



Article

Image-Based Deep Learning for Brain Tumour Transcriptomics: A Benchmark of DeepInsight, Fotomics, and Saliency-Guided CNNs

Ali Alyatimi ^{1,2,*} , Vera Chung ¹ , Muhammad Atif Iqbal ¹ and Ali Anaissi ¹

¹ Faculty of Engineering, School of Computer Science, The University of Sydney, Sydney, NSW 2008, Australia

² Department of Computer and Information Technology, Jazan College of Technology, Technical and Vocational Training Corporation (TVTC), Riyadh 12613, Saudi Arabia

* Correspondence: ali.alyatimi@sydney.edu.au

Abstract

Classifying brain tumour transcriptomic data is crucial for precision medicine but remains challenging due to high dimensionality and limited interpretability of conventional models. This study benchmarks three image-based deep learning approaches, DeepInsight, Fotomics, and a novel saliency-guided convolutional neural network (CNN), for transcriptomic classification. DeepInsight utilises dimensionality reduction to spatially arrange gene features, while Fotomics applies Fourier transforms to encode expression patterns into structured images. The proposed method transforms each single-cell gene expression profile into an RGB image using PCA, UMAP, or t-SNE, enabling CNNs such as ResNet to learn spatially organised molecular features. Gradient-based saliency maps are employed to highlight gene regions most influential in model predictions. Evaluation is conducted on two biologically and technologically different datasets: single-cell RNA-seq from glioblastoma GSM3828672 and bulk microarray data from medulloblastoma GSE85217. Outcomes demonstrate that image-based deep learning methods, particularly those incorporating saliency guidance, provide a robust and interpretable framework for uncovering biologically meaningful patterns in complex high-dimensional omics data. For instance, ResNet-18 achieved the highest accuracy of 97.25% on the GSE85217 dataset and 91.02% on GSM3828672, respectively, outperforming other baseline models across multiple metrics.

Keywords: deep learning; single-cell RNA-seq; brain tumour classification; medulloblastoma; glioblastoma; image-based omics; saliency maps; CNN; transcriptomics



Academic Editors: Jaroslaw Krzywanski, Marcin Sosnowski, Karolina Grabowska, Dorian Skrobek, Anna Zylka, Agnieszka Kijoleczkowska, Bashar Shboul and Tomasz Czakiert

Received: 6 September 2025

Revised: 9 October 2025

Accepted: 9 October 2025

Published: 15 October 2025

Citation: Alyatimi, A.; Chung, V.; Iqbal, M.A.; Anaissi, A. Image-Based Deep Learning for Brain Tumour Transcriptomics: A Benchmark of DeepInsight, Fotomics, and Saliency-Guided CNNs. *Mach. Learn. Knowl. Extr.* **2025**, *7*, 119. <https://doi.org/10.3390/make7040119>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Single-cell RNA sequencing (scRNA-seq) is a technique that enables transcriptomic profiling at the level of individual cells, providing a high-resolution view of cellular heterogeneity in complex tissues. It captures the unique expression patterns of thousands to millions of cells, making it possible to identify rare cell types and novel cellular states that may be obscured in bulk data [1]. The classification and understanding of diseases such as cancer, neurodegeneration, and immune disorders involve cell-type composition and state dynamics, which play key roles in pathogenesis.

Brain tumours, such as glioblastoma (GBM) and medulloblastoma, are highly aggressive and heterogeneous cancers. GBM is the most common malignant primary brain tumour in adults, marked by significant molecular and cellular diversity that complicates

treatment and heightens prediction [2,3]. Contrarily, medulloblastoma is the most prevalent malignant pediatric brain tumour, classified into four distinct molecular subgroups, each with differing prognostic and therapeutic implications [4]. Transcriptomic profiling using bulk RNA sequencing and single-cell RNA sequencing (scRNA-seq) is crucial for revealing gene expression signatures associated with tumour development, progression, and treatment response [1]. Transforming omics data into image formats allows for the application of convolutional neural networks (CNNs) in effective classification [5,6]. Techniques such as DeepInsight and Fotomics convert gene expression data into 2D images using dimensionality reduction and Fourier transforms, respectively [7,8]. While these methods are effective, they often lack built-in interpretability. To address this issue, we propose a saliency-guided CNN pipeline that highlights the most influential gene regions through gradient-based saliency maps, providing biologically meaningful insights [9,10].

Motivation and Objectives

Classifying brain tumors from transcriptomic data is challenging due to high dimensionality and lack of interpretability. Image-based methods like DeepInsight and Fotomics enable CNNs to process gene expression as images but offer limited explanation of predictions. We propose a saliency-guided CNN to improve both accuracy and interpretability. The objectives of this study are as follows:

- Transform brain tumor transcriptomic profiles into image representations using dimensionality reduction and spectral methods.
- Benchmark DeepInsight, Fotomics, and our proposed saliency-guided CNN on GBM and medulloblastoma datasets.
- Evaluate classification performance and extract biologically relevant gene features using saliency maps

2. Related Work

Recent research has emphasised transforming non-image omics data into image-like representations to harness the strengths of convolutional neural networks (CNNs) for classification tasks. Sharma et al. introduced DeepInsight, a method that converts non-image data into structured 2D images based on feature similarity, enabling the application of standard image-based CNNs to genomic and transcriptomic data [7]. Extensions of DeepInsight have since evolved, such as GeneViT, which leverages vision transformers for improved performance [11], and Fotomics, which maps features through Fourier transformations to capture global frequency-based patterns [8]. Other models, such as OmicsMapNet and KEGG-based mapping strategies, incorporate biological knowledge into spatial arrangements to reflect functional ontologies [12].

