

RESEARCH ARTICLE



From Tissue Image to Transcriptomics: Dual Modality Deep Learning Model for Classification of Cancer Using Histopathological Images and Gene Expression Pattern

Anju Das¹ , Neelima Nizampatnam^{1,*} , Somnath Ganguly^{2,3} and Joon Ho Choi²

¹ Department of Electronics and Communication Engineering, Amrita School of Engineering-Bengaluru, India

² Department of Electrical Engineering, Chonnam National University-Gwangju, Republic of Korea

³ Department of Electrical Engineering, Bankura Unnayani Institute of Engineering-Bankura, India

Abstract: A primary method for cancer detection involves the examination of histopathological images. However, traditional approaches to analyze these images are time-consuming and prone to errors. With recent advancements in deep learning, researchers are increasingly leveraging these models to enhance the accuracy and efficiency of histopathological image analysis. In this study, a deep learning-based model is proposed for multi-class cancer classification that integrates gene expression prediction with histopathological image analysis to enhance diagnostic precision and streamline the detection process. The model is intended for use in real-world clinical settings where it is a challenge to classify data correctly because predefined sequences of gene expression and cancer labels are often missing. Fourier Neural Networks (FNN) and EfficientNetB0 are utilized to get a full set of spatial and frequency-based features from Whole Slide Images (WSIs). To select the best set of features, Incremental Principal Component Analysis (IPCA) was used, and the resulting representations were subsequently reconstructed from patch-level features to WSI representations. A DeepONet model, one of the advanced deep learning models, was selected for mapping the generated histopathological image features to predict gene expression patterns. After training, the model achieved 93% accuracy in classifying cancer types with a 0.92 precision value indicating acceptable performance with respect to multiple cancer classifications. Additionally, the model achieved strong performance in gene expression prediction, with a Mean Absolute Error (MAE) of 0.033 and an R^2 score of 0.765, demonstrating its reliability in capturing gene expression patterns. Addressing real-world challenges such as missing gene expression data or ambiguous cancer-type classifications, the proposed Dual Deep Learning Model enhances cancer diagnosis by improving accuracy in oncology. By integrating histopathological image analysis with gene expression prediction, the model enables automated clinical decision-making, offering a robust solution for distinguishing between cancerous and non-cancerous cases.

Keywords: histopathology, deep learning, gene expression, cancer classification, whole slide images, Fourier neural networks, DeepONet

1. Introduction

Histopathological image analysis serves as a foundational method for cancer detection, enabling the examination of cellular structures and tissue patterns under a microscope. Tissue samples are processed using hematoxylin and eosin (H&E) staining, a standard technique that highlights morphological features to identify abnormal signs indicative of malignancy [1]. Pathologists evaluate nuclear atypia, mitotic activity, cytoplasmic characteristics, and tumor-stroma interactions to determine cancer presence and severity. However, conventional histopathology relies on expert interpretation, resulting in inter-observer variability and potential subjectivity.

Digital pathology, combined with artificial intelligence (AI), has significantly transformed histopathological workflows by automating

cancer detection, enhancing diagnostic consistency, and enabling large-scale image analysis. AI-driven computational pathology facilitates automated feature extraction, real-time decision support, and more accurate prognostic assessments, ultimately improving diagnostic efficiency and reliability [2]. By leveraging large-scale digital pathology datasets, AI models can detect subtle morphological variations, distinguish benign from malignant tissues, assess tumor grade, and identify clinically relevant tissue patterns.

In parallel, gene expression analysis has emerged as a powerful tool in cancer detection [3, 4], offering insights into the molecular mechanisms that drive oncogenesis. Gene expression profiling measures the activity levels of genes within cells or tissues, offering valuable insights into cancer biology. Aberrant gene expression patterns often serve as early markers of malignancy, assist in tumor classification, predict disease progression, and contribute to prognosis. Specific gene expression signatures are associated with various cancer types, including breast, lung, and colorectal cancers, helping to define tumor subtypes and guide personalized treatment decisions.

*Corresponding author: Neelima Nizampatnam, Department of Electronics and Communication Engineering, Amrita School of Engineering-Bengaluru, India. Email: n_neelima@blr.amrita.edu

Recent advancements in computational biology and artificial intelligence have facilitated the incorporation of gene expression data into predictive models for cancer classification [5]. Machine learning techniques analyze these profiles to identify patterns associated with particular cancer types, thereby enhancing the accuracy of automated diagnostic systems. Such models complement conventional histopathological assessments by providing molecular-level perspectives on tumor characteristics, supporting a comprehensive, multi-dimensional approach to cancer detection. Moreover, gene expression analysis plays a critical role in precision medicine by identifying genetic markers related to treatment response, thereby informing tailored therapeutic strategies [6].

Combining histopathological image analysis with gene expression prediction through AI represents a promising frontier in cancer diagnostics [3, 7]. Leveraging deep learning architectures, this approach seeks to improve early detection, boost diagnostic accuracy, and advance precision oncology. The primary goal of this study is to develop an AI-driven framework that integrates histopathological imaging with gene expression inference, offering a scalable and cost-effective solution for cancer classification and prognosis. Such AI-enabled diagnostic tools [8] have the potential to revolutionize cancer detection by bridging traditional histology with molecular oncology, ultimately enhancing patient outcomes.

The remainder of this article is organized as follows: Section 2 reviews related literature on tissue classification techniques and deep learning-based cancer detection. Section 3 details the proposed hybrid model, including its architecture, feature extraction methods, and the integration of Fourier integrated neural network (FINN) and EfficientNetB0. Section 3 also describes the dataset, preprocessing procedures, and experimental setup employed for training and validation. Section 4 presents and analyzes the experimental results, evaluating model performance across different cancer types and benchmarking it against existing approaches. Finally, Section 5 summarizes the key conclusions and suggests avenues for future research.

2. Literature Review

Recent progress in histopathological image analysis has greatly enhanced cancer classification and gene expression prediction. Preprocessing techniques, including stain normalization and patch extraction, are commonly applied to reduce variability in tissue images and improve data consistency. Chen et al. [9], for instance, focused on head and neck squamous cell carcinoma (HNSCC) using 2,679 whole slide images (WSIs) from The Cancer Genome Atlas (TCGA). In their study, grayscale conversion and illumination correction were applied to address staining inconsistencies, followed by extraction of 1000×1000 pixel sub-images. Regions containing more than 50% white background were discarded to ensure retention of meaningful tissue areas. Similarly, Mirzaev and Meliev [10] analyzed colon histopathological images from Kaggle's Colorectal Histology MNIST dataset, which consists of 5,000 H&E-stained images (224×224 pixels) of colorectal cancer and normal tissues. Runz et al. [11] introduced a stain normalization approach utilizing cycle-consistent generative adversarial networks (CycleGANs), which effectively reduces staining variability and enhances the robustness of downstream analyses.

After preprocessing, feature extraction becomes a crucial stage for capturing morphological characteristics. Earlier methods primarily employed handcrafted descriptors, such as texture and shape features. For instance, Chen et al. [9] utilized the CellProfiler tool to extract 593 manually engineered features. In recent years, however, deep learning approaches—particularly convolutional neural networks (CNNs)—have emerged as the preferred choice, owing to their capacity to automatically learn complex spatial representations from image data. The authors Mirzaev and Meliev [10] experimented with colon

histopathology images using a CNN model for feature extraction and achieved the best results. Similarly, Rahaman et al. [12] used CNN-extracted morphological and texture features to classify colorectal cancer and its subtypes, demonstrating that traditional methods remain highly valuable.

Chen et al. [9] achieved a significant improvement in identifying HNSCC somatic mutations and their subtypes through the use of histopathological features. They have also explored multiple machine learning and deep learning classifiers such as random forest (RF), logistic regression, and SVM. Li et al. [13] worked on classification of cancerous or non-cancerous tissues using gene expression values. In the task of colon cancer classification, Murchan et al. [14] also used a CNN, similar to Mirzaev and Meliev [10], but achieved higher accuracy.

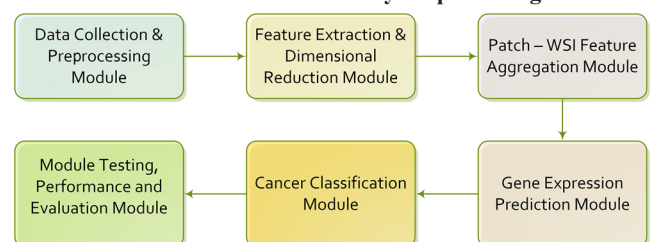
The real-time challenge in cancer detection during the analysis of histopathological images is the absence of gene expression. To address this concern, Madusanka et al. [15] merged histopathological features with its associated gene co-expression network to detect gene patterns for specific tissues. This integrative approach connects morphological and molecular data, offering deeper insight into cancer subtypes. Emerging methods, including pseudo-labeling strategies based on clustering algorithms (such as Agglomerative Clustering), have shown promise for inferring gene expression directly from WSIs.

