Briefings in Bioinformatics, 2025, **26(3)**, bbaf275
https://doi.org/10.1093/bib/bbaf275
Case Study

Gene Swin transformer: new deep learning method for colorectal cancer prognosis using transcriptomic data

Yangyang Wang[1,‡], Xinyu Yue[1,‡], Shenghan Lou[1], Peinan Feng[2], Binbin Cui[1,*], Yanlong Liu[1,*]

[1]Department of Colorectal Surgery, Harbin Medical University Cancer Hospital, No.150 Haping Road, Harbin, Heilongjiang 150081, China
[2]School of Computer Science and Engineering, Northeastern University, No.195 Innovation Road, Hunnan District, Shenyang, Liaoning 110169, China
*Corresponding authors. Yanlong Liu, Harbin Medical University Cancer Hospital, Department of Colorectal Surgery, No. 150 Haping Road, Harbin, Heilongjiang 150081, P.R. China. E-mail: 2355@hrbmu.edu.cn; Binbin Cui, Harbin Medical University Cancer Hospital, Department of Colorectal Surgery, No. 150 Haping Road, Harbin, Heilongjiang 150081, P.R. China. E-mail: 13351112888@163.com
‡Yangyang Wang and Xinyu Yue have equal contributors and co-first authors.

Abstract

Transcriptome sequencing has become essential in clinical tumor research, providing in-depth insights into the biology and functionality of tumor cells. However, the vast amount of data generated and the complex relationships between gene expressions make it challenging to effectively identify clinically relevant information. In this study, we developed a method called Gene Swin Transformer to address these challenges. This approach converts transcriptomic data into Synthetic Image Elements (SIEs). We utilized data from 12 datasets, including GSE17536-GSE103479 datasets ($n = 1771$) and The Cancer Genome Atlas ($n = 459$), to generate SIEs. These elements were then classified based on survival time using deep learning algorithms to predict colorectal cancer prognosis and build a reliable prognostic model. We trained and evaluated four deep learning models—BeiT, ResNet, Swin Transformer, and ViT Transformer—and compared their performance. The enhanced Swin-T model outperformed the other models, achieving weighted precision, recall, and F1 scores of 0.708, 0.692, and 0.705, respectively, along with area under the curve values of 80.2%, 72.7%, and 76.9% across three datasets. This model demonstrated the strongest prognostic prediction capabilities among those evaluated. Additionally, the PEX10 gene was identified as a key prognostic marker through both visual attention matrix analysis and bioinformatics methods. Our study demonstrates that the Gene Swin model effectively transforms Ribonucleic Acid (RNA) sequencing data into SIEs, enabling prognosis prediction through attention-based algorithms. This approach supports the development of a data-driven, unified, and automated model, offering a robust tool for classification and prediction tasks using RNA sequencing data. This advancement presents a novel clinical strategy for cancer treatment and prognosis forecasting.

Keywords: RNA sequencing; colorectal cancer; deep learning; survival prediction; synthetic image elements; Gene Swin transformer

Introduction

Colorectal cancer (CRC) is the fourth deadliest cancer globally, with nearly 900 000 deaths annually [1, 2]. This complex disease requires early detection and accurate assessment of tumors to improve prognosis and ensure the long-term survival of patients. Typically, CRC patients are evaluated for treatment options and survival outcomes through Tumor, Node, Metastasis (TNM) staging, mismatch repair protein (MMR) testing, and microsatellite stability testing, all of which provide critical prognostic information [3]. However, these current evaluation methods rely on postsurgical pathology and genetic testing, which necessitates careful analysis by experienced pathologists. This process can often lead to differing opinions owing to subjective interpretations. Additionally, technical factors, such as data quality, sample fixation, and varying levels of proficiency in interpreting tumor pathology can result in inconsistent outcomes. Therefore, achieving consistency and reproducibility, increasing speed and efficiency, and developing automated methods to handle large-scale data at minimal cost are essential goals for improving CRC prognosis and treatment.

In recent years, transcriptome sequencing technology has matured significantly, and its applications have expanded across various diseases. Its impact on clinical practice has grown as well. Researchers have used Ribonucleic Acid (RNA) sequencing to discover biological markers associated with tumor prognosis and to explore the functions of known biomarkers. With the accumulation of transcriptome sequencing data accumulates, data-driven bioinformatics approaches and deep learning (DL) methods have become valuable tools for studying and predicting patient prognosis in tumors [4, 5]. However, predicting patient outcomes using gene sequencing data often relies on known or newly discovered biomarkers as intermediaries for analyzing survival and prognosis [6–8], or it employs pathological markers for prognostic predictions [9].

A significant challenge in using biomarkers or pathological markers for prediction is that researchers are often limited to identifying only one or a few markers to forecast outcomes. This restriction arises from the limitations of traditional modeling techniques, which struggle to manage a large number of variables, especially when the number of variables far exceeds the number of observations or sample sizes. Consequently, this complexity

Received: January 24, 2025. **Revised:** April 27, 2025. **Accepted:** May 19, 2025

impedes thorough data analysis and the discovery of new insights or patterns in large datasets. Modern computational methods and DL algorithms offer promising solutions to these challenges. Compared to traditional machine learning, DL provides significant advantages by automatically identifying useful features in data through multiple levels of abstraction, eliminating the need for manually designed or selected features.

DL is especially well-suited for handling unstructured data, such as text, images, and sounds. These algorithms have evolved into powerful tools across various fields, including computer vision, pattern recognition, natural language processing, and biomedical research, particularly for the management of genomic data [10–14].

In recent years, transformer networks, a new and powerful deep-learning architecture, have emerged prominently in various artificial-intelligence tasks [15]. In this study, we developed a new attention mechanism algorithm, Gene Swin, which is specifically designed for handling gene-related image data. This algorithm is an improvement on the Swin transformer algorithm [16], which has been proven effective in predicting tumor mutation status from radiomics [17] and in classifying tumor pathology [18]. Although Gene Swin excels in processing image-format data, RNA sequencing inherently quantifies gene expression levels through a series of continuous numerical data. We developed a technique that initially converts transcriptomic data into Synthetic Image Elements (SIEs) and then uses DL algorithms, such as the Gene Swin transformer, to classify these SIEs based on survival time. This approach enables the survival prognosis prediction analysis, outputs risk values, and constructs a prognostic prediction model. Finally, we associated the model's prognostic risk values with the patients' clinical date using bioinformatics methods. This allowed us to study the bioinformatics mechanisms related to prognostic risk scoring, thereby providing biological interpretability for this deep-learning-based prognostic prediction model. This report summarizes the results of the present study.

Materials and methods

RNA sequencing data

We obtained 11 RNA sequencing datasets from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) including GSE17536 (n = 177) [8], GSE38832 (*n* = 122) [19], GSE39084 (*n* = 70) [20], GSE39582 (*n* = 586) [21], GSE41258 (*n* = 156) [22], GSE106584 (*n* = 156) [23], GSE17537 (*n* = 55) [24], GSE24551 (*n* = 160) [25], GSE29621 (*n* = 65) [26], GSE72968 (*n* = 68) [27], and GSE103479 (*n* = 156) [28]. These datasets provide patient clinical information, including survival data, and RNA sequencing data as described by the original authors. Expression data for these datasets were quantified using Fragments Per Kilobase of transcript per million mapped reads (FPKM). Table 1 summarizes the details of these datasets. Additionally, an RNA sequencing dataset containing information on 459 patients with CRC was downloaded from The Cancer Genome Atlas (TCGA) database (https://www.cancer.gov/ccg/research/genome-sequencing/tcga), which serves as an external validation set.

Using data produced by different teams are beneficial not only because a large dataset can enhance the accuracy and generalization capability of DL models, but also to assess the impact of data generated by different teams on model performance. After obtaining sequencing data from the GEO database, we performed a logarithmic (log2) transformation of all transcriptomic data. Subsequently, the expression levels of the transcripts were adjusted to a range from 0 to 1. Gene expression data from the 11 datasets were converted using official gene symbols, and the shared genes among these datasets were used for the research. Because the majority of noncoding RNAs have not been assigned official gene symbols, the selected genes were predominantly protein-coding genes. Consequently, we organized the RNA sequencing data according to the chromosomes on which they were located and the positions of their transcription start sites, resulting in a final usable gene list. From this list, the first 16 889 genes were used to create a square SIE for each patient. After scaling these 16 889 genes, they were configured into a 224×224 pixel grayscale SIE.

Clinical data

In this study, we used clinical information from GEO and TCGA databases to create outcome indicators or labels for model training and prediction. Table 1 summarizes the clinical information of the subjects included in these datasets. For patient prognosis, we used the patient's survival time as the outcome indicator: a survival time of <2 years is considered a poor prognosis [29], 2–10 years is considered an intermediate prognosis, and more than 10 years is considered a good prognosis. These categories were labeled as 0, 1, and 2, respectively, in our model training. The performance of our model was evaluated by predicting survival time intervals.