Building on previous foundations, this approach introduces a novel pipeline that combines GPU-accelerated image generation (using pyDeepInsight) with multiple dimensionality reduction techniques, t-SNE, UMAP, and PCA, to create RGB image representations of scRNA-seq data. Each colour channel captures distinct structural information: t-SNE emphasises local cell neighbourhoods [13], UMAP preserves the global manifold topology [14], and PCA accounts for data variance. This RGB encoding offers a richer visual representation than traditional grayscale methods, enabling enhanced pattern learning and interpretability.

Our method builds on prior work by applying CNN architectures such as ResNet and EfficientNet directly to these RGB images for tumour classification. In addition, we incorporate saliency maps to identify gene importance at the class level, adding an interpretable layer to deep learning predictions. Although recent reviews highlight the role of deep learning in single-cell and spatial transcriptomics [15], few existing approaches

provide a unified, scalable framework that combines image transformation, dimensionality reduction, CNN-based classification, and biological interpretability. This work addresses this gap by systematically benchmarking DeepInsight, Fotomics, and our saliency-guided CNN across two brain tumour datasets, glioblastoma (scRNA-seq) and medulloblastoma (bulk expression), to evaluate both performance and biological insight.

3. Datasets

This study utilised two publicly available transcriptomic datasets, GSM3828672 and GSE85217, retrieved from the Gene Expression Omnibus (GEO) repository. These datasets were selected to evaluate image-based deep learning methods across both single-cell and bulk transcriptomic modalities, thereby reflecting biological complexity and supporting the assessment of model generalisability.

- GSM3828672 Smart-seq2 GBM offers high-resolution single-cell RNA-seq profiles from 28 glioblastoma tumors, allowing for the study of intra-tumoral heterogeneity and facilitating cell-level classification.
- GSE85217 (medulloblastoma) consists of a large cohort $n = 763$ with established molecular subtype annotations, serving as a benchmark dataset for assessing model performance on bulk-level data.

These datasets enable us to evaluate the robustness and interpretability of image-based CNN models across varying resolutions, tissue types, and classification levels.

4. Methods

This study evaluates DeepInsight, Fotomics, and a saliency-guided CNN on brain tumour transcriptomic data by converting gene expression into images for classification and interpretability. The following subsections outline the theoretical foundations and implementation of each method in the analytical framework.

4.1. Preprocessing and Quality Control for Single-Cell RNA-Seq

Single-cell RNA sequencing data from GSM3828672 glioblastoma (Smart-seq2 platform) and GSE85217 medulloblastoma were processed consistently using the Scanpy pipeline. Quality control (QC) metrics were computed using `scanpy.pp.calculate_qc_metrics`, and cells exhibiting either unusually low or high gene counts, or an elevated proportion of mitochondrial gene expression—based on dataset-specific thresholds—were excluded. This filtering step was performed to eliminate low-quality cells, potential doublets, and apoptotic or stressed cells.

Following QC, gene expression counts were normalised on a per-cell basis to a total count of 10,000 using `scanpy.pp.normalize_total`, and log-transformed using the `log1p` function. Highly variable genes (HVGs) were identified using `scanpy.pp.highly_variable_genes`, and the top 100 HVGs were selected for downstream analyses, including visualisation and clustermap generation.

Dimensionality reduction was performed using Principal Component Analysis (PCA), followed by construction of a neighbourhood graph and clustering using the Leiden algorithm. The resulting clusters were embedded in a two-dimensional space using Uniform Manifold Approximation and Projection (UMAP) for visual interpretation. Differential gene expression between clusters was assessed using a t-test approach, implemented via `scanpy.tl.rank_genes_groups`.

4.2. DeepInsight

DeepInsight converts high-dimensional RNA-seq data into image representations, enabling the use of convolutional neural networks (CNNs) for classification tasks. It utilises

dimensionality reduction techniques such as t-SNE and kernel PCA to organise genes based on their similarities, resulting in pixel layouts that maintain biological relevance. Gene expression values are mapped onto these layouts, generating either RGB or greyscale images. These images are then input into a ResNet-based parallel CNN architecture, which features multiple convolutional layers to facilitate effective feature extraction and classification. The study applied a TCGA RNA-seq dataset consisting of 6216 samples and 60,483 genes across 10 cancer types, and DeepInsight achieved an impressive 99% classification accuracy, outperforming traditional methods [7].

4.3. Fotomics

Fotomics applies Fast Fourier Transform (FFT) to single-cell RNA-seq data, converting gene expression profiles into 2D frequency-based images. By mapping real and imaginary FFT components onto a Cartesian plane and applying convex hull rotation, it generates spatially structured images that encode gene relationships. These images are fed into a CNN architecture for classification. As shown in the workflow in their paper Figure (1) [8], the transformed omics images undergo convolution, pooling, and fully connected layers to predict cell types. Fotomics demonstrated strong classification performance and faster runtimes compared to DeepInsight across multiple scRNA-seq datasets.

4.4. Saliency Guided CNN

In our proposed Saliency Guided CNN framework, we introduce an interpretable deep learning pipeline for single-cell RNA-seq classification by mapping gene expression profiles into RGB images using manifold learning techniques. Each cell is represented as an image of size $S \times S$, where pixel colour encodes gene coordinates from t-SNE (R), PCA (G), and UMAP (B) embeddings. As detailed in Algorithm 1, the resulting images are used to train a convolutional neural network under a k -fold cross-validation strategy. Classification performance is evaluated using standard metrics such as accuracy, precision, recall, F1-score, and AUC. To enhance interpretability, we compute saliency maps based on concept relevance, defined by the gradient of a concept score S_{concept} with respect to each input pixel [16]:

$$M_{ij} = \frac{\partial S_{\text{concept}}}{\partial p_{ij}}$$

This highlights the contribution of each pixel and thus each gene coordinate to the representation of high-level biological concepts, offering insight into cell types specific features.