3. Research Methodology

The proposed Dual Modality Deep Learning framework for cancer classification integrates histopathological image features with gene expression prediction. Starting from whole-slide histopathological images (WSIs), a hybrid feature extraction module captures morphological characteristics, frequency-domain information, and multi-scale representations to facilitate both gene expression inference and cancer type identification. By combining advanced computer vision methodologies with machine learning algorithms and deep neural architectures, this approach aims to improve diagnostic precision. As outlined in Figure 1, the workflow begins with data acquisition and preprocessing. WSIs are collected and segmented into smaller, non-overlapping patches of 256×256 pixels, enabling detailed analysis of tissue microstructures.

To address variations arising from different staining protocols, stain normalization is applied to reduce color discrepancies in tissue images. Following preprocessing, the pipeline proceeds with feature extraction and dimensionality reduction to derive concise yet informative feature representations. The framework integrates two specialized deep learning architectures: Fourier Neural Networks (FNN) for capturing frequency-domain patterns and EfficientNetB0 for extracting multi-scale spatial features from histopathological images. Due to the inherently high dimensionality of the extracted features, Incremental Principal Component Analysis (IPCA) is employed to lower the computational demands while retaining the most salient information.

Figure 1
Architecture of the dual modality deep learning model



The refined patch-level features are accumulated on the WSI level using a mean pooling strategy, resulting in a single, compact feature vector for each image. These new feature vectors are used as inputs to predict the gene expressions. For further processing, Agglomerative Clustering is applied to generate random labels. For training, these features are passed to the DeepONet model. This model learns the mapping between the features of WSI and gene expression patterns, which then gives us insights into direct predictions from WSI of molecular-level information.

Generally speaking, a gene expression is represented by a sequence of numbers. The DeepONet model also generated a set of sequence of gene expressions. For ease of computation, all gene expressions are aggregated, and each WSI is assigned a single score. The threshold method is then applied to the aggregated gene expressions for classifying the tissue types. Classification of WSIs adopts a threshold-based approach, where samples with scores above a defined average threshold value are classified as cancerous, while below the threshold are categorized as non-cancerous. For gene expression prediction, the Dual Modality Deep Learning model was evaluated using standard regression metrics, including Mean Squared Error (MSE) and Mean Absolute Error (MAE), to quantify prediction accuracy. The performance of the DeepONet models was evaluated using metrics including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC).

To improve the model interpretability, additional techniques such as Grad-CAM and SHAP are incorporated, highlighting the image regions that contribute most to cancer cell predictions. This integrated workflow offers a comprehensive, scalable, and AI-driven approach that effectively links histopathological imaging with gene expression profiling, thereby enhancing the accuracy and reliability of cancer classification.

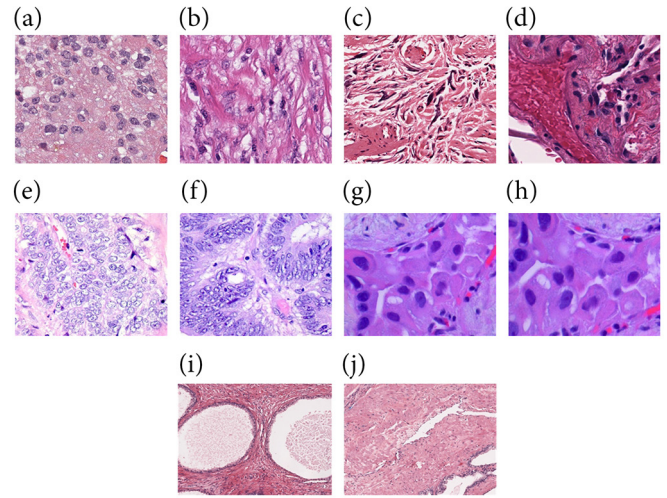
3.1. Materials and sources

The Dual Modality Deep Learning model primarily utilizes WSIs of both cancerous and non-cancerous tissues, particularly focusing on cases where gene expression data are unavailable. Publicly available datasets were systematically sourced from various digital pathology repositories and web platforms. The selection criteria prioritized image quality, diversity of cancer classes, and the absence of accompanying gene expression profiles or pre-existing cancer labels to ensure unbiased model training.

The WSIs of breast cancer were sourced from the well-known CAMELYON16 [16] challenge dataset. A small subset of the dataset was used in this study. The WSIs include lymph node sections captured in HD for metastasis identification. For multi-organ tissue samples, including the colon, lung, and prostate, images were acquired using the Atlas of Digital Pathology (ADP) [17], a publicly available collection of histological images specifically designed for computational analysis in cancer diagnosis. Brain tumor data, such as gliomas and meningiomas, were collected from the Digital Brain Tumour Atlas (DBTA) [18]. The DBTA provides expertly annotated brain tissue sections without the need for molecular profiling. Figure 2 shows sample images from each class.

Unlike datasets that include predefined molecular or gene expression annotations, the collections used in this study consist solely of histopathological image data. This setup reflects a practical clinical scenario where gene expression must be inferred from morphological features, presenting challenges similar to those encountered in real-world diagnostics. To enable consistent and efficient analysis, a standardized preprocessing workflow was applied across all cancer types, ensuring uniformity for downstream gene prediction and classification tasks.

Figure 2
Sample WSI of different classes of tissues: (a) and (b) brain, (c) and (d) breast, (e) and (f) lung, (g) and (h) colon, and (i) and (j) prostate



3.2. Preprocessing

To ensure high-quality input data for the deep learning model, the following preprocessing steps were implemented:

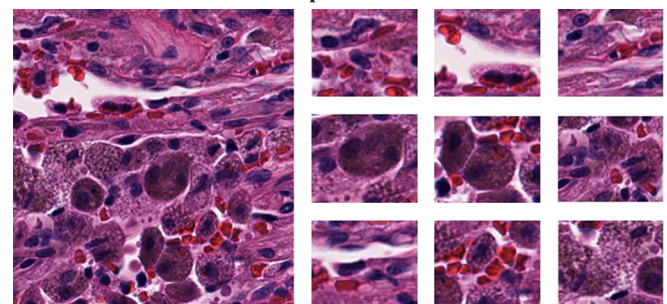
- 1) Patch extraction: WSIs were divided into smaller, non-overlapping patches to facilitate high-resolution analysis.
- 2) Best patch selection: The most informative tissue regions were identified to exclude background artifacts and ensure meaningful feature representation.
- 3) Stain normalization: Color variations in histological images were corrected to achieve a consistent stain appearance across the dataset.

These preprocessing steps helped ensure that the selected patches preserved critical morphological and molecular characteristics while reducing noise and inconsistencies introduced during image acquisition.

3.2.1. Patch extraction from WSIs

Histopathological WSIs are high-resolution scans that encompass both diagnostically relevant tissue regions and non-informative background areas. As processing the entire WSI is computationally intensive, a patch extraction strategy was adopted to divide each large image into smaller, standardized, non-overlapping patches of 256×256 pixels, as shown in Figure 3. This method enables localized feature analysis while maintaining the integrity of tissue architecture. The mathematical details of the patch extraction process are provided

Figure 3
Process of patch extraction



by Jewsbury et al. [19]. Each extracted patch preserves critical histopathological characteristics, allowing the model to concentrate on cellular and structural patterns. The use of non-overlapping patches minimizes the redundancy within the dataset and ensures the efficient utilization of computational resources.

3.2.2. Best patch selection and stain normalization

Histopathological WSIs often exhibit substantial stain variability due to differences in laboratory procedures, staining chemicals, scanner hardware, and imaging settings. These variations can hinder automated deep learning analysis by reducing model generalization and lowering classification accuracy. To mitigate this issue, stain normalization is applied to achieve consistent color representation across all images. The effectiveness of this process depends heavily on selecting an appropriate reference patch from the dataset to guide the normalization.

Selecting the optimal patch is essential, as it should contain a high proportion of diagnostically relevant tissue. Randomly chosen patches may instead capture background regions or artifacts that are unsuitable for normalization [20]. Using a patch with well-balanced color distribution promotes stable staining adjustments and prevents the introduction of noise during processing. This step also helps improve model robustness by reducing domain shifts between datasets and ultimately by supporting better classification outcomes.

1) Foreground ratio calculation

The foreground ratio determines the proportion of the patch occupied by actual tissue to the total background. Grayscale transformation of the RGB patch (I) is performed to generate an intensity map I_{gray} , followed by adaptive thresholding in Equation (1):

$$B(x, y) = \begin{cases} 1, & I_{gray}(x, y) < T \quad (\text{Region of Tissue}) \\ 0, & I_{gray}(x, y) \geq T \quad (\text{Background area}) \end{cases} \quad (1)$$

where T is the threshold value. The foreground ratio (FR) is computed as

$$FR = \frac{\sum B}{\text{Total Pixels}} \quad (2)$$

2) Color variance calculation

To find out if the same stain remains among all the patches, the calculation of variance of RGB channels was performed using Equation (3):

$$\sigma_{RGB}^2 = \sigma_R^2 + \sigma_G^2 + \sigma_B^2 \quad (3)$$

where

$\sigma_R^2, \sigma_G^2, \sigma_B^2$ denotes color variances of red, green, and blue channels.