Converting RNA sequencing data into Synthetic Information Elements

The concept of SIEs technology involves transforming non-image data, such as one-dimensional sequence data, into images, thereby enabling the processing and analysis of these data using image analysis methods and tools. Deep Insight [30] does not reorganize the input feature vectors using domain-specific information but instead adopts a general approach by initially mapping the preprocessed data to the pixel values of an image. This step is the most critical part of the process. The mapping method can be selected based on the characteristics of the data and the requirements. A common approach involves using dimensionality reduction techniques, such as t-SNE or Principal Component Analysis, to project high-dimensional data into two or three-dimensional spaces, which are then mapped onto an image.

In the next phase, the image is rotated by a specific angle to align it with the Cartesian coordinate axes. After this rotation, the adjusted Cartesian coordinates were mapped to the pixel coordinates of the image. Subsequently, the element values are mapped to pixel positions to form an image of the feature vector. Research has shown that this method can be effectively integrated with convolutional neural networks (CNNs) to classify gene expression data [31].

To address this flaw, this study proposes a technique called SIEs, a method for converting non-image format data into an image format. This method maintains the independence and integrity of the original feature data without averaging feature information. The SIEs technique uses a multichannel extension process instead of simply averaging feature values.

We developed a CNN-based multi-omics analysis framework that utilizes SIEs to uncover hidden nonlinear relationships within complex, high-dimensional data. In this framework, multi-omics data are converted into multichannel images, with each gene represented as a pixel in a genome-wide image. Different types of omics data, including mutations, gene expression, DNA methylation, and copy number variations, were integrated into these

Table 1. Descriptive overview of the datasets used in thin study.

Dataset	TNM staging	Microsatellite instability	RAS and BRAF gene mutations	Survival day	Survival event	Sequencing platform ID
GSE955465	T4N0M0 StageII	MSI-S	KRAS: WT BRAF: WT	3.049315 years	death: 0	GPL570
GSE955466	T4N1M0 StageIII	MSI-H	KRAS: M BRAF: WT	4.133247 years	death: 0	GPL570
GSE955467	T2N0M0 StageI	MSI-H	KRAS: M BRAF: WT	9.126027 years	death: 0	GPL570
GSE955468	T3N0M0 StageII	MSI-H	KRAS: WT BRAF: WT	4.446575 years	death: 0	GPL570

images as separate layers. This conversion establishes spatial relationships among the data, enabling convolutional layers to effectively explore and analyze interactions between genes. Compared to traditional nonspatial data transformations, this spatial approach offers greater efficiency. Our method organizes genomic data into gene images sorted by chromosome position, exemplifying the SIEs technology introduced in this study.

This study improves upon the aforementioned methods by first preprocessing the raw transcriptomic data—including noise filtering, normalization, and feature selection—to ensure the data have a high signal-to-noise ratio and statistical stability.

Feature Ordering: First, we order the gene/transcript features in the transcriptomic data. This ordering is typically based on biologically or statistically meaningful criteria, such as a gene's physical location on the chromosome, its membership in a specific biological pathway, or its differential expression significance (e.g. *P*-value or fold change) under different experimental conditions—similar to the feature ordering strategy that might be employed in DeepFeature. It is crucial to ensure that all samples use exactly the same gene/feature order.

Value Mapping: For each sample, we normalize the expression values of each gene/transcript (e.g. FPKM, TPM, or normalized counts) using min-max scaling to map them between 0 and 1 interval. These normalized values serve as the intensity values of the corresponding pixels.

Constructing a 2D Structure: The sorted and normalized features (i.e. the "pixel" intensity values) of each sample are then filled into a two-dimensional matrix according to a predetermined sequence, forming a synthetic "image". The most common approach is raster scanning, where the matrix is filled row by row until all features have been arranged, resulting in an SIE that is approximately square or rectangular. For example, if there are N features, they can be arranged into an approximately sqrt(N) × sqrt(N) square image.

Generating the SIE: Ultimately, each sample's transcriptomic profile is transformed into a unique SIE. In this SIE, pixel positions are determined by the gene/feature ordering, and the brightness (intensity) of the pixels represents the normalized expression level of each gene/feature.

Through this transformation process, the original one-dimensional or tabular transcriptomic data is endowed with a two-dimensional spatial structure, enabling us to apply CNNs to effectively learn and extract latent "spatial" patterns potentially introduced by the feature ordering, thereby enhancing the performance of subsequent analyses such as classification, prediction, or feature selection. To further improve the transparency and reproducibility of the method, we have added two algorithm tables of the model in Fig. S1A and Fig. S1B, which outline each step of the SIE generation process along with its specific parameters.

This method compensates for lossy compression, allowing the generated SIEs to contain more genetic information. In this study, we applied this technology to transcriptomic sequence analysis. When selecting genes or transcripts for analysis, we log2-normalized the expression levels of each gene. For a specific patient, the renormalized expression levels correspond to the pixel brightness in the patient's SIE. From the genes across the 11 datasets used in this study, we selected 16 889 genes based on gene function and chromosome position to create an SIE for each patient in all datasets. The 16 889 genes were arranged into a 224×224 (height×width) grayscale SIE.

The process of converting gene expression data into a grayscale SIE is illustrated in Fig. 1a. Using SIEs technology and scaling techniques, the 16 889 genes were placed in a 224×224 pixel image. More specifically, for the 224×224 configuration, the top 224 of the sorted 16 889 genes were used to build the first channel (layer) of the image, with the next 224 genes forming the second channel. In these arrangements, the same genes from different individuals were located at the same coordinates on the SIE, maintaining the inter-gene associations from the original datasets. Therefore, the analysis results based on the SIEs classification were consistent with the results obtained from the original expression data.

Using the Gene Swin Transformer model for classification and prognosis prediction of Synthetic Image Elements

A Vision Transformer (ViT) [32] is a network architecture entirely based on transformers networks [15], effectively processing image data through self-attention layers. It can capture complex image patterns, enhancing the performance of visual tasks such as image classification. Its embedded multi-layer perceptron (MLP) improves the ability to generalize and learn spatial correlations [33]. However, ViT lacks the inductive biases of CNNs, such as locality and translational invariance, making its training dependent on large-scale datasets and pretrained models. In contrast, the Swin Transformer [16], which adopts a CNN-like structure, performs more powerfully in image feature extraction and classification and prediction tasks.

To address these challenges, we developed a new attention mechanism named Gene Swin Transformer. Gene Swin was designed to efficiently utilize SIEs for survival analysis and prognosis prediction, with its structure illustrated in Fig. 1. In this study, we employed the Gene Swin architecture to classify and predict SIEs derived from selected gene expression data.

Survival analysis

This study utilized R 4.3.3 for data analysis and visualization. The specificity and sensitivity of the model based on the Swin Transformer were assessed by calculating the area under the receiver operating characteristic (ROC) curve (AUC). To evaluate the clinical prognosis of CRC patients, Kaplan–Meier (KM) curve analysis and COX regression analysis were performed, focusing on the risk score derived from the Gene Swin Transformer, with results

