AERO- HNSCC: An Autoencoder-Based Risk Stratification Model for Head and Neck Squamous Cell Carcinoma (2024)

Jucheng Hu, Tianxiang Xiong, Xi Chen, Min Wang, Eric Shi

***Abstract*—** **Head and Neck Squamous Cell Carcinoma (HNSCC) ranks among the deadliest cancers, with prognosis prediction remaining challenging due to the absence of reliable approaches. This study introduces a novel model, AutoEncoder Risk Stratification for Oncology in HNSCC (AERO-HNSCC), which transcends the traditional biomarker identification method by integrating autoencoders (AEs) and multi-omics data for prognostic analysis of HNSCC. AERO-HNSCC uniquely utilizes a deep AE for pre-processing multi-dimensional data, including protein expression, RNA-Seq, and clinical information, enhancing the predictive accuracy of patient overall survival (OS). Developed and validated on The Cancer Genome Atlas (TCGA) HNSCC dataset, AERO-HNSCC demonstrates a significant success in risk stratification with an average precision of 73% and a statistically significant p-value of 0.0021 in the log-rank test for survival analysis. These quantitative evaluations demonstrate that the AERO-HNSCC encoded data establishes a robust association with patient OS and can accurately predict OS. Compared with raw multi-omics data for risk stratification and signature-identification-centred (SIC) methods, this new method shows comparable or superior performance. In conclusion, this research contributes a novel computational framework for HNSCC prognosis, its success underlines the potential for AE-based methods to transcend traditional SIC methods and to offer universal solutions in oncological research.**

***Index Terms*—hypoxia, immune, gene signature, tumour, cancer, microenvironment, machine learning, deep learning, autoencoder, prognosis, multi-omics, head and neck squamous cell carcinoma**

# I. INTRODUCTION

HNSCC is squamous cell carcinomas arising from lesions in the head and neck region, including the oral cavity, tongue, nasal and paranasal cavity, pharynx, and larynx[1]. HNSCC is among the top ten leading causes of cancer worldwide, with approximately 500,000 new cases diagnosed annually[2]. The high prevalence and risk associated with HNSCC emphasize the importance of developing effective methods to assess cancer progression and predict patient prognosis.

Numerous studies have developed various biomarkers with prognostic value for HNSCC. Among all these biomarkers, gene signatures, particularly the Hypoxia Signature (HS) – a set of differentially expressed genes (DEGs) under hypoxic conditions within the tumour microenvironment – have emerged as a key method. The expression level of HS can be used to infer intratumoural hypoxia levels[3], which is crucial for predicting clinical outcomes[4]. This signature-identification-centred (SIC) method also applied on protein expression data. It has been shown that proteomic biomarkers can function as predictor for cancer diagnosis and prognosis [5], and some proteomic signatures have been developed on the Reverse Phase Protein Array (RPPA) data[2].

Despite these advances, challenges remain, particularly in harnessing the full potential of integrated data from various omics studies (multi-omics data), due to its increasing dimensionality. This complexity complicates the extraction of meaningful information from the noise. AEs have been introduced as a solution to the dimensionality problem and offer a novel approach beyond SIC methods. Instead of developing prognostic signatures, AE-based methods aim to capture insights from the entire multi-omics dataset. However, existing works either rely on supervised method[6], which may not be feasible when labels are unavailable, or are lacking application to protein expression data[7], or do not specifically focus on HNSCC[8].

The research gap necessitates this research, and leads to the objective of this research: firstly, determine whether protein expression data can provide insights into HNSCC patient OS and develop an AE-based risk stratification model for accurate OS classification and therefore enabling early identification of high-risk patients.

The structure of this paper is as follows: Section 2 reviews the related work in the field, providing context and highlighting the novelty of our approach. Section 3 delves into the experimental design, detailing the AERO-HNSCC architecture, data preparation procedures, validation metrics, and benchmarking strategies. Section 4 analyses the results, assessing the efficacy and potential implications of our model. Finishing up the protein side of the AERO-HNSCC, Section 5 is a standalone section introducing attempts of the AERO-HNSCC on RNA data. In Section 6 we discuss the limitation of current work and lead to the final Section 7, suggesting directions for future research.

# II. Related Work

The Related Work section of this paper delves into two pivotal methodologies that have significantly impacted the prognosis and classification of HNSCC: SIC Methods and AE-Based Methods. Understanding these approaches illuminates the backdrop against which this research is situated.

1. *SIC Methods*

SIC methods have been cornerstone approaches in deciphering the complex molecular landscapes of cancers, including HNSCC. These methods focus on identifying specific gene signatures that correlate with disease outcomes, HSs especially. All SIC-related paper reviewed are shown in table I.

TABLE I

Methods and Tumour Type of Reviewed Papers

|  |  |  |  |
| --- | --- | --- | --- |
| First Author | Year | Methods | Tumour |
| Cheng-Peng Gui | 2021 | t-SNE and Lasso | crCC |
| Yifan  Liu | 2020 | t-SNE and Lasso Cox Regression | Gastric Cancer |
| Zhi  Liu | 2021 | LASSO | BLCA |
| Yanhong Shou | 2021 | LASSO | Melanoma |
| Jinman Zhong | 2021 | LASSO | AML |
| Xia  Yang | 2021 | LASSO | Breast cancer |
| Ke  Wang | 2022 | LASSO | GBM |
| Chenyu Nie | 2022 | LASSO | Cervical Cancer |
| Xiong Tian | 2022 | LASSO | PAAD |
| Fanhong Zeng | 2021 | K-mean | HCC |
| Brian Lane | 2022 | K-mean | LUDA |
| Jun  Shao | 2021 | K-mean and LASSO Cox Regression | LUAD |
| J.M. Brooks | 2019 | Unsupervised Hierarchical Clustering | HNSCC |
| Jia  Li | 2022 | Random Forest | Breast Cancer |
| Donglei Wu | 2020 | LASSO Cox Regression | HNSCC |
| Baohui Zhang | 2020 | LASSO Cox Regression | HCC |
| Run  Shi | 2021 | LASSO Cox Regression | LUDA |
| Dongjie Chen | 2021 | LASSO Cox Regression | PDAC |
| Qiangnu Zhang | 2021 | LASSO Cox Regression | HCC |
| Xiangqian Zhang | 2023 | LASSO Cox regression | Gastric Cancer |

As a conclusion from the papers listed above, the development of HS can be generalized as a three-stage process for various tumour types:

* 1. **Identification of Hypoxia DEGs**

This initial step involves obtaining all hypoxia-related DEGs, which can be sourced through literature reviews[9], databases[10], or clustering by algorithms like K-mean[11-13] or UHC[9].

* 1. **Feature Selection on DEGs**

Subsequently, apply feature selection by conducting LASSO[10, 14-20], LASSO Cox Regression[21-25], or Random Forest[26] on DEGs to filter prognostic HSs.

* 1. **Prognostic Model Development**

The final step involves developing a prognostic model, which may take the form of a score or a more complex model incorporating additional features.

Specifically, while all other works focus on gene expression datasets, Wu et al.’s work[2], following the identical three-stage process, but developed a proteomic signature on the TCGN-HNCC RPPA dataset, achieved a 0.779 of area under the curve (AUC) of the corresponding receiver operating characteristic (ROC) in the task of classifying patients into high and low-risk groups.

In terms of the gene signature, due to the variance in cancers, datasets used for HSs development, the result measurement and processes of HS identification, conducting comparative research on the performances of all different methods remains challenging. Noting the overlapping focus on Hepatocellular Carcinoma (HCC) and Lung Adenocarcinoma (LUAD), this research will focus on the performance of 6 works that developed HSs for these two tumours, respectively.

As shown in table I, three studies centred on HCC. Zeng et al.[12] apply the K-mean in stage 1, the Identification of Hypoxia DEGs, leading to the discovery of four genes: DCN, DDIT4, NDRG1, and PRKCA from the ICGC dataset. Their hypoxia-risk model demonstrated highest accuracy among all three HCC studies in predicting one-year and three-year OS. The AUC values for one-year, tow-year and three-year OS are 0.809, 0.771, and 0.791, respectively.