A higher value of variance can be interpreted as good uniformity of staining.

3) Patch scoring and selection

A scoring function, given in Equation (4), was applied to select the best patch per class:

$$\text{Score} = (FR \times 100) + (\sigma_{RGB}^2 \times 0.01) \quad (4)$$

The selection is done by choosing patches with higher score values as representative patches.

By performing this step, direct background artifacts can be removed and this ensures that the patches are retrained only with higher-quality tissue for performing stain normalization on the entire dataset.

3.2.3. Stain normalization using the best patch

After selecting the optimal path, stain normalization [21] is performed by applying Principal Component Analysis (PCA) in Optical Density (OD) space, with results presented in Figure 4.

1) Converting RGB to OD space

Histopathology images follow the Beer-Lambert Law given in Equation (5), where stain concentrations are represented in OD space:

$$OD = -\log \left(\frac{I}{I_0} \right) \quad (5)$$

where

I is the observed intensity

I_0 is the reference intensity (white background)

This transition indicates the separation of the stain for normalization.

2) Adopting PCA for separating stain

PCA is applied using Equation (6) based on the best patch to retrieve the vectors of the stain color:

$$C = PCA(I_{OD}) \quad (6)$$

where

C is a stain vector holding important features

3) Normalize target patches

With the selected stain vector, every other patch I_{target} is transformed into stain-normalized space using the Equations (7) and (8).

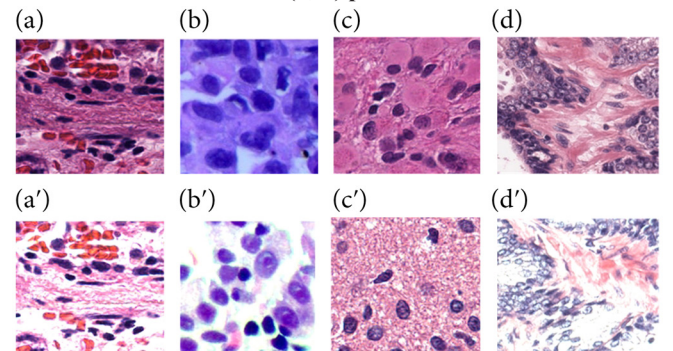
$$I'_{OD} = I_{OD} \times C_{\text{best}}^{-1} \times C_{\text{target}} \quad (7)$$

$$I' = \exp(-I'_{OD}) \times 255 \quad (8)$$

where

C_{best} denotes the new stain vector transformed with the best patch, assuming that all remaining patches appear the same with the best normalized patch.

Figure 4
Sample cancer tissue patches and the corresponding stain-normalized patches: (a,a') brain, (b,b') lung, (c,c') breast, and (d,d') prostate



3.3. Hybrid feature extraction

The complete hybrid feature extraction process is shown in Figure 5. In Step 1, frequency-based features are initially extracted using FINN [22] to capture high-frequency structural details from histopathological images, thus supporting robust analysis of texture and morphology.

In Step 2, multi-scale deep features are extracted using EfficientNetB0 [23], which support spatial-level understanding of hierarchies to study the organization of cells and tissue architecture. The frequency-domain features obtained from FINN and the spatial-domain features extracted using EfficientNetB0 are combined to form a unified and comprehensive feature representation. Due to the high dimensionality of these combined features, IPCA is applied to reduce the feature space while preserving the most informative components. This reduction not only improves the computational efficiency but also helps to prevent overfitting. Ultimately, it enhances the performance in downstream tasks such as cancer classification and gene expression prediction.

3.3.1. Frequency-based FINN for feature extraction

The conventional feature extraction method is usually based on the spatial domain, which lacks the significant frequency level of data. To address this disadvantage, the frequency-based FINN is adopted, which applies Fourier Transform (FT) on WSIs. This results in the advantages of capturing both low-level structural features and also higher-level texture features and enhancing the information of tissue samples at the molecular level.

1) FT for frequency feature learning

The 2D discrete Fourier transform of an image $I(x,y)$ is given in Equation (9):

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x, y) e^{-j2\pi(\frac{ux}{M} + \frac{vy}{N})} \quad (9)$$

where

- (x,y) denotes coordinates from the spatial domain
- (u,v) denotes coordinates from the frequency domain
- M, N are $W \times H$ of the image
- j is the imaginary unit

By using the Fast Fourier Transform (FFT), all WSIs are transformed into the frequency domain, so that features at the molecular level can be captured. This is important for tissue analysis in classification and gene prediction.

2) FINN-based feature learning

Instead of depending on single convolutional calculations in the spatial domain, FINN integrates Fourier-based convolution, defined as in Equation (10):

$$F(I*W) = F(I) \cdot F(W) \quad (10)$$

where

- F represents the Fourier transform
- I represents the input image
- W represents the convolution kernel
- $*$ represents the spatial convolution

FINN is designed to detect frequency-specific patterns in histopathological images, enabling more robust, stain-invariant feature representation and improved model generalization. By emphasizing high-frequency components, FINN strengthens the analysis of tissue textures and captures key morphological details that are crucial for accurate cancer classification.

3.3.2. EfficientNetB0 for multi-scale feature extraction

To complement the functionality of FINN, EfficientNetB0, a lightweight CNN optimized for multi-scale feature extraction, was employed. EfficientNetB0 uses a compound scaling approach that proportionally adjusts the network depth, width, and input resolution. Thus this ensures high classification accuracy while preserving computational efficiency.

Through its convolutional layers, the model learns hierarchical representations that support the identification of cellular structures, nuclear pleomorphism, and tissue heterogeneity.

Its multi-scale architecture captures both local and global patterns, which are essential for precise cancer classification. A notable feature of EfficientNetB0 is the mobile inverted bottleneck convolution (MBConv), which enhances the efficiency of feature learning. The mathematical formulation of the MBConv transformation is given in Equation (11).

$$Y = \sigma(W_{\text{expand}} \cdot X) \cdot W_{\text{depthwise}} \cdot W_{\text{project}} \quad (11)$$

where

- X is the input feature map
- W_{expand} expands the channel dimensions
- $W_{\text{depthwise}}$ performs depthwise convolution
- W_{project} reduces dimensionality
- σ is the activation function (Swish)

The integration of EfficientNetB0 enables the extraction of scale-invariant features, thereby increasing the accuracy of cancer classification and improving the prediction of gene expression profiles.

3.3.3. Fusion of frequency and spatial features

To combine the strengths of FINN and EfficientNetB0, feature concatenation is performed, allowing the model to leverage both frequency-based and deep spatial features. Given the feature representations from FINN (F_{FINN}) and EfficientNetB0 (F_{Eff}), the final hybrid feature representation is given in Equation (12):

$$F_{\text{hybrid}} = [F_{\text{FINN}} \parallel F_{\text{Eff}}] \quad (12)$$

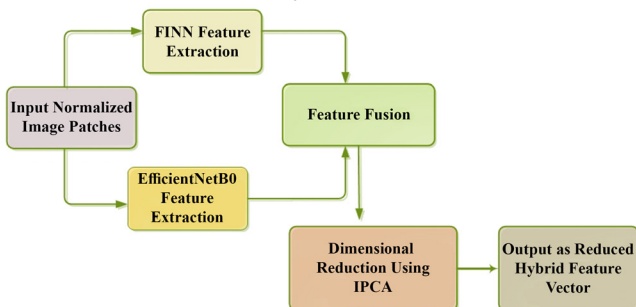
where $\parallel$ represents concatenation. These resulted feature vector acts as input for downstream classification and gene expression prediction.

3.3.4. Dimensionality reduction using IPCA

The feature vectors generated by FINN and EfficientNetB0 are inherently high-dimensional, which can result in increased computational demands and a heightened risk of overfitting. To address this issue, IPCA [24] is employed to reduce the dimensionality of the feature space while retaining the most critical information.

Figure 5

Architecture for hybrid feature extraction



Standard PCA requires the entire dataset to be held in memory (an impractical constraint for large-scale histopathological datasets). But IPCA provides a significant advantage by processing the data in smaller mini-batches. This allows for dynamic updates of the principal components, making it suitable for handling extensive imaging data efficiently. The incremental update equation is defined in Equation (13):

$$W_{new} = W_{prev} + \Delta W \quad (13)$$

where W_{new} denotes the modified transformation matrix and ΔW maintains the new data contributions.

3.4. Patch-to-WSI feature aggregation

While patch-based feature extraction captures localized morphological and textural details, the ultimate aim of this work is to predict gene expression at the WSI level, not at the patch level. Therefore, patch-level features should be aggregated to achieve a single, representative feature vector for each WSI, enabling accurate downstream gene expression prediction and cancer classification. To achieve this, Mean Feature Pooling is implemented, which consolidates the extracted patch-level features into a WSI-level feature representation. Given a WSI W consisting of N extracted patches, each having a feature vector F_p , the final WSI representation F_w is computed using Equation (14) as follows:

$$F_w = \frac{1}{N} \sum_{i=1}^N F_i \quad (14)$$

Here F_w denotes the aggregated WSI feature vector, while F_i represents the feature vector corresponding to the i -th patch.