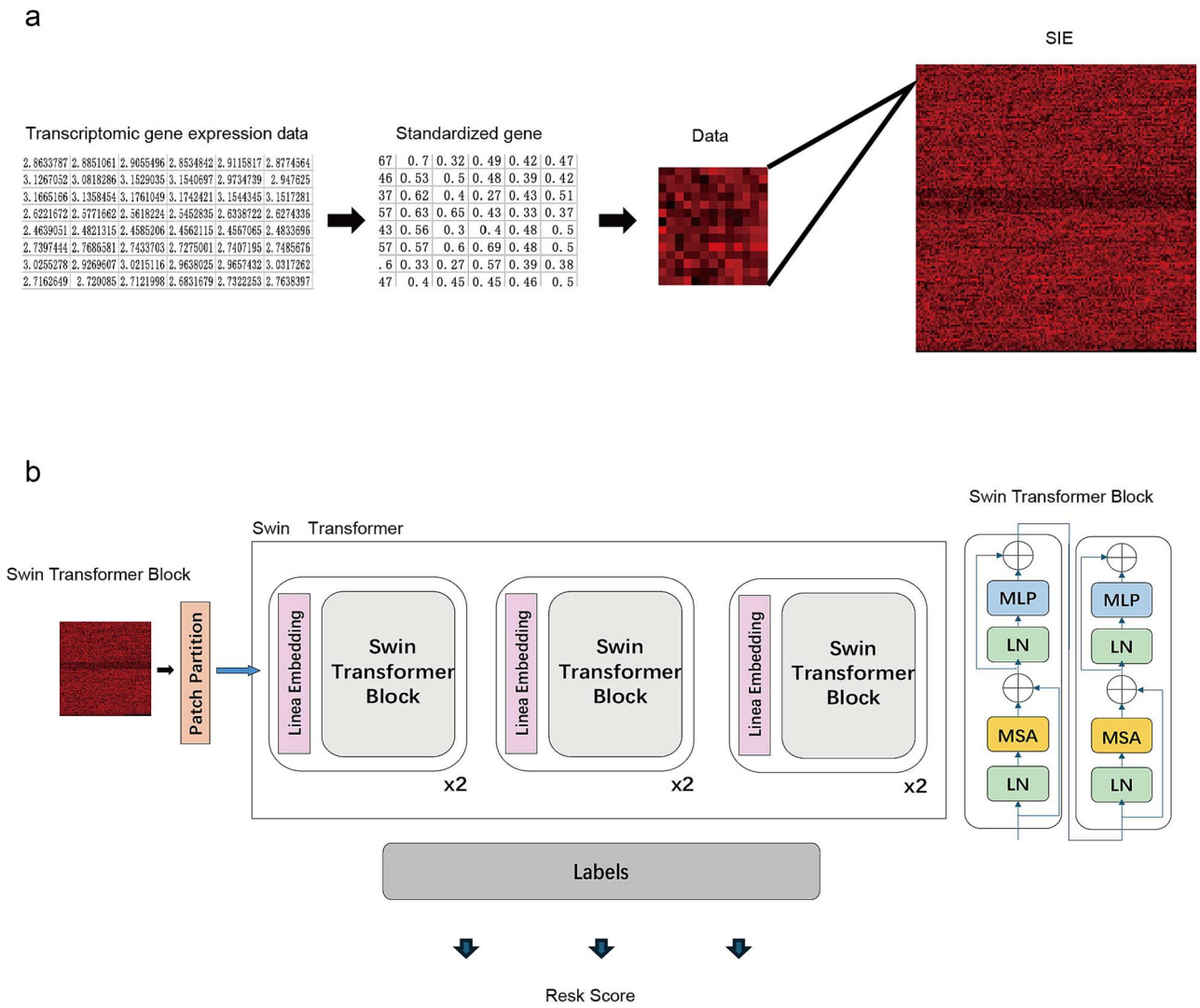

Figure 1. The diagram illustrates the conversion of gene expression data into SIEs and the architecture of the Gene Swin model proposed in this paper. (a) 1. Tabular format of transcriptomic gene expression data. 2. Log2 normalization and rescaling to fit the range of a digital image (0 to 1). 3. Organization of expression data into SIEs. 4. SIEs are color images that integrate multi-layered data. (b) Gene Swin comprises two parts: SIEs and a modified Swin transformer. The Swin transformer uses "shifting windows" and a hierarchical approach. It processes images within local non-overlapping windows and captures dependencies across windows by shifting them at each stage. This structure excels in dense prediction tasks like object detection and bioinformatics image analysis.

reported as hazard ratios (HR) and 95% confidence intervals (CI). Additionally, univariate and multivariate COX regression analyses were conducted to identify significant prognostic markers for CRC patients.

Results

Model performance evaluation via metrics and five-fold cross-validation

Combining the aforementioned 11 datasets, we tested four models, BeiT [34], ResNet [35], Swin Transformer [16], and ViT Transformer [36], to determine the most suitable DL model for this study. After evaluating the accuracy, stability, adaptability, and generalization capability of each model, we found that the Gene Swin model, which included Shifted Window Multi-head Self-Attention (SW-MSA), Hierarchical Representation, Patch Merging, Linear Embedding, Layer Normalization, and MLP, demonstrated an excellent testing accuracy. Detailed information on the model can be found in the Python script published on our website (https://github.com/). We compared the prognostic performance of four models by examining changes in model loss, accuracy, validation accuracy, and F1 score. Additionally, the ROC results for the Swin model (Fig. 3), further highlight its performance in detail. Table 2 consolidates and compares all results, providing a comprehensive overview of outcomes obtained with a 224×224 configuration. This table includes the accuracy, recall, F1 score, and AUC for each of the four models. In terms of prognostic prediction, the Swin model demonstrated a weighted average AUC of 0.7837 with corresponding accuracy, recall, and F1 scores of 0.70793, 0.69198, and 0.7052, respectively.

Model validation and sample testing using GEO and the Cancer Genome Atlas databases

In this study, we utilized a database comprising 11 GEO datasets and one TCGA dataset, which together included clinical information and transcriptome data from 2325 patients. After removing

Table 2. Comparison of the predictive performance of Gene -Swin and three published models on the prognosis of CRC patients using the TCGA-GEO dataset.

DL-model		AUC	Precision	Recall	F1score
	Training set	0.802	0.7112	0.6914	0.7043
Swin-T	Test set	0.727	0.6993	0.6894	0.7048
	Validation set	0.769	0.7032	0.7012	0.7123
	Training set	0.667	0.6812	0.6892	0.6912
BeiT	Test set	0.659	0.6912	0.6851	0.6321
	Validation set	0.621	0.6711	0.6754	0.6881
	Training set	0.627	0.6712	0.6812	0.6912
ResNet	Test set	0.636	0.6786	0.6874	0.6754
	Validation set	0.614	0.6613	0.6743	0.6612
	Training set	0.648	0.6812	0.6712	0.6712
ViT transformer	Test set	0.694	0.6936	0.6943	0.6987
	Validation set	0.657	0.6712	0.6812	0.6832

combined batch effects, we shuffled the data from 2169 patients and randomly divided them into three parts: 70% of the samples were allocated to train the model (training set), 20% were reserved to validate model accuracy (test set), and the remaining 10% were designated as an external validation set. This data classification method helps prevent performance errors in the training model that could arise from technical measurements of gene expression correlation reproducibility across different datasets. Additionally, the use of a test set and an external validation set is intended to more accurately assess and measure the model's generalization capabilities. Based on the results of five-fold cross-validation, we trained and optimized the Swin model to enhance its suitability for analyzing SIE images and predicting prognosis. The improved Gene Swin transformer was tested more than five times on a test set comprising 20% of the total data; the other three models underwent the same testing and validation procedures. The results, including three-class clinical survival data and corresponding SIE images, are presented in Table 2.When the test set was used to evaluate the Gene Swin model, the performance closely matched the five-fold cross-validation results, with an AUC of 0.727, and precision, recall, and F1 scores of 0.6993, 0.6894, and 0.7048, respectively. For the ResNet model, the AUC value for prognosis prediction in the test set was 0.614 with precision, recall, and F1 scores of 0.6613, 0.6743, and 0.6612, respectively. The BeiT model achieved an AUC value of 0.621 in the test set, with precision, recall, and F1 scores of 0.6711, 0.6754, and 0.6881, respectively. The ViT model demonstrated an AUC value of 0.694 in the test set, with precision, recall, and F1 scores of 0.6936, 0.6943, and 0.6987, respectively. The performance of all four models in the validation set is summarized in Table 2.

The loss changes of the four deep models with an increase in the number of training samples (Fig. 2a). In addition, significant differences in precision, recall, and F1 scores among the four models on the test set, as well as the trends in accuracy as training samples increased (Figs. 2b and c). Notably, the accuracy of the Swin transformer was ∼70.5%, demonstrating superior performance compared to the other three models. Furthermore, the loss variation results indicate that the Swin model exhibits greater stability and generalization ability than the other models. To further assess the capability of the model structure and the use of SIE in predicting the prognosis of CRC patients, we utilized an external validation set and performed five-fold cross-validation using the same model structure, achieving an average AUC of 0.769. This outcome closely aligns with the test set results, confirming that the adapted SIE and model structure are effective in classifying the National Health Group (NHG). Overall, the Swin transformer demonstrated outstanding performance across the training set, test set, and external validation set. The accuracy of the prognosis predictions and a comparative analysis of the four models' performance are presented in Fig. 2.

Survival and COX regression analysis based on gene Swin-T risk score in an independent patient cohort

The ROC curve analysis illustrates the performance of the Gene Swin transformer prognosis prediction model across different cohorts: the training set, the testing set, and the external validation cohort at 1 year, 3 years, and 5 years (Fig. 3a–c). The AUC values were 80.2% for the training set, 72.7% for the testing set, and 76.9% for the external validation set, demonstrating the stability and consistent predictive performance of the model across various datasets.

We used the R package maxstat [37] to calculate the optimal cut-off value for the Risk Score in the Gene Swin training set, with the minimum group sample size set to more than 25% and the maximum group sample size set to <75%. The final optimal cut-off value obtained was 0.689908, allowing us to categorize patients in each cohort into high Risk Score or low Risk Score groups (Fig. 3d). The significance of prognostic differences between these groups was assessed using the logrank test, which revealed significant differences (HR = 0.20, 95% CI: 0.14–0.29, $P < .0001$). Similarly, KM survival analyses of overall survival stratified by risk scores in the testing and validation cohorts of CRC patients are presented in Fig. 3e and f. In the testing cohort (HR = 0.36, 95% CI: 0.22–0.57, $P < .0001$, Fig. 3f) and validation cohort (HR = 4.99, 95% CI: 2.48–10.03, $P < .0001$), the differences in overall survival between high and low risk score groups were statistically significant. These findings highlight the significance of DL attention mechanism techniques for risk scoring in predicting clinical outcomes in CRC patients.

We utilized the R package glmnet to integrate survival time, survival status, and gene expression data, performing regression analysis on the training set using the Least Absolute Shrinkage and Selection Operator (LASSO)-Cox method (Fig. 5a and b). To optimize the model, we conducted 10-fold cross-validation, setting the Lambda value at 0.07, which led to the identification of 24 genes associated with the survival of CRC samples in the test cohort. Simultaneously, using the R package survival, we combined the survival time, survival status, and expression data of these 24 genes to evaluate their prognostic significance in the training set using the cox method (Fig. 5c).