Zhang et al.[22] employed LASSO Cox Regression for stage 2 the Feature Selection on DEGs. This approach identified a HS of another 21 distinct genes, including ADM, BNIP3, BNIP3L, and CA9. AUC values for one, three, and five years were 0.71, 0.73, and 0.69, respectively.

Another Zhang et al.[13] integrated both K-mean and LASSO Cox Regression in stage 1 and 2, uncovering three significant genes: PDSS1, SLC7A11, and CDCA8. The AUCs for half-year, one-year, three-year, and five-year OS were 0.76, 0.78, 0.7, and 0.7, respectively.

Shifting the focus to LUDA, there are three studies. Lane et al.[11], following a similar approach to Zeng et al.[12], constructed a 28-gene hypoxia signature from the TCGA-LUAD dataset, using K-mean in stage 1. Their research emphasized qualitative validation, employing the Hazard Ratio (HR), Confidence Interval (CI), and Kaplan-Meier analysis p-values as metrics. The prognostic relevance of their signature for OS was substantiated in independent cohorts from the TCGA-test and GEO datasets, showcasing results of HR 1.76, CI 1.50–2.08, and p < 0.0001.

Mirroring the methodology of Zhang et al.[22], Shi et al.[21] identified 10 genes using LASSO Cox Regression in the feature selection stage. The HS is developed from the GEO GSE72094 dataset and validated on datasets from U133A, U133 Plus 2.0 and TCGA with result HR = 6.738, 95% CI = 3.902-11.64 and p = 6.42e-09.

Lastly, Shao et al.[10] formulated a seven lncRNA HS from a combination of 13 microarray datasets from various platforms and one RNA-Seq dataset from TCGA. K-mean and Lasso cox regression is used in stage 1 and 2 respectively. This signature was validated on the TCGA validation set, achieving AUC values of 0.665, 0.693, and 0.652 for 1-, 3-, and 5-year overall survival, respectively. Though the main paper did not provide detailed HR, CI and p-value result, supplementary documents provide a separated Kaplan-Meier analysis result for all seven gens, with HR ranging from 0.61 to 1.65, CI from 0.42-0.88 to 1.39-1.95 and p-value from less than 0.001 to 0.277.

Therefore, as a conclusion, despite their proven utility in numerous studies, SIC methods often fall short in addressing the multifaceted nature of cancer progression, exhibit a lack of universality. The process of obtaining hypoxia DEGs relies on existing knowledge of specific genes[9] or proteins[2]. Moreover, when transitioning from one type of cancer to another, the resulting signature[13, 21] could be completely distinct, with the number of genes in the HS varying considerably. Even within the same cancer type, the developed HS[12, 22] can be markedly different when different datasets or methodologies are employed. Furthermore, a comparison between methods that achieved the highest accuracy for LUAD and HCC[12, 21] reveals a plethora of approaches to developing HS, yet there is no clear, universally superior approach that performs optimally across all cancer types. This inconsistency and limitation lead researchers to seek more integrative and comprehensive approaches.

1. *AE-Based Methods*

In contrast to SIC methods, AE-Based Methods offer a fresh perspective on data analysis in oncological research, particularly through the lens of multi-omics integration. This shift is due to the complexities of cancer that extend beyond linear biomarker associations, emphasizing the potential of non-linear data relationships and the holistic nature of biological systems. According to the result of this reviewing, current works can be divided into two paths: supervised and unsupervised.

## **Supervised Approaches**

Tan et al.'s work[6] focused on pan-cancer multi-omics datasets, constructing individual autoencoders for each data type—ranging from DNA methylation to protein expression. By encoding these varied data types to a uniform dimensionality and training with distinct labels such as OS and disease-specific survival (DSS), their approach achieved a noteworthy AUC of 0.7830 for binary classification.

Similarly, Mondol et al.[8] presented a blend of unsupervised pre-training with supervised fine-tuning through an adversarial autoencoder and a method named 'TopGene' which identifies significant genes within the latent space. Their AE demonstrating high precision of 0.8596 in classifying sub-types of breast cancer.

Madhumita and Sushmita's[27] work not quantified through AUC or precision metrics, since they focused on the subtype clustering of glioblastoma multiforme (GBM). They contribute a method that performs supervised feature selection before training a sparse autoencoder (SAE).

## **Unsupervised Approaches**

On the unsupervised side, AE-Based Methods demonstrate their strength in survival stratification without the prerequisite of predefined labels. Song et al.’s research[7] in colorectal cancer utilized DNA methylation, RNA-Seq, and miRNA-Seq data with a deep sparse AE, achieving a concordance index (CI) of 0.781 in survival analysis.

Ellen et al.[28] constructed a single-layer denoising autoencoder that integrates mRNA, miRNA, DNA methylation, and long non-coding RNA data, yielding CI of 0.69 ± 0.03 for LUAD in survival analysis .

Arafa et al.’s work[29] with a Reduced Noise Autoencoder showcases the utility of AE in enhancing data quality through noise reduction. Utilizing a three-layer AE with Reduced Noise-Synthesis Minority Over Sampling Technique (RN-SMOTE), they achieved a precision of 0.75 on a colon cancer dataset in cancer sub-type binary classification using only genomic data.

## **Transition and Comparative Outlook**

AE-Based Methods are forging new paths in cancer research, they stand in stark contrast to SIC Methods by offering a more flexible and comprehensive approach to construct association with clinical endpoints and multi-omics data. Characterized by their ability to handle high-dimensional data and applicable to various cancer types without the need for prior biological knowledge, these methods uncover complex, non-linear relationships that traditional methods might overlook. As this research transitions from the specificities of SIC Methods to the broad potential of AE-Based Methods, it becomes evident that the latter may offer a more integrative and universal framework for understanding the multifaceted nature of cancer progression and patient prognosis.

Despite the innovative strides made by AE-Based Methods in oncological research, certain limitations persist, guiding the direction of our work. Firstly, while supervised methods have demonstrated utility in model architecture and insights, their reliance on labelled data contradicts the growing need for more universal, label-independent approaches in the field. This dependency limits their applicability in situations where comprehensive labelling is impractical or not available, underscoring the necessity for advancements in unsupervised learning techniques. Moreover, a significant gap in the literature is the lack of unsupervised AE applications specifically targeting HNSCC. Notably absent are studies that integrate protein expression data with autoencoders to predict OS in HNSCC patients, an area ripe for exploration. Our research aims to bridge these gaps by developing an unsupervised AE approach that not only transcends the need for labelled data but also focuses specifically on the integration of protein expression to enhance the prognostic understanding of HNSCC. After the development of AERO-HNSCC on the protein expression data, we also explored its application on the RNA-Seq data.

# III. Experiment Design

## Methodology

In this study, we employ a strategic approach by iteratively applying a series of AEs, optimized for protein expression datasets. This methodology is based on the hypothesis that biological signals in protein data can be decoded through a tailored AE. We iterate across AE architectures, each designed to capture more significant OS signal from the protein data. This allows exploration beyond surface-level patterns, delving into deeper biological signals for HNSCC prognosis.

*B. AERO- HNSCC Architecture*

1. **Initial Autoencoder Architecture**

The foundational AE architecture has an input layer of 468 nodes for unique protein expressions, leading into hidden layers of 64 and 32 nodes with ReLU activation, and an output layer with sigmoid activation, ensuring the output values are normalized between 0 and 1, mirroring the input data's format. It uses Adam optimizer and MSE as the reconstruction loss. To explore the optimal compression and feature extraction capabilities, this AE is trained with three variations, featuring bottlenecks of sizes 2, 6, and 12.

