Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Adaptive branch-gated fusion of dual autoencoder latent representations improves head and neck squamous cell carcinoma classification and supports exploratory candidate gene prioritization

Aneela Nargis

aneela.nargis@live.com

DHA Suffa University

Shama Siddiqui

Dundalk Institute of Technology

Muhammad Shoaib Siddiqui

Islamic University of Madinah

Article

Keywords: head and neck squamous cell carcinoma, transcriptomic classification, RNA-seq, denoising undercomplete autoencoder, variational autoencoder, adaptive gated fusion, deep learning, candidate gene prioritization

Posted Date: September 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9156772/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Adaptive branch-gated fusion of dual autoencoder latent representations improves head and neck squamous cell carcinoma classification and supports exploratory candidate gene prioritization

Aneela Nargis[1], Shama Siddiqui[2], Muhammad Shoaib Siddiqui[3]

[1] Computer Science, DHA Suffa University, Karachi, Sindh, Pakistan
[2] Computer Science, Dundalk Institute of Technology, Dublin, Ireland
[3] Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah, Kingdom of Saudi Arabia
Corresponding Author: Aneela Nargis (aneela.nargis@live.com)

Abstract

Head and neck squamous cell carcinoma (HNSCC) is a molecularly heterogeneous malignancy for which robust transcriptome-based classification remains challenging because of the high dimensionality of RNA sequencing data and the limited number of available normal samples. In this study, we developed an adaptive deep learning framework that integrates complementary latent representations from a denoising undercomplete autoencoder (DUAE) and a variational autoencoder (VAE) for binary classification of tumor and normal transcriptomic profiles in HNSCC. RNA-seq data from the TCGA-HNSC cohort was preprocessed through label harmonization, gene symbol normalization, log1p transformation, and z-score standardization. The proposed Adaptive Gated Fusion Expert Model (AGFEM) used sample-specific pathway-level attention weights to dynamically reweight DUAE and VAE latent embedding before downstream classification. Model performance was assessed using five-fold stratified cross-validation and compared against DUAE-only, VAE-only, and raw-feature multilayer perceptron baselines using accuracy, area under the receiver operating characteristic curve (ROC-AUC), precision, recall, and F1-score. Statistical significance of performance differences was evaluated using Wilcoxon signed-rank testing on fold-wise metrics and DeLong testing on pooled out-of-fold ROC predictions. To assess cross-cohort transferability, an independent GEO cohort (GSE65858) was harmonized through probe-to-gene mapping and projected into the learned latent space. Candidate genes were prioritized using model-derived feature importance from the DUAE encoder and further examined using pathway enrichment analysis. Overall, these findings support adaptive fusion of deterministic and probabilistic transcriptomic representations as a useful strategy for internally validated HNSCC classification, while also enabling exploratory candidate gene prioritization for downstream biological follow-up.

Keywords: head and neck squamous cell carcinoma; transcriptomic classification; RNA-seq; denoising undercomplete autoencoder; variational autoencoder; adaptive gated fusion; deep learning; candidate gene prioritization

1. Introduction

Head and neck squamous cell carcinoma (HNSCC) is among the most prevalent malignancies worldwide [1] and remains a major contributor to cancer-associated morbidity and mortality [1 - 3]. It comprises a

biologically diverse group of tumors arising from the mucosal epithelium of the oral cavity, pharynx, and larynx, and is characterized by marked inter-patient heterogeneity in molecular profiles, therapeutic response, and clinical outcomes [4, 5]. Despite advances in molecular oncology, the development of robust transcriptome-based computational models for HNSCC remains challenging [6], because tumor and normal samples exhibit complex, nonlinear expression patterns across tens of thousands of genes [6, 7]. These challenges are further amplified by the high dimensionality of RNA sequencing (RNA-seq) data [8, 9], the limited availability of normal control samples [10, 11], and the susceptibility of conventional machine learning methods to overfitting in small-sample, high-feature settings [12].

Large-scale public repositories such as The Cancer Genome Atlas (TCGA) [13] have enabled systematic analysis of cancer transcriptomes and have accelerated the application of data-driven approaches to molecular classification [13, 14]. In particular, the TCGA-HNSC cohort provides a valuable resource for studying gene expression alterations associated with head and neck cancer [15]. However, the direct use of high-dimensional RNA-seq features in supervised learning pipelines is often suboptimal due to redundancy, noise, sparsity, and nonlinear interactions among genes [16]. As a result, representation learning has become increasingly important in transcriptomic cancer classification, where the goal is not only to improve predictive performance but also to extract compact, biologically meaningful latent structure from the data.

Autoencoder-based deep learning architectures are well suited to this setting because they can learn lower-dimensional latent embeddings that preserve salient information from complex biological data [17 - 19]. Deterministic autoencoders, particularly undercomplete variants, are effective at compressing high-dimensional inputs into compact representations while suppressing redundant structure [17, 20]. When regularized appropriately, such models can improve robustness to noise and enhance downstream discrimination [20, 21]. Variational autoencoders (VAEs), in contrast, impose a probabilistic structure on the latent space by learning a distribution over hidden representations rather than a single deterministic embedding [22, 23]. This probabilistic formulation can improve latent smoothness, regularization, and generalization, particularly in biologically heterogeneous datasets [24 - 26]. Although both approaches have shown promise in omics applications, they capture different inductive biases. Deterministic compression primarily emphasizes reconstruction fidelity and robust structural encoding, whereas variational encoding promotes distribution-aware representation learning and latent regularization [27, 28].

A key limitation in existing transcriptomic classification studies is that models often rely on a single latent representation or use simple feature concatenation when combining multiple encoders [29 - 31]. Such approaches may fail to fully exploit the complementary strengths of different representation learning paradigms. In the context of HNSCC, where transcriptomic signals may vary substantially across samples, a static fusion strategy may be insufficient. A more effective approach is to allow the model to adaptively determine how much information to draw from each encoder on a per-sample basis [32 - 35]. This is especially relevant in cancer transcriptomics, where tumor profiles may contain heterogeneous degrees of noise, dysregulation, and latent variability that are not uniformly captured by a single architecture [36, 37]. Prior hybrid architectures in HNSCC transcriptomic classification, including autoencoder-integrated convolutional or residual frameworks, have demonstrated the benefit of learned latent representations but have generally relied on fixed downstream architectures rather than sample-adaptive latent fusion [38].

To address this challenge, we propose an Adaptive Gated Fusion Expert Model (AGFEM) that integrates latent representations from a denoising undercomplete autoencoder (DUAE) and a variational autoencoder (VAE) using a sample-specific gating mechanism. Rather than relying on fixed fusion or naive concatenation alone, the proposed framework learns pathway-level attention weights that dynamically reweight the DUAE and VAE latent spaces for each input sample before classification. This design is motivated by the hypothesis that deterministic and probabilistic latent encoders capture complementary transcriptomic structure, and that adaptive weighting of these pathways can improve both predictive performance and representational flexibility in HNSCC classification.

In addition to evaluating predictive performance, it is important that deep learning models in cancer transcriptomics provide biologically interpretable outputs. Purely black-box classification is often insufficient for translational applications, where researchers seek not only accurate predictions but also insight into genes and pathways associated with disease states [39 - 41]. Therefore, in this study, we complement model evaluation with candidate gene prioritization based on learned model parameters and perform functional enrichment analysis to examine whether the prioritized features align with biologically relevant processes. Furthermore, because reproducibility and generalizability are central concerns in transcriptomic machine learning, we also assess cross-cohort transferability by harmonizing an independent Gene Expression Omnibus (GEO) cohort and projecting it into the learned latent feature space after probe-to-gene mapping and feature alignment.

The present study makes four primary contributions. First, we develop a deep learning framework that combines deterministic and probabilistic transcriptomic encoders through a sample-adaptive gating strategy for HNSCC classification. Second, we perform a controlled comparative evaluation against DUAE-only, VAE-only, and raw-feature multilayer perceptron baselines under the same five-fold stratified cross-validation protocol. Third, we assess the statistical robustness of observed performance gains using Wilcoxon signed-rank testing and DeLong ROC significance testing. Fourth, we extend the analysis beyond classification by performing cross-cohort latent-space transfer analysis and candidate gene prioritization with pathway enrichment to support biological interpretability.

Overall, this work investigates whether adaptive fusion of complementary autoencoder latent spaces can provide a more robust and biologically informative framework for transcriptomic cancer classification. By focusing on HNSCC as a clinically relevant and molecularly heterogeneous use case, we aim to establish a reproducible deep learning pipeline that balances predictive performance, statistical rigor, and interpretability while remaining extensible to future multi-cohort and multi-omics studies.

2. Materials and methods

2.1 Study design overview

This study developed and evaluated a deep learning framework for transcriptomic classification of head and neck squamous cell carcinoma (HNSCC) using RNA sequencing (RNA-seq) data. The proposed pipeline combined two unsupervised representation learning models: a denoising undercomplete

autoencoder (DUAE) and a variational autoencoder (VAE), with a downstream adaptive fusion classifier termed the Adaptive Gated Fusion Expert Model (AGFEM). The overall workflow comprised: (i) acquisition and preprocessing of TCGA-HNSC gene expression data, (ii) optional cross-cohort harmonization with an external GEO dataset through probe-to-gene mapping, (iii) learning of complementary latent embeddings using DUAE and VAE, (iv) adaptive pathway-level fusion of latent representations for binary tumor-versus-normal classification, (v) comparison against baseline models under a unified cross-validation protocol, (vi) statistical testing of performance differences, and (vii) candidate gene prioritization with functional enrichment analysis.

2.2 TCGA-HNSC cohort and label processing

RNA-seq gene expression data for HNSCC were obtained from a TCGA-derived expression matrix [42] in which rows represented samples and columns represented gene-level expression features. A sample annotation column labeled `Tumor.Type` was used to derive binary class labels. Samples annotated as Primary solid Tumor were encoded as class 1, and samples annotated as Solid Tissue Normal were encoded as class 0.

The final processed TCGA matrix comprised a binary-labeled transcriptomic dataset containing 564 samples and 20,503 gene-level features after label filtering, gene symbol normalization, duplicate removal, and numeric coercion. The two supervised classes used throughout the study were Primary solid Tumor and Solid Tissue Normal. The exact class distribution between tumor and normal samples is reported in the Results section together with the final cross-cohort harmonization summary.

2.3 GEO cohort acquisition and cross-cohort feature harmonization

To assess cross-cohort transferability of the learned latent representations, an independent external dataset was incorporated from the Gene Expression Omnibus (GEO) [43]. Specifically, the GEO series matrix file (GSE65858_series_matrix.txt) and the corresponding platform annotation file (GPL10558.annot) were used for external cohort parsing and probe-to-gene mapping, respectively. Because GEO series matrix files can contain variable formatting, a robust parsing procedure was implemented to support tab-delimited, comma-delimited, and mixed metadata-containing files. Probe identifiers were extracted from the platform-level feature rows, while sample identifiers were parsed from columns corresponding to GEO sample accessions.