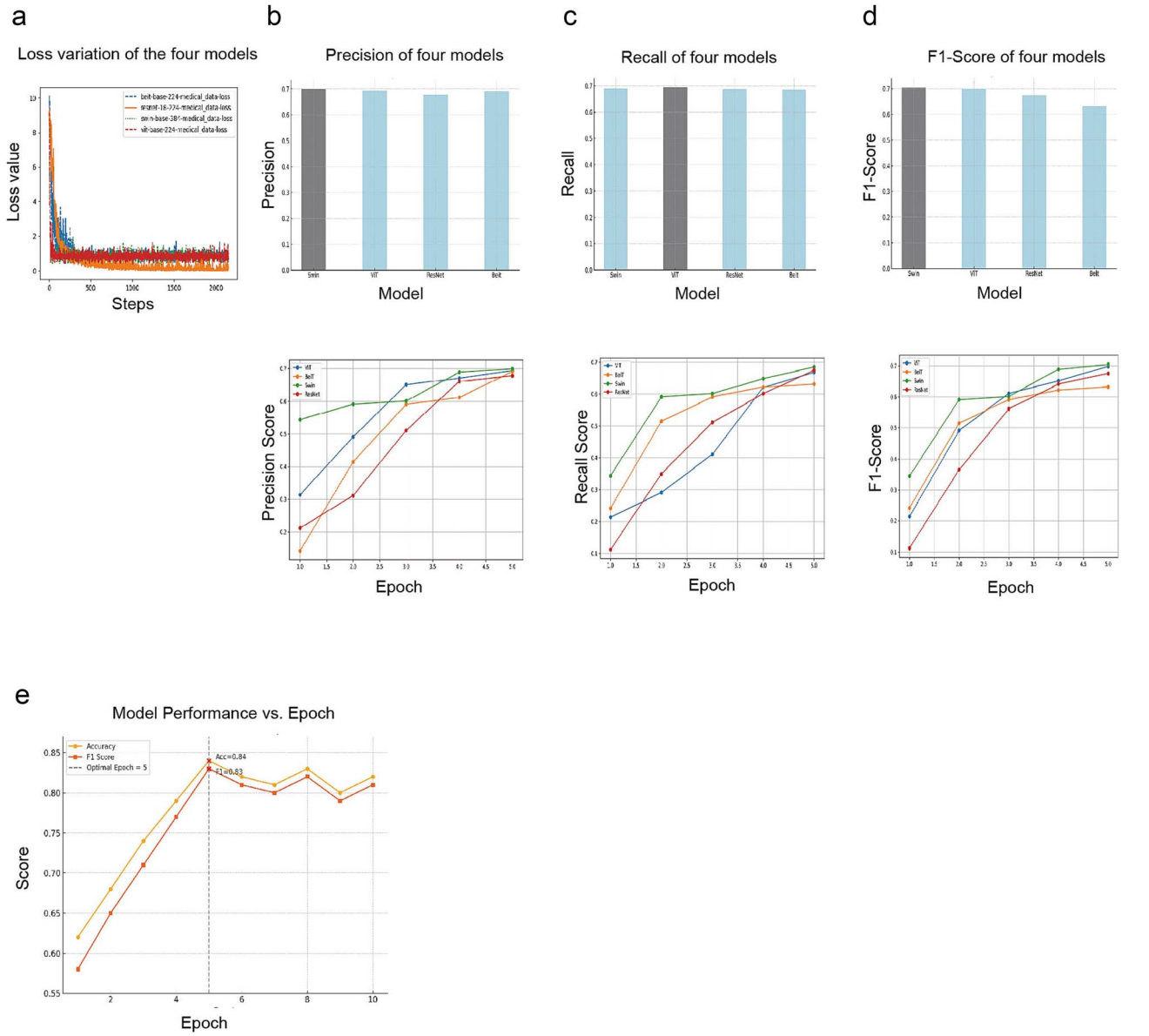

Figure 2. Comparison of performance in predicting prognosis across four models. (a) Variations in loss across the four deep learning models. (b) Comparison of the test set accuracy for the four models, along with the trend in accuracy as the number of training samples increases. (c) Comparison of the recall rates on the training set, including the trend in recall rate changes with increasing training samples. (d) Comparison of F1 scores on the training set, and the trend of F1 score changes as training samples increase. Detailed numerical comparisons are provided in Table 2. (e) Price chart of the performance of DL models with the number of training epochs.

Multivariate survival regression confirms the predictive accuracy of the gene Swin model for colorectal cancer patient survival

Currently, survival predictions for patients with CRC are primarily based on a range of known clinical and pathological factors, including clinical tumor staging, tumor gene typing, and patient age. However, our multivariate survival regression analysis demonstrated the high accuracy of the Gene Swin model in predicting the survival of CRC patients. To validate and improve the current accuracy of survival predictions, we constructed an integrated nomogram across the entire patient cohort based on the risk score from the Gene Swin transformer and predictable clinical and pathological factors (Fig. 4a).

ROC analysis revealed that the AUC value for predicting Disease-Free Survival (DFS) in CRC patients using the nomogram was 63.7% (Fig. 4b), which was notably higher than the predictive performance of TNM staging (AUC = 57.7%). This finding highlights the superior predictive accuracy of the integrated nomogram compared with traditional clinical and pathological factors. Moreover, the Gene Swin-based predictive model demonstrated robust effectiveness in forecasting the prognosis of patients with CRC, underscoring its potential for further clinical validation and widespread application. Supporting these results, both the decision curve analysis (DCA) curve and the calibration curve of multivariate survival regression reinforced that the risk score generated by the Gene Swin transformer

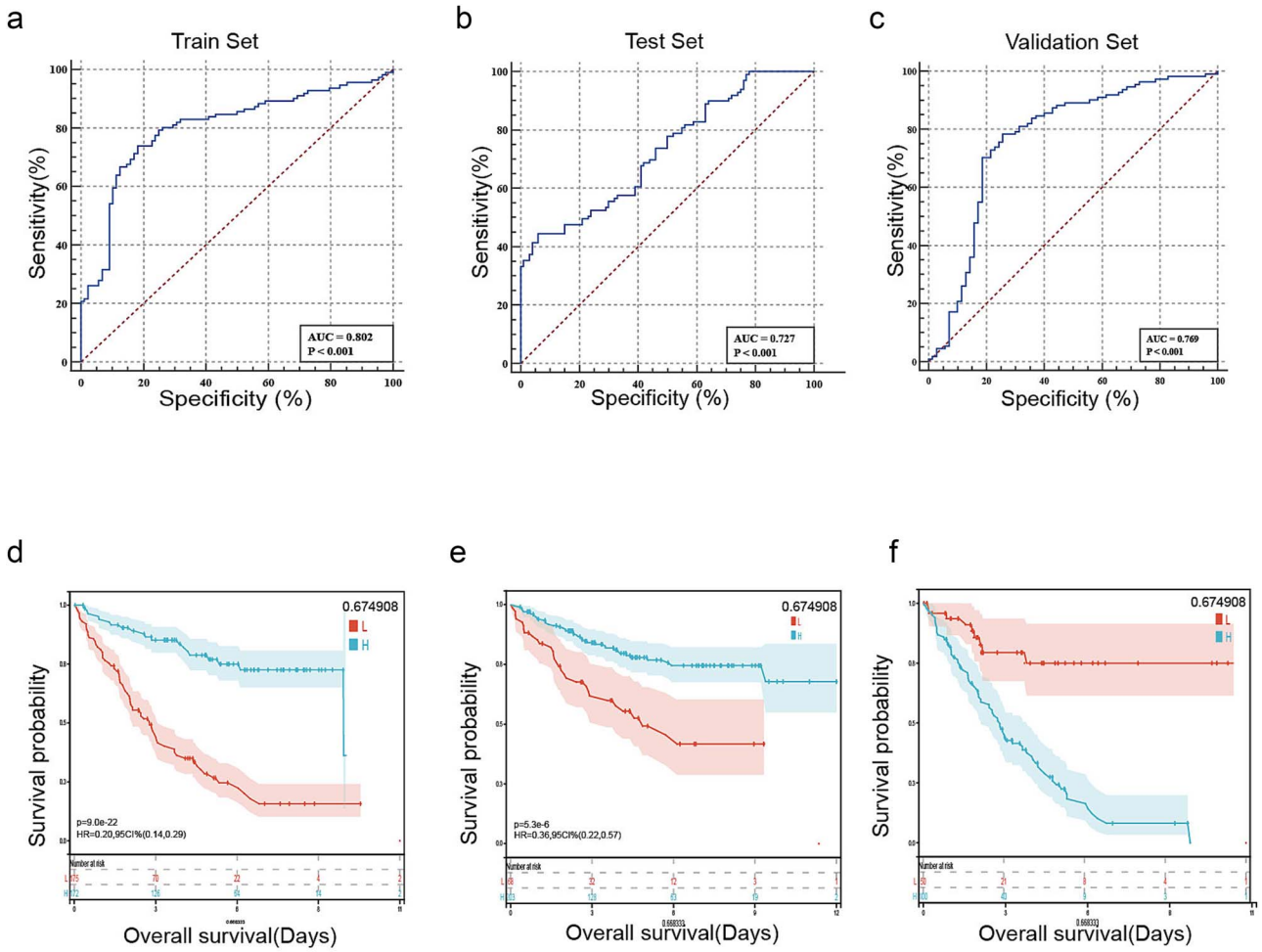

Figure 3. ROC curve and KM survival analyses for Gene Swin's prediction of disease-free survival in CRC patients. (a–c) ROC curves showing the prognostic prediction scores from Gene Swin in the training, test, and validation sets. (d–f) KM survival analyses, stratified by Gene Swin risk scores, demonstrate overall survival outcomes for CRC patients in the training, test, and validation cohorts. (AUC, Area under the curve; CRC, Colorectal cancer; ROC, Receiver operating characteristic).

provided highly accurate prognostic predictions for CRC patients (Fig. 4c and d).