Algorithm 1 Saliency-Guided scRNA-seq Classification via RGB Mapping

Require: Gene expression matrix $X \in \mathbb{R}^{n \times g}$, labels $y \in \{1, \dots, C\}$, image size $S \times S$

Ensure: Trained CNN model, evaluation metrics, concept-based saliency maps

- 1: Normalise X and select top variable genes
 - 2: Compute PCA(X), t-SNE(X), and UMAP(X) embeddings
 - 3: **for** $i = 1$ to n **do**
 - 4: Normalise coordinates of PCA, t-SNE, and UMAP to $[0, 1]$
 - 5: Map t-SNE coordinates to pixel location $(x, y) \leftarrow X_{\text{tsne}}[i] \cdot (S - 1)$
 - 6: Construct RGB vector: $R \leftarrow X_{\text{tsne}}[i]$, $G \leftarrow X_{\text{pca}}[i]$, $B \leftarrow X_{\text{umap}}[i]$
 - 7: Assign RGB value to pixel (x, y) in blank $S \times S$ image
 - 8: Save image to directory associated with class label $y[i]$
 - 9: **end for**
-

Algorithm 1 *Cont.*

```

10: Initialise  $k$ -fold cross-validation
11: for each fold  $f = 1$  to  $k$  do
12:   Train CNN model on training images
13:   Evaluate model on validation images
14:   Record performance metrics: accuracy, precision, recall, F1-score, AUC
15: end for
16: Save best-performing model as model_final.pth
17: for each class  $c = 1$  to  $C$  do
18:   Select representative image  $I_c$ 
19:   Forward pass  $I_c$  through CNN to get concept score  $S_{\text{concept}}$ 
20:   Compute concept-based saliency:  $M_{ij} = \frac{\partial S_{\text{concept}}}{\partial p_{ij}}$ 
21:   Visualise and save saliency heatmap for  $I_c$ 
22: end for

```

5. Experimental Setup

Preprocessing and Image Generation

Raw gene expression matrices were processed using the Scanpy library. Preprocessing included log-normalisation, followed by dimensionality reduction using a combination of Principal Component Analysis (PCA), Uniform Manifold Approximation and Projection (UMAP), and t-distributed Stochastic Neighbour Embedding (t-SNE). PCA was first applied to reduce the feature space to 50 components. A k -nearest neighbour graph was then constructed in the PCA space and used to compute 2D UMAP embeddings. Additionally, t-SNE was computed using the top 50 principal components.

The resulting 2D embeddings (X_{pca} , X_{umap} , and X_{tsne}) were normalised to the $[0, 1]$ range and mapped to RGB channels: t-SNE to red, PCA to green, and UMAP to blue. Each sample was plotted as a coloured blob on a 128×128 pixel grid using the t-SNE coordinates for spatial positioning (with a radius of 3 pixels to simulate local density). The final images were saved in label-specific folders for downstream CNN-based classification.

6. Model Architecture and Training

To evaluate the classification performance of transcriptomic image representations, two convolutional neural network (CNN) architectures were employed: ResNet-18 and EfficientNet-B0. Both models were implemented using the PyTorch deep learning framework and trained on a system configured to utilise GPU acceleration via CUDA when available, with automatic fallback to CPU execution otherwise.

Both models were initialised with ImageNet-pretrained weights, and their final classification layers were replaced with task-specific linear layers to match the number of output classes:

- ResNet-18: The original fully connected (fc) layer was replaced with a new linear layer whose output dimension matched the number of target classes.
- EfficientNet-B0: The final layer in the classifier block was similarly substituted with a linear layer adapted to the classification task.

Models were deployed to the appropriate computational device (`cuda` or `cpu`) using PyTorch.

The study architectures, ResNet-18 and EfficientNet-B0, were applied exclusively to the custom RGB image representations derived from PCA, UMAP, and t-SNE embeddings of the transcriptomic data. In contrast, DeepInsight and Fotomics were each used as self-contained pipelines, which include their own built-in CNN classifiers. These models were used without modification, as provided in their respective frameworks.

To ensure a fair and consistent comparison across all evaluated methods, including DeepInsight and Fotomics, which employ relatively short training cycles, training was limited to five epochs. Both ResNet-18 and EfficientNet-B0 reached performance stability within this range, and extending the training duration primarily resulted in overfitting due to the dataset size and feature dimensionality. To further ensure robust performance estimation and minimise variance arising from data partitioning, a five-fold stratified cross-validation strategy was employed, maintaining balanced class distributions across folds. Each model was trained independently for each fold using the defined hyperparameters:

- Batch size: 32;
- Epochs: 5;
- Optimizer: Adam;
- Learning rate: 1×10^{-4} ;
- Loss function: Cross-entropy loss.

6.1. Evaluation Protocol

Model performance was evaluated using multiple metrics to ensure both global and class-specific assessment. These included classification accuracy, macro-averaged one-vs-one (OvO) ROC-AUC, and the confusion matrix to measure per-class prediction quality. Additionally, training loss curves were plotted for each fold to monitor convergence behaviour and detect potential overfitting.

The confusion matrix was computed and visualised after each fold, providing insights into class-level misclassifications. An averaged confusion matrix across all folds was also generated to summarise the model’s generalisation capability. These visualisations are included in the Figures 1–4.



Figure 1. Training loss per fold for the GSE85217 ResNet-18.

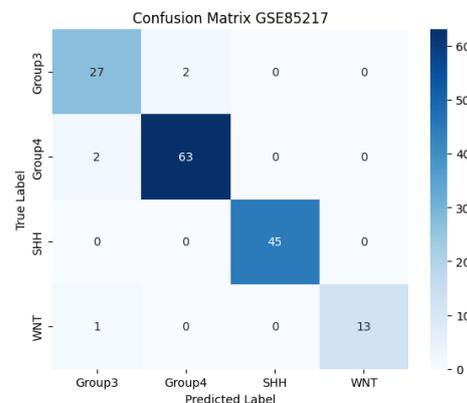


Figure 2. Confusion matrix for GSE85217.