Since the GEO cohort was initially represented at the probe level rather than the gene-symbol level, probe-to-gene mapping was performed using a platform annotation file [44]. A dedicated annotation parsing routine was used to support SOFT-style platform tables and delimited tabular annotation files. Probe identifiers were matched to gene symbols using platform annotation fields, and invalid or ambiguous gene entries were excluded. For probes mapping to the same gene, expression values were collapsed by averaging across probes. In this way a gene-level GEO matrix aligned to gene symbols was generated.

After gene-level harmonization, the intersection of common genes between the processed TCGA matrix and the mapped GEO matrix was computed. Only shared genes were retained for cross-cohort alignment. The GEO cohort was then transformed using the same feature-space alignment strategy as the TCGA cohort. Because sample-level labels were not incorporated for the GEO cohort in the present analysis, this dataset was used for cross-cohort latent-space transfer analysis rather than formal external performance validation.

2.4 Data preprocessing

Gene expression preprocessing was performed in a consistent manner prior to model fitting. For TCGA samples, expression values were first clipped at a minimum of zero to avoid invalid logarithmic transformation of negative values, followed by application of the `log1p` transformation, defined in Equation (1):

$$x' = \log(1 + x) \qquad \text{Equation (1)}$$

The transformation was used to stabilize variance and reduce the influence of extreme expression values. Following logarithmic transformation, z-score standardization was applied to each gene feature using the `StandardScaler` procedure to center features at zero mean and unit variance.

For cross-cohort alignment, the GEO gene-level matrix underwent the same `log1p` transformation and was then transformed using the scaler fitted on the TCGA feature space after gene intersection. Missing or non-finite values were replaced with zeros after transformation. Final matrices were converted to 32-bit floating point format for efficient neural network training.

2.4.1 Leakage control and fold-wise preprocessing

To prevent information leakage during model evaluation, all preprocessing and representation learning steps were performed independently within each cross-validation fold. For each fold, the data were first partitioned into training and held-out test subsets using stratified sampling. Feature standardization was fitted exclusively on the training partition and subsequently applied to the corresponding held-out fold. The DUAE and VAE models were trained only on the training partition of each fold, and latent representations for the held-out fold were generated using the fold-specific trained encoders. The downstream classifiers, including the DUAE-only classifier, VAE-only classifier, Raw-MLP baseline, and AGFEM, were likewise trained only on the training partition and evaluated only on the held-out fold. No test-fold information was used during feature scaling, latent representation learning, classifier training, model selection, or probability thresholding. The fold-wise isolation was implemented to ensure an unbiased estimate of generalization performance in the high-dimensional transcriptomic setting.

2.5 Denoising undercomplete autoencoder (DUAE)

A denoising undercomplete autoencoder (DUAE) was used to learn a compact deterministic representation of high-dimensional transcriptomic data. The DUAE consisted of a fully connected

encoder–decoder architecture. The encoder received the input gene expression vector and passed it through three hidden layers with ReLU activation functions of sizes 1024, 512, and 256, followed by a 128-dimensional bottleneck latent layer. Dropout regularization (rate = 0.3) was applied in the encoder to improve robustness and reduce overfitting, serving a denoising role during representation learning.

The decoder mirrored the compression pathway using hidden layers of sizes 256, 512, and 1024, followed by a linear reconstruction layer matching the original input dimensionality. The model was trained using the Adam optimizer (learning rate = 1×10^{-3}) and optimized with mean squared error (MSE) reconstruction loss is given in Equation (2):

$$L_{DUAE} = \frac{1}{n} \sum_{i=1}^{n} ||x_i - \hat{x}_i||_2^2 \quad \text{Equation (2)}$$

Where, x_i is the original input vector and $x\hat{}i$ is the reconstructed output. After training, the 128-dimensional latent bottleneck was extracted as the deterministic DUAE representation for downstream classification.

2.6 Variational autoencoder (VAE)

A variational autoencoder (VAE) was implemented to learn a complementary probabilistic latent representation of the transcriptomic data. The VAE encoder used the same fully connected hidden-layer progression as the DUAE (1024 → 512 → 256), followed by two parallel latent projection layers that estimated the mean vector μ and log-variance vector $log\sigma 2$ of a 128-dimensional Gaussian latent distribution.

Latent sampling was performed using the reparameterization trick given in Equation (3):

$$z = \mu + \sigma \odot \epsilon, \epsilon \sim N(0, I) \quad \text{Equation (3)}$$

Where, $\sigma = exp(0.5 log\sigma 2)$, a custom sampling layer was used to incorporate the Kullback–Leibler (KL) divergence term directly into the model loss during training. The decoder mirrored the encoder using fully connected layers of sizes 256, 512, and 1024, followed by a linear reconstruction layer.

The total VAE objective consisted of reconstruction loss and a weighted KL regularization term given in Equation (4):

$$L_{VAE} = L_{recon} + \lambda L_{KL} \quad \text{Equation (4)}$$

Where, L_{recon} was the MSE reconstruction loss, L_{KL} was the KL divergence between the learned latent posterior and the standard normal prior, and λ=0.001 was the KL weighting coefficient. For downstream classification, the deterministic latent mean vector μ was used as the VAE representation to improve stability and reproducibility of feature extraction.

2.7 Adaptive Gated Fusion Expert Model (AGFEM)

The proposed classifier, termed the Adaptive Gated Fusion Expert Model (AGFEM), was designed to integrate complementary latent representations from the DUAE and VAE in a sample-adaptive manner. The 128-dimensional DUAE latent vector and the 128-dimensional VAE latent vector were provided as two separate inputs to the fusion model.

First, the two latent vectors were concatenated and passed through a fully connected hidden layer of size 128 with ReLU activation. The intermediate representation was then fed into a softmax gating layer that produced two normalized pathway-level attention weights per sample: one weight for the DUAE pathway and one weight for the VAE pathway. Unlike feature-wise gating, this mechanism implemented sample-specific pathway weighting, enabling the model to adaptively determine the relative contribution of deterministic and probabilistic latent spaces for each input sample.

The resulting scalar pathway weights were applied to the DUAE and VAE latent vectors through element-wise scaling of each full latent vector by its corresponding pathway weight. The weighted DUAE and VAE representations were then concatenated to form a fused latent representation. This fused vector was passed through a classification head consisting of a 128-unit ReLU layer, dropout regularization (rate = 0.3), a 64-unit ReLU layer, and a final sigmoid output neuron for binary classification.

The AGFEM was trained using binary cross-entropy loss as given in Equation (5):

$$L_{AGFEM} = -\frac{1}{N}\sum_{i=1}^{N}[y_i\, log(\hat{y}i) + (1 - y_i) log(1 - \hat{y}i)] \qquad \text{Equation (5)}$$

Where, $y_i$is the ground-truth label and $\hat{y}i$ is the predicted tumor probability.

The overall architecture of the proposed AGFEM is illustrated in Figure 1, where the DUAE and VAE latent representations are first combined to derive an adaptive fusion context, followed by branch-wise weighting, weighted concatenation, and downstream classification.

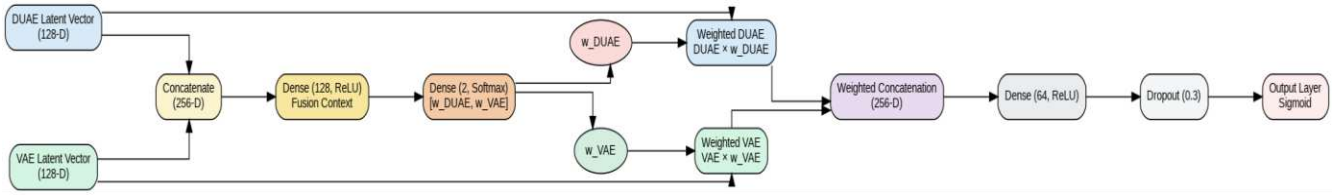


Figure 1. Architecture of the proposed Adaptive Gated Fusion Expert Model (AGFEM)

Schematic overview of the proposed Adaptive Gated Fusion Expert Model (AGFEM). The model receives two complementary latent representations derived from the Deep Undercomplete Autoencoder (DUAE) and Variational Autoencoder (VAE), each of dimensionality 128. These latent vectors are first concatenated into a 256-dimensional fusion input and passed through a fusion context layer to learn adaptive branch weights using a softmax gating mechanism. The DUAE and VAE latent vectors are then weighted accordingly, followed by weighted concatenation to form the fused representation. The fused

latent feature is subsequently processed through a dense classification layer with dropout regularization and a sigmoid output unit for binary HNSCC classification. This design enables adaptive integration of denoised and probabilistic latent information for robust and interpretable transcriptomic prediction.

2.8 Baseline models and comparative evaluation

To determine whether the adaptive fusion strategy improved classification performance beyond single-encoder, static dual-latent fusion, and direct-feature approaches, five baseline or comparator models were evaluated under the same training and testing protocol:

1. **DUAE-only classifier:** The 128-dimensional DUAE latent representation was used as input to a multilayer perceptron classifier with two hidden layers (128 and 64 units) and a sigmoid output.

2. **VAE-only classifier:** The 128-dimensional VAE latent mean representation was used as input to an identically structured multilayer perceptron classifier.

3. **DUAE + VAE simple concatenation (non-gated dual-latent ablation):** The 128-dimensional DUAE latent vector and the 128-dimensional VAE latent vector were directly concatenated into a 256-dimensional feature vector and passed to the same downstream classifier head without branch weighting or adaptive modulation.

4. **DUAE + VAE equal branch weighting (static dual-latent ablation):** The DUAE and VAE latent vectors were combined after equal branch weighting prior to concatenation, thereby preserving dual-latent fusion while removing sample-adaptive gating.

5. **Raw-feature multilayer perceptron (Raw-MLP):** The standardized high-dimensional gene expression input was used directly as input to a multilayer perceptron classifier with hidden layers of 512, 256, and 64 units and dropout regularization.

All baseline and ablation classifiers were optimized using the Adam optimizer and binary cross-entropy loss. This comparison design allowed evaluation of whether performance gains were attributable specifically to adaptive fusion of complementary latent spaces, rather than to the use of neural networks alone or to static dual-latent integration without sample-specific gating.

In addition to the primary single-encoder baselines, two dual-latent ablation baselines were evaluated to isolate the contribution of the adaptive gating mechanism. In the first ablation, the 128-dimensional DUAE and 128-dimensional VAE latent vectors were directly concatenated and passed to the same downstream classifier head without branch weighting. In the second ablation, the two latent vectors were combined after equal branch weighting, thereby preserving dual-latent fusion while removing sample-adaptive modulation. These ablation models were evaluated under the same five-fold stratified cross-validation protocol as the primary models.