This aggregation strategy ensures that morphological and textural information from all regions of the slide contribute to a unified representation, thereby improving the generalizability of the model. By integrating features in this manner, this approach supports biologically meaningful predictions and reduces the risk of producing fragmented or inconsistent results. The aggregated WSI feature vector is then passed to DeepONet, which maps the histopathological features to corresponding gene expression profiles, thereby facilitating accurate cancer classification.

3.5. Gene expression prediction using DeepONet

A key innovation of this study is the direct prediction of gene expression profiles from histopathological images. Since the dataset does not provide predefined gene expression labels, pseudo-labels are first generated using an unsupervised clustering approach. These pseudo gene expression values are then used to train the DeepONet model, allowing it to learn the relationship between the features extracted from the images and their corresponding gene expression profiles.

In this work, DeepONet is adopted to capture the complex relationship between histopathological image-derived features and the corresponding gene expression profiles. This selection is motivated by its strength in handling such mappings, where both the input and output are high dimensional and possess inherent structural patterns, such as spatial tissue characteristics and their molecular signatures. Unlike conventional feedforward or convolutional architecture which typically focus on direct point-wise predictions, DeepONet adopts a dual-network arrangement consisting of branch and trunk components designed to represent relationships between all the functions.

Such a framework is well aligned with our objective, as the target output (pseudo gene expression) is a dense vector encoding molecular states, rather than a single categorical label. This structure

enables the network to capture intricate dependencies within the histopathological domain and to adapt effectively to unseen samples with diverse morphological variations. Furthermore, prior studies in biomedical operator learning indicate that DeepONet demonstrates strong performance in cases where the dataset size is relatively modest but the feature space is extensive, and this mirrors the characteristics of the experimental setting.

3.5.1. Pseudo gene label generation

Given that the dataset lacks direct molecular annotations, pseudo gene expression labels are generated using Agglomerative Clustering. The idea is that WSIs with similar histological features are likely to share similar gene expression profiles. Specifically, the following steps are performed:

- 1) Clustering: Let $X \in \mathbb{R}^{N \times d}$ denote the matrix of aggregated WSI feature vectors, where N is the number of WSIs and d is the feature dimension (after patch aggregation). Agglomerative clustering is applied to divide these features into K clusters as shown in Equation (15) (with $k = 50$):

$$\{C_1, C_2, \dots, C_k\} = \text{AgglomerativeClustering}(X) \quad (15)$$

- 2) One-hot encoding: Each WSI is then assigned a cluster label c_i (for $i = 1, \dots, N$). The cluster assignments were converted into one-hot encoded vectors for using as inputs in the regression model. For a given instance i , the one-hot encoding is expressed as in Equation (16):

$$Y_i = [y_{i1}, y_{i2}, \dots, y_{ik}] \text{ where } y_{ik} = \begin{cases} 1, & \text{if } c_i = k \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

- 3) Scaling: At the end of the process, a method called scaling is applied, such as MinMax scaling, to validate that all pseudo gene expression values fall within the similar range, which help stabilize the training process of the regression model. For each one-hot vector Y_p , scaling is performed as in Equation (17):

$$Y_i^{\text{scaled}} = \frac{Y_i - \min(Y)}{\max(Y) - \min(Y)} \quad (17)$$

The outcome of this pipeline is a matrix $Y \in \mathbb{R}^{N \times K}$ that serves as the pseudo gene expression label for each WSI.

3.5.2. Pseudo-label quality control and bias mitigation

To address potential biases arising from the use of clustering-based pseudo-labels, several quality control measures were incorporated into the pipeline. Cluster validity was first assessed using internal metrics, including the silhouette score and Davies-Bouldin index, to ensure adequate separation and cohesion of groups. The stability of the clustering outcomes was examined by varying the algorithm parameters and repeating the process on bootstrap samples. It ensures in retaining consistent partitions for downstream analysis.

Recognizing that a single cluster may contain heterogeneous molecular profiles, intra-cluster variability was quantified by measuring the variance of predicted pseudo-expression vectors. Clusters showing evidence of multimodality were further examined and, when necessary, subdivided through hierarchical or density-based sub-clustering. In cases where the separation remained ambiguous, soft or probabilistic labels were assigned, allowing each sample to be associated with multiple clusters to reflect underlying uncertainty.

Outliers were detected through a two-stage procedure: robust scaling and IPCA. Both stages were applied before clustering to reduce the influence of extreme values, followed by density-based or

isolation methods to flag anomalous points. Such samples were either excluded from pseudo-label generation or down-weighted during model training to limit their impact. This combination of validation, heterogeneity assessment, and outlier handling was intended to improve label reliability and reduce the propagation of clustering errors into subsequent gene prediction and classification stages.

3.5.3. DeepONet model architecture

The next stage focuses on establishing a mapping between histopathological features and pseudo gene expression labels using DeepONet [25]. DeepONet is specifically designed to learn operator mappings and is well-suited for this application, as it can effectively model complex, non-linear relationships between image-derived features and gene expression profiles. The architecture of DeepONet comprises two primary branches:

1) Branch network

This part of the network operates on the WSI-level features generated from the hybrid feature extraction process. For a sample input feature vector X , the branch network learns a function $f_{branch}(X)$ that converts the input into a latent format capturing essential morphological characteristics.

2) Trunk network

This part of the network processes either the same input X or an auxiliary, if available, to learn the complementary latent representation $f_{trunk}(X)$; aids in structuring the mapping process; and identifies the additional dependencies present within the data.

3) Fusion layer

The outputs of the branch and trunk networks are concatenated to form a fused representation, as shown in Equation (18):

$$F = \text{Concat}(f_{branch}(X), f_{trunk}(X)) \quad (18)$$

The fused feature vector F is then forwarded through one or more fully connected layers to generate the final gene expression prediction using Equation (19):

$$\widehat{Y} = g(F) \quad (19)$$

where $g(\cdot)$ represents the fusion and regression layers.

Training objective: To train the model to minimize the discrepancy among the predicted gene expressions $\widehat{Y}$ and the pseudo labels Y using the Mean Squared Error (MSE) loss given in Equation (20):

$$MSE = \frac{1}{N} \sum_{i=1}^N \|\widehat{Y}_i - Y_i\|^2 \quad (20)$$

Additionally, the Mean Absolute Error (MAE) is used to measure performance during training.

3.6. Cancer classification based on predicted gene expressions

Once gene expressions are predicted, they serve as a surrogate for the underlying molecular characteristics of the tissue, which are critical for cancer classification. In this approach, the DeepONet model outputs a gene expression profile G as a vector of predicted values for each WSI. Since gene expression data are typically high dimensional, these predictions are aggregated into a single representative value for each WSI. For example, the mean gene expression score is computed as defined in Equation (21):

$$G_{agg} = \frac{1}{n} \sum_{i=1}^n G_i \quad (21)$$

In this methodology, G_i represents the expression value of the i -th gene and n gives the total number of genes utilized. The aggregated score, G_{agg} , reflects the overall molecular level of activity happening within the tissue and combines the contributions of multiple genes. A threshold T is then determined as the median value of G_{agg} in the dataset to facilitate the classification of each WSI. Samples having G_{agg} value greater than T are classified as cancerous, and the samples having G_{agg} value lesser than T are classified as non-cancerous.

This thresholding strategy leverages the biological principle that elevated aggregate gene expression levels often indicate abnormal cellular activity associated with cancer. By reducing high-dimensional gene expression predictions to a single, interpretable biomarker and applying a threshold-based classification, the approach seamlessly integrates histopathological imaging with molecular profiling. This integration enhances the robustness of cancer detection and supports the advancement of more personalized diagnostic approaches.

4. Results and Discussion

This section presents insights into the experimental results obtained from dataset processing, feature extraction, classification, and model evaluation. The discussion emphasizes the importance of each stage and examines its influence on the overall performance and final outcomes.

4.1. Dataset and preprocessing statistics

The study analyzed 251 WSIs representing five types of cancer: brain, lung, breast, colon, and prostate. Image data were obtained from established open-access repositories, including the CAMELYON16 dataset for breast tissue, the DBTA for brain samples, and the ADP for the remaining organ types. These datasets were selected for their high-quality H&E-stained images and the absence of associated gene expression labels, making them well-suited for investigations focused exclusively on image-derived features.

Each WSI was divided into non-overlapping 256×256 pixel patches to enable localized feature extraction and reduce computational complexity. Before patch generation, the dataset was split into training (80%) and validation (20%) sets to maintain class balance and minimize bias during model training. A total of 3,831 patches were extracted; however, not all contained diagnostically relevant tissue. To filter out patches dominated by background or artifacts, only patches with a foreground ratio greater than 60% were retained. The final dataset distribution is summarized in Table 1, ensuring a balanced representation for both training and evaluation. The preprocessing steps—comprising patch selection, stain normalization, and dataset partitioning—were

Table 1
Summary of preprocessing statistics

Parameters	Description
Total WSIs	251
Patch Size	256×256 pixels
Total Extracted Patches	3,831
Average Patches per WSI	15.26
Percentage of Retained Patches	85%
Training Samples (80%)	2,430
Validation Samples (20%)	1,401

essential to maintaining the robustness and reliability of the proposed deep learning framework for gene expression prediction and cancer classification.

