrary deterioration in model performance. Given the complexity of regression tasks for ANNs and associated structural disparities, focus shifted towards binary classification. Moreover, delineating between "long" and "short" survivors proved more clinically relevant in treatment decision-making scenarios. In binary outcome prediction, the model achieved up to 72% accuracy for OS and up to 71% for PFS, accurately classifying patients into respective categories. However, the evaluation of additional metrics such as specificity and sensitivity is crucial, especially considering the modest size of the test split. Upon thorough evaluation, the binary classification model demonstrated promising accuracy rates, with specificity notably outperforming sensitivity in predicting short-term survivors.

Table II. BINARY PREDICTION USING OS AS THE ENDPOINT FOLLOWING FEATURE SELECTION ON THE LIU ET AL. DATASET.

Training/ Test Split	Dropout Rate	L_1 Reg	L_2 Reg	Learning Rate	Neurons	Hidden Layers	Optimizer	Accuracy	Sensitivity	Specificity	PPV	NPV
1	0.0	0.0	0.0	0.01	3	1	SGD	0.71	0.48	0.85	0.66	0.73

 doi:10.20944/preprints202405.1160.v1

2	0.2	0.0	0.2	0.01	5	3	SGD	0.52	0.33	0.63	0.35	0.61
3	0.4	0.0	0.4	0.01	3	3	SGD	0.53	0.21	0.72	0.31	0.60
4	0.0	0.0	0.0	0.01	3	1	Adam	0.65	0.46	0.77	0.56	0.71
x	0.4	0.0	0.4	0.01	5	3	SGD	0.59	0.22	0.81	0.42	0.63

To address the challenge of dealing with large amounts of data, a method called feature selection is employed to narrow down the variables to 50 before the actual training of the ANN as shown in Table III. In the context of predicting PFS, a greater number of features were initially considered. Notably, the selected features demonstrated superior performance in predicting PFS, where none of the initially considered features were present across all five datasets. The feature selection outcomes, detailed in Table 7, highlight consistent identification of genes like HIKESHI and LIN28A across all five datasets, with several others recurring in four out of five datasets. Post-selection, separate training of the ANN for each dataset using these features and clinical variables yielded binary prediction accuracies for PFS ranging from 68% to 86%, surpassing those for OS (49% to 79%). This improvement, alongside a notable boost in PFS prediction accuracy compared to the model without feature selection, underscores the efficacy and adaptability of the approach across diverse dataset characteristics.

Table III. BINARY PREDICTION USING PFS AS THE ENDPOINT FOLLOWING FEATURE SELECTION ON THE LIU ET AL. DATASET.

Best Parameter	Metrics
Dropout rate:0.0 L1 reg:0.0 L2 reg:0.0 Learning rate:0.01 Number neurons: 3 Number hidden: 3 Optimizer: SGD	Accuracy 0.76 +/-(0.04) Sensitivity 0.47 +/-(0.13) Specificity 0.84 +/-(0.04) PPV 0.44 +/-(0.09) NPV 0.86 +/-(0.03)
Dropout rate:0.4 L1 reg:0.0 L2 reg:0.2 Learning rate:0.01 Number neurons: 10 Number hidden: 2 Optimizer: Adam	Accuracy 0.68 +/-(0.03) Sensitivity 0.19 +/-(0.05) Specificity 0.81 +/-(0.04) PPV 0.21 +/-(0.07) NPV 0.79 +/-(0.01)

4.3. Predictions on Combination of Cohorts

In dealing with the challenge of the limited dataset, we chose to expand the training dataset. To achieve this, we combine the heterogeneous sets, resulting in a pooled dataset of 322 observations. To ensure consistency, we first followed [1], as the TPM cannot accommodate variations in libraries or different levels of rRNA contamination in different clusters. Next, we apply normalization between samples using the mean method of ratios using DESeq2 to reduce systematic and technical

 doi:10.20944/preprints202405.1160.v1

biases resulting from different experimental conditions and sequencing protocols. Given the absence of clinical or genomic characteristics beyond OS in all cohorts, our predictions are based solely on gene expression data. Clinical data for each patient in each group are combined using binary coding. For the binary prediction of OS, we train and evaluate in five different partitions using grid search parameters as follows: number of neurons (3, 5, 10), number of hidden layers (1, 2, 3), Number of L_1 regularization (0.0, 0.2, 10). 0.4), the number of L_2 routines (0.0, 0.2, 0.4), and the number of classes (0.0, 0.2, 0.4), including the ADAM optimizer. The predictive accuracy results for OS ranged from 61% to 66% (refer to Table IV). Specificity is consistently greater than 70% for all separations, while sensitivity is high at 55%, indicating better performance in predicting shorter lifetimes than longer ones. Although the metric changes are relatively small compared to the forecasts based on the Liu group alone, overall the combination of the different groups does not significantly improve the forecast results.