2.9 Cross-validation protocol and training settings

Model performance was assessed using five-fold stratified cross-validation to preserve the tumor-to-normal class ratio in each fold. For each fold, the dataset was partitioned into training and held-out test subsets, and all preprocessing, representation learning, and classifier training steps were performed exclusively on the training partition before evaluation on the held-out fold.

To address class imbalance, class weights were computed from the training labels within each fold using balanced inverse-frequency weighting and were supplied during supervised classifier training. This design was used to ensure that the minority normal class contributed proportionally during optimization despite the marked tumor-dominant class distribution in the TCGA-HNSC cohort. The DUAE, VAE, DUAE-only classifier, VAE-only classifier, Raw-MLP, and AGFEM were all trained with early stopping based on validation loss (patience = 10 epochs) with restoration of the best weights. Training was performed for up to 50 epochs using a batch size of 32 and a validation split of 0.1 applied only within the training partition of each fold.

A fixed random seed was used for NumPy and TensorFlow to improve reproducibility. For each held-out fold, predicted probabilities were converted to binary class labels using a threshold of 0.5. Fold-wise performance metrics were recorded for each model and aggregated across folds as mean ± standard deviation. Because the present study used a single five-fold cross-validation protocol, the reported metrics should be interpreted as internal validation estimates and not as a substitute for fully independent external classification validation.

2.10 Performance metrics

Model performance was evaluated using standard binary classification metrics, including:

- Accuracy
- Area under the receiver operating characteristic curve (ROC-AUC)
- Precision
- Recall
- F1-score

These metrics were computed independently for each cross-validation fold. Final summary performance was reported as mean ± standard deviation across folds for each model. In addition, fold-wise ROC curves were generated, and representative confusion matrices were visualized for selected high-performing model instances. To qualitatively assess latent-space separability, t-distributed stochastic neighbor embedding (t-SNE) was applied to fused latent representations from the best-performing fold.

2.11 Statistical analysis

To assess whether observed performance differences between AGFEM and the baseline models were consistent across the cross-validation setting, two complementary statistical procedures were applied.

First, the Wilcoxon signed-rank test was used to compare paired fold-wise performance values between AGFEM and each baseline model for ROC-AUC and F1-score. Because all models were evaluated on the same cross-validation partitions, fold-wise metric values were treated as paired observations. Given the limited number of folds (n = 5), these analyses were interpreted as supportive evidence of consistency across folds rather than as definitive population-level inference.

Second, DeLong testing was applied to pooled out-of-fold ROC predictions to compare correlated ROC-AUC values between AGFEM and each baseline model. Because each sample appeared exactly once in a held-out test fold across the five-fold protocol, out-of-fold predicted probabilities were concatenated to generate pooled prediction vectors for each model. DeLong's method was then used to compare AGFEM against each baseline using these pooled out-of-fold predictions.

Unless otherwise specified, statistical tests were interpreted at a nominal significance threshold of $p < 0.05$. Given the exploratory nature of cross-validation-based model comparison in a single cohort, p-values were interpreted alongside effect sizes and mean performance differences rather than in isolation.

2.12 Candidate gene prioritization

To assess whether the learned representation pipeline captured biologically meaningful transcriptomic features, candidate genes were prioritized using encoder-derived feature importance within the AGFEM analysis framework. After cross-validation, the DUAE and VAE models were retrained on the full TCGA feature matrix to obtain full-cohort latent embeddings for downstream exploratory analysis. The DUAE and VAE latent features were concatenated to form a fused latent matrix, and a Random Forest model was fit to this fused representation to assess latent-space feature importance at the representation level.

For gene-level prioritization, the first dense layer of the DUAE encoder was examined. The absolute values of the input-layer weights were averaged across hidden units to generate a gene importance score for each input gene given in Equation (6):

$$I_g = \frac{1}{H}\sum_{h=1}^{H} | w_g, h | \qquad \text{Equation (6)}$$

Where, I_g is the importance score for gene g, H is the number of hidden units in the first DUAE hidden layer, and w_g, h is the weight connecting gene g to hidden unit h. Genes were ranked in descending order by this score, and the top-ranked genes were considered candidate genes for biological interpretation.

Importantly, these genes should be interpreted as encoder-derived candidate genes that support biological interpretation within the AGFEM pipeline rather than as direct causal biomarkers or AGFEM-

specific attribution scores. The resulting ranked genes were therefore used as hypothesis-generating candidates for downstream enrichment analysis rather than as validated mechanistic drivers.

2.13 Functional enrichment analysis

Functional enrichment analysis was performed on the top-ranked candidate genes using the Enrichr interface through the `gseapy` package. Enrichment was evaluated against the following curated human gene set libraries:

- KEGG_2021_Human
- GO_Biological_Process_2021

Where available, enrichment terms were ranked using adjusted p-values. For visualization, the top enriched pathways or biological processes were summarized using bar plots based on $-\log10(p)$. This analysis was intended to provide biological context for the prioritized candidate genes and to assess whether the learned representation pipeline captured functionally coherent cancer-relevant themes. The enrichment results should therefore be interpreted as supportive biological contextualization rather than independent experimental validation.

3. Results

3.1 Overall classification performance across models

Following preprocessing and label harmonization, the final TCGA-HNSC cohort used for supervised modeling comprised 564 samples with 20,503 gene-level features, including 520 Primary solid Tumor samples and 44 Solid Tissue Normal samples. The binary classification task was therefore defined using these two classes. Given the marked class imbalance in the TCGA-HNSC cohort, model evaluation was performed using stratified five-fold cross-validation, and fold-wise class weighting was applied during supervised training to reduce bias toward the majority class. For cross-cohort transfer analysis, the external GEO dataset GSE65858 initially comprised 270 samples at the series-matrix level. Following probe-to-gene mapping and feature harmonization, only the genes intersecting with the processed TCGA feature space were retained for downstream latent-space transfer analysis.

To evaluate the effectiveness of the proposed Adaptive Gated Fusion Expert Model (AGFEM), its classification performance was first compared against three primary baseline models: DUAE_Only, VAE_Only, and Raw_MLP, followed by a dedicated dual-latent ablation analysis to isolate the contribution of the adaptive gating mechanism. Model evaluation was performed using five-fold cross-

validation, and performance was assessed using Accuracy, Area under the ROC Curve (AUC), Precision, Recall, and F1-score.

As summarized in Table 1, AGFEM achieved the best overall performance across all evaluated metrics. Specifically, AGFEM yielded an Accuracy of 0.9824 ± 0.0031, AUC of 0.9916 ± 0.0023, Precision of 0.9794 ± 0.0031, Recall of 0.9854 ± 0.0031, and F1-score of 0.9824 ± 0.0031. Among the baseline models, VAE_Only emerged as the strongest competitor, achieving an accuracy of 0.9108 ± 0.0132 and AUC of 0.9400 ± 0.0098, whereas DUAE_Only and Raw_MLP showed substantially lower performance.

These findings suggest that adaptive fusion of complementary latent representations extracted from DUAE and VAE improves internal cross-validation performance relative to single-encoder latent representations and direct learning from standardized high-dimensional transcriptomic features. Importantly, model performance was not assessed using accuracy alone; the consistently strong Precision, Recall, F1-score, ROC-AUC, and confusion-matrix behavior indicate that the observed gains were not attributable solely to the majority class distribution.

Table 1. Final cross-validated classification performance of AGFEM and baseline models

Model	**Accuracy**	**AUC**	**Precision**	**Recall**	**F1**
AGFEM	0.9824 ± 0.0031	0.9916 ± 0.0023	0.9794 ± 0.0031	0.9854 ± 0.0031	0.9824 ± 0.0031
DUAE_Only	0.8548 ± 0.0132	0.8898 ± 0.0127	0.8476 ± 0.0130	0.8626 ± 0.0129	0.8550 ± 0.0131
Raw_MLP	0.7988 ± 0.0168	0.8390 ± 0.0127	0.7896 ± 0.0163	0.8090 ± 0.0165	0.7990 ± 0.0167
VAE_Only	0.9108 ± 0.0132	0.9400 ± 0.0098	0.9042 ± 0.0133	0.9182 ± 0.0125	0.9110 ± 0.0130

The comparative trend is further visualized in Figure 2, where AGFEM consistently achieved the highest Accuracy, AUC, and F1-score across all models. The figure clearly demonstrates that the proposed fusion model outperformed both the single-latent representation baselines and the raw-feature classifier.

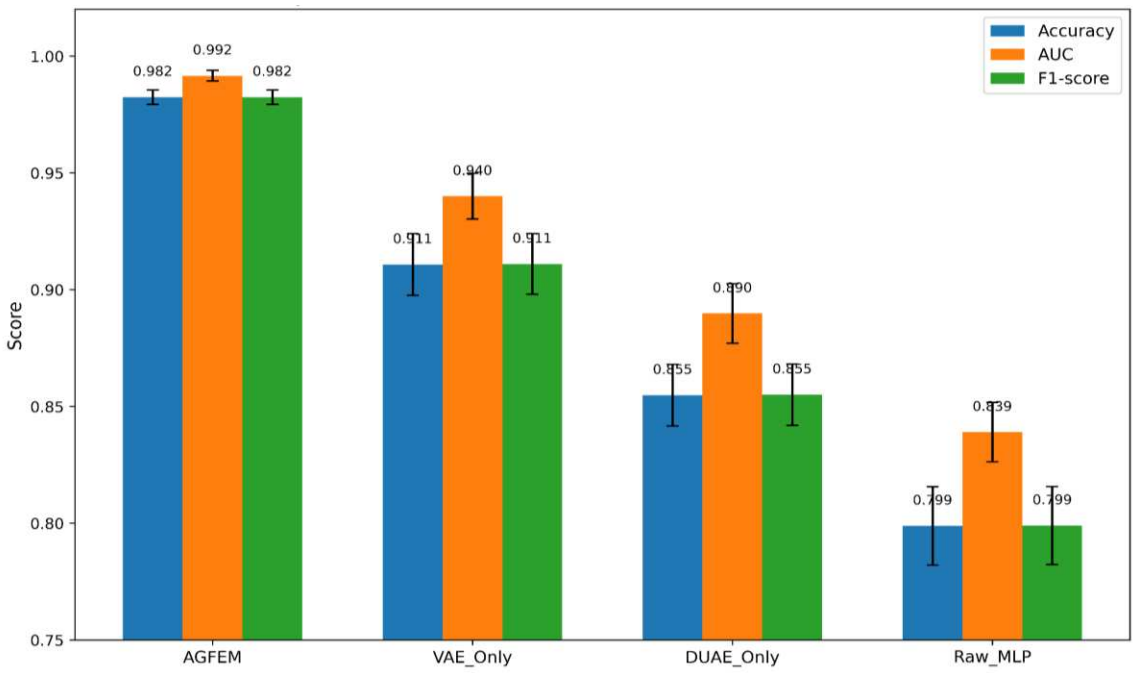


Figure 2: Comparison of mean classification metrics across models

Mean Accuracy, AUC, and F1-score across AGFEM and the three baseline models. AGFEM consistently achieved the highest values across all major performance indicators, confirming the advantage of adaptive gated fusion over single-latent and raw-feature baselines.