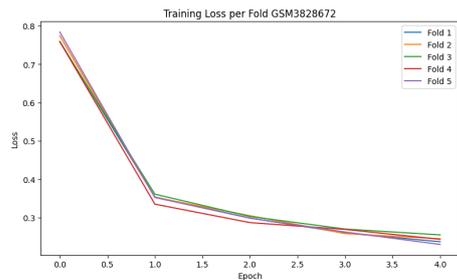


Figure 3. Training loss per fold for the GSM3828672 ResNet-18.

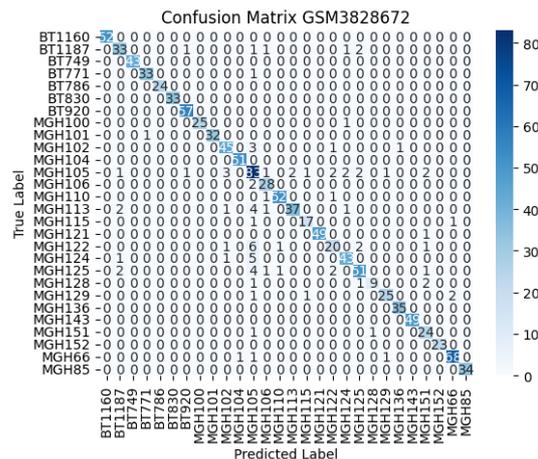


Figure 4. Confusion matrix for GSM3828672.

6.2. Saliency-Based Gene Interpretation

To support interpretability, gradient-based saliency mapping was applied to the trained CNN models, following the method of Simonyan et al. [9]. Each input image $I \in \mathbb{R}^{3 \times 128 \times 128}$ was represented as an RGB tensor derived from independent embeddings of the omics data: PCA (green), UMAP (blue), and t-SNE (red). The classifier is defined as a function $f : \mathbb{R}^{3 \times 128 \times 128} \rightarrow \mathbb{R}^C$, where C denotes the number of tumour classes. For a given input image I , the predicted class is

$$\hat{c} = \arg \max f(I).$$

To compute saliency, we calculated the gradient of the logit corresponding to the predicted class with respect to the input:

$$S = \frac{\partial f(I)_{\hat{c}}}{\partial I} \in \mathbb{R}^{3 \times 128 \times 128}.$$

The saliency map was obtained by taking the maximum absolute gradient across the RGB channels:

$$M_{i,j} = \max_{k \in \{R,G,B\}} |S_{k,i,j}|, \quad M \in \mathbb{R}^{128 \times 128},$$

highlighting those pixels most influential for the classification decision. Saliency maps were computed in PyTorch using the `.backward()` operation and visualised as heatmaps. One representative image per class was selected for illustration. Notably, the red channel (t-SNE) frequently dominated high-saliency regions, suggesting that local transcriptomic structure contributed substantially to classification.

As the RGB channels correspond to independent embeddings rather than coherent colour channels, pixel-level saliency was not interpreted in isolation. Instead, each pixel

was mapped back to its associated gene identity using the coordinate transformations defined during image generation. This procedure ensured that interpretability remained anchored in the original transcriptomic feature space.

Genes were ranked by their aggregated saliency values and examined for biological relevance. Several top-ranked features corresponded to recognised tumour markers, including *OLIG2* and *SOX2* in glioblastoma, and *MYC* and *MYCN* in medulloblastoma. To evaluate functional coherence, enrichment analysis of the top 200 saliency-ranked genes was performed using Gene Ontology (GO) and pathway annotations.

6.3. Implementation and Hardware Details

All components were implemented in Python using PyTorch version 1.12.0 for deep learning and Scanpy for omics preprocessing. Image rendering and visualisation were done using PIL (version 8.0.0), Matplotlib (version 3.8.0), and Seaborn (version 0.11.0). Training and evaluation were carried out on a workstation equipped with an NVIDIA CUDA-enabled GPU, Intel CPU, and 32 GB of RAM.

6.4. Statistical Analysis

To assess whether the observed differences in classification performance between ResNet-18 and EfficientNet-B0 were statistically significant, the Wilcoxon Signed-Rank Test was employed [17,18]. This non-parametric test was chosen as a robust alternative to the paired t-test, as it does not assume normality and is well-suited for small-sample comparisons across cross-validation folds.

The test was conducted separately on the fold-wise accuracy scores from two datasets.

From Table 1, the results indicate no statistically significant difference in classification accuracy between the two architectures ($p > 0.05$), although the GSM3828672 Smart-seq2 result approaches marginal significance. These findings suggest that both ResNet-18 and EfficientNet-B0 exhibit similar and reliable performance on the evaluated tasks.

Considering its lower computational complexity and faster training time, ResNet-18 was selected for subsequent saliency analysis and interpretation. Full classification results, including loss curves, confusion matrices, and ROC visualisations, are provided in Section 7.

Table 1. Wilcoxon Signed-Rank Test results comparing ResNet-18 and EfficientNet-B0 classification accuracy.

Dataset	Statistic	<i>p</i> -Value
GSE85217 (Number of Classes: 4)	3.0000	0.6250
GSM3828672 (Number of Classes: 30)	0.0000	0.0625

7. Results

This section presents a comprehensive evaluation of the proposed saliency-guided CNNs framework, implemented with both ResNet-18 and EfficientNet-B0 architectures, across two transcriptomic datasets: GSE85217 (bulk RNA-seq, 4-class molecular subtype classification) and GSM3828672 (single-cell RNA-seq, 30-class cell type-level classification). The performance of these deep learning models was benchmarked against two conventional image-based bioinformatics approaches: Fotomics and DeepInsight, providing a comparison across modern CNN architectures and traditional handcrafted feature-based methods.

