

Brain tissue classification in hyperspectral images using multistage diffusion features and transformer

Neetu Sigger^{1,2} | Tuan T. Nguyen³  | Gianluca Tozzi²

¹School of Computing, University of Buckingham, Buckingham, UK

²School of Engineering, University of Greenwich, Greenwich, UK

³School of Computing & Mathematical Sciences, University of Greenwich, Greenwich, UK

Correspondence

Tuan T. Nguyen, School of Computing, and Mathematical Sciences, University of Greenwich, Greenwich, United Kingdom.
Email: tuan.nguyen@greenwich.ac.uk

Abstract

Brain surgery is a widely practised and effective treatment for brain tumours, but accurately identifying and classifying tumour boundaries is crucial to maximise resection and avoid neurological complications. This precision in classification is essential for guiding surgical decisions and subsequent treatment planning. Hyperspectral (HS) imaging (HSI) is an emerging multidimensional optical imaging method that captures detailed spectral information across multiple wavelengths, allowing for the identification of nuanced differences in tissue composition, with the potential to enhance intraoperative tissue classification. However, current frameworks often require retraining models for each HSI to extract meaningful features, resulting in long processing times and high computational costs. Additionally, most methods utilise the deep semantic features at the end of the network for classification, ignoring the spatial details contained in the shallow features. To overcome these challenges, we propose a novel approach called MedDiffHSI, which combines diffusion and transformer techniques. Our method involves training an unsupervised learning framework based on the diffusion model to extract high-level and low-level spectral-spatial features from HSI. This approach eliminates the need for retraining of spectral-spatial feature learning model, thereby reducing time complexity. We then extract intermediate multistage features from different timestamps for classification using a pretrained denoising U-Net. To fully explore and exploit the rich contextual semantics and textual information hidden in the extracted diffusion feature, we utilise a spectral-spatial attention module. This module not only learns multistage information about features at different depths, but also extracts and enhances effective information from them. Finally, we employ a supervised transformer-based classifier with weighted majority voting (WMV) to perform the HSI classification. To validate our approach, we conduct comprehensive experiments on in vivo brain database data sets and also extend the analysis to include additional HSI data sets for breast cancer to evaluate the framework performance across different types of tissue. The results demonstrate that our

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Journal of Microscopy* published by John Wiley & Sons Ltd on behalf of Royal Microscopical Society.

framework outperforms existing approaches by using minimal training samples (5%) while achieving state-of-the-art performance.

KEYWORDS

bioinformatics, brain tumour, cancer surgery, deep learning, diffusion model, hyperspectral imaging, precision medicine

1 | INTRODUCTION

The global cancer burden is projected to surpass 35 million new cases by 2050, a 77% rise from the estimated 20 million cases in 2022.¹ In addition to radiotherapy and chemotherapy, surgery is a primary treatment option for brain tumours. However, distinguishing tumour tissue from normal brain tissue during surgery is challenging due to the infiltrative nature of brain tumours. Several image guidance tools, such as intraoperative neuro-navigation, intra-operative magnetic resonance imaging, and fluorescent tumour markers, are used in neurosurgery to identify and resect brain tumours. However, these tools have limitations. Neuronavigation systems, which rely on preoperative computed tomography (CT) or Magnetic Resonance Imaging (MRI) data, lose accuracy during surgery due to brain shift. Intraoperative MRI mitigates brain shift issues but prolongs surgery and requires specialised equipment, posing challenges for patient safety and monitoring. Additionally, these methods may not provide sufficient contrast for all tumour types, sometimes leading to false positives.^{2,3}

Hyperspectral imaging (HSI) is a promising noninvasive and nonionising technique that enables rapid acquisition and analysis of diagnostic information across various fields, such as remote sensing,^{4–6} object detection,⁷ environmental monitoring,⁸ forensics,^{9,10} defence and security,¹¹ among many others. HSI is an emerging multidimensional optical imaging method, which allows objective identification of tissues and conveys information from spectral bands beyond the capabilities of the human visual system or conventional cameras. Recently, HSI has emerged as a powerful tool for medical imaging^{12,13} due to its ability to capture detailed spectral information across numerous narrow bands. This capability is particularly advantageous in the classification and diagnosis of cancer, where distinguishing between different types of tissues is crucial. Several studies have demonstrated the efficacy of HSI in the classification of various cancers using deep learning (DL) techniques. These include gastric and colon cancer,^{14,15} breast cancer,^{16,17} head and neck cancer,¹⁸ and brain cancer.^{19,20}