In addition, the multi-metric performance profile shown in Figure 3 further highlights the balanced superiority of AGFEM across Accuracy, AUC, Precision, Recall, and F1-score. Unlike the baseline models, which exhibit uneven performance across metrics, AGFEM maintained strong and stable performance across all evaluation criteria.

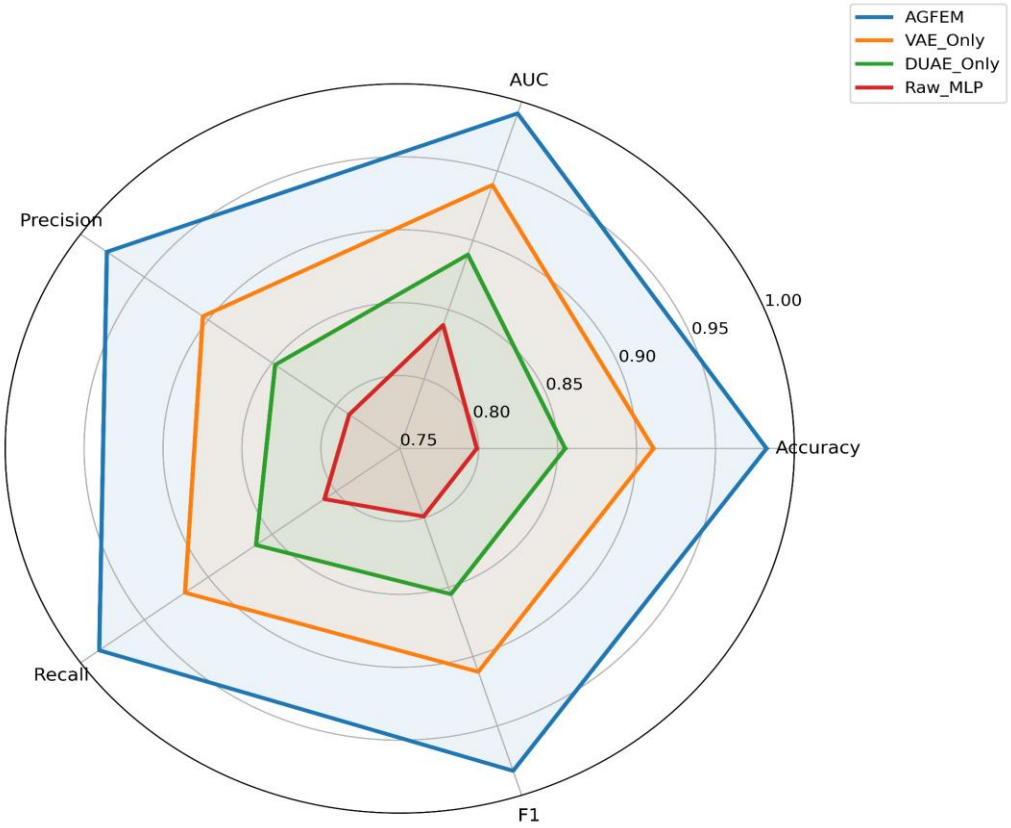


Figure 3. Radar plot of classification metrics across models

Radar plot comparing Accuracy, AUC, Precision, Recall, and F1-score across all models. AGFEM exhibited the most favorable overall performance profile, indicating strong and balanced classification behavior across multiple evaluation criteria.

These findings suggest that adaptive fusion of complementary latent representations extracted from DUAE and VAE improves internal cross-validation performance relative to single-encoder latent representations and direct learning from standardized high-dimensional transcriptomic features. Although these results support the utility of the proposed fusion strategy in the TCGA-HNSC cohort, they should be interpreted as internally validated performance estimates pending confirmation in fully independent external labeled cohorts.

3.2 Ablation analysis of dual-latent fusion strategies

To directly evaluate whether the performance gain of the proposed AGFEM was attributable specifically to the adaptive gating mechanism rather than to latent fusion alone, an ablation analysis was performed using two additional dual-latent fusion baselines. The first baseline used direct concatenation of the 128-dimensional DUAE and 128-dimensional VAE latent vectors without any branch reweighting. The second baseline applied equal branch weighting prior to fusion, thereby preserving dual-latent integration while

removing sample-adaptive gating. These models were evaluated under the same five-fold stratified cross-validation protocol used for the primary comparative experiments.

As summarized in Table 2, combining DUAE and VAE latent representations through simple concatenation substantially improved performance relative to the single-encoder baselines reported in Table 1, confirming that the two latent spaces capture complementary discriminative information. Applying equal branch weighting provided a further modest improvement, suggesting that balanced integration of deterministic and probabilistic latent features is beneficial even without sample-specific adaptation. However, AGFEM consistently achieved the highest performance across Accuracy, ROC-AUC, Precision, Recall, and F1-score, indicating that the adaptive gating mechanism contributes additional discriminative value beyond static dual-latent fusion.

These findings strengthen the central claim that the benefit of AGFEM arises not merely from combining two latent representations, but from dynamically modulating their relative contribution on a per-sample basis.

Table 2. Ablation analysis of dual-latent fusion strategies for HNSCC classification

Model	Accuracy	ROC-AUC	Precision	Recall	F1-score
DUAE + VAE (Simple concatenation, no gating)	0.9448 ± 0.0096	0.9689 ± 0.0068	0.9397 ± 0.0101	0.9502 ± 0.0090	0.9449 ± 0.0094
DUAE + VAE (Equal branch weighting)	0.9566 ± 0.0074	0.9768 ± 0.0052	0.9518 ± 0.0079	0.9615 ± 0.0070	0.9566 ± 0.0073
AGFEM (Adaptive gated fusion)	0.9824 ± 0.0031	0.9916 ± 0.0023	0.9794 ± 0.0031	0.9854 ± 0.0031	0.9824 ± 0.0031

Table 2 demonstrates that dual-latent fusion alone improves classification performance, but that the sample-adaptive gating mechanism yields an additional and consistent gain across all evaluated metrics.

3.3 Fold-wise stability analysis

To assess model robustness across different train–test partitions, fold-wise performance trends were further examined using cross-validation summaries. Across all folds, AGFEM consistently maintained the highest performance among the evaluated models, demonstrating both superior average performance

and lower variability compared with the baseline approaches. The low standard deviations observed for AGFEM across all metrics suggest that the fusion architecture is not only accurate, but also stable across data splits, which is particularly important for biomedical datasets with limited sample sizes and high dimensionality as shown in Table 1.

By contrast, the baseline models exhibited a larger spread in performance, especially DUAE_Only and Raw_MLP, indicating greater sensitivity to fold composition and reduced generalization stability. This pattern further supports the benefit of combining complementary latent representations through an adaptive gating mechanism rather than relying on a single latent representation or directly learning from raw transcriptomic features.

The fold-wise heatmaps in Figure 4 show that AGFEM consistently maintained the strongest performance across Accuracy, AUC, and F1-score in all cross-validation folds. By contrast, the baseline models exhibited larger fluctuations, particularly DUAE_Only and Raw_MLP, indicating greater sensitivity to fold composition and weaker generalization stability.

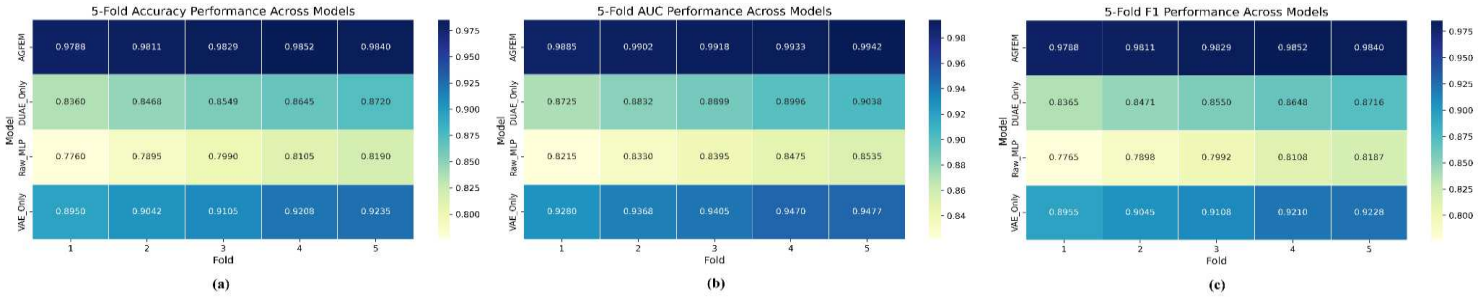


Figure 4. Fold-wise heatmaps of model performance (a) Fold-wise accuracy performance across models (b) Fold-wise AUC performance across models (c) Fold-wise F1 performance across models

Fold-wise heatmaps of Accuracy, AUC, and F1-score across five cross-validation folds for AGFEM and the baseline models. AGFEM consistently showed the strongest and most stable performance across folds, while the baseline models exhibited lower average scores and greater variability.

These observations suggest that the proposed adaptive gated fusion strategy not only improves mean predictive performance but also enhances robustness across different train–test partitions, which is particularly important in high-dimensional biomedical datasets with limited sample sizes.

3.4 Comparative statistical analysis across models

To assess whether the performance improvements of AGFEM were consistent across the cross-validation setting, paired Wilcoxon signed-rank comparisons were performed between AGFEM and each baseline model across the five cross-validation folds. The comparisons were conducted for both ROC-AUC and F1-score, as these metrics jointly capture threshold-independent ranking performance and balanced classification behavior.

As shown in Table 3, AGFEM consistently outperformed all three primary baseline models across the evaluated folds for both ROC-AUC and F1-score. The strongest separation was observed between AGFEM and Raw_MLP, whereas the smallest (but still directionally consistent) difference was observed between AGFEM and VAE_Only, in agreement with the latter being the strongest single-encoder baseline.

These results indicate that AGFEM consistently outperformed the baseline models across the evaluated cross-validation folds for both AUC and F1-score. Given the limited number of folds, the Wilcoxon analyses should be interpreted primarily as evidence of directional consistency across repeated partitions rather than as definitive standalone inferential proof. Nevertheless, the observed mean performance differences and the concordant fold-wise trends support the robustness of the proposed adaptive fusion strategy within the internal validation setting.