Identification of prognosis-related high-contribution genes via attention matrix visualization

Currently, survival prediction for CRC patients is primarily based on established clinical and pathological factors [38, 39], such as TNM staging of tumor tissues, tumor markers CEA and CA199, and results of immunohistochemical tests. High expression of tumor markers or pMMR status is often associated with the overexpression or underexpression of certain genes, which we refer to as high-contribution genes related to prognosis. By visualizing the attention matrix, we highlighted the regions with significant attention weights in the Gene Swin model within the SIE. The brighter the color of a region, the higher the attention score assigned by the model to that region, indicating a greater contribution to the model's output prognostic risk score. From the SIE images, we extracted high-contribution genes related to prognosis, as illustrated in Fig. 5a. Compared with the features obtained through LASSO-COX analysis, our Gene Swin prognostic model identified 29 prognosis-related genes that were not previously detected by traditional bioinformatics analysis. Table 3 presents the genes identified by both methods. Through simultaneous visualization of the attention matrix and bioinformatics analysis, we identified the overlapping high-contributing gene PEX10. A literature review further revealed that enzalutamide inhibits PEX10 function, sensitizing prostate cancer cells to ROS activators [40]. However, no studies have reported a similar effect in CRC.

This discovery opens new possibilities for exploring prognosis-related gene pathways and mechanisms, as well as providing potential new targets for clinical translation and the development of novel therapeutic drugs.

Discussion

The rapid development of high-throughput DNA sequencing technology has revolutionized the field of clinical research, leading to the rapid accumulation of large-scale transcriptome data. This presents significant opportunities for applying DL algorithms to address clinical challenges, particularly because these algorithms can handle nonlinear and high-dimensional data, making them well-suited for interpreting complex biological datasets, such as transcriptome data. However, most studies on transcriptomic genes and survival rely on biomarkers to predict prognosis [38, 39], which may result in the loss of valuable gene expression information. To overcome this limitation, our study innovatively employed DL algorithms to convert gene expression data into SIEs,

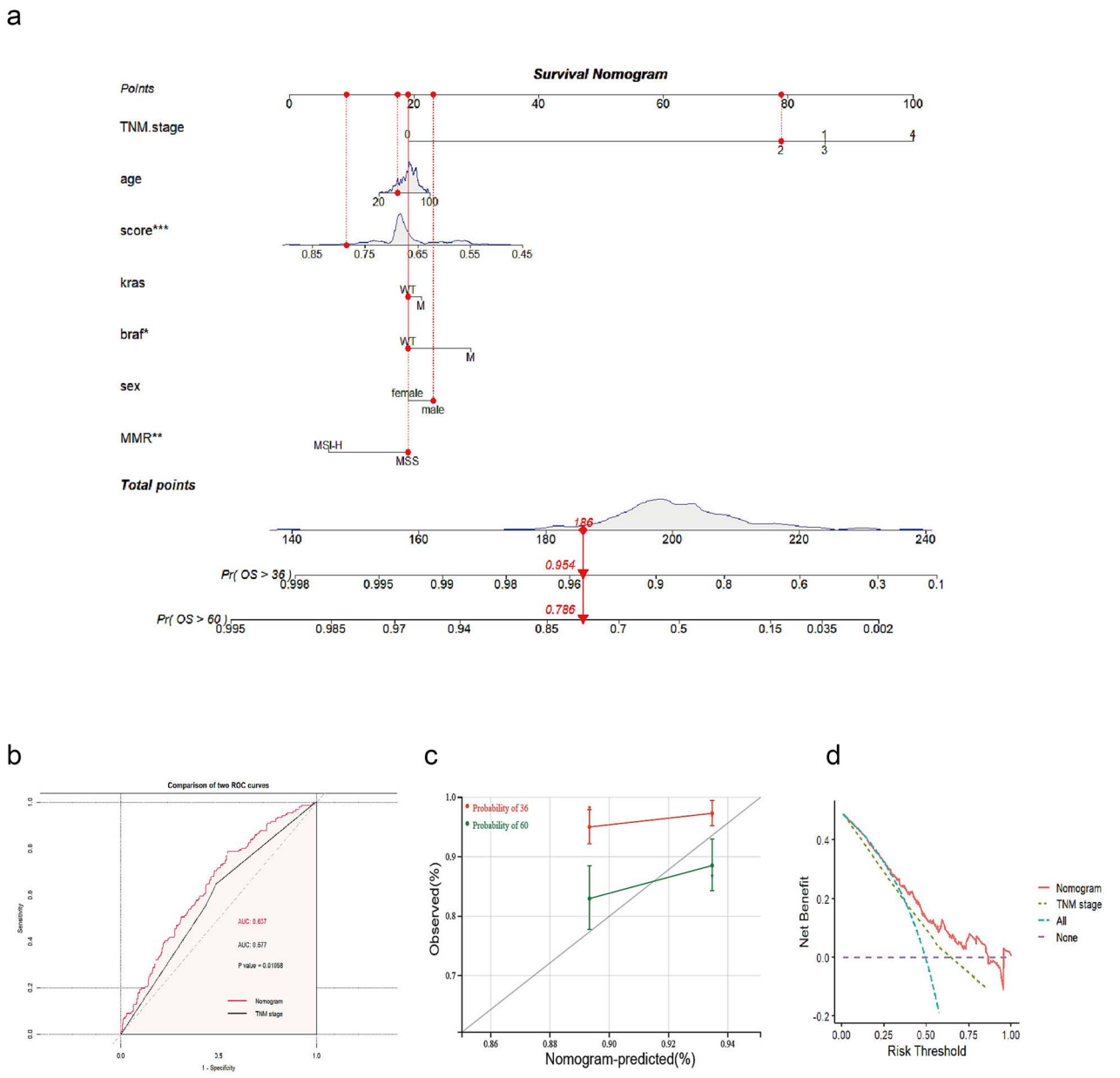

Figure 4. Multivariate survival regression analysis for the Gene Swin model and construction of the nomogram model. (a) Nomogram model predicting 3-year, and 5-year survival rates. (b) ROC curve for the nomogram model's survival prediction. (c) DCA of the nomogram (a) for survival prediction. (d) Calibration plot of the nomogram, showing the correlation between predicted probabilities and actual values.

enabling a more intuitive understanding of their characteristics. The challenge lies in effective use of SIEs to solve clinical problems. The introduction of artificial intelligence has significantly transformed traditional diagnostic methods and prognosis prediction approaches [38]. The emergence of attention mechanism models, such as the Swin Transformer, addresses this challenge by optimizing the standard Transformer architecture with a sliding window approach, making it more suitable for structured data, such as images and potentially three-dimensional medical imaging data. The Swin Transformer possesses strong feature extraction capabilities, enabling it to accurately identify and localize disease-related features when processing structured image data such as histological images, pathological slides, or other biomedical imaging data. However, very few studies have reported the application of attention mechanism models to transcriptomic date of CRC.

In this study, we applied DL algorithms to RNA sequencing data of CRC and compared the prognostic prediction capabilities of the four models in patients with CRC. Based on the training results, we selected and improved the most suitable model for analyzing SIEs. Our study has two main objectives: first, to optimize the representation of SIE technology in the transcriptomic data of CRC patients, preserving the independence and integrity of the original feature data while making the transcriptomic data expression more intuitive; second, to evaluate the performance of our Gene Swin Transformer model using SIEs for prognostic prediction and validate the clinical value of the improved Swin Transformer model in enhancing prognostic accuracy.