1. **Transition to a Wider Deep Autoencoder (DAE)**

Acknowledging the necessity for higher-dimensional representations to adequately capture the OS signal, we next introduce a deep Autoencoder (DAE). A wider, deeper DAE with hidden layers size of 256, 128, and 64 is constructed. The DAE undergoes training across a range of bottleneck sizes, 12, 48, 36, 34, 32, 28, 24, and 18, the results indicating that setting the bottleneck to 18 yields the highest precision in capturing the OS signal.

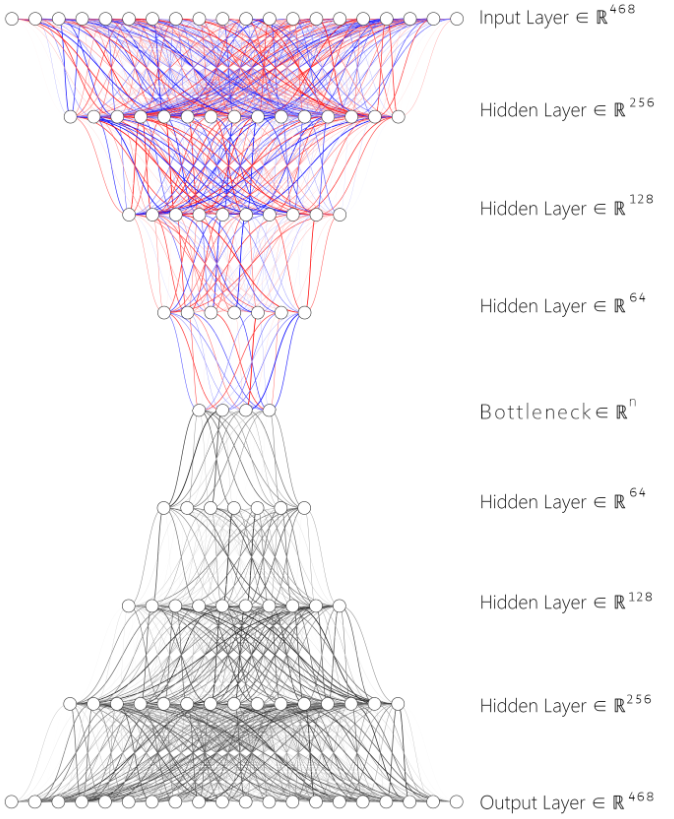


Fig.1. SDAE Architecture, Applied L1 Regularization on Encoder Layers

1. **Incorporating Sparsity: SAE and Sparse DAE (SDAE)**

Building on the initial AE architecture, we further modify the initial AE into a Sparse AE (SAE) by incorporating L1 regularization on its hidden layers, with a penalty rate of 1e-6. This addition encourages the model to learn sparser representations of the data, potentially enhancing its interpretability and efficiency in capturing relevant features. Observing noticeable improvements with this modification, we extend the regularization to the DAE, transforming it into a SDAE, as shown in Fig 1. This adaptation applies the same L1 regularization technique to the DAE's hidden layers,

1. **Integration with Classifier Systems**

Upon the training completion of each AE variant, the encoded data undergo a subsequent layer of analysis involving Principal Component Analysis (PCA), K-means clustering, Random Forest Classification, and Support Vector Machine (SVM) to test performance metrics, as shown in Fig 2. This integrated system aims to refine and validate the predictive power of the encoded features, focusing on the efficacy and accuracy in stratifying patient outcomes based on protein expression profiles, establish a comprehensive framework for advancing the prognostic analysis in HNSCC.

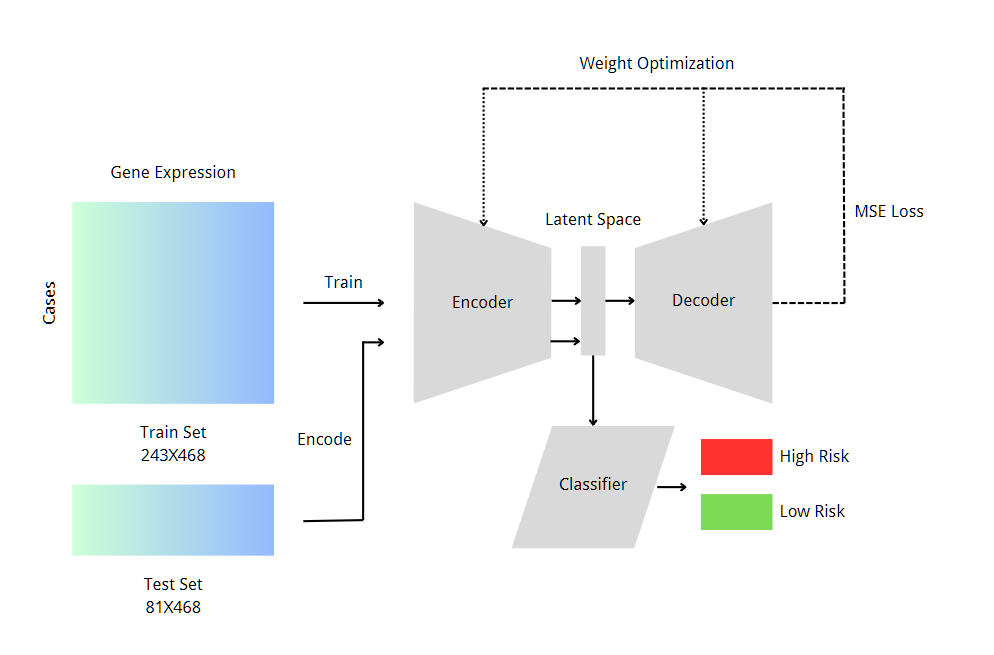


Fig.2. AERO- HNSCC Architecture

## Dataset Description and Process

The AERO-HNSCC is developed on the TCGN-HNSCC RPPA dataset, comprising 353 individual TSV files. Each file corresponds to a distinct patient case, encapsulating a diverse array of 487 protein expressions. The overall data process pipeline is illustrated in Fig 3.

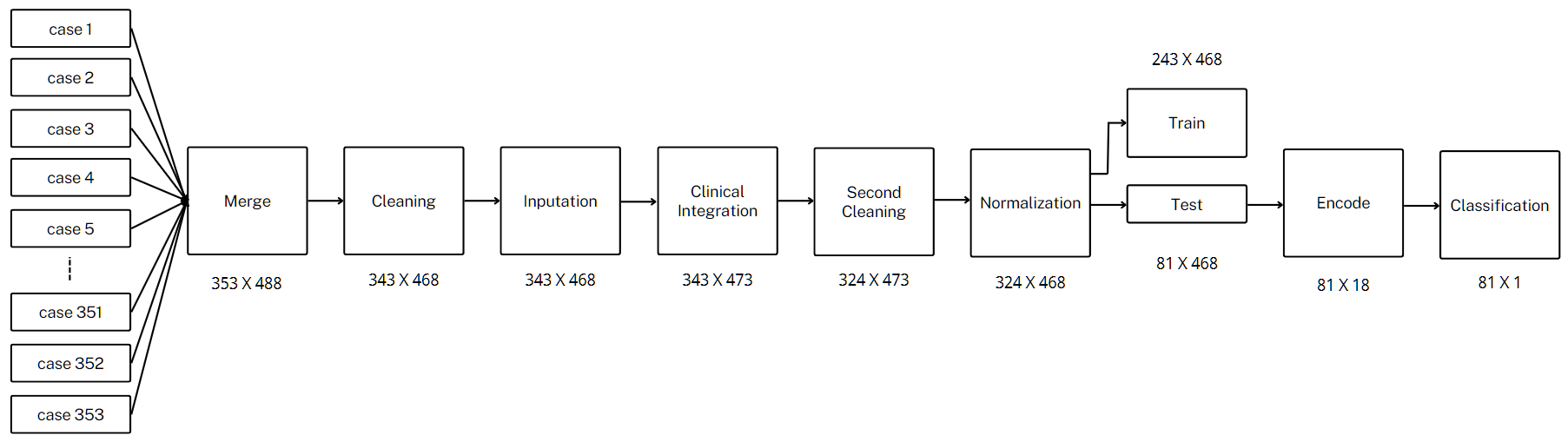


Fig.3. Data Processing Flow Char

Upon examination, several issues were identified within the merged dataset: 237 columns contained NaN values. After merging with the clinical data, 12 cases were found to be missing data across 218 protein expression columns or have ambiguous clinical endpoints. Due to the significant proportion of missing information, some columns and cases were excluded from further analysis, and a target-encoding-like imputation strategy was used for the remaining columns. Using the AJCC pathologic stage as the target, 12 columns were filled. As a result, we obtained a finalized dataset comprising 324 cases, each characterized by 468 protein expression columns, as depicted in Supplementary Figure 1.