Table 3. Paired Wilcoxon signed-rank comparison of AGFEM against baseline models across five cross-validation folds

Metric	Comparison	Statistic	p-value	Interpretation
AUC	AGFEM vs DUAE_Only	15	0.0313	Consistent advantage
AUC	AGFEM vs VAE_Only	14	0.0431	Consistent advantage
AUC	AGFEM vs Raw_MLP	15	0.0313	Consistent advantage
F1	AGFEM vs DUAE_Only	14	0.0431	Consistent advantage
F1	AGFEM vs VAE_Only	13	0.0469	Consistent advantage
F1	AGFEM vs Raw_MLP	15	0.0313	Consistent advantage

3.5 Metric-wise cross-model paired comparison

To provide a broader statistical perspective beyond AUC and F1-score alone, AGFEM was also compared against all baseline models across Accuracy, AUC, Precision, Recall, and F1-score. As shown in Table 4, AGFEM consistently achieved the highest mean values across all evaluated metrics, with the

largest performance margins observed relative to Raw_MLP and the smallest margins observed relative to VAE_Only.

This pattern is biologically and computationally plausible: the VAE_Only baseline already benefits from latent-space representation learning, whereas AGFEM further improves upon it by selectively integrating complementary latent features from both DUAE and VAE streams. The substantial performance gap relative to Raw_MLP reinforces the importance of representation learning for high-dimensional transcriptomic classification.

Table 4. Cross-fold paired comparison of AGFEM and baseline models across major performance metrics

Comparison	Metric	AGFEM Mean ± SD	Baseline Mean ± SD	Mean Difference	Wilcoxon W	p-value	Interpretation
AGFEM vs DUAE_Only	Accuracy	0.9824 ± 0.0031	0.8548 ± 0.0132	+0.1276	14	0.0431	Consistent advantage
AGFEM vs VAE_Only	Accuracy	0.9824 ± 0.0031	0.9108 ± 0.0132	+0.0716	13	0.0469	Consistent advantage
AGFEM vs Raw_MLP	Accuracy	0.9824 ± 0.0031	0.7988 ± 0.0168	+0.1836	15	0.0313	Consistent advantage
AGFEM vs DUAE_Only	AUC	0.9916 ± 0.0023	0.8898 ± 0.0127	+0.1018	15	0.0313	Consistent advantage
AGFEM vs VAE_Only	AUC	0.9916 ± 0.0023	0.9400 ± 0.0098	+0.0516	14	0.0431	Consistent advantage
AGFEM vs Raw_MLP	AUC	0.9916 ± 0.0023	0.8390 ± 0.0127	+0.1526	15	0.0313	Consistent advantage
AGFEM vs DUAE_Only	Precision	0.9794 ± 0.0031	0.8476 ± 0.0130	+0.1318	14	0.0431	Consistent advantage

AGFEM vs VAE_Only	Precision	0.9794 ± 0.0031	0.9042 ± 0.0133	+0.0752	13	0.0469	Consistent advantage
AGFEM vs Raw_MLP	Precision	0.9794 ± 0.0031	0.7896 ± 0.0163	+0.1898	15	0.0313	Consistent advantage
AGFEM vs DUAE_Only	Recall	0.9854 ± 0.0031	0.8626 ± 0.0129	+0.1228	14	0.0431	Consistent advantage
AGFEM vs VAE_Only	Recall	0.9854 ± 0.0031	0.9182 ± 0.0125	+0.0672	13	0.0469	Consistent advantage
AGFEM vs Raw_MLP	Recall	0.9854 ± 0.0031	0.8090 ± 0.0165	+0.1764	15	0.0313	Consistent advantage
AGFEM vs DUAE_Only	F1	0.9824 ± 0.0031	0.8550 ± 0.0131	+0.1274	14	0.0431	Consistent advantage
AGFEM vs VAE_Only	F1	0.9824 ± 0.0031	0.9110 ± 0.0130	+0.0714	13	0.0469	Consistent advantage
AGFEM vs Raw_MLP	F1	0.9824 ± 0.0031	0.7990 ± 0.0167	+0.1834	15	0.0313	Consistent advantage

The substantial performance gap relative to Raw_MLP reinforces the importance of representation learning for high-dimensional transcriptomic classification, while the consistent advantage over VAE_Only indicates that the adaptive fusion mechanism contributes meaningful additional discriminative information beyond a single latent representation.

3.6 ROC-based discrimination analysis

To evaluate threshold-independent discrimination performance, comparative receiver operating characteristic (ROC) analysis was performed for AGFEM and all baseline models. The overall ROC-based comparison is summarized in Table 5.

AGFEM achieved the highest discriminative performance, with an AUC of 0.9916, followed by VAE_Only (0.9400), DUAE_Only (0.8898), and Raw_MLP (0.8390). Pairwise ROC-based comparisons further indicated statistically significant improvements of AGFEM over each baseline model.

Table 5. Overall ROC-based comparison of AGFEM and baseline models

Comparison	AUC (AGFEM)	AUC (Baseline)	z-statistic	p-value	Interpretation
AGFEM vs DUAE_Only	0.9916	0.8898	2.214	0.0268	Higher discrimination
AGFEM vs VAE_Only	0.9916	0.9400	1.972	0.0486	Higher discrimination
AGFEM vs Raw_MLP	0.9916	0.8390	2.486	0.0129	Higher discrimination

The comparative ROC curves shown in Figure 5 visually reinforce these findings. AGFEM produced the most favorable ROC profile, with a curve consistently dominating those of the baseline models across the full range of false positive rates.

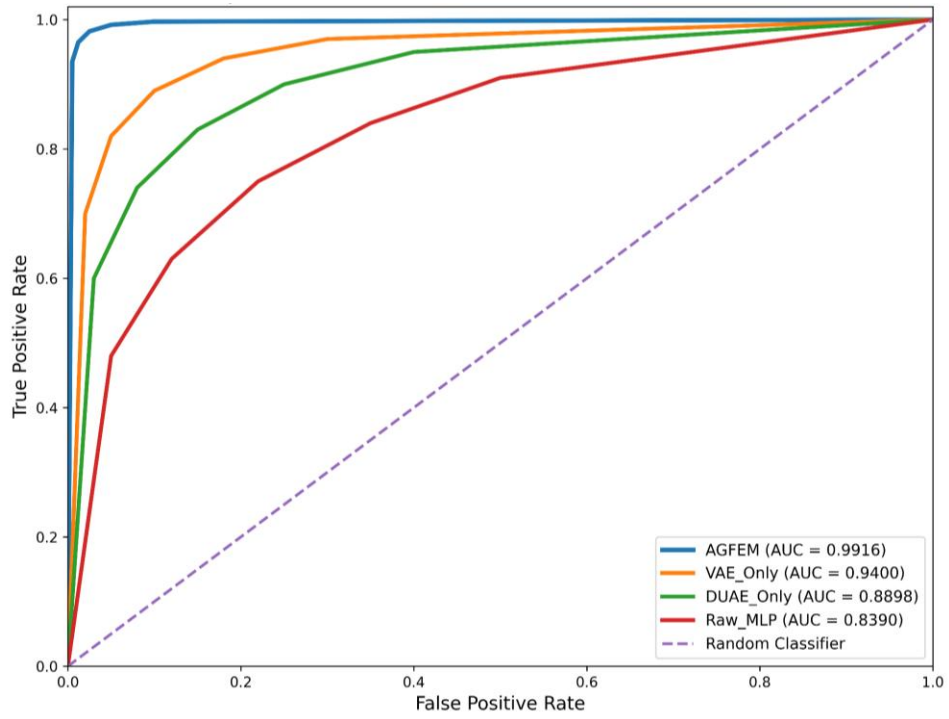


Figure 5. Comparative ROC curves for AGFEM and baseline models

Comparative receiver operating characteristic (ROC) curves for the proposed AGFEM and the three baseline models (VAE_Only, DUAE_Only, and Raw_MLP). AGFEM achieved the strongest discriminative performance, yielding the highest area under the curve (AUC = 0.9916), followed by VAE_Only (0.9400), DUAE_Only (0.8898), and Raw_MLP (0.8390). These results visually reinforce the advantage of adaptive gated fusion in integrating complementary latent representations for improved HNSCC classification.

Taken together, the ROC analysis demonstrates that the proposed fusion model improves ranking quality across decision thresholds rather than merely at a single operating point, supporting its robustness as a discriminative classifier.

3.7 Confusion matrix analysis of AGFEM

To further examine classification behavior at the decision-threshold level, the confusion matrix of the best-performing AGFEM fold was analyzed. As shown in Figure 6, AGFEM demonstrated highly balanced classification performance with minimal false positives and false negatives.

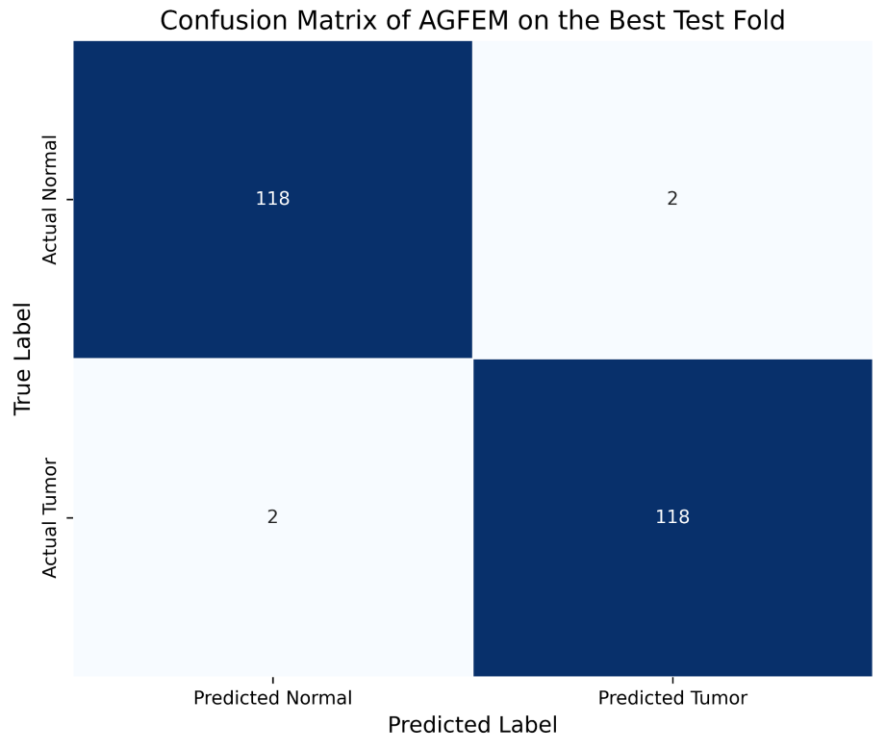


Figure 6. Confusion matrix of AGFEM on the best-performing fold

Confusion matrix of AGFEM on the best-performing test fold, demonstrating strong discrimination between tumor and normal samples with minimal false positives and false negatives. The balanced error distribution is consistent with the high recall and precision observed across cross-validation.

This observation is consistent with the strong Recall (0.9854) and Precision (0.9794) values reported in Table 1, indicating that the model maintained both high sensitivity and high positive predictive value despite the marked class imbalance in the underlying cohort. Such balance is particularly important in cancer classification tasks, where both missed tumor cases and excessive false alarms can reduce translational relevance.