Vivo-HSI has been utilised to create a human-brain imaging database¹³ for Glioblastoma (GBM) tumour

surgery. The primary objective of this database is to develop an approach for the qualitative delineation of tumour margins, with the goal of providing noninvasive surgical tools capable of accurately identifying tumour locations. Since HSI is designed to store a large amount of information, some challenges still exist. HSI data provide a combination of spatial and spectral information, which is crucial for accurate tissue classification. The data contain a vast number of interconnected bands, leading to high dimensionality, significant illumination variance, and imbalanced class distribution. Additionally, the presence of mixed materials in the background, such as medical devices and/or other objects labelled as one class, complicates the classification process and reduces accuracy.

To address these challenges, various efforts have been made, with most tumour identification models using HSI based on traditional machine learning (ML) algorithms. Fabelo et al.²¹ explored the Vivo-HSI data set, using spatial–spectral data to differentiate between tumour cells. They performed the classification of different tissue samples using both supervised methods (support vector machine (SVM) classifier, K-nearest neighbour (KNN) filtering, and fixed reference t-distributed stochastic neighbour embedding (FR-tSNE) dimensional reduction) and unsupervised methods (hierarchical K-means clustering algorithm). However, traditional classifiers like KNN and SVM primarily utilise spectral information and often disregard spatial variability. The performance of these tumour detection methods often depends on manual feature extraction, which is time-consuming and prone to subjective bias. Therefore, seeking efficient automatic feature extraction and classification methods is of significant research importance. The rapid advancement of DL techniques has led to improvements in the spectral and spatial properties of HSI data. The use of a 2D convolutional neural network (CNN) and a 1D CNN model, as investigated in Ref. [19], provided more accurate results in classifying the Vivo-HSI data set compared to the traditional SVM classifier. Additionally, a 3D-2D hybrid CNN-based approach²² is employed, to extract spectral–spatial features for brain tissue classification, effectively distinguishing between tumours, normal tissue, and blood vessels in the human brain. Compared with hand-crafted feature

extraction methods, DL-based methods can automatically extract informative features from the original images through a series of hierarchical layers. Furthermore, a 1D-deep neural network (DNN) model is used for deep spectral HSI classification, and a 2D-CNN model is adopted for spectral-spatial HSI classification in Ref. [20]. Decision fusion strategies, edge-preserving filtering, and a fully convolutional network (FCN) are then employed to optimise and combine classification and segmentation results for the final classification, achieving competitive tumour identification performance.

However, research on using DL techniques for tumour identification with *in vivo* human brain HSI is very limited. Collecting labelled samples of brain tumour tissue is either expensive or time-consuming, resulting in a limited number of training samples in the medical field. Given the variations in the appearance of different levels and types of brain tumours, effective feature extraction is crucial for practical applications. Developing accurate and robust feature extraction and classification methods for tumour identification is a significant research direction. Different deep networks can extract various types of features, such as spectral, spatial, shallow, and deep features, each contributing differently to classification accuracy. Leveraging the complementarity of different networks to utilise these diverse features fully is essential for improving tumour recognition results. The DL have been shown to be effective in HSI classification; however, there are still many challenges in this field. For instance, convolutional operations handle a local neighbourhood,²³ so spatial-spectral information is not correlated across different layers, leading to information loss. CCR-Net²⁴ introduces a cross-channel reconstruction (CCR) module for more effective fusion of multimodal features in remote sensing image classification. Similarly, DC-Net²⁵ proposes a subpixel-level HS super-resolution (HS-SR) framework that fuses HS and multispectral (MS) images. Although it addresses the distribution gap between data sources, spectral degradation remains a challenge in real-world HS-SR applications. The HighDAN²⁶ network proposes a high-resolution domain adaptation approach for cross-city semantic segmentation in multimodal remote sensing images. The network²⁶ effectively preserves the spatial topology of remote sensing images through parallel high- and low-resolution fusion, leading to improved segmentation performance. The SpectralGPT²⁷ model presents a novel 3D generative pretrained transformer (GPT) designed for spectral remote sensing data. Although it effectively captures spatial-spectral couplings and preserves spectral characteristics, current generalisability may be constrained by the need for a larger and more diverse training data set.