Table IV. BINARY FORECASTING USING THE SET OF COHORTS WITH OS AS THE ENDPOINT.

Split	Dropout Rate	L_1 Reg	L_2 Reg	Learning Rate	Neurons	Hidden Layers	Optimizer	Accuracy	Sensitivity	Specificity	PPV	NPV
1	0.0	0.0	0.4	0.01	3	2	Adam	0.65	0.33	0.82	0.49	0.71
2	0.0	0.0	0.2	0.01	5	1	SGD	0.66	0.35	0.82	0.49	0.71
3	0.0	0.0	0.2	0.01	3	3	Adam	0.61	0.25	0.80	0.39	0.68
4	0.2	0.0	0.0	0.01	3	3	Adam	0.60	0.39	0.71	0.41	0.70

5. Discussion

The quest for personalized therapy response prediction poses significant biological and computational challenges, constrained by data accessibility and algorithmic limitations. Utilizing neural networks offers a promising avenue for forecasting therapy responses, particularly based on microarray data. Ensuring the reproducibility of predictive models is paramount for developing clinically relevant techniques, necessitating a flexible neural network architecture and robust data workflow. Applied to the context of predicting responses to ICB therapy, neural networks hold promise for personalized treatment of melanoma patients. However, challenges arise from limited sample sizes in standardized datasets, necessitating innovative approaches to address cohort deficiencies. Notably, diverse treatment modalities further complicate the analysis, highlighting the need for robust computational workflows tailored to biological nuances. Predicting PFS and OS for patients undergoing ICB therapy entails complex analyses integrating gene expression and clinical variables. The developed workflow demonstrates efficacy in leveraging these data, achieving competitive prediction accuracies for both PFS and OS. However, challenges such as overfitting and model variance underscore the importance of regularization techniques and comprehensive parameter optimization. Feature selection plays a crucial role in enhancing prediction accuracy and interpreting model outcomes. Various methods, including RFE and Principal Component Analysis (PCA), offer avenues to address high-dimensional data challenges. Notably, the efficacy of feature selection varies depending on the predictive task, highlighting the need for tailored approaches. Presently, the amalgamation of ICB treatment combinations and limited available data pose challenges to attaining conclusive outcomes, necessitating further investigation, particularly

 doi:10.20944/preprints202405.1160.v1

stratifying for treatment combinations. Strides toward enhancing the reproducibility of ICB treatment response prediction have been realized through the formulation of a flexible workflow, the amalgamation of diverse melanoma cohorts predicated on disparate treatments, and the amalgamation of available clinical and genomic features. The adaptability of the workflow renders it amenable to diverse and heterogeneously structured datasets, while the GridSearch function alleviates the human workload by streamlining parameter optimization, fostering clinical applicability. In essence, this study delineates a workflow with demonstrated efficacy provided the available data furnishes pertinent information for regression or classification tasks. By amalgamating several methodologies such as regularization and feature selection, the workflow grapples with high-dimensional, low-sample-size data, which is pivotal for therapy response prediction predicated on transcriptomic data. Moreover, this master study elucidates the challenges stemming from class imbalance and data heterogeneity engendered by therapy prediction grounded on transcriptomic data, underscoring the exigency for further research addressing low-sample-size, high-dimensional data. The evolved workflow, spearheaded in this study, fills the lacuna of DL-based prediction of clinical benefit from ICB therapy for melanoma patients, thus serving as a springboard for personalized treatment decision-making for melanoma patients undergoing ICB therapy.

6. Conclusion and Future Works

Issues persist in predicting response to ICB treatment, largely due to limited melanoma patient samples. Nonetheless, incorporating publicly accessible datasets like those from the Gene Expression Omnibus (GEO) database could bolster predictive models by expanding training sets. Another strategy involves pretraining or transfer learning, where algorithms leverage knowledge from related tasks or unlabeled data to improve performance. Convolutional Neural Networks (CNNs) excel in transfer learning, particularly when faced with challenges in accessing extensive omics data. While this study focused on predicting PFS and OS, exploring alternative endpoints like RECIST criteria presents intriguing possibilities. However, challenges persist in standardizing response measurement for preclinical models, hindering fair model comparisons. Additionally, selecting input features requires careful consideration, with methods incorporating biological knowledge yielding superior results. Enhancing personalized treatment response prediction entails incorporating diverse methodologies like transfer learning and graph-convolutional networks into automated workflows. With access to more data and related datasets, DL has the potential to revolutionize clinical decision-making for ICB treatment in melanoma patients.