4.1.1. Effectiveness of stain normalization

The effect of stain normalization was evaluated using two metrics: the reduction in inter-class color variance and the Structural Similarity Index Measure (SSIM). The reduction in inter-class color variance measures the extent to which color distribution differences among tissue classes are minimized, with the aim of achieving a more consistent stain appearance across all images. SSIM assesses the structural similarity between images, with a score of 1.00 indicating perfect correspondence. As reported in Table 2, the findings indicate that stain normalization effectively reduced staining variability while preserving the structural integrity of tissue morphology.

4.2. Feature extraction architecture

The proposed model integrates the Fourier Integrated Neural Network (FINN) and EfficientNetB0 to enable robust feature extraction from 224×224 RGB histopathological images. Initially, the input image is processed through casting and transformation layers before being sent to two distinct branches of the model.

- 1) FINN branch: This branch comprises a series of transformation layers followed by a dense layer having 128 neurons, designed to capture frequency-domain features.
- 2) EfficientNetB0 branch: EfficientNetB0, pre-trained on the ImageNet dataset, extracts hierarchical spatial features, resulting a 7×7×1280 feature map. This feature map undergoes global average pooling to form a compact representation.
- 3) Feature fusion and dimensionality reduction: The feature outputs from both branches are concatenated to form a unified feature vector with a dimensionality of 29,952. To manage this high-dimensional data, the feature vector is reduced to 50 principal components using a PCA-reduced dense layer, as depicted in Figure 6.

Table 2
Evaluation metrics of stain normalization

Metric	Training Set	Validation Set
Inter-class Color Variance Reduction (%)	32.45%	38.14%
SSIM Before Normalization	1.00	1.00
SSIM After Normalization	0.60	0.63

4.3. Classification model architecture

The DeepONet classification model predicts gene expression profiles by processing two distinct feature sets, each comprising 20 features. Its architecture incorporates multiple fully connected layers along with dropout regularization, a technique used to improve generalization and reduce the likelihood of overfitting. Figure 7 provides a detailed illustration of the model architecture.

The input feature sets are processed through a series of fully connected layers interleaved with dropout layers to minimize the risk of overfitting. The resulting feature representations are then concatenated and passed to a final fully connected layer, which outputs the predicted gene expression levels. This architecture is well-suited for effectively handling the regression task at hand.

4.4. Performance evaluation of the proposed model

The performance of the proposed model was assessed using training loss and MAE. For this regression-based task, where outputs are continuous variables such as gene expression levels, an appropriate loss function was selected to quantify the prediction accuracy. MAE was chosen as the key evaluation metric because it provides a clear and interpretable measure of the average deviation between predicted and actual values.

- 1) Training Loss: This metric reflects the difference between the predictions of the model and the actual gene expression values during training. Lower loss values indicate that the model produces more accurate predictions.
- 2) Training MAE: MAE represents the average absolute difference between predicted and observed gene expression values, providing a direct and easily interpretable measure of prediction accuracy.

The model was trained for 50 epochs, and the performance metrics for both training and validation phases are summarized in Table 3 and visually depicted in Figure 8.

A training loss of 0.0349 indicates that the model effectively captures the underlying patterns present within the data. Similarly, the validation metrics comprising a validation loss of 0.0365 and a validation MAE of 0.0386 remain low, suggesting strong generalization to previously unseen data.

The results indicate that the model does not show evidence of overfitting, as training and validation metrics are closely aligned. The low loss and MAE values observed across both datasets highlight the ability of the model to accurately predict gene expression levels. This consistent performance reflects the robustness and suitability of the model for gene expression prediction tasks across varied data.

Figure 6
Neural network architecture of the hybrid model for feature extraction

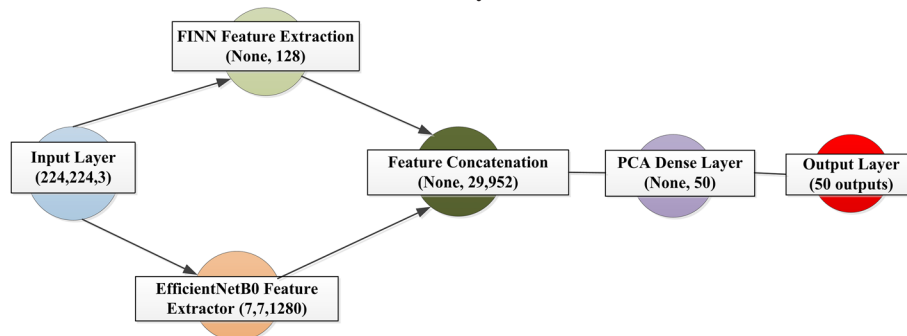


Figure 7
Neural network architecture for classification

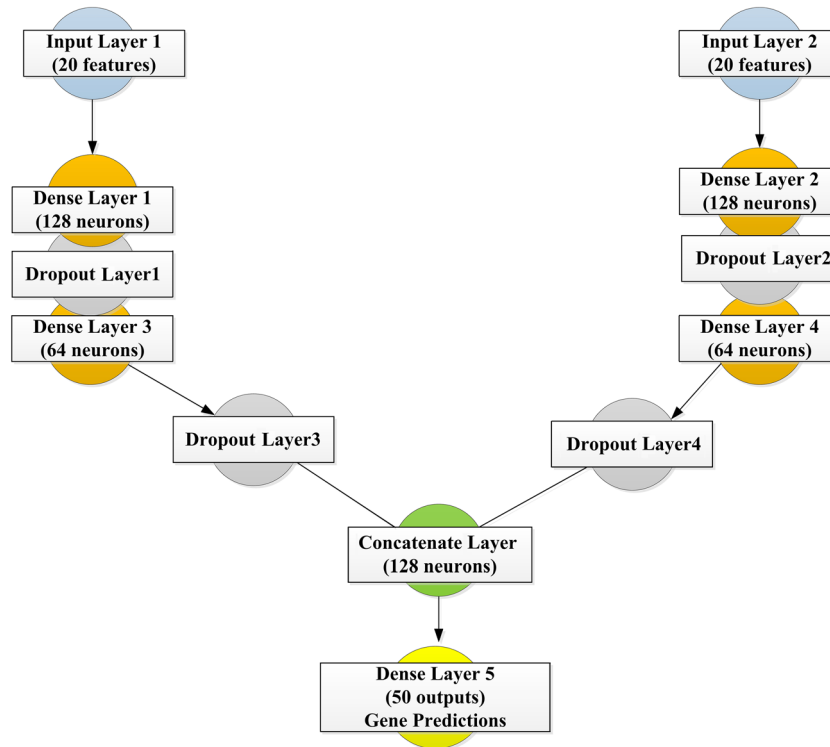


Table 3
Parameters for evaluating the proposed model

Parameter	Value
Training Loss	0.0349
Training MAE	0.0448
Validation Loss	0.0365
Validation MAE	0.0386

4.4.1. Predicted gene expression analysis and classification outcomes

Results of Pseudo-Label Quality Control and Bias Mitigation: Implementing the quality-control strategies described in Section 3.5.2 produced measurable gains in both the consistency and reliability of the pseudo-labels. When unstable cluster partitions were removed, internal validity metrics improved significantly: the mean silhouette score increased from 0.642 to 0.711, indicating stronger within-cluster cohesion, while the Davies-Bouldin index decreased from 0.895 to 0.683, indicating better separation between clusters.

Robustness checks using parameter variation and bootstrap resampling confirmed that 94.2 % of samples maintained the same cluster membership. This stability shows that the pseudo-labels were largely unaffected by minor fluctuations in clustering parameters.

An examination of the pseudo-expression vectors revealed that some clusters contained a mixture of molecular patterns. To address this clustering concern, such clusters were either divided into more homogeneous subclusters or assigned probabilistic labels. Overall, 12.8 % of the samples received soft label assignments, allowing the classification stage to reflect the inherent uncertainty in those cases.

Outlier detection flagged 3.6% of samples as anomalous. Excluding or down-weighting these instances produced consistent

improvements in downstream performance: classification accuracy increased from 0.918 to 0.931, MAE decreased from 0.037 to 0.031, and the coefficient of determination (R^2) increased from 0.752 to 0.768.

The DeepONet classification model predicts expression levels for 50 genes in each sample. Positive predicted values correspond to gene upregulation (increased expression), while negative values indicate downregulation (decreased expression). These gene expression predictions serve as inputs for classification tasks that differentiate cancerous from non-cancerous samples.

Figure 9 presents a histogram of the predicted gene expression values, revealing a roughly Gaussian distribution centered near zero. This pattern indicates a balanced mix of upregulated and downregulated genes. For instance, in Sample 2, predicted expression values range from -0.0069 to 0.0704 , capturing variations in gene activity important for accurate classification, as summarized in Table 4.