Effectively learning rich representations and addressing the complexities of spectral-spatial relations in high-dimensional data are crucial for achieving optimal HSI classification results. Recently, the denoising diffusion probabilistic model (DDPM)²⁸ has emerged as a groundbreaking class of generative models, adept at modelling complex relationships and effectively learning high-level and low-level visual features. In our previous work,⁴ we evaluated the use of HSI, specifically focusing on diffusion models and transformer algorithms, for effective classification in remote sensing images. The diffusion model effectively extracts both low and high-level features, aiding in efficient HSI classification.

In this study, a diffusion-transformer-based framework, MedDiffHSI, was developed to generate classification maps identifying tumour, healthy, and blood vessel tissues using *in vivo* human brain hyperspectral images. Unlike conventional approaches that rely on traditional classifiers, such as SVM and KNN, which often neglect spatial information and require manual feature extraction,²¹ MedDiffHSI integrates diffusion models architectures to extract both spectral and spatial features. We made significant improvements to MedDiffHSI compared to our previous model, DiffSpectralNet.⁴ Firstly, we eliminated the need for retraining the diffusion model, significantly reducing time complexity. Secondly, intermediate hierarchical features were extracted from different timestamps using a pretrained denoising U-Net, enabling more efficient and accurate classification. To better utilise the timestep-wise features extracted from the diffusion model, a spectral-spatial attention mechanism was employed. This mechanism learns multistage information at various depths, enhancing sensitivity to relevant features while adaptively retaining critical information. Finally, a supervised transformer-based classifier was applied for HSI classification, and postprocessing using weighted majority voting (WMV) was introduced to further refine the classification maps, particularly improving classification maps by leveraging spatial information.

We evaluated the effectiveness of the proposed method on seven hyperspectral (HS) images from five patients with brain tumours, as initially established by Ref. [13], and we extended our evaluation beyond brain tumour classification to include additional HSI data sets for breast cancer. Our results clearly demonstrate that MedDiffHSI significantly improves classification results and outperforms other advanced HSI classification methods. Moreover, this study also opens the way for further investigations into the potential of diffusion models in processing complex, high-dimensional hyperspectral data, opening up promising prospects for diverse applications. This framework could assist neurosurgeons in the critical task of

identifying cancer tissue during brain surgery with higher accuracy.

2 | METHODS

In this section, we describe a novel method called MedDiffHSI, a diffusion-based feature learning framework designed to effectively and efficiently explore multi-stage diffusion features for comprehensively modelling spectral–spatial relations for classification. The framework is illustrated in Figure 1. The following section provides a detailed introduction to the proposed MedDiffHSI model.

2.1 | Diffusion-based unsupervised spectral–spatial feature learning

In order to capture complex spectral–spatial relations and label-agnostic information of HSI data effectively, the first step of our proposed approach is to train a diffusion model in an unsupervised manner, as shown in Figure 1A. We introduce the detailed formulation of our unsupervised feature learning procedure, which involves diffusion-based forward and backward processes with the HSI data.

- **Forward Diffusion Process.** DDPM represents a category of models based on likelihood estimations. In the forward process, Gaussian noise is added to the original training data. Our proposed model aims to learn spectral–spatial features effectively in an unsupervised manner by training a DDPM using unlabelled patches randomly cropped from the HSI data set. Given an unlabelled patch $x_0 \in \mathbb{R}^{P \times P \times B}$, where P denotes the height and width of the patch and B represents the number of spectral channels, Gaussian noise is gradually added to the HSI patch according to the variance schedule $\{\beta_t\}_{t=0}^T$ in the diffusion process, where T is the total number of timesteps. This follows a Markov chain²⁸ process:

$$x_t = \sqrt{\alpha_t}x_0 + \sqrt{1 - \alpha_t}\epsilon, \quad \epsilon \sim \mathcal{N}(0, I), \quad (1)$$

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t$ represents the product of α_1 to α_t . This formulation provides the probability distribution of the HSI at a given time step t .