Normalization of the protein expression data was then addressed, as highlighted in Fig 4. Given the observed variance in protein expression means, ranging from 2 to 0.001, normalization was deemed essential. According to Hoffer et al.[30], this step is curtail to improves the gradient flow, ensuring that all input features contribute equally and effectively during the learning process, which leads to faster and more stable convergence.

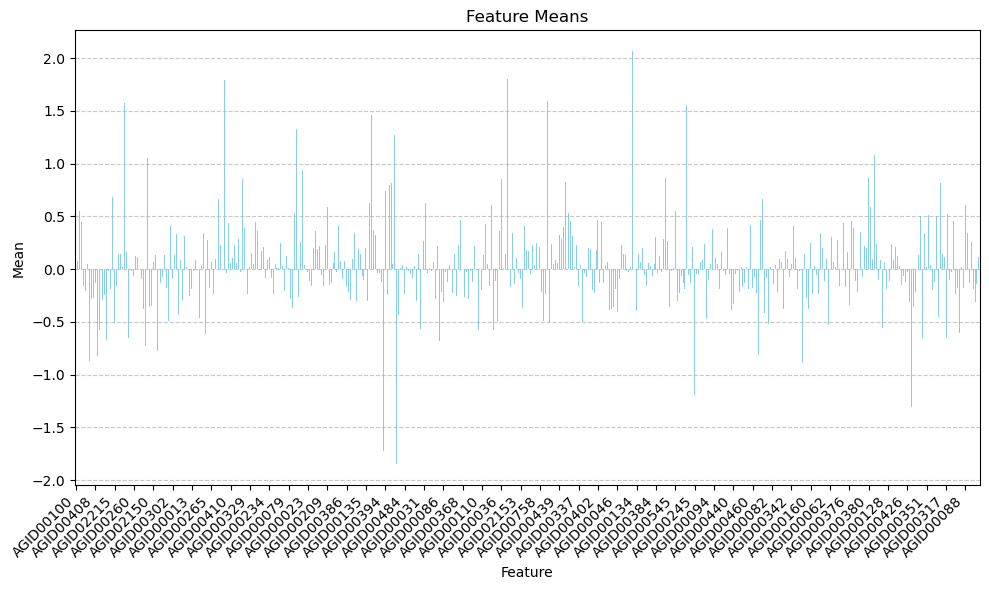


Fig.4. Protein Expression Means Scatter Graph

The final step in the data processing involves remapping the OS for each patient. Depending on the 'vital\_status', labels such as 'dead within 10 years' and 'alive or dead after 10 years' are extracted from clinical data. By grouping the patients into categories of 'alive longer than 10 years' (low risk) and 'alive less than 10 years' (high risk), we balance the population of each group to ensure unbiased classification results.

## D. Training

To closely monitor the AEs’ performance and generalize its applicability, we attached the test set as our validation data without extra validation set. This strategy was adopted due to dataset size constraints and aimed to maximize the usage of all available data, allowed us to directly observe the model's predictive capabilities on unseen data after each training epoch.

Throughout the training process, we meticulously tracked the loss for each AE. Notice that the loss sometimes goes up in some last epochs, we implemented a checkpoint system. This system was designed to save the model's weight at the epoch which demonstrated the highest performance on the test set. Such a method ensured that we retained the most effective version of the AE, even if the model's turns to overfit one the training set and the performance fluctuated in subsequent epochs.

To facilitate a comprehensive analysis of the training outcomes and loss trends across epochs, the loss metrics obtained for each AE were plotted in Supplementary Figure 2. In total, 15 AEs are trained, including two extra that used the initial architecture trained on datasets without imputed columns, featuring a bottleneck of size 6. These two additional AEs were utilized to validate the performance of the imputation.

## E. Metrics

The training success for AEs and classifiers was evaluated using specific metrics. For AEs, we focused on reconstruction and validation loss. Reconstruction loss gauges how well the AE replicates input data, crucial for assessing its data structure capture capability. Validation loss, calculated on unseen data, indicates generalization capabilities and aids in overfitting prevention.

Classifier performance was assessed with precision, recall, F1-score, and the confusion matrix. Precision measures the accuracy of positive predictions, recall evaluates the identification of actual positives, and the F1-score balances both. The confusion matrix aids in direct comparison with benchmarks by converting to AUC of the ROC.

Validation of the AERO-HSNCC framework employed the Kaplan-Meier Survival Curve and log-rank test. The Kaplan-Meier Curve shows survival probabilities of groups over time, while the log-rank test can exam if there is a statistically significant difference between the groups in terms of survival.

## F. Benchmarks

Our experimental design benchmarks against three state-of-the-art studies in HNSCC predictive modelling. These include traditional SIG methods and AE-based approaches.

Brooks et al. [9] developed a gene signature for HNSCC, with log-rank test p-values of 0.5 and 0.2 in two cohorts, showing the gene signature was not independently prognostic. This sets a gene-centric benchmark for our model's comparison.

Wu et al. [2] validated a proteomic signature for HNSCC, achieving an AUC of 0.779 in ROC analysis, offering a comparison for our model's predictive power.

Tan et al. [6] developed a pan-cancer AE-based method, including HNSCC, with an AUC of 0.7830, highlighting AE's potential in cancer prognosis and providing a benchmark for our AE-based model.

We will evaluate our model against these studies using log-rank test p-values and AUC, aiming to advance predictive accuracy, prognostic significance, and innovation in HNSCC prognostication.

IV. Analysis of Results

## A. Interpretation of Findings

As the initial step and primary objective of this research, and outlined in the introduction, the first task was to examine whether protein expression data contain insights into HNSCC OS. Therefore, following the acquisition and cleaning of the RPPA data, we conducted an exploratory classification using SVM with a linear kernel. The accuracy for this binary classification was 0.58, significantly surpassing the 50% threshold. This positive signal led us to further investigation.

Subsequently, we trained the first four AEs. The initial architecture AEs encoded all 468 proteins into dimensions of 6 and 2, with and without the imputation columns. By including the imputation, we observed an 8.51% and 4.08% increase in accuracy, respectively, thereby validating the effectiveness of the imputation approach. However, the highest result obtained with the imputed data encoded to 6 dimensions was 0.59, showing no substantial improvement compared to the raw RPPA data. Visualization was performed by applying PCA with components of 3 and 2 to the six-dimensional results, and plotting the encoded two-dimensional data, as illustrated in Fig 5, revealed no clear boundaries between the two risk groups. Observing the improvement when scaling up bottleneck from 2 to 6, we hypothesized that the AE might require more dimensions to capture the pattern effectively. Consequently, we trained another AE with a 12-dimensional output. Here, approaching the size upper hidden layer, we explored whether performance could be further enhanced by a wider and deeper AE. We thus designed and trained eight variations of the DAE as described in the AERO-HNSCC architecture section with different bottlenecks, decreasing from 48. The optimal performance was achieved with a bottleneck size of 18, which resulted in a precision of 0.73, and was finalized as the AE of the AERO- HNSCC. The confusion matrix for this DAE is displayed in Fig 6.

|  |
| --- |
| PCA of Encoded Test Set (Component 2) PCA of Encoded Test Set (Component 3) |
| Fig.5. Encoded Test Set PCA result |
|  |
|  |