We first transformed the transcriptomic gene sequencing data into SIEs (Fig. 1) and then divided all patients' SIEs in the dataset into three cohorts: the training cohort (70%), test cohort (20%), and external validation cohort (10%). We employed four distinct CNN

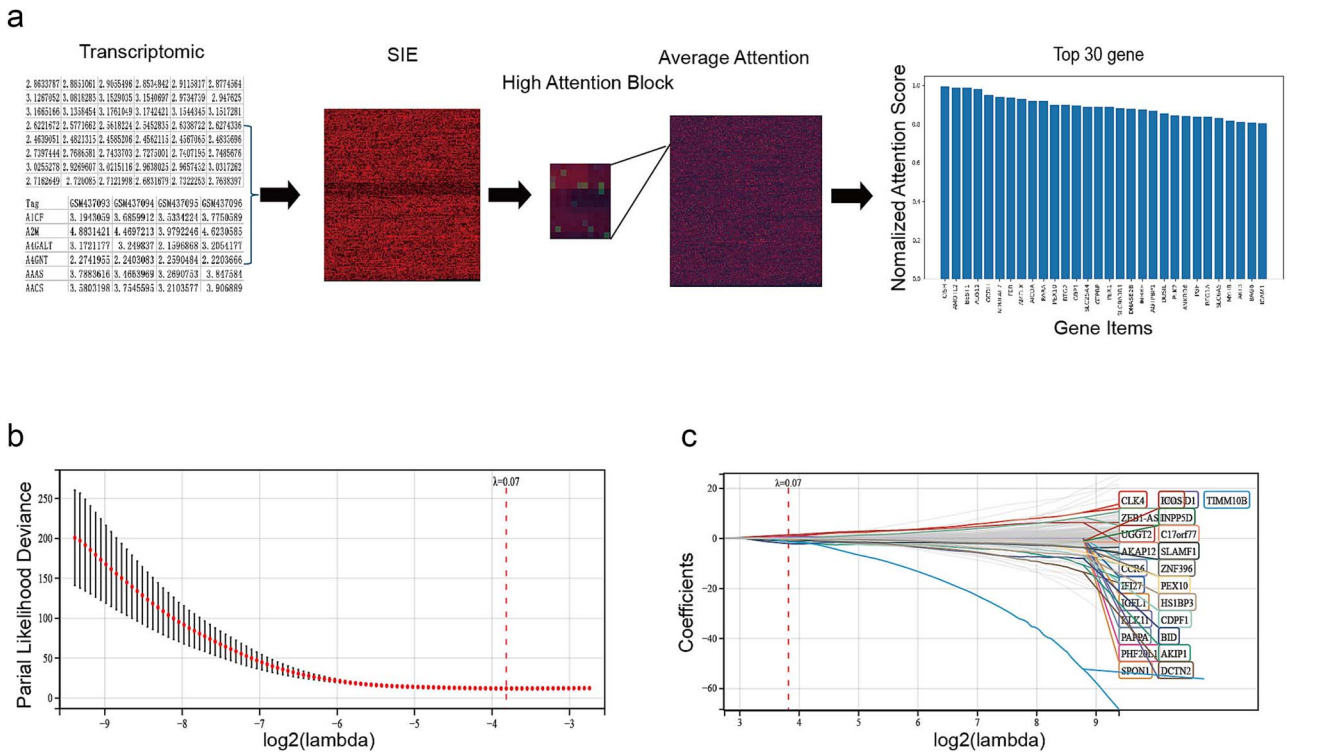

Figure 5. Comparison of high-contribution prognosis-related genes identified by DL and bioinformatics analysis. (a) Genes identified by Gene Swin using visual attention matrix algorithms (highlighting high attention weights). (b) Ten-fold cross-validation error in LASSO analysis. (c) LASSO-COX analysis identifying 24 survival-related genes. The overall prognostic significance was robust (logtest = 5.12921649608594e-25, sctest = 1.72339363408518e-25, waldtest = 1.19757559188477e-18), with a C-index of 0.855861304304002.

architectures—ViT-base, Swin-base, BEiT-base, and ResNet-50—to evaluate their performance on the training dataset. Each model was trained for five epochs under identical conditions to ensure fair comparison. The reason for selecting five training epochs is to achieve optimal performance: fewer epochs may lead to underfitting, while more than five epochs may result in overfitting. We chose five epochs to ensure that the model achieves the best balance between underfitting and overfitting. Underfitting occurs when the model fails to adequately learn the data features, while overfitting happens when the model excessively adapts to the training data, thus reducing its ability to generalize to new data. By selecting five epochs, we ensure that the model can effectively learn the data features without memorizing irrelevant details. This decision is supported by the performance curve during model training, which shows that the model's accuracy improves as the number of epochs increases, but decreases after a certain point due to overfitting. Additionally, using the same number of training epochs for all models ensures consistency in evaluation, enabling a fair comparison of the feature extraction capabilities of different models. This approach also helps in assessing the convergence behavior of each model and its ability to consistently extract features under identical training settings. The results in the chart clearly reflect the impact of training rounds on model performance. The results are provided in Fig. 2e. By maintaining consistent training parameters across all models, we aimed to directly compare their architectural strengths and explore their potential applications in bioinformatics, particularly for tasks that require robust feature representation. The training cohort was used to train and tune the performance of the four deep-learning models for prognostic prediction, and the tuned models were then applied to the test set. By comparing the precision, recall, and F1 scores of the four models (Table 2), we concluded that the Gene Swin model based on the SIE technology proposed in this study outperformed the other three advanced models in prognostic prediction across the three patient cohorts, demonstrating high predictive accuracy.

We developed an innovative gene ViT DL model named "Gene Swin", which achieves prognostic prediction using gene expression data through three main steps: (i) data preprocessing, (ii) converting transcriptomic sequencing data into SIEs, and (iii) inputting the SIEs into the improved Swin-T model for prediction, ultimately outputting a risk score associated with prognosis. The model consists of three core components. The first component is responsible for data preprocessing, starting with data augmentation to expand the data volume and reduce class imbalance. Next, the min-max normalization method was used to normalize the augmented dataset, ensuring stable convergence of the model's weights and biases. Dimensionality reduction is then performed using stacked autoencoders to select the genes most relevant to the study. The second component utilizes the SIE technique with a channel extension algorithm, transforming the data samples into an image format, i.e. leveraged for feature extraction and classification. This structure not only improves the data processing efficiency but also enhances the model's ability to understand features. The third component is the improved Swin-T model, which processes the SIEs to output the risk score associated with prognosis.

This study demonstrates that the Swin Transformer (Swin-T) model outperforms BeiT, ResNet, and ViT in predictive tasks, primarily due to its unique structural design, which is particularly suited for handling latent patterns within our generated SIEs.

Enhanced multi-scale feature capture: Transcriptomic data complexity implies critical information may manifest in local details (fine interactions between a few genes) or global structures

Table 3. Presents the high-contribution genes identified through both bioinformatics and DL algorithm analyses. Prognosis-related high-contribution genes.

	Bioinformatics-analysis	Deep learning algorithm
1	AKAP12	AXELX
2	AKIP1	ALG12
3	BID	AICDA
4	C17orf77	AMOTL2
5	CCR6	AKT3
6	CDPF1	AGTPBP1
7	CLK4	ANKRD6
8	DCTN2	BTG2
9	HS1BP3	BAG6
10	ICOS	BEST1
11	IFI27	CISH
12	IGFL1	DANSE2B
13	INPP5D	DUSIL
14	KLK11	FER
15	PAPPA	GCDH
16	PEX10	GBP1
17	PHF20L1	GTPBP
18	SLAMF1	MYH8
19	SPON1	NDUFAF7
20	TIMM10B	ICAM1
21	UGGT2	INHBE
22	ZEB1-AS1	PGF
23	ZFAND1	PEX1
24	ZNF396	PEX10
25		PLK2
26		REG1A
27		RARA
28		SLC6A5
29		SLC25A4
30		SLC9A3R1

Note: PEX10 is the overlapping gene and has been highlighted.

(coordinated changes across gene sets). Swin-T constructs feature maps of varying resolutions through hierarchical patch merging, naturally and effectively capturing these multi-scale details. In contrast, ViT and BeiT typically use single-resolution patches and global attention, potentially less efficient. ResNet's convolutional approach may struggle with flexible capture of distant gene interactions.

Efficient local correlation handling: Mapping gene expression data to SIEs, adjacent "pixels" or regions may represent genes or features with biological associations. Swin-T's windowed self-attention restricts computations to local windows, reducing complexity while focusing on local gene modules or interaction patterns. In contrast, ViT/BeiT's global attention may prematurely handle all patch relations, leading to computational overload.

Balancing local and global information: The gene expression regulatory network is highly complex, where long-distance gene interactions are equally critical. Swin-T's shifted window mechanism effectively propagates information across window boundaries between layers, maintaining computational efficiency while modeling local and global dependencies. In contrast, ResNet expands receptive fields through stacked layers, and ViT/BeiT's global attention has inherent limitations.

In summary, Swin-T, with its layered design, efficient local window attention, and innovative shifted window mechanism, efficiently captures and integrates complex gene expression patterns at multiple scales in SIEs. This explains its superior performance over BeiT, ResNet, and ViT in predictive tasks.

Finally, we integrated survival time, survival status, and gene expression data from the test and validation cohorts and conducted a regression analysis using the LASSO-COX method, which ultimately identified 24 prognosis-related genes (Fig. 5c). Additionally, by visualizing the self-attention matrix, we extracted high-contribution, prognosis-related genes from the SIE images (Fig. 5a). A comparison between the two methods showed that our Gene Swin model identified 29 prognosis-related genes that were not detected by conventional bioinformatics analysis. Although these genes are linked to prognosis, their mechanisms of action remain unclear. To investigate the molecular mechanisms by which these genes affect CRC cells, various techniques such as proteomics, transcriptomics, and single-cell sequencing can be employed to analyze the downstream signaling pathways they regulate. The relationship between these genes and prognosis can also be validated through *in vivo* or *in vitro* experiments, for instance, by testing their effects using CRC cell lines. The overexpression or knockout of these genes can be used to observe their impact on cell proliferation, migration, invasion, and apoptosis. Alternatively, animal models, such as nude mice or transgenic mouse models, can be developed to study the effects of gene overexpression or knockout on CRC growth. Finally, analyzing the expression levels of these genes in CRC patient tissue samples and correlating them with clinical data, such as patient prognosis and treatment response, can evaluate their potential as therapeutic targets and facilitate the translation from research to clinical practice.