- **Reverse Diffusion Process.** In the reverse diffusion process, a spectral–spatial U-Net²⁹ denoising network is employed is trained to predict the noise added on x_{t-1} , taking noisy patch x_t and timestep t as inputs. And x_{t-1} is calculated by subtracting the predicted noise from x_t . DDPM uses a Markov chain process to remove the noisy

sample x_T to x_0 step by step. Under large T and small β_t , the probability of reverse transitions is approximated as a Gaussian distribution and is predicted by a U-Net as follows:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sigma_\theta(x_t, t)), \quad (2)$$

where the reverse process can be reparameterised by estimating $\mu_\theta(x_t, t)$ and $\sigma_\theta(x_t, t)$. $\sigma_\theta(x_t, t)$ is set to $\sigma_t^2 I$, where σ_t^2 is not learned. To obtain the mean of the conditional distribution $p_\theta(x_{t-1}|x_t)$, we need to train the network to predict the added noise. The mean of $\mu_\theta(x_t, t)$ is derived as follows:

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(x_t, t) \right), \quad (3)$$

where $\epsilon_\theta(\cdot, \cdot)$ denote the spectral–spatial denoising network whose input is the timestep t and the noisy hyperspectral instance x_t at timestep t .

The denoising network takes in the noisy hyperspectral instance along with the timestep to produce the predicted noise. The U-Net denoising model $\epsilon_\theta(x_t, t)$ is optimised by minimising the loss function of the spectral–spatial diffusion process can be expressed as follows:

$$\mathcal{L}(\theta) = \mathbb{E}_{t, x_0, \epsilon} \left[\left(\epsilon - \epsilon_\theta \left(\sqrt{\alpha_t}x_0 + \sqrt{1 - \alpha_t}\epsilon, t \right) \right)^2 \right]. \quad (4)$$

2.2 | Feature extraction and enhancement

After training the network using Unsupervised spectral–spatial methods, we start extracting useful diffusion features from the pretrained DDPM. During the feature extraction step, we utilise the U-Net denoising network to extract a spectral–spatial timestep-wise feature. The pre-training of DDPM enables it to capture rich and diverting information from the input data during the reverse process. As a result, we extract features from the intermediate hierarchies of DDPM at various timesteps to create robust representations that encapsulate the salient features of the input HSI. The parameters of the pretrained DDPM remain constant, as shown in Figure 1B. Subsequently, x_t is fed into the pretrained spectral–spatial denoising U-Net to derive hierarchical features from the U-Net decoder. Diffusion features from various decoder layers are collectively upsampled to $P \times P$ and then merged to form the feature f_t in $\mathbb{R}^{P \times P \times L}$ at timestep t , where P represents the