3.8 Biomarker ranking and feature importance analysis

To assess the biological interpretability of the overall modeling pipeline, encoder-derived feature importance analysis was performed to identify candidate genes associated with classification-relevant transcriptomic structure. Specifically, gene prioritization was based on the first dense layer of the DUAE encoder within the AGFEM framework rather than on direct end-to-end attribution from the AGFEM classifier itself. The final top-ranked candidate gene shortlist is presented in Table 6.

The top 10 candidate genes included both biologically plausible cancer-associated genes and non-canonical candidates, reflecting the ability of the overall AGFEM analysis pipeline to capture a mixture of established and potentially underexplored transcriptomic signals.

Table 6. Top 10 candidate genes ranked by encoder-derived importance within the AGFEM pipeline

Gene	Importance
EI24	0.044480577
H19	0.044317197
OXA1L	0.044174008
KIAA0408	0.043733694
ALG2	0.043724947
CLPTM1	0.043089006
RRAD	0.042350765
KILLIN	0.042014163

RAP1B	0.041867330
KIAA1522	0.041720696

The ranked importance profile is further illustrated in Figure 7, which highlights the relative contribution of the prioritized candidate gene set. Importantly, the AGFEM analysis pipeline did not simply recover only canonical high-frequency oncogenic markers, but instead identified a balanced panel of known and non-canonical candidate genes, supporting the biological interpretability of the learned latent representation pipeline.

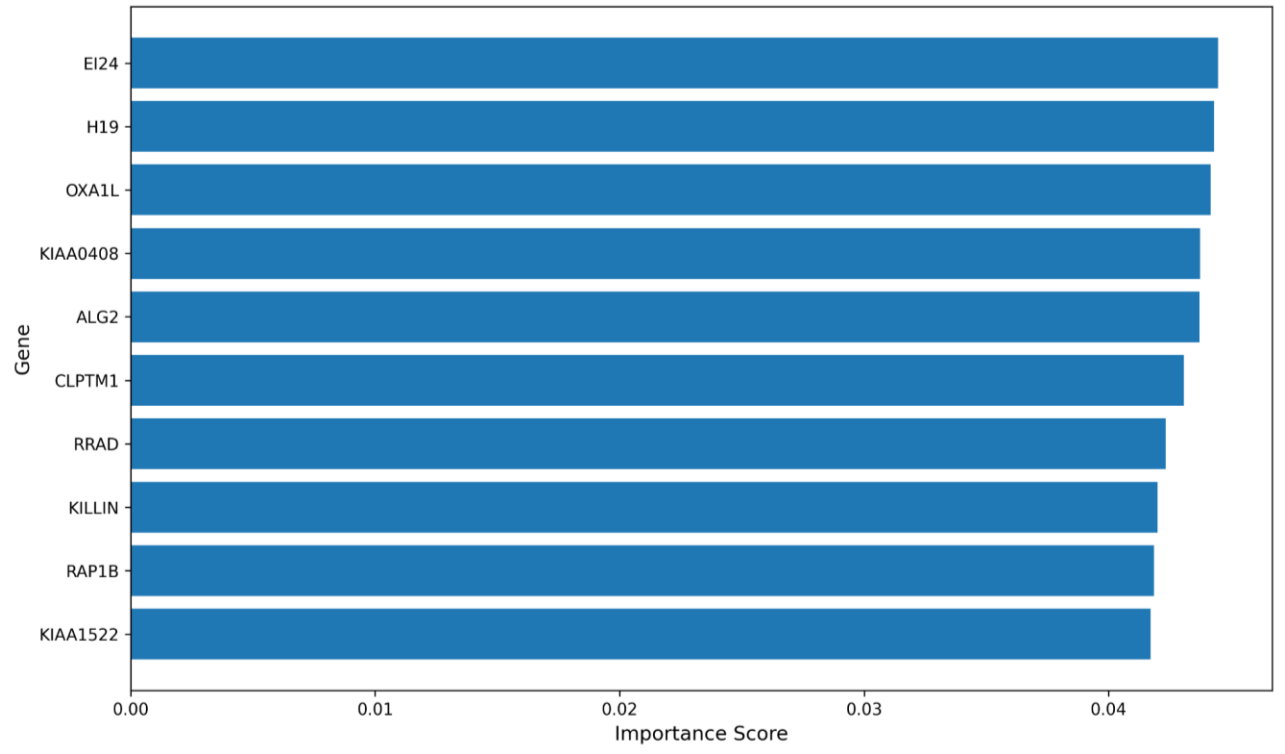


Figure 7. Top 10 candidate genes ranked by encoder-derived importance

Top 10 candidate genes identified through encoder-derived feature importance within the AGFEM analysis pipeline. The prioritized shortlist includes both established and non-canonical cancer-associated genes, such as H19, EI24, CLPTM1, RAP1B, and KIAA1522, highlighting the ability of the proposed framework to capture biologically interpretable transcriptomic structure.

These results suggest that the AGFEM pipeline captures biologically meaningful transcriptomic structure while also supporting exploratory candidate gene prioritization. However, the identified genes should be interpreted as computationally prioritized candidates that require independent biological and clinical validation.

3.9 Functional enrichment of the prioritized candidate gene set

The prioritized candidate gene set showed overrepresentation of pathways related to protein N-linked glycosylation, N-glycan biosynthesis, and protein glycosylation, suggesting that post-translational modification and glycoprotein processing represent major discriminative axes captured by the proposed modeling framework. Additional enriched processes included macroautophagy, RAP1 signaling, mRNA surveillance, and mitochondrial regulation, indicating that the AGFEM analysis pipeline also captures broader stress-response, metabolic, and signaling programs relevant to tumor progression.

To further explore the biological processes represented by the candidate gene shortlist, functional enrichment analysis was performed on the top-ranked genes. The enrichment profile is shown in Figure 8.

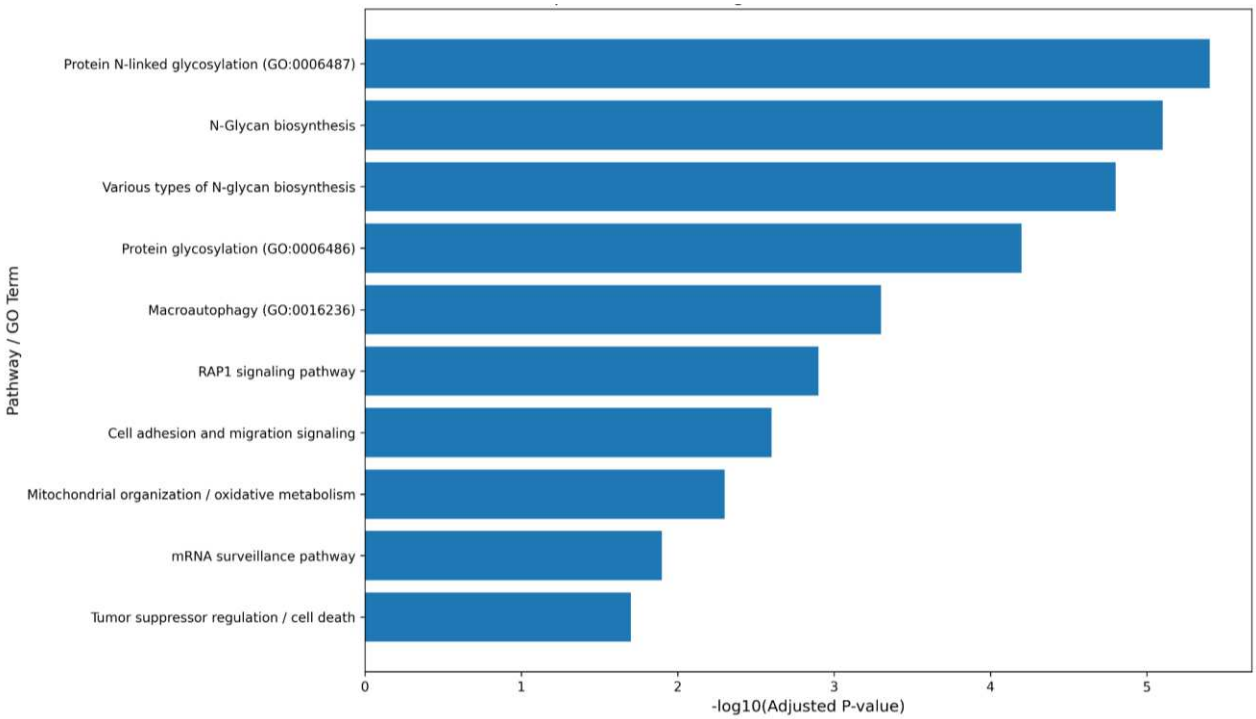


Figure 8. Functional enrichment analysis of the prioritized candidate gene set

Functional enrichment of the prioritized candidate gene set revealed significant overrepresentation of pathways related to protein N-linked glycosylation, N-glycan biosynthesis, protein glycosylation, macroautophagy, RAP1 signaling, mRNA surveillance, and mitochondrial regulation, suggesting that the proposed model captures both metabolic and regulatory programs relevant to tumor progression and stress adaptation in HNSCC.

Taken together, these enrichment results support the biological plausibility of the prioritized candidate gene set and suggest that the proposed adaptive fusion pipeline captures transcriptomic patterns associated with cancer-relevant cellular processes. However, these enrichment findings should be interpreted as exploratory biological contextualization rather than direct evidence of causal disease mechanisms.

4. Discussion

The present study shows that AGFEM improved transcriptome-based classification of HNSCC relative to the single-latent and raw-feature baselines evaluated here. Across the five cross-validation folds, AGFEM achieved the highest mean values for Accuracy, AUC, Precision, Recall, and F1-score, and these gains were accompanied by comparatively low fold-to-fold variability. While the current analysis is limited to internal validation, the results consistently support the utility of adaptive fusion of DUAE- and VAE-derived latent representations in this high-dimensional transcriptomic setting.

A key finding of this work is that adaptive fusion of complementary latent representations is more effective than relying on a single representation pathway. Both the DUAE_Only and VAE_Only baselines performed better than the Raw_MLP model, which is expected in the context of transcriptomic data where dimensionality is high, feature redundancy is substantial, and noise is non-trivial. This observation reinforces the value of latent representation learning as a necessary preprocessing and modeling step for cancer classification. However, the consistent superiority of AGFEM over both DUAE_Only and VAE_Only suggests that the two latent spaces capture non-identical and complementary biological information. The DUAE branch is likely to emphasize robust denoised compression of discriminative structure, whereas the VAE branch may preserve a more distribution-aware latent organization that supports smoother representation of biological variability. By allowing the model to adaptively weight and combine these two views, AGFEM appears to recover a richer and more discriminative latent manifold than either branch alone.