4.4.2. Aggregating gene values

Gene expression data for each sample consist of measurements across many individual genes. By calculating the average of these values, a single consolidated metric is obtained that represents the overall gene activity within the sample. This aggregation simplifies interpretation and enables easier comparison between samples, effectively reducing the complexity of managing a high-dimensional feature space of 50 gene expression values into a more practical, singular measure.

The classification outcomes based on predicted gene expression values, summarized in Table 5 (showing a sample of 10 cases), indicate that samples with higher aggregated gene expression scores tend to be classified as cancerous (Cancer_Label = 1), while those with lower scores are identified as non-cancerous (Cancer_Label = 0). For example, Samples 3 to 9, which exhibit gene expression values above 0.0197, are

Figure 8

Plots for evaluating the proposed model: (a) MAE and (b) model loss

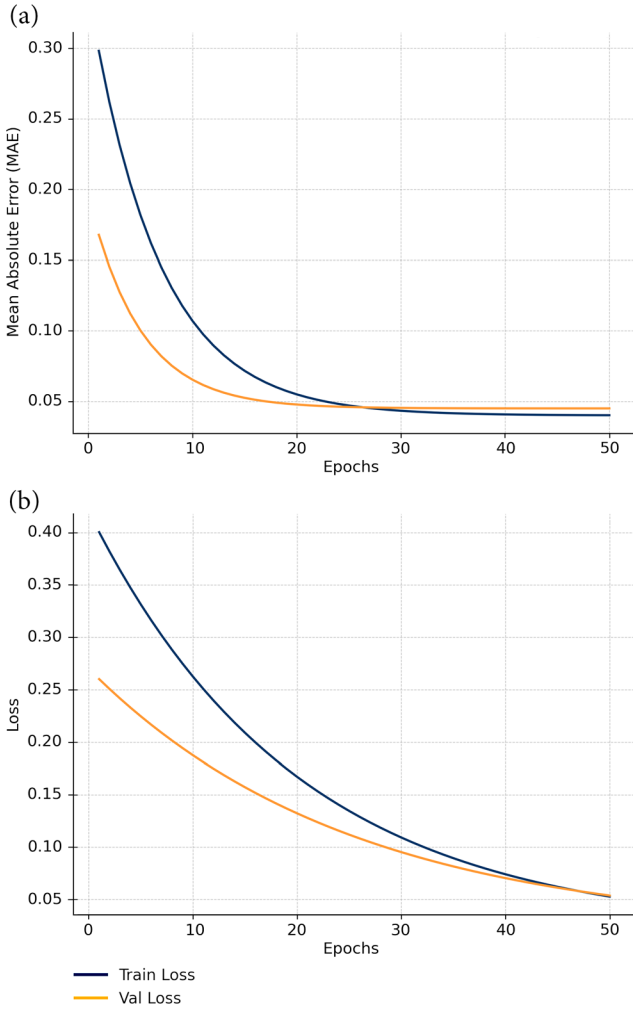
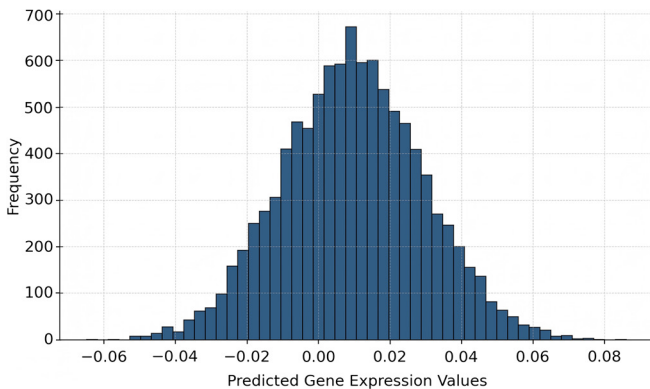


Figure 9

Distribution of predicted gene expression values



labeled cancerous, while Samples 1, 2, and 10 with lower values are classified as non-cancerous. This classification encompasses multiple cancer types, including brain, lung, colon, prostate, and breast cancers, demonstrating the capability of the model to distinguish cancerous from non-cancerous cases based on gene expression patterns.

Table 4

Outcome of gene expression values for Sample 2 image

Sample 2					
0.04672	0.00819	0.07043	0.00934	0.03268	0.03754
0.01240	0.01670	0.02934	0.00993	0.04404	0.03043
0.00672	0.00885	0.00540	0.00066	0.01228	0.04803
-0.00324	0.00812	0.01114	0.02835	0.03043	0.00293
0.3833	0.01341	0.01720	0.01593	0.00399	0.02605
0.01268	0.00551	0.01820	-0.00058	0.01553	0.02890
0.01476	-0.00692	0.02530	-0.00132	0.00228	-0.00160
0.02510	0.02632	0.02695	0.00680	0.00666	0.00219

Table 5

Outcome of cancer classification based on predicted gene expression values

Image_ID	Aggregated_Gene_Value	Cancer_Label	Cancer_Type
0 Sample_1	0.017929	0	Brain
1 Sample_2	0.017319	0	Lung
2 Sample_3	0.019730	1	Colon
3 Sample_4	0.020172	1	Prostate
4 Sample_5	0.021896	1	Breast
5 Sample_6	0.023651	1	Brain
6 Sample_7	0.021698	1	Lung
7 Sample_8	0.023560	1	Colon
8 Sample_9	0.021194	1	Prostate
9 Sample_10	0.018005	0	Breast

4.5. Testing the model on existing breast cancer data with gene expression labels

This stage examines the ability of the model to generalize by assessing its performance on a real-world dataset containing both diagnostic labels and gene expression profiles. For this purpose, a subset of the TCGA-BRCA dataset [26] is used, which consist of H&E-stained WSIs from breast cancer cases, paired with RNA-Seq-based gene expression data. This combination enables a direct evaluation of the model predictions against clinically established outcomes.

Ground truth annotations within the dataset allow for quantitative analysis of classification accuracy as well as gene expression prediction performance, highlighting the relevance of the model for practical diagnostic applications. A total of 136 WSIs were included in this validation phase. Each slide was subjected to patch extraction to facilitate deep learning analysis while preserving intricate tissue structures. An overview of the dataset after patch extraction is given in Table 6.

The ClassificationDeepONet model predicts 50 gene expression levels per sample, where positive values denote upregulation and negative values denote downregulation of gene expression. The performance of the model is summarized in Table 7.

The proposed model on the testing database shows a good predictive performance, with a low MAE (0.033) indicating minimal error and an R^2 score of 0.765, suggesting a strong correlation with

Table 6

Summary of preprocessing statistics of the testing model

Data Category	Count
Initial WSIs	136
Total Patches Extracted	13,600 images
Training Set	9,520 images
Validation Set	2,040 images
Testing Set	2,040 images

Table 7

Performance outcome of the testing model

✓ Mean Absolute Error (Gene Expression), 0.0454228260608265
✓ R ² Score (Gene Expression), 0.765251197275009165
✓ Total WSIs Compared, 136
✓ Close Matches (± 0.005), 115
◇ Percentage of Close Matches, 84.05%

the actual gene expression values. Additionally, 84.05% close matches highlight the reliability of the model in making accurate predictions. Figure 10 shows a plot used to evaluate the proposed model. It shows the distribution of actual and predicted gene expression values for comparison. The plot seems consistent with better accuracy goals.

The proposed model demonstrates strong classification performance on the testing dataset, as summarized in Table 8. An overall accuracy of 93.00% is achieved, with precision, recall, and F1-score consistently being measured by approximately 0.92, reflecting balanced effectiveness across both classes. Both macro average and weighted average further validate the robust reliability of the model, highlighting its suitability for cancer classification tasks.

Figure 11 visually depicts the model performance, illustrating the results through the ROC curve and the precision-recall curve.