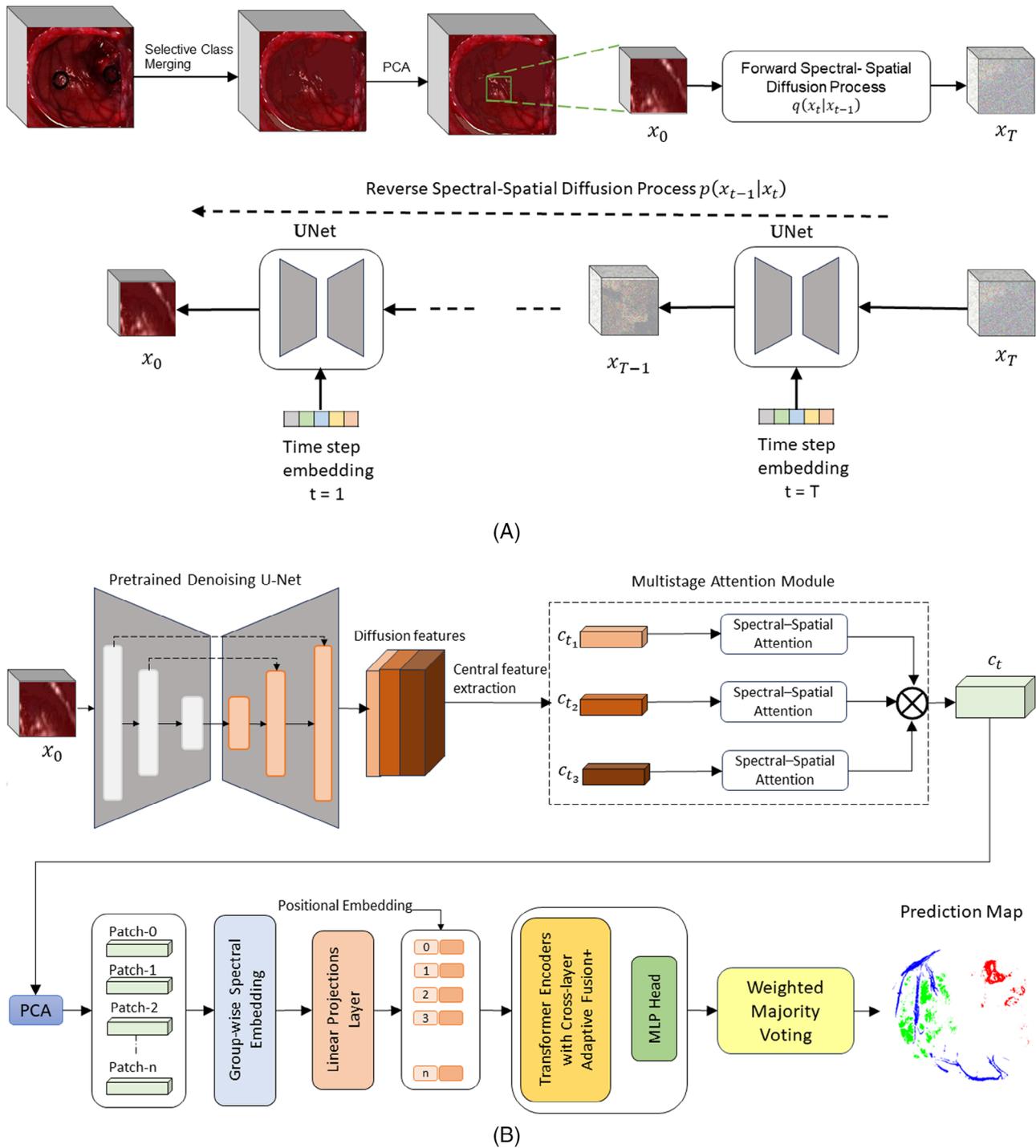


FIGURE 1 Overview of our proposed MedDiffHSI. (A) Unsupervised Spectral-Spatial Feature Learning Network. x_0 and x_T represent HSI patches of timestep 0 and timestep T . $q(x_t | x_{t-1})$ and $p(x_{t-1} | x_t)$ represent forward and reverse spectral-spatial diffusion processes, respectively. (B) Supervised Classification: (1) Extracting multistage features from the pretrained denoising U-Net decoder at different timesteps t . To better utilise these features, we employ a spectral and spatial attention mechanism to adaptively retain relevant information. (2) Using the patch-wise feature vectors to train a cross-layer transformer for HSI classification. (3) Using weighted majority voting to smooth the local neighbourhood.

height and width of the patch and L denotes the number of feature channel. As for each timestep, we extracted multistage features from three different layers represent the extracted low-level, mid-level, and high-level features,

respectively. For each feature $f_{ti} \in \mathbb{R}^{P \times P \times L}$, we retain only the vector associated with the centre pixel, indexed as $C_i \in \mathbb{R}^{1 \times 1 \times L}$, which significantly reduces computational cost by decreasing the number of parameters.