Our research did not adopt SHAP (SHapley Additive exPlanations) values to enhance model interpretability. While SHAP values are indeed a powerful tool in many machine learning scenarios, they are not suitable for our research context for several key reasons:

SHAP values require evaluating the contribution of each feature individually. For high-dimensional transcriptome data, which contains thousands of genes, the computational cost is extremely high. Our model handles complex SIEs, and SHAP analysis may require significant computational resources, with results that may be difficult to directly translate into biological insights. SHAP values are typically more suited for traditional machine learning models (such as random forests or gradient boosting machines) or neural networks with fewer features. In large transformer models, like the Swin Transformer that we used, the interpretation of SHAP values might not be intuitive because the decision-making process in these models heavily depends on nonlinear interactions between features. In contrast, the attention mechanism directly reflects the internal workings of the model, providing a more natural and model architecture-aligned way to explain the model. Our input data consist of converting transcriptome data into SIEs, a process that increases the complexity of the data. The attention mechanism effectively analyzes the importance of these synthetic elements, whereas SHAP values may struggle to adapt to this nonstandard input format.

Our interpretability approach is highly consistent with the latest trends in interpretability of transformer models in the field of bioinformatics. The attention mechanism has been widely recognized as an effective tool for uncovering the decision-making process of DL models in biological sequence analysis. For instance, Zhang *et al.* [41] pointed out in a review that the interpretability and adaptability of transformer models have led to their widespread application in bioinformatics, especially in gene sequence analysis and drug discovery. They emphasized that self-attention mechanisms not only improve model accuracy but

also enhance model transparency by revealing key parts of the input data.

Similarly, Choi and Lee [42] discussed in a review of transformer architectures for genomic data analysis how attention mechanisms integrate domain knowledge (such as known biological relationships) to make the model's representations more interpretable. They noted that attention weights could be used to identify biologically important features, such as key genes or regulatory elements, which aligns with our method of identifying the PEX10 gene using attention matrices.

Furthermore, Wu *et al.* [43] proposed a transformer-based genomic prediction model called GPformer, which utilizes attention mechanisms to capture global relationships between single nucleotide polymorphisms. Their research further validates the interpretability potential of attention mechanisms when handling complex biological data, although their primary focus was genomic prediction rather than prognostic analysis.

From this analysis, we can see that the computational complexity, model specificity, and data characteristics of SHAP values make them less advantageous in our study. On the other hand, the attention mechanism not only fits large transformer models better but also handles complex data formats effectively, and is highly consistent with the latest trends in bioinformatics. Relevant studies support the use of visualized attention mechanism matrices to enhance model interpretability, which further validates our choice.

Despite these promising results, this study had several limitations. First, although our predictive model performed well in distinguishing between patients with high and low survival risk, the study was retrospective in nature. Therefore, prospective validation is necessary before the model can be applied to clinical practice. Second, the prognosis-related genes identified by Gene Swin have not been systematically validated and their mechanisms of action remain unexplored. This indicates that significant work remains before these findings can be translated into clinical application.

This study presents a new method called Gene Swin Transformer, which transforms transcriptomic data into SIEs and uses DL models to predict the prognosis of CRC patients. This data-driven, automated prognostic modeling method offers new possibilities for personalized medicine, assisting clinical decision-making through more accurate patient outcome predictions. Converting complex transcriptomic data into a visual form and processing it using advanced DL algorithms represents a significant advancement in the fields of bioinformatics and clinical oncology.

However, there are several limitations in this study. First, the datasets used in the study (such as GSE17536-GSE103479 and TCGA) may have biases, such as sample source (mainly North American patients) or an imbalance in the distribution of clinical variables. Although we mitigated these biases by integrating multiple datasets and stratifying by clinical variables, further validation on broader, independent datasets is needed to confirm the mode's generalization ability and robustness.

Second, training Transformer-based DL models (such as Swin Transformer) requires substantial computational resources. Our study used an NVIDIA RTX 3090 GPU, with each training session taking ~10–15 hours. Although the popularity of cloud computing and efficient hardware has made these demands manageable, they may still pose a barrier for researchers with limited resources. However, once model training is completed, the inference time is relatively short (usually a few seconds to minutes), making the model practical for clinical applications.

Third, while this study focuses on CRC, the method of converting transcriptomic data into SIEs is theoretically applicable to other cancer types. However, the gene expression characteristics of different cancers (such as the number of differentially expressed genes or the complexity of regulatory networks) may affect model performance. Therefore, although we believe this method has broad applicability, it needs to be validated on datasets from other cancer types (such as breast cancer or lung cancer) to confirm its effectiveness.

Additionally, our method assumes that SIEs can effectively capture the biological information in transcriptomic data. While this assumption has been validated in this study, different data representation methods may be required for other tasks or cancer types. Moreover, this study focuses solely on survival prediction and does not address other clinical outcomes, such as treatment response or recurrence risk. Exploring the potential of this method in other clinical applications is an important direction for future research.

Based on these considerations, future research directions include: (i) validating the model's performance on independent datasets with greater geographical and clinical diversity to confirm its generalization ability; (ii) applying the Gene Swin Transformer method to data from other cancer types to assess its applicability; (iii) optimizing the model's computational efficiency through model distillation or more lightweight architectures to improve accessibility; (iv) exploring the model's performance on other clinical endpoints (such as treatment response or recurrence risk) to further elucidate its clinical value.

Conclusion

In this paper, we present a technique that transforms genomic data into artificial images and compares the performance of four models using these SIEs to predict patient prognosis. Our results show that the enhanced Gene Swin model outperforms the other models. Specifically, the risk indicators generated by the Gene Swin Transformer proved to be reliable prognostic markers for CRC patients, highlighting the clinical utility of SIE technology in forecasting patient outcomes. These findings suggest that SIE technology has significant potential for widespread use in both clinical and genomic research, paving the way for more effective diagnostic and therapeutic strategies for cancer patients.

However, it is important to note that these experimental conclusions require further validation through prospective studies. Additionally, by visualizing the SIE attention matrix in CRC patients, we identified high-contribution genes related to prognosis that traditional bioinformatics methods have overlooked. Despite this promising discovery, these genes have not yet undergone systematic validation, and substantial work remains before they can be translated into clinical treatment.

Key Points

- A novel deep learning model, Gene Swin Transformer, is developed to predict colorectal cancer prognosis using transcriptomic data.
- The model achieves superior accuracy compared to existing methods by integrating transformer-based architecture with transcriptomic features.
- Identifies key gene signatures associated with colorectal cancer prognosis, providing potential biomarkers for clinical applications.

- Demonstrates robustness across multiple independent datasets highlighting its generalizability and translational potential.
- Offers insights into the biological mechanisms underlying colorectal cancer progression.

Acknowledgements

Thanks for the data provided by Harbin Medical University Cancer Hospital.

Author contributions

Yangyang Wang and Xinyu Yue designed and conceptualized the project, drafted the manuscript, and contributed equally. Shenghan Lou and Peinan Feng were responsible for data collection. Yanlong Liu and Binbin Cui conducted the final review of the manuscript. All authors contributed to the article and approved the final version.

Conflict of interest

The authors declare no competing of interests.

Funding

None declared.

Data access statement

The research data supporting this publication can be accessed through the TCGA and GEO databases, with the following URLs: https://portal.gdc.cancer.gov/ and https://www.ncbi.nlm.nih.gov/geo/.

Consent for publication

Not applicable.