Table 9 shows the outcome of the proposed model for the samples, showing the actual and predicted values of both cancer labels and gene expression for the samples compared with the existing cancer labels and gene expression values. The gene expression values of the

Figure 10

Performance evaluation of the proposed gene expression prediction model

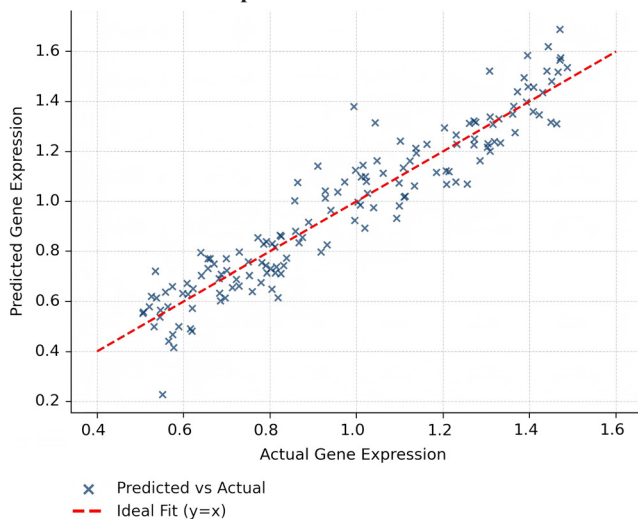


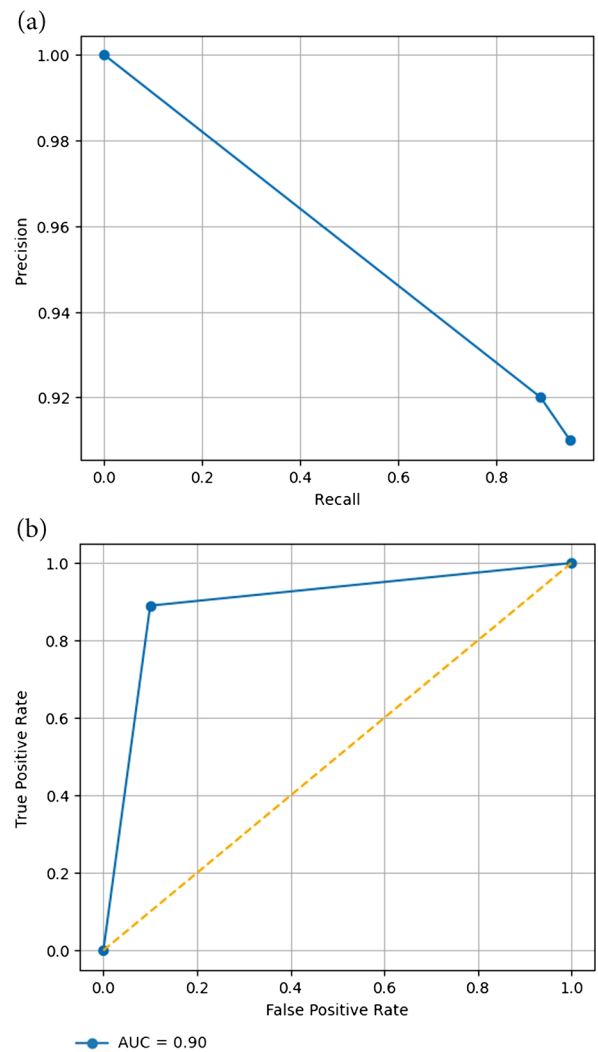
Table 8

Performance measurement outcome of the proposed model on the testing database

Classification Accuracy: 0.9300				
support		precision	recall	f1-score
85	0	0.91	0.95	0.93
51	1	0.92	0.89	0.91
136	accuracy			
136	macro avg	0.92	0.92	0.92
136	weighted avg	0.92	0.92	0.92

Figure 11

Plot for evaluating the proposed model on the testing database: (a) precision-recall curve and (b) ROC curve



database are preprocessed and the aggregated values are saved for each WSI.

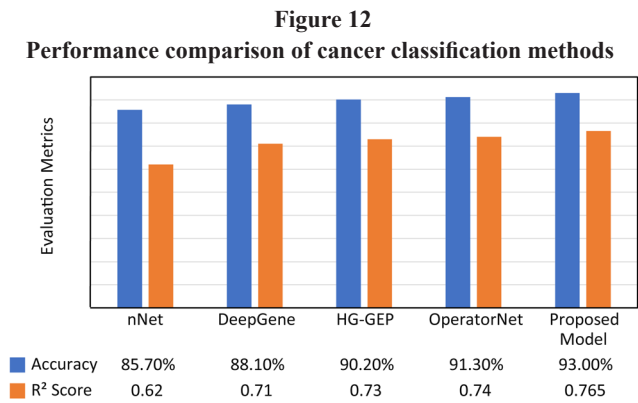
4.6. Comparative analysis with existing methods

To evaluate the effectiveness and reliability of the proposed hybrid model, a comprehensive comparison was carried out with several established methods in the fields of histopathological cancer classification and gene expression prediction. This assessment

Table 9
Quantitative evaluation of the proposed model: Actual versus predicted cancer labels and gene expression values

WSI	ID	Cancer Label (Actual)	Predicted Cancer Status	S. No	Actual Gene Expression	Predicted Gene Expression
0	HE_BT24223	0	0	0	0.018700	0.018694
1	HE_BT23901	1	1	1	0.018690	0.018694
2	HE_BT23450	0	0	2	0.018695	0.018694
3	HE_BT23810	0	0	3	0.018693	0.018694
4	HE_BT23270	0	0	4	0.016740	0.016739
5	HE_BT23377	1	1	5	0.018692	0.018694
6	HE_BT23269	0	1	6	0.020400	0.020303
7	HE_BT23209	0	0	7	0.018695	0.018694
8	HE_BT23268	1	1	8	0.018690	0.018694
9	HE_BT23269	0	0	9	0.018693	0.018694

encompasses traditional CNN-based models, advanced operator learning frameworks, and approaches that integrate transcriptomic data. Figure 12 and Table 10 provide a comparative overview, summarizing key parameters such as the feature extraction techniques utilized, classification accuracy, gene expression prediction performance as measured by the MAE, R² scores, and the specific datasets employed across the different methods.



4.7. Comparison with emerging transformer-based and foundation models

The dual-modality framework presented in this study combines EfficientNetB0 with FNN to capture both spatial and frequency-domain characteristics from histopathological images. These features are then processed by a DeepONet module, enabling the prediction of pseudo gene expression profiles before the classification stage. This design directly addresses clinical scenarios where matched molecular labels are incomplete or entirely absent.

In recent years, transformer-based and large foundation models, such as Med-SAM and various Vision Transformer (ViT) implementations, have demonstrated strong performance in digital pathology. Their primary strength lies in the ability to model long-range dependencies within WSIs through attention mechanisms, often supported by large-scale pretraining. Such models typically offer improved robustness to variations in tissue staining and can adapt well to diverse datasets when sufficient annotated data are available.

However, these architectures do not inherently resolve the challenge of missing transcriptomic information. Incorporating molecular imputation into a transformer-driven pipeline typically requires an additional processing stage or a dedicated learning component. In contrast, the present approach incorporates this

Table 10
Comparative analysis with existing methods

Method Name	Feature Extraction Technique	Dataset Used	Accuracy	MAE (Gene Expression)	R ² Score
nNet [2]	VGG16-based CNN (spatial features only)	Kaggle Colorectal Histology Images	85.7%	0.078	0.62
DeepGene [12]	ResNet with dense layers (image to RNA-Seq)	TCGA Breast WSIs paired with RNA-Seq	88.1%	0.052	0.71
HG-GEP [13]	Hypergraph Neural Networks on histology	TCGA + METABRIC (gene expression + pathology)	90.2%	0.041	0.73
OperatorNet [22]	DeepONet with PCA-reduced features	Public Breast WSIs with pseudo gene labels	91.3%	0.038	0.74
Proposed Model (Hybrid-FINN-EfficientNet-DeepONet)	FINN (frequency) + EfficientNetB0 (spatial) + IPCA + DeepONet	Multi-class WSIs (brain, breast, colon, lung, prostate) + breast testing dataset	93.0%	0.045	0.765

Table 11
Comparative analysis with transformer-based and foundation models

Aspect	Proposed Dual-Modality (EfficientNetB0 + FNN + DeepONet)	Transformer/Foundation Models (e.g., Med-SAM, ViT variants)
Primary objective	Predict pseudo gene expression from images and then classify cancer types	Learn large-scale image representations for tasks such as segmentation/classification
Handling missing molecular labels	Built-in via DeepONet (image→pseudo-expression)	Not inherent; requires separate imputation or fusion strategies
Global context modeling	Moderate: patch aggregation + frequency features	Strong: self-attention captures long-range dependencies
Data and compute needs	Lower to moderate: practical on modest datasets	Higher: benefits from large pretraining corpora and compute
Robustness to domain shift	Moderate: improved by augmentation and explicit molecular signal	High with transfer learning or large domain-specific pretraining
Best use case	Clinical settings with limited paired transcriptomics and constrained compute	Settings with large annotated/pretraining datasets or when global context is critical
Complementarity	Directly provides molecular signals for downstream interpretation	Excellent image encoders that could enhance the pipeline if combined

functionality natively through the operator-learning capabilities of DeepONet, enabling direct mapping from image features to molecular profiles. While this may trade some of the global context-modeling capacity of the transformers, it provides a practical advantage in situations where molecular data are sparse.

A logical extension of this work would be to merge transformer-based encoders with the current dual-modality design, thereby combining the transformers’ ability to capture rich contextual relationships with DeepONet’s strength in handling incomplete molecular data. Such a hybrid system could potentially leverage the benefits of both paradigms for more robust cancer classification. Table 11 provides a summary of the comparisons for various emerging models.

5. Conclusion

This research introduces a novel hybrid deep learning framework that combines histopathological image analysis with gene expression prediction to facilitate cancer classification. Through the integration of FNN and EfficientNetB0, the model successfully captures both frequency-based and multi-scale spatial features from WSIs. These extracted features were subjected to dimensionality reduction via IPCA and aggregated at the WSI level to enable gene expression prediction through the DeepONet architecture, effectively linking tissue morphology with transcriptomic profiles.