7.1. Medulloblastoma Subtype Classification (GSE85217)

In the analysis of the GSE85217 dataset, Table 2 summarises the 5-fold cross-validation performance of the ResNet-18-based saliency-guided CNN model, achieving a mean accuracy of 97.20%, F1-score of 97.00%, and macro-AUC of 99.18%, indicating high consistency across folds. The training loss curves across five folds shown in Figure 1 indicate rapid and stable convergence within the first few epochs. Furthermore, the confusion matrix in Figure 2 illustrates the model's effectiveness in correctly classifying all four medulloblastoma subtypes (Group 3, Group 4, SHH, and WNT), with minimal confusion.

Table 2. Five-fold cross-validation results on GSE85217 ResNet-18.

Fold	Accuracy	Precision	Recall	F1-Score	ROC-AUC Macro
1	0.9600	0.9500	0.9700	0.9600	0.9720
2	0.9900	0.9900	0.9900	0.9900	0.9960
3	0.9700	0.9700	0.9800	0.9700	0.9970
4	0.9600	0.9400	0.9500	0.9500	0.9970
5	0.9800	0.9800	0.9800	0.9800	0.9970
Mean	0.9720	0.9660	0.9740	0.9700	0.9918
SD	0.0130	0.0207	0.0152	0.0158	0.0111

7.2. GSM3828672 Smart-Seq2 GBM Classification

For the GSM3828672 glioblastoma dataset, the saliency-guided CNN with ResNet-18 demonstrated the highest classification performance among all evaluated methods. The model's 5-fold cross-validation results Table 3 show consistent performance with low standard deviation, indicating stable generalisation across folds. The training loss curves Figure 3 confirm effective convergence within a few epochs. Furthermore, the confusion matrix Figure 4 illustrates the model's capability to correctly classify a diverse set of 28 cell type-derived GBM samples with minimal misclassification, underscoring its robustness in handling highly heterogeneous single-cell transcriptomic data.

Table 3. Five-fold cross-validation results on GSM3828672 ResNet-18.

Fold	Accuracy	Precision	Recall	F1-Score	ROC-AUC (Macro)
1	0.9000	0.9100	0.9100	0.9100	0.9930
2	0.9100	0.9300	0.9000	0.9100	0.9935
3	0.9200	0.9400	0.9100	0.9200	0.9935
4	0.9100	0.9200	0.9000	0.9100	0.9935
5	0.9100	0.9100	0.9200	0.9100	0.9935
Mean	0.9100	0.9220	0.9080	0.9120	0.9934
SD	0.0063	0.0117	0.0075	0.0040	0.0002

7.3. Comparative Summary and Model Insights

As illustrated in Table 4, the saliency-guided CNN using ResNet-18, trained on RGB images generated from PCA, UMAP, and t-SNE embeddings, achieved the highest performance across both datasets, with maximum accuracies of 97.25% on GSE85217 and 91.02% on GSM3828672, and F1-scores of 96.96% and 91.26%, respectively. While EfficientNet-B0, also trained on the same RGB embeddings, performed well, it slightly lagged behind ResNet-18 in precision and interpretability. Fotomics and DeepInsight, evaluated as end-to-end pipelines using their default CNN classifiers, achieved competitive results. Fotomics, which leverages frequency-domain transformations, achieved 96.07% accuracy on the bulk dataset, showcasing its ability to capture global expression patterns. DeepInsight, which preserves local structure through spatial feature mapping, performed

relatively better in recall and AUC (Macro), especially on GSM3828672, indicating its strength in modelling localised expression signals. However, both transformation-based methods lack inherent interpretability. In contrast, the proposed saliency-guided CNN integrates accurate classification with class-level feature attribution, making it particularly suited for biologically meaningful and explainable predictions in both bulk and single-cell transcriptomic contexts.

Table 4. Comparative performance summary across datasets. Metrics are averaged over 5-fold cross-validation.

Dataset	Model	Accuracy (%)	Precision	Recall	F1-Score	AUC
GSE85217	ResNet-18	97.25	97.24	96.77	96.96	99.61
	EfficientNet-B0	96.72	96.96	95.50	96.13	99.36
	Fotomics	96.07	96.16	95.61	95.64	99.92
	DeepInsight	85.62	87.50	85.62	81.77	99.03
GSM3828672	ResNet-18	91.02	92.29	90.85	91.26	99.35
	EfficientNet-B0	89.94	91.22	90.10	90.19	99.30
	Fotomics	71.60	72.13	69.73	70.36	95.25
	DeepInsight	70.43	72.27	70.43	67.89	97.35

7.4. Interpretability via Saliency Maps

To improve the interpretability of the classification model, class-wise saliency maps were generated for the GSE85217 dataset using gradient-based visualisation techniques. As illustrated in Figure 5, saliency maps were generated for representative tumours from the Group 3, Group 4, SHH, and WNT medulloblastoma subtypes. Genes with saliency scores ≥ 0.2 are annotated, representing those with the strongest influence on the model's predictions. The saliency score quantifies the local sensitivity of the model's output to perturbations in individual gene expression levels. Notable examples include *CXCR4* (0.41) in Group 3, *TOP2A* (0.39) in Group 4, *CDK6* (0.25) in SHH, and *WNT5A* (0.20) in WNT. These visualisations provide interpretable, model-derived evidence supporting subtype-specific gene relevance. Figure 6 further demonstrates the effective separation of tumour subtypes through a unified pixel map, confirming the model's ability to learn subtype-specific spatial features. Moreover, Figure 7 shows the gene density per pixel, indicating that highly salient areas correspond to biologically relevant gene clusters. Collectively, these visualisations validate the biological plausibility of the learned representations and support the interpretability of the saliency-guided CNN framework.