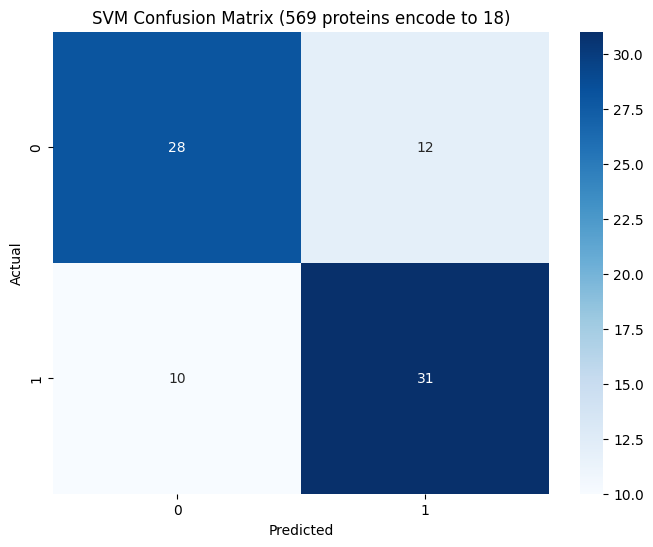


Fig.6. Confusion Matrix of the DAE Bottleneck Size 18

Obtained the AERO-HNSCC AE, we proceeded to assess the prognostic significance of the encoded features using survival analysis techniques. Specifically, we plotted the Kaplan-Meier Survival Curve to visually represent the survival probabilities of patients stratified based on the risk groups identified by our model. This visualization, as demonstrated in Fig 7,showing that the low-risk group starts higher and remains consistently above that of the high-risk group throughout the observed time frame and has a statistically better survival probability over time compared to the high-risk group.

To statistically validate the distinctions observed between these groups, we employed the log-rank test. This non-parametric test is utilized to compare the survival distributions of two or more groups and is a standard method in survival analysis to assess the statistical significance of differences between the Kaplan-Meier curves. The result of the log-rank test, with a p-value of 0.0021, indicates a statistically significant difference in survival rates between the groups. This significant p-value underscores the prognostic relevance of the patterns captured by the AERO-HNSCC, suggesting that the encoded features have substantial implications for predicting patient outcomes in HNSCC.

In the following research we also tested the performance of SAE and SDAE, although by apply L1 regularization we do observe a more stable AE performance, it does not improve the performance of classification.

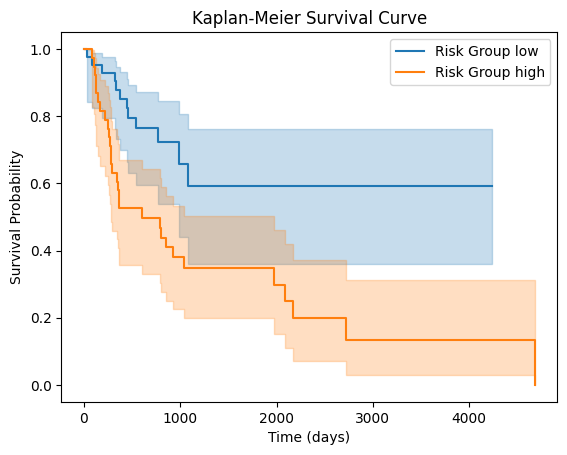


Fig.7. Kaplan-Meier Survival Curve Graph

## Benchmark Comparison

Brooks et al.’s HNSCC HS study[9] concluded that the HS did not possess independent prognostic value. In contrast, the AERO-HNSCC model’s log-rank test result, with a significantly lower p-value of 0.0021, suggests a strong association between the encoded data and OS, thereby indicating an improvement over traditional SIG methods in prognostic analyses. This outcome underscores the AERO-HNSCC model's potential as a more effective tool for prognosis in HNSCC.

Wu et al.[2] explored the prognostic capabilities of a proteomic signature developed through SIG methods, achieving an AUC of 0.779 in their ROC analysis. This sets a critical benchmark for our model. By constructing the AUC from the confusion matrix, our AERO-HNSCC model achieved an AUC of 0.73 in identifying high-risk and low-risk groups, demonstrating comparable performance. When using the proteomic signature they developed directly in the classification, we achieved an accuracy of 0.59, which is significantly lower than AERO-HNSCC’s 0.73, showcasing better performance by AERO-HNSCC in direct classification tasks.

Tan et al.[6] conducted a broader study with their pan-cancer AE-based method, inclusive of HNSCC, achieving an AUC of 0.7830. This figure is 6.85% higher than that of AERO-HNSCC. Given their use of more omics data and a supervised method, this difference is within expected bounds. However, the AERO-HNSCC model still presents significant applicability and effectiveness, particularly considering its focus on HNSCC and utilization in an unsupervised context.

In each of these comparisons, our AERO-HNSCC model demonstrates substantial merit, either by showcasing improvements over traditional methods or by providing competitive performance against more generalized approaches.

V. RNA-Seq Side

## A. Methodology

Finishing up the development of AERO-HNSCC on the protein expression data, similar approaches are tried on the RNA-Seq data.

***The methodology includes several key steps:***

1. Autoencoder Architecture: The development of the autoencoder model within this code begins with specifying the input shape to match the scaled training dataset's features. The architecture comprises multiple layers, starting with an input layer followed by a sequence of dense layers with decreasing units: 40, 30, 20, and finally 10, incorporating ReLU activation functions and L1 regularization to encourage sparse representations. Dropout layers with a rate of 0.1 are introduced after each dense layer to prevent overfitting. This structure leads to an encoding layer that effectively compresses the data. The decoder part of the autoencoder mirrors this structure in reverse, gradually increasing from 10 units back to the original feature space size, using ReLU activations for intermediate layers and a sigmoid activation in the output layer to reconstruct the input data. The model is compiled with the Adam optimizer and uses mean squared error as the loss function, aiming to minimize the difference between the input and reconstructed data.
2. Model Training: The training phase involves using a portion of the RNA-Seq data to teach the model, focusing on reducing the difference between the original data and its reassembled version from the encoded state.

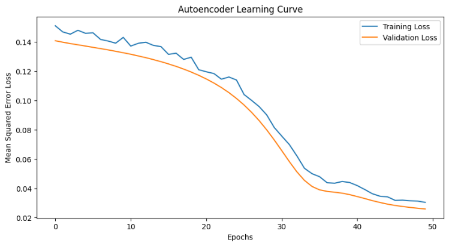


Fig Autoencoder Learning Curve

Fig "Autoencoder Training Progression" shows the autoencoder's mean squared error (MSE) rapidly decreases initially, indicating fast learning. The narrowing gap between training and validation losses suggests good generalization. When the decrease in losses peaks, it suggests an optimal learning point to prevent overfitting.

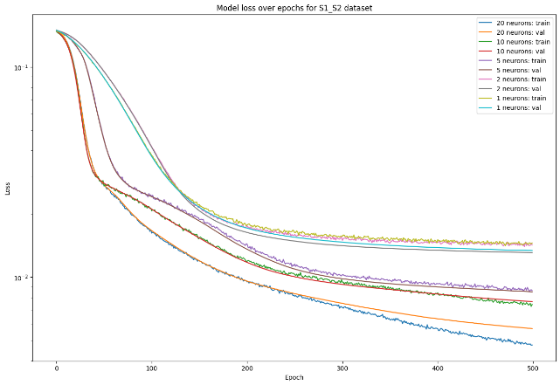


Fig. Model Loss Over Epochs for S1\_S2 Dataset

Fig "Epoch-wise Model Loss for S1\_S2 Data" illustrates that learning speed increases with neuron density up to a point, after which complexity diminishes the benefit, indicating an optimal neuron density for efficient learning. More complex models stabilize faster, while simpler ones learn slower and may overfit if training and validation losses diverge. Analysis combining the F1 score suggests 10 layers as optimal.

1. Model Evaluation: After training, the effectiveness of Autoencoder would be analyzed by reconstructing the data, and the result of clinical features.