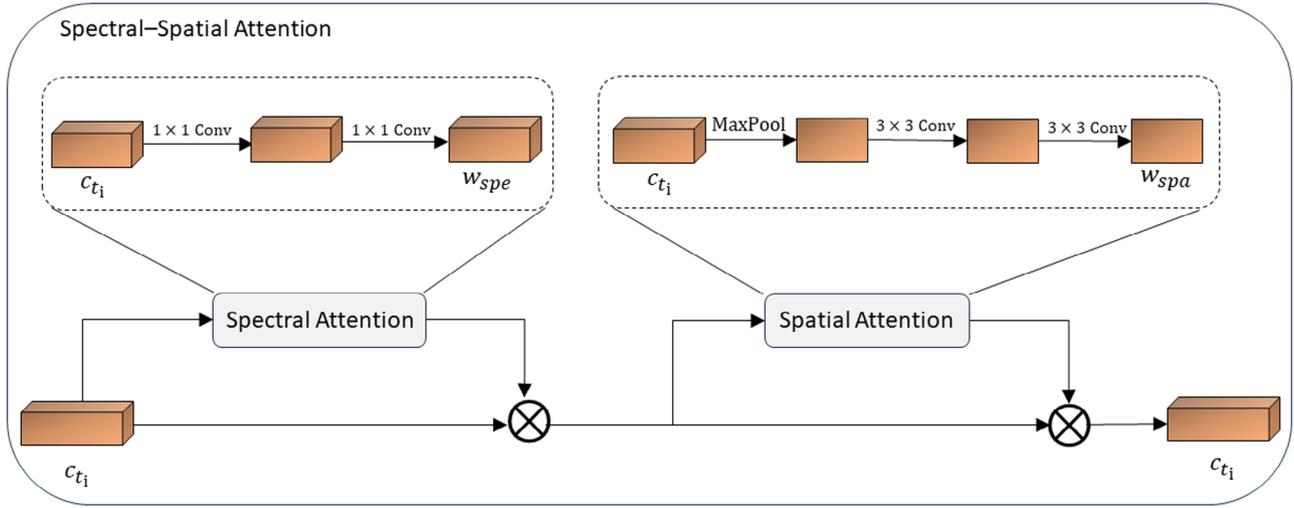


FIGURE 2 The spectral–spatial attention module in MedDiffHSI, which extracts effective and efficient information from multistage features c_{t_i} ($i \in \{1, 2, 3\}$) from the diffusion model, resulting in the combined feature c_{t_i} .

In this work, we enhanced our previous model, DiffSpectralNet,⁴ by introducing a spectral–spatial attention module. Attention modules are effective in improving feature extraction and enhancing relevant information.^{30–32} These modules selectively focus on the most informative features, amplifying important details while suppressing irrelevant data. Inspired by the design of spectral–spatial attention modules in Ref. [30], we employed this approach to extract and enhance effective information from the features derived by the diffusion model, where spectral and spatial attention were cascaded, as shown in Figure 1B and detailed in Figure 2. The feature weight $w_{spe} \in \mathbb{R}^{1 \times 1 \times L}$ was obtained through the two convolutional layers with a kernel size of 1×1 , as shown in Equation (5).

$$w_{spe} = \sigma(w_2 * (\delta(w_1 * p_i))), \quad (5)$$

where σ and δ represent the sigmoid and ReLU activation functions, respectively. w_1 and w_2 are the weight parameters of the two convolutional layers, and $*$ denotes the convolution operation. Finally, as shown in Figure 2, the band weight w_{spe} was used to recalibrate the bands in the feature c_{t_i} to highlight the useful spectral information, using Equation (6).

$$c_{t_i} = w_{spe} \otimes c_{t_i}, \quad (6)$$

where \otimes represents element-wise multiplication.

Spatial attention aims to enhance the spatial information for pixels belonging to the same class as the central pixel, while suppressing pixels of other classes. Firstly,

global max pooling was applied to the input feature c_{t_i} along the channel direction, as shown in Equation (7).

$$c_{t_i, \max} = \max_c(c_{t_i}(i, j)), \quad (7)$$

where $c_{t_i}(i, j)$ represents the value at position (i, j) in the feature $c_{t_i} \in \mathbb{R}^{p \times p \times L}$, \max_c represents taking the maximum value along the channel direction c , and $c_{t_i, \max} \in \mathbb{R}^{p \times p}$ is the feature map after global max pooling. This pooled feature map is then passed through two 2D convolutional layers to generate the spatial weight $w_{spa} \in \mathbb{R}^{p \times p}$, as shown in Equation (8).

$$w_{spa} = \sigma(\delta(c_{t_i, \max} * w_1) * w_2), \quad (8)$$

where w_1 and w_2 are the weight parameters of the two convolutional layers, σ and δ represent the sigmoid and ReLU activation functions, respectively, and $*$ denotes the convolution operation. Finally, as shown in Figure 2, the spatial weight w_{spa} is used to recalibrate the spatial information in the feature c_{t_i} and highlight the useful spatial information, using Equation (9).

$$c_{t_i} = w_{spa} \otimes c_{t_i}, \quad (9)$$

where \otimes represents element-wise multiplication.