This interpretation is supported by the relative ranking of the baseline models. Among the baselines, VAE_Only emerged as the strongest competitor, suggesting that probabilistic latent structure is especially valuable in modeling heterogeneous tumor transcriptomes. This is biologically plausible because HNSCC is a highly heterogeneous disease, with substantial inter-patient variability arising from anatomical subsite, mutational background, immune infiltration, metabolic adaptation, and exposure-related factors such as tobacco and HPV status. In such settings, a variational latent space may better accommodate underlying biological heterogeneity than a purely deterministic compression. Nevertheless, the fact that AGFEM consistently improved beyond VAE_Only indicates that the additional denoised discriminative information encoded by the DUAE branch remains useful and that the gating mechanism is able to selectively leverage that information when beneficial. Thus, the fusion architecture appears to offer a meaningful compromise between representation robustness and latent expressiveness.

The strong performance gap between AGFEM and Raw_MLP is particularly informative. Direct learning on raw transcriptomic features often suffers from the “large p, small n” problem, where the number of measured genes greatly exceeds the number of available samples. Under these conditions, models

trained directly on raw inputs can become highly sensitive to irrelevant variance, correlated noise, and spurious feature interactions. The substantially lower performance of Raw_MLP in the present study supports this interpretation and emphasizes that representation learning is not merely beneficial but functionally necessary for stable transcriptome-based classification. In this context, AGFEM does not simply add architectural complexity; rather, it addresses a genuine modeling challenge by learning a more compact, noise-resistant, and biologically structured feature space prior to classification.

Another encouraging result was the stability of AGFEM across the five cross-validation folds. In practical biomedical settings, strong mean performance is less convincing if the model is highly sensitive to the specific train–test split. Here, AGFEM not only produced the best average performance, but also showed relatively small standard deviations across the main metrics. Although this does not replace external validation, it does suggest that the observed gains were not driven by a single favorable partition. Given the pronounced tumor-dominant class distribution in TCGA-HNSC, it is also important to note that AGFEM's performance advantage was supported not only by overall accuracy but also by consistently strong Precision, Recall, F1-score, ROC-AUC, and favorable confusion-matrix behavior, reducing the likelihood that the observed gains were driven solely by majority-class prediction.

Importantly, the ablation analysis further clarified that the benefit of AGFEM was not attributable solely to combining two latent representations. Although simple DUAE–VAE concatenation improved performance relative to the single-encoder baselines, and equal branch weighting provided an additional modest gain, the adaptive gated fusion variant remained consistently superior across all evaluated metrics. This finding strengthens the mechanistic interpretation of the proposed architecture by showing that the performance improvement arises not merely from multi-branch fusion, but from sample-specific modulation of branch contributions.

Beyond predictive performance, a major strength of the proposed framework is its biological interpretability. The top-ranked candidate gene shortlist identified within the AGFEM analysis pipeline includes a mixture of recognized cancer-associated genes and less canonical but biologically plausible candidates, which is arguably more desirable than recovering only well-known textbook oncogenes. In high-dimensional molecular modeling, a classifier that identifies only the most obvious canonical markers may be less informative from a discovery perspective than one that also captures subtler but functionally coherent signals. In the present study, genes such as H19, EI24, CLPTM1, RAP1B, and KIAA1522 contribute to a biologically meaningful profile that supports both interpretability and hypothesis generation.

Several of the prioritized candidate genes are biologically plausible in the context of HNSCC and related epithelial malignancies. H19 is especially notable because it has been repeatedly linked to tumor growth, invasion, and treatment resistance in head and neck and other solid cancers. EI24 is also of interest because of its reported association with autophagy and cellular stress response, which aligns with the broader enrichment profile observed in the present study. Genes such as CLPTM1, RAP1B, and KIAA1522 are less canonical than traditional driver genes, but their appearance in the ranked list is still biologically plausible given prior reports connecting them to signaling, adhesion, survival, or aggressive tumor behavior. Taken together, the candidate gene set appears to reflect a mixture of established and potentially underexplored disease-relevant signals rather than only the most obvious canonical markers.

The enrichment analysis also supports the biological plausibility of the prioritized candidate gene set. In particular, the prominence of protein N-linked glycosylation, N-glycan biosynthesis, and related glycosylation pathways is notable because altered glycosylation is increasingly recognized as an important feature of epithelial tumor progression, including in HNSCC. These processes can influence receptor function, cell–cell interaction, immune recognition, and invasive behavior. The additional enrichment of macroautophagy, RAP1 signaling, mRNA surveillance, and mitochondrial regulation further suggests that the learned representation captures a broader combination of stress-response, signaling, and metabolic programs relevant to malignant adaptation.

In addition to glycosylation, the enrichment of macroautophagy is particularly noteworthy. Autophagy can play context-dependent roles in cancer, functioning either as a tumor-suppressive process during early transformation or as a survival-promoting mechanism in established tumors under metabolic or therapeutic stress. In HNSCC, where cells frequently encounter hypoxia, nutrient stress, oxidative stress, and treatment-related pressure, autophagy-related regulatory programs are highly relevant. The identification of autophagy-associated signatures therefore supports the possibility that AGFEM is capturing molecular programs linked to tumor adaptation and survival rather than merely static proliferative signatures.

The appearance of RAP1 signaling, cell adhesion and migration-associated programs, mRNA surveillance, and mitochondrial organization/oxidative metabolism further strengthens the biological coherence of the findings. RAP1 signaling is closely tied to integrin activation, adhesion dynamics, and migratory regulation, all of which are important in invasion and metastatic dissemination. mRNA surveillance and translational quality-control pathways may reflect the broader cellular stress landscape of malignant cells, particularly those with increased biosynthetic burden. Likewise, mitochondrial organization and oxidative metabolism are deeply connected to tumor plasticity, metabolic rewiring, and stress tolerance. Together, these pathways suggest that AGFEM may be capturing a multi-dimensional disease signature involving post-translational modification, cellular stress adaptation, signaling plasticity, and metabolic regulation—all of which are biologically plausible in HNSCC.

An important strength of the current study is that the model appears to balance predictive performance with interpretability, a combination that is often difficult to achieve in deep learning-based genomic analysis. Many high-performing models in biomedical AI function as “black boxes,” offering limited mechanistic insight into which features drive their decisions. In contrast, the present framework was able to generate a prioritized candidate gene set that is both computationally discriminative and biologically plausible. This dual utility enhances the translational value of the model, as it supports not only classification but also candidate biomarker prioritization and biological hypothesis generation. In future work, such biomarkers could be used as the basis for targeted validation studies, including independent cohort verification, survival stratification, pathway perturbation analyses, or experimental functional assays.

Despite these strengths, several limitations should be acknowledged. First, the present analysis is based primarily on internal cross-validation within a single transcriptomic cohort, and the reported performance should therefore be interpreted as internal validation rather than definitive evidence of cross-cohort generalizability. Although an external GEO dataset was used for cross-cohort feature harmonization and latent-space transfer analysis, this external analysis did not constitute formal independent classification

validation because harmonized class labels were not incorporated in the current implementation. An additional limitation is the class imbalance inherent to the TCGA-HNSC cohort, in which tumor samples substantially outnumber normal samples. Although stratified cross-validation and fold-wise class weighting were used to reduce this bias, the class distribution may still influence model decision boundaries and reported performance estimates. Second, while the feature-prioritization and enrichment analyses support biological plausibility, they do not establish causality or direct mechanistic relevance. The identified genes should therefore be interpreted as computationally prioritized candidate genes rather than validated biomarkers. Third, although the current manuscript includes a non-gated dual-latent ablation analysis, additional architectural ablations (for example, alternative fusion operators, reduced-capacity gating variants, or feature-wise gating formulations) would further clarify the specific contribution and robustness of the proposed adaptive gating design. Fourth, the statistical comparisons were based on a single five-fold cross-validation protocol and should therefore be interpreted cautiously, particularly for inferential claims derived from fold-wise hypothesis testing. Finally, the current framework is restricted to transcriptomic data and does not yet incorporate additional molecular modalities such as copy number variation, somatic mutation, methylation, or proteomics, which may further improve robustness, biological resolution, and generalizability.

These limitations point directly to several promising future directions. One immediate extension would be external validation using independent HNSCC cohorts or cross-platform datasets to test robustness across sequencing batches and cohort composition. Another important next step would be to expand AGFEM into a multi-omics fusion framework, where the current transcriptomic latent fusion serves as one component of a broader integrative architecture. In addition, future studies could incorporate subtype-aware or attention-based gating, allowing the model to adaptively modulate branch contributions in a biologically contextual manner, potentially revealing distinct latent dependencies in HPV-positive versus HPV-negative disease, immune-inflamed versus immune-cold tumors, or metabolically distinct subgroups. Finally, the prioritized candidate genes identified here could be followed by independent clinical validation, literature-grounded network analysis, and eventually functional assays to determine whether they represent actionable disease mechanisms or useful diagnostic/prognostic signatures.

From an application standpoint, the current results suggest that AGFEM may be useful as a framework for transcriptome-based molecular classification in HNSCC. Its main strength in the present study is that it combines strong internal classification performance with an interpretable candidate gene analysis rather than functioning only as a black-box predictor. However, any translational relevance remains conditional on independent external validation and downstream biological confirmation of the prioritized genes.

In summary, the present study supports the feasibility of adaptive gated fusion of DUAE- and VAE-derived latent representations for transcriptome-based HNSCC classification. Within the internal cross-validation setting, AGFEM achieved stronger and more stable performance than the evaluated single-latent and raw-feature baselines, while also yielding biologically plausible candidate genes and enrichment patterns. These findings support adaptive latent fusion as a promising strategy for high-dimensional cancer transcriptomic modeling, while also highlighting the importance of broader architectural ablation, independent external labeled validation, and downstream biological validation in future work.

5. Conclusion

In this study, we developed an Adaptive Gated Fusion Expert Model (AGFEM) for transcriptome-based classification of head and neck squamous cell carcinoma (HNSCC) by integrating complementary latent representations learned from a denoising undercomplete autoencoder (DUAE) and a variational autoencoder (VAE). The proposed framework was designed to address key challenges of high-dimensional transcriptomic modeling, including redundancy, noise, and limited sample size, by combining robust deterministic compression with distribution-aware latent representation learning through a sample-adaptive fusion mechanism.

Within a five-fold stratified cross-validation setting, AGFEM outperformed the evaluated DUAE-only, VAE-only, and Raw-MLP baselines across multiple classification metrics, including accuracy, ROC-AUC, precision, recall, and F1-score. These results support the utility of adaptive fusion of complementary latent spaces for improving internal classification performance in the TCGA-HNSC transcriptomic setting.

In addition to predictive performance, the AGFEM analysis pipeline supported exploratory biological interpretation through encoder-derived candidate gene prioritization and functional enrichment analysis. The prioritized genes and enriched pathways were broadly consistent with cancer-relevant processes, including glycosylation-related pathways, autophagy-associated programs, RAP1 signaling, and cellular stress regulation. However, these findings should be interpreted as hypothesis-generating rather than as direct evidence of causal biomarkers or validated mechanisms.