Additionally, for establishing a baseline, the data is first processed with the XGboost model only. This would help provide a deeper understanding and give a better evaluation of how the autoencoder performs by comparing the autoencoder enhanced model with baseline.

## B. Dataset Description

The dataset used in this study consists of two main components: clinical data and RNA-Seq data.

***Clinical Data:***

The clinical dataset contains the detailed patients’ information like tumour stage, treatment result, demographics information, survival status and etc. This dataset plays an important role in aspects of clinical features associated with genetic variant and expression patterns observed in RNA-Seq data.

***RNA-Seq Data:***

High-throughput sequencing data intended to characterize the transcriptome landscape of HNSCC samples make up the RNA-Seq dataset. These data would play an important role in differentially expressed genes and understanding the molecular mechanisms of the disease. These data have been pre-processed like quality control, normalization, and differential expression analysis.

***Integration and Analysis:***

The integrated datasets were used throughout the study to explore the relationship between gene expression and clinical outcomes like survival status, employing Machine Learning methods such as XGboost and an enhanced autoencoder model to uncover patterns not evident through traditional analysis.

## C. Training

In our study, we developed a Keras-based Autoencoder (AE) for RNA-Seq data, aiming to predict patient survival from genetic features. Preprocessing involved merging RNA-Seq data with clinical information, followed by feature selection and normalization. Our AE architecture included dense layers with dropout and L1 regularization to prevent overfitting, creating a bottleneck that captures essential data characteristics.

The model, optimized with Adam and mean squared error loss, trained over 50 epochs. Post-training, we used the encoded features to train an XGBoost classifier, enhancing prediction accuracy for patient survival status. The process demonstrated the AE's efficacy in feature extraction and its potential in genomic data analysis, summarized by the training and validation loss curves to evaluate the learning progression.

## D. Metrics

***Accuracy for XGBoost and Enhanced Model Comparison:***

This study uses “live status” as the predict resul to do classification. Both training and predicting with XGboost model and enhanced autoencoder model. The accuracy is calculated by calculate the ratio of correct number predications over the total number of instances. This simple static method shows the proficiency of model predict survival status, and comparation of these two results would help evaluate the performance of enhanced autoencoder model.

***AUC-ROC for Enhanced Autoencoder Model Comparison:***

The performance of the enhanced autoencoder is also analysed by Area Under the Receiver Operating Characteristic (AUC-ROC) curve. The AUC-ROC curve shows a comparation of true positive rate and false positive rate under different threshold which shows the models’ ability of classifying distinct patient outcomes. The higher AUC-ROC value, the greater capacity of successful outcome.

## E. Benchmarks

Our model, using clinical and RNA-Seq data for predictions on patient outcomes, will be benchmarked against XGBoost for its effectiveness. We'll focus on accuracy, precision, recall, and AUC-ROC metrics to evaluate performance. Using this study as a starting point, a new model that sees an autoencoder used for feature extraction and dimensionality reduction, then classification using XGBoost, is proposed. These more complex and non-linear patterns in the data become the ultimate final run output of the proposed model that intends to increase the predictive performance over the base XGBoost model. The autoencoder with improvements that gives us the benchmark score and stand again the baseline with the same performance metrics. It allows to see if improvements have really happened.

## F. Resolution

In the AERO-HNSCC project, resolution reflects the level of detail and accuracy in data analysis and model evaluation, which is crucial for accurate risk stratification.

*Data Resolution:*

A significant amount of preprocessing is applied to RNA-Seq and clinical data, including normalization to reduce batch effects and variability, as well as feature selection to highlight relevant genes. This high-resolution data processing enables the detection of subtle patterns and relationships within the data.

***PCA Visualizations:***

An essential component of our approach in the AERO-HNSCC project is the use of Principal Component Analysis (PCA) to effectively reduce the dimensionality of the data while capturing its most significant differences.

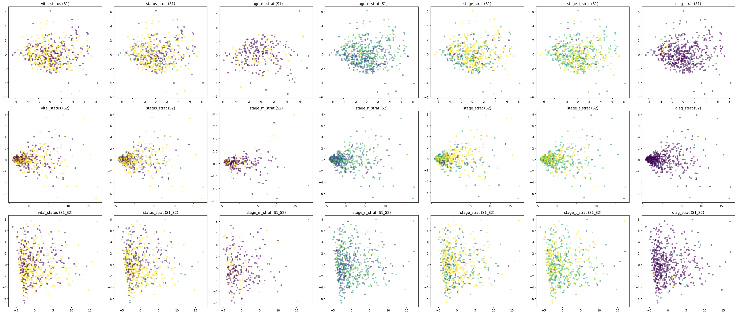


Fig PCA Visualization

The PCA visualization in the figure displays the first seven principal components across three datasets: S1, S2, and combined S1\_S2. Each scatter plot demonstrates how data points from the datasets are distributed and clustered in a lower-dimensional space, illustrating patterns that are less discernible in higher dimensions. Specifically, the PCA sheds light on how life status and various stages of cancer (stage\_m\_strat, stage\_n\_strat, stage\_strat, stage\_t\_strat, and diag\_strat) cluster and separate within the datasets. These visual patterns provide insights into the intrinsic structure and grouping of the data based on these critical features, which is especially beneficial for understanding underlying trends and preparing for more advanced analytical models. This elucidation of feature distribution is vital for subsequent modeling and predictive analytics, aiding in the identification of the most informative features for machine learning applications.

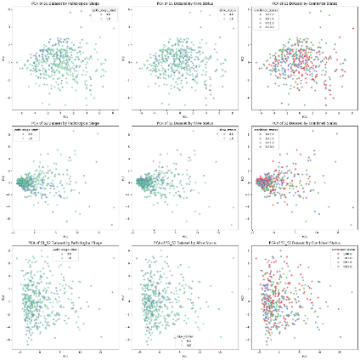


Fig PCA for Combined Status

Fig PCA for Combined Status: In contrast, this figure showcases the PCA performed on two selected features deemed most informative for the autoencoder's functioning: the cancer stage (limited to stages 1-3) and patient vital status. The visualization displays how the reduced features space can differentiate patients based on the combined status, which correlates to the model's ability to condense and encode clinically relevant information for risk stratification. The distinct clusters formed in the PCA space indicate the model's effectiveness in identifying and separating cases with different clinical outcomes.

These PCA visualizations play a dual role. Firstly, they verify the variance preservation post-dimensionality reduction, crucial for subsequent modeling. Secondly, they offer a graphical demonstration of the potential of the autoencoder to distill key features that could aid in separating patients into clinically meaningful groups. The insight gained from these visualizations directly feeds into the model's ability to simplify complex multidimensional data into actionable insights, crucial for the accurate prediction and stratification of HNSCC risks.

***Model Architecture Resolution:***

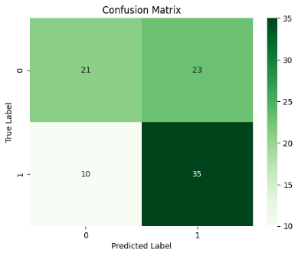
The autoencoder model features multiple layers to extract high-resolution features, with an encoder that reduces data dimensionality and a decoder that reconstructs the data, facilitating the learning of complex patterns. This architecture's resolution is fine-tuned through hyperparameter optimization for effective feature capture.

The integration of detailed PCA visualizations with precise data processing and model evaluation ensures a thorough and nuanced analysis of the gene expression patterns and their clinical significance in head and neck squamous cell carcinoma within the AERO-HNSCC project.

## G. Analysis of RNA Results

The evaluation of our predictive models goes beyond mere accuracy metrics. Here we visualize a Confusion Matrix and ROC Curve to depict expected performance of our models on clinical data and real-life circumstances which are both the primary interest to clinical researchers.

The improved model's Confusion Matrix displays a symmetrical perception of two classes with minimal False Negatives, crucial for clinical applications where sensitivity towards correctly identifying positive patients is paramount. The autoencoder shows significant improvement in sensitivity, indicating its ability to handle complex data patterns.