To improve the interpretability of the GSM3828672 dataset, detailed gene-level saliency annotations for selected tumour samples Figure 8 identified high-impact genes, including *LEMD1* (0.73), *SERPINA3* (0.60), *SNORD72* (0.92), and *MIR3671* (0.61), demonstrating the model's ability to extract biologically meaningful features from transcriptomic profiles. In Figure 9, a purple box indicates a specific pixel region in the original image, potentially marking a spatial feature or annotated input. The saliency map (right) highlights regions with the greatest influence on the model's prediction (class 8), showing spatial clustering that may correspond to cell type-specific profiles. Spatial clustering of cell type profiles Figure 10 and the corresponding gene density distributions Figure 11 depict biologically meaningful encodings that support the interpretability of the learned representations.

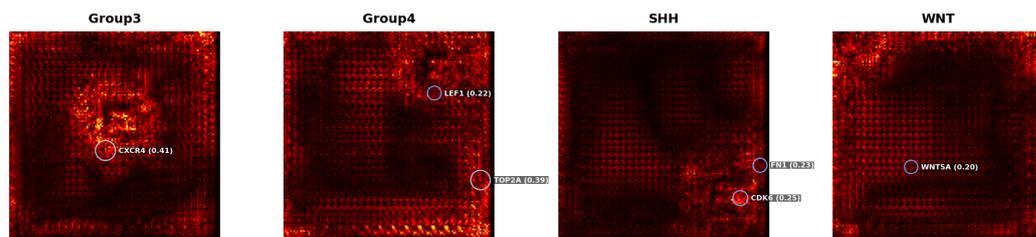


Figure 5. Saliency maps annotate high-impact genes across each tumour subtype for GSE85217.

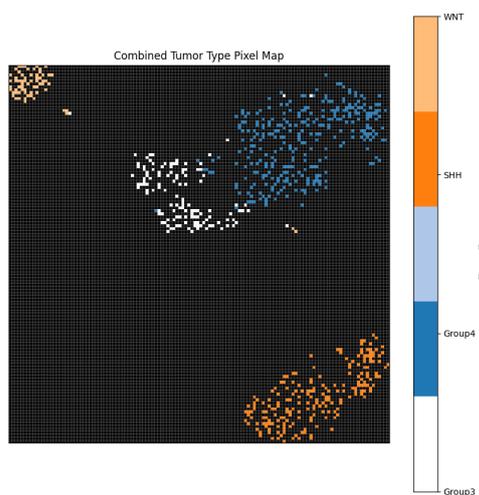


Figure 6. Combined tumour type pixel map with tumour type for GSE85217.

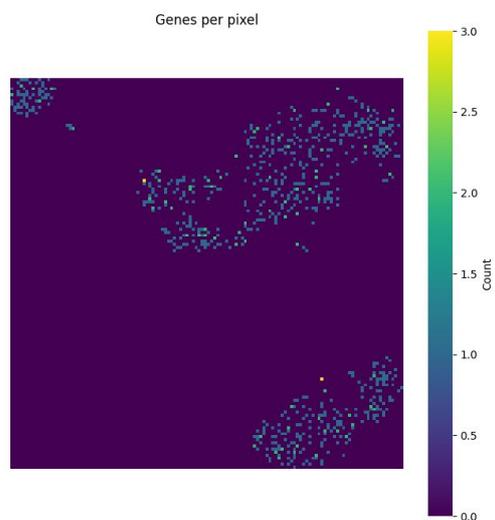


Figure 7. Map of the number of genes contributing per pixel for GSE85217.

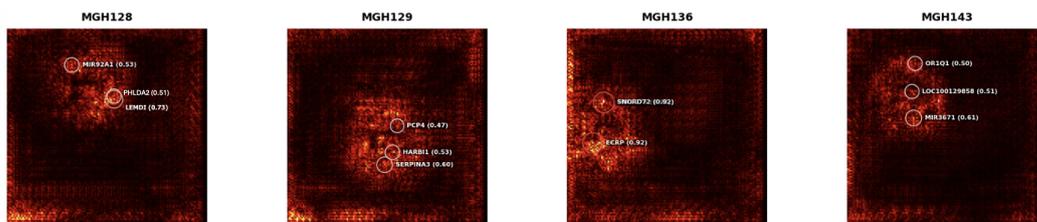


Figure 8. Saliency maps annotate high-impact genes across tumour samples for GSM3828672.

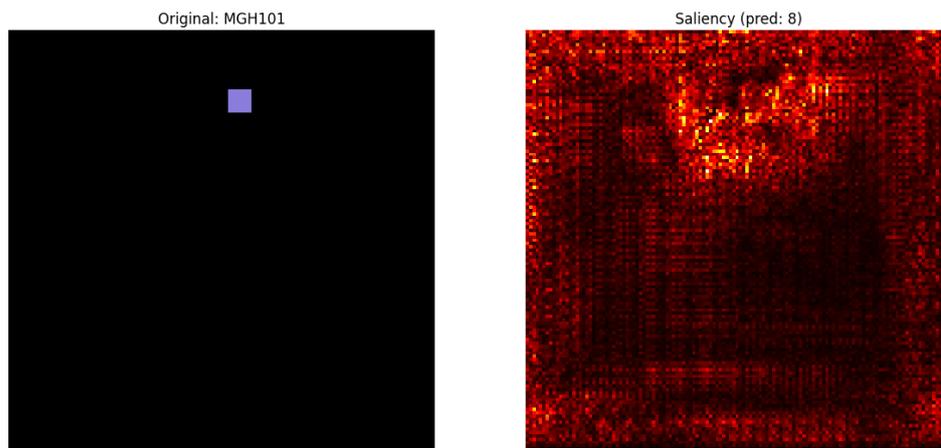


Figure 9. RGB image pixel vs. saliency for GSM3828672.