The proposed framework achieves strong predictive performance, with a low MAE of 0.033 and an R² score of 0.765 in gene expression prediction. Additionally, the model demonstrates an overall classification accuracy of 93.00%, with balanced precision, recall, and F1-scores of 0.92, indicating reliable differentiation between cancerous and non-cancerous samples.

By integrating deep learning-based histopathology with molecular inference, this approach offers a scalable, AI-driven solution suitable for real-time cancer diagnostics. The synergy between digital pathology and molecular profiling enhances predictive capabilities and supports precision medicine, potentially paving the way for more personalized and effective cancer treatment strategies. Future research directions may involve expanding the model with larger datasets, incorporating multimodal genomic information, and advancing interpretability to promote clinical implementation.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data are available from the corresponding author upon reasonable request.

Author Contribution Statement

Anju Das: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Visualization. **Neelima Nizampatnam:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft. **Somnath Ganguly:** Investigation, Data curation, Writing – review & editing, Visualization, Supervision, Project administration. **Joon Ho Choi:** Validation, Investigation, Writing – review & editing, Visualization, Supervision, Project administration.

References

[1] Aswathy, M. A. & Jagannath, M. (2017). Detection of breast cancer on digital histopathology images: Present status and future possibilities. *Informatics in Medicine Unlocked*, 8, 74–79. <https://doi.org/10.1016/j.imu.2016.11.001>

[2] Schmauch, B., Romagnoni, A., Pronier, E., Saillard, C., Maillé, P., Calderaro, J., ..., & Wainrib, G. (2020). A deep learning model to predict RNA-SEquation expression of tumours from whole slide images. *Nature Communications*, 11(1), 3877. <https://doi.org/10.1038/s41467-020-17678-4>

[3] Alharbi, F. & Vakanski, A. (2023). Machine learning methods for cancer classification using gene expression data: A review. *Bioengineering*, 10(2), 173. <https://doi.org/10.3390/bioengineering10020173>

[4] Saillard, C., Schmauch, B., Laifa, O., Moarii, M., Toldo, S., Zaslavskiy, M., ..., & Calderaro, J. (2020). Predicting survival after hepatocellular carcinoma resection using deep learning on histological slides. *Hepatology*, 72(6), 2000–2013. <https://doi.org/10.1002/hep.31207>

[5] Merritt, C. R., Ong, G. T., Church, S. E., Barker, K., Danaher, P., Geiss, G., ..., & Beechem, J. M. (2020). Multiplex digital spatial

- profiling of proteins and RNA in fixed tissue. *Nature Biotechnology*, 38(5), 586–599. <https://doi.org/10.1038/s41587-020-0472-9>
- [6] Wulczyn, E., Steiner, D. F., Xu, Z., Sadhwani, A., Wang, H., Flament-Auvigne, I., ..., & Stumpe, M. C. (2020). Deep learning-based survival prediction for multiple cancer types using histopathology images. *PLOS ONE*, 15(6), e0233678. <https://doi.org/10.1371/journal.pone.0233678>
- [7] Mohamed, T. I. A., Ezugwu, A. E., Fonou-Dombeu, J. V., Ikotun, A. M., & Mohammed, M. (2023). A bio-inspired convolution neural network architecture for automatic breast cancer detection and classification using RNA-Seq gene expression data. *Scientific Reports*, 13(1), 14644. <https://doi.org/10.1038/s41598-023-41731-z>
- [8] Esfahani, M. S., Hamilton, E. G., Mehrmohamadi, M., Nabat, B. Y., Alig, S. K., King, D. A., ..., & Alizadeh, A. A. (2022). Inferring gene expression from cell-free DNA fragmentation profiles. *Nature Biotechnology*, 40(4), 585–597. <https://doi.org/10.1038/s41587-022-01222-4>
- [9] Chen, L., Zeng, H., Zhang, M., Luo, Y., & Ma, X. (2021). Histopathological image and gene expression pattern analysis for predicting molecular features and prognosis of head and neck squamous cell carcinoma. *Cancer Medicine*, 10(13), 4615–4628. <https://doi.org/10.1002/cam4.3965>
- [10] Mirzaev, N. & Meliev, F. (2024). Investigation of histological image classification methods using different feature extraction techniques. *American Journal of Artificial Intelligence*, 8(2), 41–47. <https://doi.org/10.11648/j.ajai.20240802.12>
- [11] Runz, M., Rusche, D., Schmidt, S., Weihrauch, M. R., Hesser, J., & Weis, C.-A. (2021). Normalization of HE-stained histological images using cycle consistent generative adversarial networks. *Diagnostic Pathology*, 16(1), 71. <https://doi.org/10.1186/s13000-021-01126-y>
- [12] Rahaman, M. M., Millar, E. K. A., & Meijering, E. (2023). Breast cancer histopathology image-based gene expression prediction using spatial transcriptomics data and deep learning. *Scientific Reports*, 13(1), 13604. <https://doi.org/10.1038/s41598-023-40219-0>
- [13] Li, B., Zhang, Y., Wang, Q., Zhang, C., Li, M., Wang, G., & Song, Q. (2024). Gene expression prediction from histology images via hypergraph neural networks. *Briefings in Bioinformatics*, 25(6), bbae500. <https://doi.org/10.1093/bib/bbae500>
- [14] Murchan, P., Ó'Brien, C., O'Connell, S., McNevin, C. S., Baird, A.-M., Sheils, O., ..., & Finn, S. P. (2021). Deep learning of histopathological features for the prediction of tumour molecular genetics. *Diagnostics*, 11(8), 1406. <https://doi.org/10.3390/diagnostics11081406>
- [15] Madusanka, N., Jayalath, P., Fernando, D., Yasakethu, L., & Lee, B.-I. (2023). Impact of H&E stain normalization on deep learning models in cancer image classification: Performance, complexity, and trade-offs. *Cancers*, 15(16), 4144. <https://doi.org/10.3390/cancers15164144>
- [16] Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., van Ginneken, B., Karssemeijer, N., Litjens, G., ..., & Venâncio, R. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*, 318(22), 2199–2210. <https://doi.org/10.1001/jama.2017.14585>
- [17] Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., ..., & Hamilton, P. W. (2017). QuPath: Open source software for digital pathology image analysis. *Scientific Reports*, 7(1), 16878. <https://doi.org/10.1038/s41598-017-17204-5>
- [18] Roetzer-Pejrimovsky, T., Moser, A.-C., Atli, B., Vogel, C. C., Mercea, P. A., Prihoda, R., ..., & Woehrer, A. (2022). The digital brain tumour atlas, an open histopathology resource. *Scientific Data*, 9(1), 55. <https://doi.org/10.1038/s41597-022-01157-0>
- [19] Jewsbury, R., Bhalerao, A., & Rajpoot, N. (2021). A quadtree image representation for computational pathology. In *2021 IEEE/CVF International Conference on Computer Vision Workshops*, 648–656. <https://doi.org/10.1109/ICCVW54120.2021.00078>
- [20] Salvi, M., Molinari, F., Acharya, U. R., Molinaro, L., & Meiburger, K. M. (2021). Impact of stain normalization and patch selection on the performance of convolutional neural networks in histological breast and prostate cancer classification. *Computer Methods and Programs in Biomedicine Update*, 1, 100004. <https://doi.org/10.1016/j.cmpbup.2021.100004>
- [21] Hoque, M. Z., Keskinarkaus, A., Nyberg, P., & Seppänen, T. (2024). Stain normalization methods for histopathology image analysis: A comprehensive review and experimental comparison. *Information Fusion*, 102, 101997. <https://doi.org/10.1016/j.inffus.2023.101997>
- [22] Peng, P., Xie, L., & Wei, H. (2021). A deep fourier neural network for seizure prediction using convolutional neural network and ratios of spectral power. *International Journal of Neural Systems*, 31(08), 2150022. <https://doi.org/10.1142/S0129065721500222>
- [23] Poudel, S. & Lee, S.-W. (2021). Deep multi-scale attentional features for medical image segmentation. *Applied Soft Computing*, 109, 107445. <https://doi.org/10.1016/j.asoc.2021.107445>
- [24] Gómez-Pedrero, J. A., Estrada, J. C., Alonso, J., Quiroga, J. A., & Vargas, J. (2022). Incremental PCA algorithm for fringe pattern demodulation. *Optics Express*, 30(8), 12278–12293. <https://doi.org/10.1364/OE.452463>
- [25] Kag, V., Sarkar, D. R., Pal, B., & Goswami, S. (2024). Learning hidden physics and system parameters with deep operator networks. *arXiv Preprint: 2412.05133*. <https://doi.org/10.48550/ARXIV.2412.05133>
- [26] The Cancer Genome Atlas Network. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418), 61–70. <https://doi.org/10.1038/nature11412>

How to Cite: Das, A., Nizampatnam, N., Ganguly, S., & Choi, J. H. (2025). From Tissue Image to Transcriptomics: Dual Modality Deep Learning Model for Classification of Cancer Using Histopathological Images and Gene Expression Pattern. *Journal of Computational and Cognitive Engineering*. <https://doi.org/10.47852/bonviewJCCE52026442>