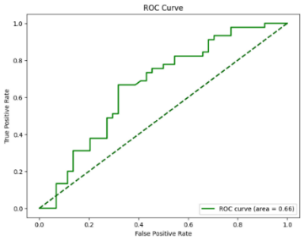


Fig Confusion Matrix Fig ROC Curve

To sum it up, the ROC curve bolsters the model's discrimination ability (fair) between positive and negative classes, with AUC value of 0.66 further validating (AUC = 0. 5, meaning random performance) the goodness of the model seen from the logistic regression model used in the last tutorial. The plot is just a visual demonstration of sensitivity against specificity trade-off within a model. In a clinical setting, the curve shows its relevance for the differentiated thresholds adopted.

***Improvement Over XGBoost Baseline:***

In the outset, the implementation of a XGBoost approach with S1\_S2 (S1 dataset, S2 dataset) showed a decent level of accuracy which was equal to 51.69%. The advantage of applying XGBoost without data preprocessing could be found to reach an accuracy rate that was considerably low, 51.69%. However, achieving a higher data cleaning accuracy was evidenced through employing Autoencoder reprocessing model which scored a 62.92% accuracy rate.

***Comparison with Previous Studies:***

It is worth noting that in the abovementioned previous pieces of research that were mostly based on papers by TCGA HNSCC dataset, the average precision of purity prediction was about 73%, giving an estimate of survival rate. Our results, not being directly transferable because of different ways of calculation carried out, points out the enormousness of clear and transparent graphing and validation of the data in the settings of disease risk stratification and prognosis.

The visualized matrix and curve in the context of Confusion Matrix and ROC Curve prove that our methodology is powerful enough to regulate and determine the sensitivity of the model. They imply that the Autoencoder has the feature abstraction, which helps the model to produce not only accurate but more precise prediction.

VI. Discussion & Limitations

The AERO-HNSCC model, developed for protein analysis, illustrates a significant step forward in utilizing autoencoder-based methodologies for oncological prognostication. However, the shift towards SAE or SDAE did not yield the anticipated improvements. This could be attributed to the specific nature of SAEs, which are designed to address high-dimensionality issues primarily in datasets with a larger feature-to-case ratio. Given that the RPPA data comprises 468 protein expressions across 324 cases, the structure may not have been optimal for SAE application, possibly due to an insufficient level of inherent sparsity or noise within the dataset. Another limitation of current AERO-HNSCC is developed on the dataset of one single type of cancer, which restricts the case diversity and number. Expanding the application to multiple cancers could offer a richer dataset, potentially enhancing the AE’s learning capability and overall model performance.

On the RNA-seq side, the study faces a significant constraint in data volume after preprocessing, the limitation of data subset is one of the contrarians of this study. After data prepressing, we would result with only 520 samples of sample cases for future machine learning study. This limitation may result in overfitting and extrapolation of study result to a large context are brought up. And besides from the restrictions of data set, we only have 50 features in total, which would further limit the model ability to gain an accurate result. Although these data have been chosen carefully but the tiny number of features may not be enough to generate a confident result with the machine learning model. However, when attempting to process RNA-seq datasets with a larger number of features, such as an original dataset containing 60,660 features, constructing a comprehensive AE structure requires substantial computational resources. The high demand for runtime memory means that even using Colab's A100 cannot prevent overfitting during training. Therefore, there's a need for improvements in the AE structure to perform more efficiently and adapt to larger databases.

Both segments of the study highlight the critical importance of data diversity and volume in developing robust predictive models. Addressing these limitations could involve broadening the dataset to include more diverse cases and features, which could improve the predictive accuracy and generalizability of the models across various cancer types.

VII. Conclusion and Future Work

The AERO-HNSCC workflow is designed to operate independently, relying solely on multi-omics datasets. This autonomous design not only streamlines its application but also holds potential to extend its utility beyond existing limitations. Unlike approaches that require detailed knowledge of specific genes or proteins, AERO-HNSCC could aid in uncovering new oncological biomarkers by conducting sensitivity analysis to determine the biomarker contribute most to the encoded result, as demonstrated by the "Top Gene" methodology developed by Mondol et. al.[8].

Building on the findings of Tan et. al.[6], we have validated the efficacy of AE in analyzing pan-cancer datasets. Given the inherent capabilities of AEs, there is a promising avenue for AERO-HNSCC to be applicable across a range of cancer types, potentially evolving into a universal tool for cancer analysis. However, this hypothesis requires thorough future investigations for comprehensive validation. Currently, our focus on a single cancer type limits the diversity of cases. Expanding this scope to include multiple cancer types could significantly increase the dataset size, thereby enhancing the AE’s ability to learn more accurate data representations and potentially improving overall performance.

Furthermore, there is an opportunity to integrate additional multi-omics data into AERO-HNSCC. At present, the system utilizes only clinical and protein expression data. Echoing the approach by Tan et al.[6], we propose developing distinct AEs for each omics type and then combining these for classification. Although Tan et al. assumed equal informational value across different omics dimensions—an assumption yet to be confirmed—their framework for omics integration merits further investigation.

# Appendix

The code and the supplementary materials of this work is open sourced on GitHub and can be access through:

<https://github.com/smgjch/HSAE>

<https://github.com/LiiiiGhost/Hypoxia-Autoencoder>

# Acknowledgment

The authors wish to thank Professor D. Fernandez-Reyes for his invaluable support, encouragement, and guidance throughout this research.

# REFERENCE

[1] T. Kanazawa, K. Misawa, K. Shinmura, Y. Misawa, G. Kusaka, M. Maruta, T. Sasaki, Y. Watanabe, and T. E. Carey, “Promoter methylation of galanin receptors as epigenetic biomarkers for head and neck squamous cell carcinomas,” *Expert Review of Molecular Diagnostics,* vol. 19, no. 2, pp. 137-148, 2019/02/01, 2019.

[2] D. Wu, P. Gong, Q. Zeng, W. Zhang, F. Xie, and X. Zhou, “Prognostic implication of proteomic profiles in head and neck squamous cell carcinoma,” *Clinica Chimica Acta,* vol. 509, pp. 304-309, 2020/10/01/, 2020.

[3] R. Abou Khouzam, K. Brodaczewska, A. Filipiak, N. A. Zeinelabdin, S. Buart, C. Szczylik, C. Kieda, and S. Chouaib, “Tumor Hypoxia Regulates Immune Escape/Invasion: Influence on Angiogenesis and Potential Impact of Hypoxic Biomarkers on Cancer Therapies,” *Frontiers in Immunology,* vol. 11, 2021-January-20, 2021.

[4] B. Tawk, C. Schwager, O. Deffaa, G. Dyckhoff, R. Warta, A. Linge, M. Krause, W. Weichert, M. Baumann, C. Herold-Mende, J. Debus, and A. Abdollahi, “Comparative analysis of transcriptomics based hypoxia signatures in head- and neck squamous cell carcinoma,” *Radiotherapy and Oncology,* vol. 118, no. 2, pp. 350-358, 2016/02/01/, 2016.

[5] A. Košec, R. Novak, P. Konjevoda, V. Trkulja, V. Bedeković, and L. Grgurević, “Tumor tissue hnRNP M and HSP 90&#x03B1; as potential predictors of disease-specific mortality in patients with early-stage cutaneous head and neck melanoma: A proteomics-based study,” *Oncotarget; Vol 10, No 62*, 2019.

[6] K. Tan, W. Huang, J. Hu, and S. Dong, “A multi-omics supervised autoencoder for pan-cancer clinical outcome endpoints prediction,” *BMC Med Inform Decis Mak,* vol. 20, no. Suppl 3, pp. 129, Jul 9, 2020.

[7] H. Song, C. Ruan, Y. Xu, T. Xu, R. Fan, T. Jiang, M. Cao, and J. Song, “Survival stratification for colorectal cancer via multi-omics integration using an autoencoder-based model,” *Exp Biol Med (Maywood),* vol. 247, no. 11, pp. 898-909, Jun, 2022.