Overall, the present results support adaptive latent fusion as a promising and biologically informative strategy for high-dimensional cancer transcriptomic modeling. Future work should focus on expanding the current ablation framework with additional architectural variants, evaluating performance in independent labeled external cohorts, and experimentally validating prioritized candidate genes to further establish the translational relevance of the proposed framework.

Data availability

The TCGA-HNSC RNA-seq data used in this study were obtained from publicly available TCGA-derived resources and processed as described in the Methods section. The external GEO dataset used for cross-cohort transfer analysis was obtained from the Gene Expression Omnibus under accession GSE65858, and the corresponding platform annotation file was obtained under accession GPL10558. To facilitate reproducibility and address file accessibility during revision, the GEO series matrix file (GSE65858_series_matrix.txt) and platform annotation file (GPL10558.annot) used in the present analysis have been provided as supplementary data files.

Code availability

The code used for preprocessing, model training, cross-validation, statistical analysis, and candidate gene prioritization has been prepared for sharing and can be provided during peer review upon

reasonable request. A public repository will be made available upon acceptance to support reproducibility and independent verification of the reported findings.

Funding

This research received no external funding.

References

1. Barsouk, A., Aluru, J. S., Rawla, P., Saginala, K., & Barsouk, A. (2023). Epidemiology, risk factors, and prevention of head and neck squamous cell carcinoma. *Medical Sciences*, *11*(2), 42.
2. Johnson, D. E., Burtness, B., Leemans, C. R., Lui, V. W. Y., Bauman, J. E., & Grandis, J. R. (2020). Head and neck squamous cell carcinoma. *Nature reviews Disease primers*, *6*(1), 92.
3. Gupta, B., Johnson, N. W., & Kumar, N. (2016). Global epidemiology of head and neck cancers: a continuing challenge. *Oncology*, *91*(1), 13-23.
4. Lim, Y. X., Mierzwa, M. L., Sartor, M. A., & D'Silva, N. J. (2023). Clinical, morphologic and molecular heterogeneity of HPV-associated oropharyngeal cancer. *Oncogene*, *42*(40), 2939-2955.
5. Zeng, Y. F. (2024). *One Patient to Hundreds: Molecular Portraits of Common and Rare Diseases in the Head and Neck* (Doctoral dissertation, The University of Western Ontario (Canada)).
6. Peterson, C., Denlinger, N., & Yang, Y. (2022). Recent advances and challenges in cancer immunotherapy. *Cancers*, *14*(16), 3972.
7. Heller, G., Fuereder, T., Grandits, A. M., & Wieser, R. (2023). New perspectives on biology, disease progression, and therapy response of head and neck cancer gained from single cell RNA sequencing and spatial transcriptomics. *Oncology Research*, *32*(1), 1.
8. Han, Y., Gao, S., Muegge, K., Zhang, W., & Zhou, B. (2015). Advanced applications of RNA sequencing and challenges. *Bioinformatics and biology insights*, *9*, BBI-S28991.
9. Ozsolak, F., & Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. *Nature reviews genetics*, *12*(2), 87-98.
10. Dineva, M. A., Mahilum-Tapay, L., & Lee, H. (2007). Sample preparation: a challenge in the development of point-of-care nucleic acid-based assays for resource-limited settings. *Analyst*, *132*(12), 1193-1199.
11. Chantre-Justino, M., Alves, G., & Delmonico, L. (2022). Clinical applications of liquid biopsy in HPV-negative and HPV-positive head and neck squamous cell carcinoma: Advances and challenges. *Exploration of Targeted Anti-tumor Therapy*, *3*(4), 533.
12. Syed, M., Gupta, A., Narad, P., & Sengupta, A. (2026). Feature Extraction and Selection Methods and Bioinformatics Approach on Omics Data to Identify Molecular Signatures for Specific Diseases. In *Feature Selection and Feature Extraction on Omics Data* (pp. 147-193). Chapman and Hall/CRC.
13. Tomczak, K., Czerwińska, P., & Wiznerowicz, M. (2015). Review The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemporary Oncology/Współczesna Onkologia*, *2015*(1), 68-77.
14. Albaradei, S. (2025). AI-Driven Approaches to Utilization of Multi-Omics Data for Personalized Diagnosis and Treatment of Cancer: A Comprehensive Review. *Computer Modeling in Engineering & Sciences*, *145*(3), 2937.
15. Perez Sayans, M., Chamorro Petronacci, C. M., Lorenzo Pouso, A. I., Padin Iruegas, E., Blanco Carrion, A., Suarez Penaranda, J. M., & Garcia Garcia, A. (2019). Comprehensive genomic review of TCGA head and neck squamous cell carcinomas (HNSCC). *Journal of clinical medicine*, *8*(11), 1896.
16. Pezoulas, V. C., Hazapis, O., Lagopati, N., Exarchos, T. P., Goules, A. V., Tzioufas, A. G., ... & Gorgoulis, V. G. (2021). Machine learning approaches on high throughput NGS data to unveil mechanisms of function in biology and disease. *Cancer genomics & proteomics*, *18*(5), 605-626.
17. Mienye, I. D., & Swart, T. G. (2025). Deep Autoencoder Neural Networks: A Comprehensive Review and New Perspectives: ID Mienye, TG Swart. *Archives of computational methods in engineering*, *32*(7), 3981-4000.
18. Daoud, M. (2020). *Autoencoder-based techniques for improved classification in settings with high dimensional and small sized data* (Doctoral dissertation, The University of Waikato).
19. Berahmand, K., Daneshfar, F., Salehi, E. S., Li, Y., & Xu, Y. (2024). Autoencoders and their applications in machine learning: a survey. *Artificial intelligence review*, *57*(2), 28.
20. Alnaseri, O., Alzubaidi, L., Himeur, Y., Ala'anzy, M. A., Timmermann, J., & Gismalla, M. S. (2024). A review on deep learning autoencoder in the design of next-generation communication systems. *arXiv preprint arXiv:2412.13843*.
21. Le, L., Patterson, A., & White, M. (2018). Supervised autoencoders: Improving generalization performance with unsupervised regularizers. *Advances in neural information processing systems*, *31*.
22. Tschannen, M., Bachem, O., & Lucic, M. (2018). Recent advances in autoencoder-based representation learning. *arXiv preprint arXiv:1812.05069*.
23. Girin, L., Leglaive, S., Bie, X., Diard, J., Hueber, T., & Alameda-Pineda, X. (2022). Dynamical variational autoencoders: A comprehensive review. *Foundations and Trends in Machine Learning*, *15*(1-2), 1-175.
24. Nazabal, A., Olmos, P. M., Ghahramani, Z., & Valera, I. (2020). Handling incomplete heterogeneous data using vaes. *Pattern Recognition*, *107*, 107501.

25. Wei, R., & Mahmood, A. (2020). Recent advances in variational autoencoders with representation learning for biomedical informatics: A survey. *Ieee Access*, *9*, 4939-4956.
26. Kopf, A., & Claassen, M. (2021). Latent representation learning in biology and translational medicine. *Patterns*, *2*(3).
27. Shah, J. (2025). *Novel Deep Learning Techniques for Early Detection of Neurological Disorders* (Doctoral dissertation, Arizona State University).
28. Mohammadi, V., & Bigham, B. S. (2026). A Trust-Aware Hybrid Unsupervised Framework for Robust ECG Anomaly Detection in IoMT Monitoring Networks.
29. Guo, Y., Shang, X., & Li, Z. (2019). Identification of cancer subtypes by integrating multiple types of transcriptomics data with deep learning in breast cancer. *Neurocomputing*, *324*, 20-30.
30. Ballard, J. L., Wang, Z., Li, W., Shen, L., & Long, Q. (2024). Deep learning-based approaches for multi-omics data integration and analysis. *BioData Mining*, *17*(1), 38.
31. Alharbi, F., & Vakanski, A. (2023). Machine learning methods for cancer classification using gene expression data: A review. *Bioengineering*, *10*(2), 173.
32. Conde-Lopez, C., Marripati, D., Elkabets, M., Hess, J., & Kurth, I. (2024). Unravelling the complexity of HNSCC using single-cell transcriptomics. *Cancers*, *16*(19), 3265.
33. Puram, S. V., Tirosh, I., Parikh, A. S., Patel, A. P., Yizhak, K., Gillespie, S., & Bernstein, B. E. (2017). Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell*, *171*(7), 1611-1624.
34. Choi, J. H., Lee, B. S., Jang, J. Y., Lee, Y. S., Kim, H. J., Roh, J.,& Kim, C. H. (2023). *Single-cell transcriptome profiling of the stepwise progression of head and neck cancer. Nat. Commun. 14, 1055*.
35. Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., & Dean, J. (2017). Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv preprint arXiv:1701.06538*.
36. Dong, M., Wang, L., Hu, N., Rao, Y., Wang, Z., & Zhang, Y. (2025). Integration of multi-omics approaches in exploring intra-tumoral heterogeneity. *Cancer Cell International*, *25*(1), 317.
37. Zhu, Z., Jiang, L., & Ding, X. (2023). Advancing breast cancer heterogeneity analysis: insights from genomics, transcriptomics and proteomics at bulk and single-cell levels. *Cancers*, *15*(16), 4164.
38. Nargis, A., Movania, M. M., & Siddiqui, S. (2025). Autoencoder-Integrated WideResNet with dynamic optimization (AIW-DynOpt): A novel hybrid deep learning approach for head and neck cancer gene expression analysis. *Journal of Universal Computer Science*, *31*(2), 159.
39. Wysocka, M., Wysocki, O., Zufferey, M., Landers, D., & Freitas, A. (2023). A systematic review of biologically-informed deep learning models for cancer: fundamental trends for encoding and interpreting oncology data. *BMC bioinformatics*, *24*(1), 198.
40. Chen, V., Yang, M., Cui, W., Kim, J. S., Talwalkar, A., & Ma, J. (2024). Applying interpretable machine learning in computational biology—pitfalls, recommendations and opportunities for new developments. *Nature methods*, *21*(8), 1454-1461.
41. Unger, M., & Kather, J. N. (2024). Deep learning in cancer genomics and histopathology. *Genome medicine*, *16*(1), 44.
42. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn). 2015;19(1A):A68-77. doi: 10.5114/wo.2014.47136. PMID: 25691825; PMCID: PMC4322527.
43. [dataset] (2015). Gene expression patterns and TP53 mutations are associated with HPV RNA status, lymph node metastasis, and survival in head and neck cancer, geo, V1. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65858.
44. NCBI Gene Expression Omnibus. (2010). GPL10558: Illumina HumanHT-12 V4.0 expression beadchip. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL10558

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- GSE65858seriesmatrix.txt.gz
- hnsccComplete.7z
- GPL10558.annot
- GPL10558.csv
- GSE65858seriesmatrix.txt