Ethical compliance

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

References

1. Arnold M, Sierra MS, Laversanne M. *et al.* Global patterns and trends in colorectal cancer incidence and mortality. *Gut* 2017;**66**:683–91. https://doi.org/10.1136/gutjnl-2015-310912.
2. Sung H, Ferlay J, Siegel RL. *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;**71**:209–49. https://doi.org/10.3322/caac.21660.
3. Li J, Ma X, Chakravarti D. *et al.* Genetic and biological hallmarks of colorectal cancer. *Genes Dev* 2021;**35**:787–820. https://doi.org/10.1101/gad.348226.120.
4. Kuntz S, Krieghoff-Henning E, Kather JN. *et al.* Gastrointestinal cancer classification and prognostication from histology using deep learning: Systematic review. *Eur J Cancer* 2021;**155**:200–15. https://doi.org/10.1016/j.ejca.2021.07.012.
5. Chaudhary K, Poirion OB, Lu L. *et al.* Deep learning-based multi-omics integration robustly predicts survival in liver cancer. *Clin Cancer Res* 2018;**24**:1248–59. https://doi.org/10.1158/1078-0432.CCR-17-0853.
6. Chen L, Lu D, Sun K. *et al.* Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. *Gene* 2019;**692**:119–25. https://doi.org/10.1016/j.gene.2019.01.001.
7. Sharma A, Yadav D, Rao P. *et al.* Identification of potential therapeutic targets associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. *Comput Biol Med* 2022;**146**:105688. https://doi.org/10.1016/j.compbiomed.2022.105688.
8. Zheng H, Liu H, Ge Y. *et al.* Integrated single-cell and bulk RNA sequencing analysis identifies a cancer associated fibroblast-related signature for predicting prognosis and therapeutic responses in colorectal cancer. *Cancer Cell Int* 2021;**21**:552. https://doi.org/10.1186/s12935-021-02252-9.
9. Ye L, Zhang T, Kang Z. *et al.* Tumor-infiltrating immune cells act as a marker for prognosis in colorectal cancer. *Front Immunol* 2019;**10**:2368. https://doi.org/10.3389/fimmu.2019.02368.
10. Khan S, Rahmani H, Shah SAA. *et al.* A guide to convolutional neural networks for computer vision. *Synthesis Lectures on Computer Vision* 2018;**8**:1–207. https://doi.org/10.1007/978-3-031-01821-3.
11. Zou J, Huss M, Abid A. *et al.* A primer on deep learning in genomics. *Nat Genet* 2019;**51**:12–8. https://doi.org/10.1038/s41588-018-0295-5.
12. Deng J, Hu S, Fan W. Privacy-preserving deep learning for grey scale image classification with pixel-based encryption. In: Xie B, ed. *Proceedings of the 2023 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC 2023); 2–4 Nov 2023; Jiangsu, China*. Piscataway (NJ): IEEE; 2023. p. 263–271. https://doi.org/10.1109/CyberC58899.2023.00049.
13. Esteva A, Robicquet A, Ramsundar B. *et al.* A guide to deep learning in healthcare. *Nat Med* 2019;**25**:24–9. https://doi.org/10.1038/s41591-018-0316-z.
14. Tran KA, Kondrashova O, Bradley A. *et al.* Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med* 2021;**13**:152. https://doi.org/10.1186/s13073-021-00968-x.
15. Vaswani A, Shazeer N, Parmar N. *et al.* Attention Is All You Need. arXiv 2017, https://doi.org/10.1021/acschemneuro.5c00203.
16. Liu Z, Lin Y, Cao Y, *et al.* Swin Transformer: hierarchical vision transformer using shifted windows. In: Damen D, Hassner T, Pal C, Sato Y, eds. *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV); 10–17 Oct 2021; Montreal, QC, Canada*. Piscataway (NJ): IEEE; 2021. p. 10012–10022. https://doi.org/10.1109/ICCV48922.2021.00986.
17. Wu J, Xu Q, Shen Y. *et al.* Swin transformer improves the IDH mutation status prediction of gliomas free of MRI-based tumor segmentation. *J Clin Med* 2022;**11**. https://doi.org/10.3390/jcm11154625.
18. Song B, Kc DR, Yang RY. *et al.* Classification of mobile-based oral cancer images using the vision transformer and the Swin transformer. *Cancers (Basel)* 2024;**16**. https://doi.org/10.3390/cancers16050987.
19. Li E, Yang X, Du Y. *et al.* CXCL8 associated dendritic cell activation marker expression and recruitment as indicators of Favorable outcomes in colorectal cancer. *Front Immunol* 2021;**12**:667177. https://doi.org/10.3389/fimmu.2021.818487.
20. Liu Z, Xu H, Ge X. *et al.* Gene expression profile reveals a prognostic signature of non-MSI-H/pMMR colorectal cancer. *Front Cell Dev Biol* 2022;**10**:790214. https://doi.org/10.3389/fcell.2022.1079548.

21. Yue T, Chen S, Zhu J. *et al.* The aging-related risk signature in colorectal cancer. *Aging (Albany NY)* 2021;**13**:7330–49. https://doi.org/10.18632/aging.202589.
22. Lee JH, Jung S, Park WS. *et al.* Prognostic nomogram of hypoxia-related genes predicting overall survival of colorectal cancer-analysis of TCGA database. *Sci Rep* 2019;**9**:1803. https://doi.org/10.1038/s41598-018-38116-y.
23. Huang X, Ke K, Jin W. *et al.* Identification of genes related to 5-fluorouracil based chemotherapy for colorectal cancer. *Front Immunol* 2022;**13**:887048. https://doi.org/10.3389/fimmu.2022.1058606.
24. Di Z, Zhou S, Xu G. *et al.* Single-cell and WGCNA uncover a prognostic model and potential oncogenes in colorectal cancer. *Biol Proced Online* 2022;**24**:13. https://doi.org/10.1186/s12575-022-00175-x.
25. Yu Q, Wang X, Yang Y. *et al.* Upregulated NLGN1 predicts poor survival in colorectal cancer. *BMC Cancer* 2021;**21**:884. https://doi.org/10.1186/s12885-021-08621-x.
26. Tang Q, Hu X, Guo Q. *et al.* Discovery and validation of a novel metastasis-related lncRNA prognostic signature for colorectal cancer. *Front Genet* 2022;**13**:704988. https://doi.org/10.3389/fgene.2022.1056389.
27. Zhang L, Xu C, Wang SH. *et al.* Cancer-associated fibroblast-related gene signatures predict survival and drug response in patients with colorectal cancer. *Front Genet* 2022;**13**:1054152. https://doi.org/10.3389/fgene.2022.1054152.
28. Qiao Y, Jiang X, Li Y. *et al.* Identification of a hypoxia-related gene prognostic signature in colorectal cancer based on bulk and single-cell RNA-seq. *Sci Rep* 2023;**13**:2503. https://doi.org/10.1038/s41598-023-29718-2.
29. Jiang Y, Yuan H, Li Z. *et al.* Global pattern and trends of colorectal cancer survival: a systematic review of population-based registration data. *Cancer Biol Med* 2021;**19**:175–86. https://doi.org/10.20892/j.issn.2095-3941.2020.0634.
30. Sharma A, Vans E, Shigemizu D. *et al.* DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture. *Sci Rep* 2019;**9**:11399. https://doi.org/10.1038/s41598-019-47765-6.
31. Sokač M, Kjær A, Dyrskjøt L. *et al.* Spatial transformation of multi-omics data unlocks novel insights into cancer biology. *elife* 2023;**12**:12. https://doi.org/10.7554/eLife.87133.3.
32. Dosovitskiy A, Beyer L, Kolesnikov A. *et al.* An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv. 2020 Oct 22. Available from: https://arxiv.org/abs/2010.11929. https://doi.org/10.48550/arXiv.2010.11929.
33. ChaudhariSneha MV. RamanathRohan: An attentive survey of attention models. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2021;**12**:1–32. 10.1145/3465055.
34. Bao H, Dong L, Piao S. *et al.* BEiT: BERT Pre-Training of Image Transformers. In: International Conference on Learning Representations (ICLR); 2022 May 2–6; Vienna, Austria. Vienna: ICLR; 2022.
35. Xu W, Fu YL, Zhu D. ResNet and its application to medical image processing: Research progress and challenges. *Comput Methods Prog Biomed* 2023;**240**:107660. https://doi.org/10.1016/j.cmpb.2023.107660.
36. Gokhale M, Mohanty SK, Ojha A. GeneViT: Gene vision transformer with improved DeepInsight for cancer classification. *Comput Biol Med* 2023;**155**:106643. https://doi.org/10.1016/j.compbiomed.2023.106643.
37. George B, Seals S, Aban I. Survival analysis and regression models. *J Nucl Cardiol* 2014;**21**:686–94. https://doi.org/10.1007/s12350-014-9908-2.
38. Foersch S, Glasner C, Woerl AC. *et al.* Multistain deep learning for prediction of prognosis and therapy response in colorectal cancer. *Nat Med* 2023;**29**:430–9. https://doi.org/10.1038/s41591-022-02134-1.
39. Chen X, Chen DG, Zhao Z. *et al.* Artificial image objects for classification of breast cancer biomarkers with transcriptome sequencing data and convolutional neural network algorithms. *Breast Cancer Res* 2021;**23**:96. https://doi.org/10.1186/s13058-021-01474-z.
40. Feng Y, Zhang Y, Li H. *et al.* Enzalutamide inhibits PEX10 function and sensitizes prostate cancer cells to ROS activators. *Cell Death Dis* 2024;**15**:559. https://doi.org/10.1038/s41419-024-06937-7.
41. Zhang S, Fan R, Liu Y. *et al.* Applications of transformer-based language models in bioinformatics: a survey. *Bioinformatics Advances*. 2023;**3**:vbad001. https://doi.org/10.1093/bioadv/vbad001.
42. Choi SR, Lee M. Transformer architecture and attention mechanisms in genome data analysis: a comprehensive review. *Biology (Basel)* 2023;**12**:1033.
43. Wu C, Zhang Y, Ying Z. *et al.* A transformer-based genomic prediction method fused with knowledge-guided module. *Brief Bioinform* 2023;**25**. https://doi.org/10.1093/bib/bbad438.

Briefings in Bioinformatics, 2025, **26(3)**, bbaf275
https://doi.org/10.1093/bib/bbaf275
Case Study