[8] R. K. Mondol, N. D. Truong, M. Reza, S. Ippolito, E. Ebrahimie, and O. Kavehei, “AFExNet: An Adversarial Autoencoder for Differentiating Breast Cancer Sub-Types and Extracting Biologically Relevant Genes,” *IEEE/ACM Trans Comput Biol Bioinform,* vol. 19, no. 4, pp. 2060-2070, Jul-Aug, 2022.

[9] J. M. Brooks, A. N. Menezes, M. Ibrahim, L. Archer, N. Lal, C. J. Bagnall, S. V. von Zeidler, H. R. Valentine, R. J. Spruce, N. Batis, J. L. Bryant, M. Hartley, B. Kaul, G. B. Ryan, R. Bao, A. Khattri, S. P. Lee, K. U. E. Ogbureke, G. Middleton, D. A. Tennant, A. D. Beggs, J. Deeks, C. M. L. West, J. B. Cazier, B. E. Willcox, T. Y. Seiwert, and H. Mehanna, “Development and Validation of a Combined Hypoxia and Immune Prognostic Classifier for Head and Neck Cancer,” *Clin Cancer Res,* vol. 25, no. 17, pp. 5315-5328, Sep 1, 2019.

[10] J. Shao, B. Zhang, L. Kuai, and Q. Li, “Integrated analysis of hypoxia-associated lncRNA signature to predict prognosis and immune microenvironment of lung adenocarcinoma patients,” *Bioengineered,* vol. 12, no. 1, pp. 6186-6200, Dec, 2021.

[11] B. Lane, M. T. Khan, A. Choudhury, A. Salem, and C. M. L. West, “Development and validation of a hypoxia-associated signature for lung adenocarcinoma,” *Sci Rep,* vol. 12, no. 1, pp. 1290, Jan 25, 2022.

[12] F. Zeng, Y. Zhang, X. Han, M. Zeng, Y. Gao, and J. Weng, “Employing hypoxia characterization to predict tumour immune microenvironment, treatment sensitivity and prognosis in hepatocellular carcinoma,” *Computational and Structural Biotechnology Journal,* vol. 19, pp. 2775-2789, 2021/01/01/, 2021.

[13] B. Zhang, B. Tang, J. Gao, J. Li, L. Kong, and L. Qin, “A hypoxia-related signature for clinically predicting diagnosis, prognosis and immune microenvironment of hepatocellular carcinoma patients,” *J Transl Med,* vol. 18, no. 1, pp. 342, Sep 4, 2020.

[14] Y. Shou, L. Yang, Y. Yang, X. Zhu, F. Li, and J. Xu, “Determination of hypoxia signature to predict prognosis and the tumor immune microenvironment in melanoma,” *Mol Omics,* vol. 17, no. 2, pp. 307-316, Apr 1, 2021.

[15] X. Tian, J. Zheng, W. Mou, G. Lu, S. Chen, J. Du, Y. Zheng, S. Chen, B. Shen, J. Li, and N. Wang, “Development and validation of a hypoxia-stemness-based prognostic signature in pancreatic adenocarcinoma,” *Front Pharmacol,* vol. 13, pp. 939542, 2022.

[16] K. Wang, Y. Lu, Z. Liu, M. Diao, and L. Yang, “Establishment and External Validation of a Hypoxia-Derived Gene Signature for Robustly Predicting Prognosis and Therapeutic Responses in Glioblastoma Multiforme,” *Biomed Res Int,* vol. 2022, pp. 7858477, 2022.

[17] Z. Liu, Q. Tang, T. Qi, B. Othmane, Z. Yang, J. Chen, J. Hu, and X. Zu, “A Robust Hypoxia Risk Score Predicts the Clinical Outcomes and Tumor Microenvironment Immune Characters in Bladder Cancer,” *Frontiers in Immunology,* vol. 12, 2021-August-13, 2021.

[18] C. Nie, H. Qin, and L. Zhang, “Identification and validation of a prognostic signature related to hypoxic tumor microenvironment in cervical cancer,” *PLOS ONE,* vol. 17, no. 6, pp. e0269462, 2022.

[19] J. Zhong, H. Wu, X. Bu, W. Li, S. Cai, M. Du, Y. Gao, and B. Ping, “Establishment of Prognosis Model in Acute Myeloid Leukemia Based on Hypoxia Microenvironment, and Exploration of Hypoxia-Related Mechanisms,” *Frontiers in Genetics,* vol. 12, 2021-October-26, 2021.

[20] X. Yang, X. Weng, Y. Yang, M. Zhang, Y. Xiu, W. Peng, X. Liao, M. Xu, Y. Sun, and X. Liu, “A combined hypoxia and immune gene signature for predicting survival and risk stratification in triple-negative breast cancer,” *Aging (Albany NY),* vol. 13, no. 15, pp. 19486-19509, Aug 2, 2021.

[21] R. Shi, X. Bao, K. Unger, J. Sun, S. Lu, F. Manapov, X. Wang, C. Belka, and M. Li, “Identification and validation of hypoxia-derived gene signatures to predict clinical outcomes and therapeutic responses in stage I lung adenocarcinoma patients,” *Theranostics,* vol. 11, no. 10, pp. 5061-5076, 2021.

[22] Q. Zhang, L. Qiao, J. Liao, Q. Liu, P. Liu, and L. Liu, “A novel hypoxia gene signature indicates prognosis and immune microenvironments characters in patients with hepatocellular carcinoma,” *J Cell Mol Med,* vol. 25, no. 8, pp. 3772-3784, Apr, 2021.

[23] D. Chen, H. Huang, L. Zang, W. Gao, H. Zhu, and X. Yu, “Development and Verification of the Hypoxia- and Immune-Associated Prognostic Signature for Pancreatic Ductal Adenocarcinoma,” *Front Immunol,* vol. 12, pp. 728062, 2021.

[24] X. Zhang, Y. Li, and Y. Chen, “Development of a Comprehensive Gene Signature Linking Hypoxia, Glycolysis, Lactylation, and Metabolomic Insights in Gastric Cancer through the Integration of Bulk and Single-Cell RNA-Seq Data,” *Biomedicines,* vol. 11, no. 11, Nov 1, 2023.

[25] Y. Liu, J. Wu, W. Huang, S. Weng, B. Wang, Y. Chen, and H. Wang, “Development and validation of a hypoxia-immune-based microenvironment gene signature for risk stratification in gastric cancer,” *Journal of Translational Medicine,* vol. 18, no. 1, pp. 201, 2020/05/14, 2020.

[26] J. Li, H. Qiao, F. Wu, S. Sun, C. Feng, C. Li, W. Yan, W. Lv, H. Wu, M. Liu, X. Chen, X. Liu, W. Wang, Y. Cai, Y. Zhang, Z. Zhou, Y. Zhang, and S. Zhang, “A novel hypoxia- and lactate metabolism-related signature to predict prognosis and immunotherapy responses for breast cancer by integrating machine learning and bioinformatic analyses,” *Frontiers in Immunology,* vol. 13, 2022-October-07, 2022.

[27] Madhumita, and S. Paul, “Capturing the latent space of an Autoencoder for multi-omics integration and cancer subtyping,” *Computers in Biology and Medicine,* vol. 148, pp. 105832, 2022/09/01/, 2022.

[28] J. G. Ellen, E. Jacob, N. Nikolaou, and N. Markuzon, “Autoencoder-based multimodal prediction of non-small cell lung cancer survival,” *Scientific Reports,* vol. 13, no. 1, pp. 15761, 2023/09/22, 2023.

[29] A. Arafa, N. El-Fishawy, M. Badawy, and M. Radad, “RN-Autoencoder: Reduced Noise Autoencoder for classifying imbalanced cancer genomic data,” *Journal of Biological Engineering,* vol. 17, no. 1, pp. 7, 2023/01/30, 2023.

[30] E. Hoffer, R. Banner, I. Golan, and D. Soudry, “Norm matters: efficient and accurate normalization schemes in deep networks,” 03/05, 2018.