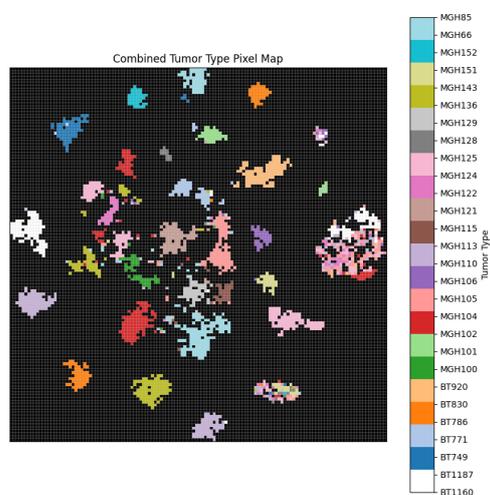


Figure 10. Combined tumour type pixel map with tumour type for GSM3828672.

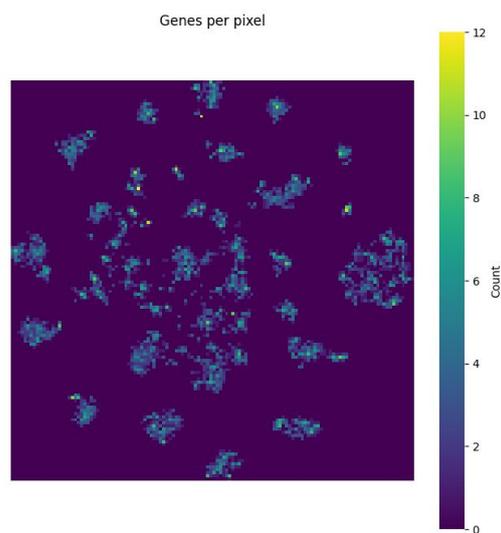


Figure 11. Map of the number of genes contributing per pixel for GSM3828672.

8. Discussion

This study presents a comparative benchmark of three image-based deep learning approaches: DeepInsight, Fotomics, and a saliency-guided CNN applied to transcriptomic data from two distinct brain tumour datasets, medulloblastoma GSE85217 and glioblastoma GSM3828672. The Results Section demonstrates that transforming gene expression

profiles into RGB image representations significantly enhances classification performance, particularly when integrated with state-of-the-art CNN architectures such as ResNet-18 and EfficientNet-B0 [19]. The saliency-guided CNN consistently produced competitive performance across both datasets, while uniquely offering model interpretability through gradient-based saliency maps. In the medulloblastoma dataset, the saliency-enhanced model achieved high classification accuracy and successfully identified spatial gene features relevant to tumour subtypes Table 4. Similarly, for the GBM cell types classification experiment, the model maintained stable accuracy across folds while enabling class-specific insight into gene contributions (Table 3). These results validate the potential of our pipeline to support both diagnostic precision and biological understanding. Compared to traditional approaches, the proposed pipeline benefits from its modular design, integrating dimensionality reduction, image transformation, and CNN classification while preserving statistical and biological structure. DeepInsight enables tabular data to be converted into spatially meaningful image formats using methods such as t-SNE, UMAP, and PCA [7], while Fotomics captures frequency-based relationships via Fourier transforms [8]. The RGB encoding further facilitates CNN feature extraction by incorporating complementary structural information. The inclusion of saliency maps provides interpretability by tracing prediction gradients back to input features, a critical requirement in AI healthcare [9,10].

To improve interpretability of the CNN model, a three-step process was employed to identify genes contributing to tumour classification. Gene expression features were first embedded into a 2D grid using PCA followed by t-SNE, enabling each gene to be mapped to a unique pixel location. The top 2000 highly variable genes were then selected for focused analysis. Saliency maps were subsequently generated using a trained ResNet-18 model, and genes with saliency scores ≥ 0.2 were annotated as high-impact contributors. This approach revealed biologically relevant genes in individual samples from the GSM3828672 dataset, including LEMD1, SERPINA3, SNORD72, and MIR3671, as well as subtype-specific markers in the GSE85217, such as CXCR4, TOP2A, CDK6, and WNT5A. Future directions include expanding this pipeline to integrate multi-omics modalities for methylation, proteomics, and spatial transcriptomics, where physical gene coordinates can complement virtual embeddings. Additionally, applying attention-based architectures such as vision transformers could further improve both accuracy and interpretability [20]. Another promising avenue involves incorporating biological priors, such as pathway information or gene ontologies, into image layout strategies to enhance biological relevance.

Despite the strong performance and interpretability of the proposed framework, several limitations remain. The current pipeline relies on pre-defined dimensionality reduction techniques, which may introduce variability depending on parameter choices. Additionally, the transformation of transcriptomic data into images can obscure original gene–gene relationships if not properly preserved. Finally, while saliency maps provide valuable insights, they do not guarantee causal interpretability and may be sensitive to network architecture or input noise.

9. Conclusions

This study presented an image-based deep learning framework for the classification of brain tumour transcriptomic data, with a particular focus on medulloblastoma GSE85217 and glioblastoma GSM3828672. High-dimensional gene expression profiles were transformed into RGB image representations using dimensionality reduction techniques including PCA, UMAP, and t-SNE. This enabled the application of convolutional neural networks specifically ResNet-18 and EfficientNet for accurate classification of tumour subtypes and cell types.

Among the approaches evaluated, the proposed saliency-guided CNN demonstrated robust classification performance alongside improved interpretability through class-wise saliency maps. Comparative analyses with existing methods, such as DeepInsight and Fotomics, validated the effectiveness of the image-based approach while offering the additional advantage of visualising gene-level contributions to model predictions.