References

1. D. Liu *et al.*, “Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma,” *Nature Medicine*, vol. 25, no. 12, pp. 1916–1927, Dec. 2019, doi: 10.1038/s41591-019-0654-5.
2. V. K. Kanaparthi, “Examining the Plausible Applications of Artificial Intelligence & Machine Learning in Accounts Payable Improvement,” *FinTech*, vol. 2, no. 3, pp. 461–474, Jul. 2023, doi: 10.3390/fintech2030026.
3. V. Kanaparthi, “Robustness Evaluation of LSTM-based Deep Learning Models for Bitcoin Price Prediction in the Presence of Random Disturbances,” Jan. 2024, doi: 10.21203/RS.3.RS-3906529/V1.
4. V. Kanaparthi, “Evaluating Financial Risk in the Transition from EONIA to ESTER: A TimeGAN Approach with Enhanced VaR Estimations,” Jan. 2024, doi: 10.21203/RS.3.RS-3906541/V1.
5. M. Daoud and M. Mayo, “A survey of neural network-based cancer prediction models from microarray data,” *Artificial Intelligence in Medicine*, vol. 97. Elsevier, pp. 204–214, Jun. 01, 2019. doi: 10.1016/j.artmed.2019.01.006.
6. G. S. Kashyap, A. Siddiqui, R. Siddiqui, K. Malik, S. Wazir, and A. E. I. Brownlee, “Prediction of Suicidal Risk Using Machine Learning Models.” Dec. 25, 2021. Accessed: Feb. 04, 2024. [Online]. Available: https://papers.ssrn.com/abstract=4709789
7. G. S. Kashyap, A. E. I. Brownlee, O. C. Phukan, K. Malik, and S. Wazir, “Roulette-Wheel Selection-Based PSO Algorithm for Solving the Vehicle Routing Problem with Time Windows,” Jun. 2023, Accessed: Jul. 04, 2023. [Online]. Available: https://arxiv.org/abs/2306.02308v1

8. S. Wazir, G. S. Kashyap, and P. Saxena, "MLOps: A Review," Aug. 2023, Accessed: Sep. 16, 2023. [Online]. Available: https://arxiv.org/abs/2308.10908v1
9. G. S. Kashyap, D. Mahajan, O. C. Phukan, A. Kumar, A. E. I. Brownlee, and J. Gao, "From Simulations to Reality: Enhancing Multi-Robot Exploration for Urban Search and Rescue," Nov. 2023, Accessed: Dec. 03, 2023. [Online]. Available: https://arxiv.org/abs/2311.16958v1
10. H. Habib, G. S. Kashyap, N. Tabassum, and T. Nafis, "Stock Price Prediction Using Artificial Intelligence Based on LSTM– Deep Learning Model," in *Artificial Intelligence & Blockchain in Cyber Physical Systems: Technologies & Applications*, CRC Press, 2023, pp. 93–99. doi: 10.1201/9781003190301-6.
11. N. Marwah, V. K. Singh, G. S. Kashyap, and S. Wazir, "An analysis of the robustness of UAV agriculture field coverage using multi-agent reinforcement learning," *International Journal of Information Technology (Singapore)*, vol. 15, no. 4, pp. 2317–2327, May 2023, doi: 10.1007/s41870-023-01264-0.
12. S. Naz and G. S. Kashyap, "Enhancing the predictive capability of a mathematical model for pseudomonas aeruginosa through artificial neural networks," *International Journal of Information Technology 2024*, pp. 1–10, Feb. 2024, doi: 10.1007/S41870-023-01721-W.
13. V. Kanaparthi, "Examining Natural Language Processing Techniques in the Education and Healthcare Fields," *International Journal of Engineering and Advanced Technology*, vol. 12, no. 2, pp. 8–18, Dec. 2022, doi: 10.35940/ijeat.b3861.1212222.
14. V. Kanaparthi, "Exploring the Impact of Blockchain, AI, and ML on Financial Accounting Efficiency and Transformation," Jan. 2024, Accessed: Feb. 04, 2024. [Online]. Available: https://arxiv.org/abs/2401.15715v1
15. V. Kanaparthi, "Credit Risk Prediction using Ensemble Machine Learning Algorithms," in *6th International Conference on Inventive Computation Technologies, ICICT 2023 - Proceedings*, 2023, pp. 41–47. doi: 10.1109/ICICT57646.2023.10134486.
16. F. S. Hodi *et al.*, "Improved Survival with Ipilimumab in Patients with Metastatic Melanoma," *New England Journal of Medicine*, vol. 363, no. 8, pp. 711–723, Aug. 2010, doi: 10.1056/nejmoa1003466.
17. C. Robert *et al.*, "Pembrolizumab versus Ipilimumab in Advanced Melanoma," *New England Journal of Medicine*, vol. 372, no. 26, pp. 2521–2532, Jun. 2015, doi: 10.1056/nejmoa1503093.
18. W. Hugo *et al.*, "Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma," *Cell*, vol. 165, no. 1, pp. 35–44, Mar. 2016, doi: 10.1016/j.cell.2016.02.065.
19. J. Larkin *et al.*, "Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma," *New England Journal of Medicine*, vol. 373, no. 1, pp. 23–34, Jul. 2015, doi: 10.1056/nejmoa1504030.
20. D. Liu, R. W. Jenkins, and R. J. Sullivan, "Mechanisms of Resistance to Immune Checkpoint Blockade," *American Journal of Clinical Dermatology*, vol. 20, no. 1. Springer, pp. 41–54, Sep. 26, 2019. doi: 10.1007/s40257-018-0389-y.
21. S. D. Brown *et al.*, "Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival," *Genome Research*, vol. 24, no. 5, pp. 743–750, May 2014, doi: 10.1101/gr.165985.113.
22. M. S. Rooney, S. A. Shukla, C. J. Wu, G. Getz, and N. Hacohen, "Molecular and genetic properties of tumors associated with local immune cytolytic activity," *Cell*, vol. 160, no. 1–2, pp. 48–61, Jan. 2015, doi: 10.1016/j.cell.2014.12.033.
23. G. Adam, L. Rampášek, Z. Safikhani, P. Smirnov, B. Haibe-Kains, and A. Goldenberg, "Machine learning approaches to drug response prediction: challenges and recent progress," *npj Precision Oncology*, vol. 4, no. 1. Nature Publishing Group, pp. 1–10, Jun. 15, 2020. doi: 10.1038/s41698-020-0122-1.
24. T. N. Gide *et al.*, "Distinct Immune Cell Populations Define Response to Anti-PD-1 Monotherapy and Anti-PD-1/Anti-CTLA-4 Combined Therapy," *Cancer Cell*, vol. 35, no. 2, pp. 238-255.e6, Feb. 2019, doi: 10.1016/j.ccell.2019.01.003.
25. E. M. Van Allen *et al.*, "Genomic correlates of response to CTLA-4 blockade in metastatic melanoma," *Science*, vol. 350, no. 6257, pp. 207–211, Oct. 2015, doi: 10.1126/science.aad0095.
26. N. Riaz *et al.*, "Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab," *Cell*, vol. 171, no. 4, pp. 934-949.e15, Nov. 2017, doi: 10.1016/j.cell.2017.09.028.
27. S. Wazir, G. S. Kashyap, K. Malik, and A. E. I. Brownlee, "Predicting the Infection Level of COVID-19 Virus Using Normal Distribution-Based Approximation Model and PSO," Springer, Cham, 2023, pp. 75–91. doi: 10.1007/978-3-031-33183-1_5.
28. M. Kanojia, P. Kamani, G. S. Kashyap, S. Naz, S. Wazir, and A. Chauhan, "Alternative Agriculture Land-Use Transformation Pathways by Partial-Equilibrium Agricultural Sector Model: A Mathematical Approach," Aug. 2023, Accessed: Sep. 16, 2023. [Online]. Available: https://arxiv.org/abs/2308.11632v1

 doi:10.20944/preprints202405.1160.v1

29. P. Kaur, G. S. Kashyap, A. Kumar, M. T. Nafis, S. Kumar, and V. Shokeen, "From Text to Transformation: A Comprehensive Review of Large Language Models' Versatility," Feb. 2024, Accessed: Mar. 21, 2024. [Online]. Available: https://arxiv.org/abs/2402.16142v1
30. G. S. Kashyap, K. Malik, S. Wazir, and R. Khan, "Using Machine Learning to Quantify the Multimedia Risk Due to Fuzzing," *Multimedia Tools and Applications*, vol. 81, no. 25, pp. 36685–36698, Oct. 2022, doi: 10.1007/s11042-021-11558-9.
31. G. S. Kashyap *et al.*, "Revolutionizing Agriculture: A Comprehensive Review of Artificial Intelligence Techniques in Farming," Feb. 2024, doi: 10.21203/RS.3.RS-3984385/V1.
32. G. S. Kashyap *et al.*, "Detection of a facemask in real-time using deep learning methods: Prevention of Covid 19," Jan. 2024, Accessed: Feb. 04, 2024. [Online]. Available: https://arxiv.org/abs/2401.15675v1