These findings underscore the promise of combining deep learning with omics imagification for developing interpretable, data-driven tools in cancer diagnostics. Future work will aim to incorporate biological priors, extend the framework to spatial and multi-omics datasets, and explore transformer-based architectures to enhance both predictive performance and biological insight.

Author Contributions: A.A. (Ali Alyatimi) and M.A.I. contributed equally to the conceptualisation, methodology, analysis, and writing. V.C. and A.A. (Ali Anaissi) supervised the project and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The code used in this study will be made publicly available upon acceptance at the GitHub repository: <https://github.com/alyatimi/Image-Based-Deep-Learning-for-Brain-Tumour-Transcriptomics.git> (accessed on 14 October 2025). The transcriptomic datasets analysed in this work are publicly accessible from the Gene Expression Omnibus (GEO) database: GSE85217, which contains bulk expression profiling data from medulloblastoma subtypes, and GSM3828672, comprising single-cell RNA-seq (Smart-seq2) data from glioblastoma multiforme (GBM) cell types.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

1. Svensson, V.; Vento-Tormo, R.; Teichmann, S.A. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* **2018**, *13*, 599–604. [[CrossRef](#)] [[PubMed](#)]
2. Stupp, R.; Mason, W.P.; Van Den Bent, M.J.; Weller, M.; Fisher, B.; Taphoorn, M.J.; Belanger, K.; Brandes, A.A.; Marosi, C.; Bogdahn, U.; et al. Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. *N. Engl. J. Med.* **2005**, *352*, 987–996. [[CrossRef](#)] [[PubMed](#)]
3. Verhaak, R.G.; Hoadley, K.A.; Purdom, E.; Wang, V.; Qi, Y.; Wilkerson, M.D.; Miller, C.R.; Ding, L.; Golub, T.; Mesirov, J.P.; et al. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **2010**, *17*, 98–110. [[CrossRef](#)] [[PubMed](#)]
4. Cavalli, F.M.; Remke, M.; Rampasek, L.; Peacock, J.; Shih, D.J.; Luu, B.; Garzia, L.; Torchia, J.; Nor, C.; Morrissy, A.S.; et al. Intertumoral Heterogeneity within Medulloblastoma Subgroups. *Cancer Cell* **2017**, *31*, 737–754.e6. [[CrossRef](#)] [[PubMed](#)]
5. Angermueller, C.; Pärnamaa, T.; Parts, L.; Stegle, O. Deep learning for computational biology. *Mol. Syst. Biol.* **2016**, *12*, 878. [[CrossRef](#)] [[PubMed](#)]
6. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)] [[PubMed](#)]
7. Sharma, A.; Vans, E.; Shigemizu, D.; Boroevich, K.A.; Tsunoda, T. DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture. *Sci. Rep.* **2019**, *9*. [[CrossRef](#)] [[PubMed](#)]
8. Zandavi, S.M.; Liu, D.; Chung, V.; Anaissi, A.; Vafae, F. Fotomics: Fourier transform-based omics imagification for deep learning-based cell-identity mapping using single-cell omics profiles. *Artif. Intell. Rev.* **2023**, *56*, 7263–7278. [[CrossRef](#)]
9. Simonyan, K.; Vedaldi, A.; Zisserman, A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. *arXiv* **2013**, arXiv:1312.6034. Version Number: 2. [[CrossRef](#)]
10. Moujahid, H.; Cherradi, B.; Al-Sarem, M.; Bahatti, L.; Bakr Assedik Mohammed Yahya Eljialy, A.; Alsaedi, A.; Saeed, F. Combining CNN and Grad-Cam for COVID-19 Disease Prediction and Visual Explanation. *Intell. Autom. Soft Comput.* **2022**, *32*, 723–745. [[CrossRef](#)]

11. Gokhale, M.; Mohanty, S.K.; Ojha, A. GeneViT: Gene Vision Transformer with Improved DeepInsight for cancer classification. *Comput. Biol. Med.* **2023**, *155*, 106643. [[CrossRef](#)]
12. Ma, S.; Zhang, Z. OmicsMapNet: Transforming omics data to take advantage of Deep Convolutional Neural Network for discovery. *arXiv* **2018**, arXiv:1804.05283. Version Number: 2. [[CrossRef](#)]
13. Kobak, D.; Berens, P. The art of using t-SNE for single-cell transcriptomics. *Nat. Commun.* **2019**, *10*. [[CrossRef](#)] [[PubMed](#)]
14. McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* **2018**. [[CrossRef](#)]
15. Ge, S.; Sun, S.; Xu, H.; Cheng, Q.; Ren, Z. Deep learning in single-cell and spatial transcriptomics data analysis: Advances and challenges from a data science perspective. *Briefings Bioinform.* **2025**, *26*. [[CrossRef](#)] [[PubMed](#)]
16. Brocki, L.; Chung, N.C. Concept Saliency Maps to Visualize Relevant Features in Deep Generative Models. In Proceedings of the 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), Boca Raton, FL, USA, 16–19 December 2019; pp. 1771–1778. [[CrossRef](#)]
17. Sayeedunnisa, S.F.; Hegde, N.P.; Khan, K.U.R. Wilcoxon Signed Rank Based Feature Selection for Sentiment Classification. In *Proceedings of the Second International Conference on Computational Intelligence and Informatics*; Bhateja, V., Tavares, J.M.R., Rani, B.P., Prasad, V.K., Raju, K.S., Eds.; Springer Singapore: Singapore, 2018; Volume 712, pp. 293–310. [[CrossRef](#)]
18. Wilcoxon, F. Individual Comparisons by Ranking Methods. *Biom. Bull.* **1945**, *1*, 80. [[CrossRef](#)]
19. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
20. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2021**, arXiv:2010.11929. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.