Then, the different layer features are concatenated to get the multistage feature c_t at timestep t_i . Thus, the most appropriate and relevant representation is automatically learned and leverages the complementary information from the different feature levels, resulting in a more robust and informative representation.

2.3 | Classification

After mapping the patch representation, we use the same classification model from our previous work⁴ to predict the classification labels. Transformer-based classifiers are trained based on the inspiration from Ref. [33], as shown in Figure 1B. The classification module combines CNN and transformer structures to form an effective classifier. The classifier takes positionally embedded feature patches as inputs and employs an MLP head to predict the final classification scores. The cross-layer skip connections introduced in the classifier are inspired by the success of similar mechanisms in U-Net³⁴ and ResNet.³⁵ This aims to minimise the loss of valuable information during layer-wise propagation and enhance information flow between layers. The classifier model incorporates skip connections, multihead attention mechanisms, and feed-forward neural networks for spectral-spatial feature mapping, as well as a transformer structure for deep feature extraction, resulting in outstanding classification performance.

2.4 | Postprocessing

In our experiments, Weighted Majority Voting (WMV)³⁶ is applied as a postprocessing correction method following transformer-based classification to enhance the accuracy of the classified maps. WMV adjusts the classification of each pixel based on the labels of its surrounding neighbours, incorporating spatial context into the final classification. Unlike other postprocessing methods, such as median filtering, which tends to oversmooth boundaries,³⁷ or conditional random fields (CRFs), which increase computational complexity,³⁸ WMV offers a balanced approach. By considering the weighted votes of neighbouring pixels, WMV reduces noise and misclassifications, following the principle that neighbouring pixels are likely to belong to the same class. This decreases the likelihood of isolated misclassifications and enhances the reliability of the results. The use of WMV significantly improves the overall classification performance, providing more accurate maps.

3 | EXPERIMENTS AND RESULTS

3.1 | Data preprocessing

The proposed framework was evaluated on seven HS images from five patients with GBM tumours, as initially established by Ref. [13]. For this study, we have only chosen those HSI(7) that contain all four labels in their ground truth information Normal Tissue (NT), Tumour Tissue (TT), hypervascularised tissue or blood vessels (HV), and

Background (BG). The HSI sensor employed had a spectral range of 400×1000 nm and utilised a pushbroom A-Series camera (Headwall Photonics Inc., Fitchburg, MA, USA). This camera could capture a total of 826 spectral bands using a 150-W QTH (quartz-tungsten-halogen) lamp. As a line scanner, it was capable of capturing a maximum of 1004×826 pixels, depending on the sample size. To mitigate the heat generated by the light bulbs, a cold light emitter was used to prevent any adverse effects on the brain due to elevated temperatures.

During surgery, the neurosurgeon utilised preoperative imaging data to distinguish normal tissue from tumour tissue by placing cylindrical rubber markers. To validate the accuracy of these markings, biopsies of the tissue within the markers were taken to confirm the GBM tumour type and grade. To detect tumours in deeper layers, HSI data was collected after the commencement of tumour resection. The Spectral Angle Mapper (SAM) algorithm was used to label the data acquired by HSI, with each labelling performed by the operating neurosurgeon to create ground truth for every patient. Typically, each patient had a set of five unique labels: unlabelled pixels, normal tissue, tumour tissue, blood vessels, and background. Then, the raw HS images were preprocessed. This preprocessing chain is based on two main steps: Selective Class Merging and Dimensionality Reduction. The goal of each preprocessing step is briefly explained:

- **Selective Class Merging.** The background class includes materials or substances present in the surgical scenario that are not relevant to the tumour resection procedure, such as skull bone, skin, cylindrical rubber markers, or surgical material. These background pixels are marked as unclassified as they are not useful in the process. From these preprocessed cubes, a specific set of pixels comprising three different classes tumour tissue, normal tissue, and hypervascularised tissue (blood vessels) is obtained. This set of pixels is used to train and test the model.
- **Dimensionality Reduction.** The original HS images consist of 826 bands, but due to hardware limitations and redundancy in HSI, it is not feasible to process all bands. Therefore, a dimensionality reduction algorithm is applied. Numerous algorithms for dimensionality reduction have been developed in the literature.³⁹ Principal Component Analysis (PCA)⁴⁰ is one of the most popular linear techniques for this purpose, as it maps the data while preserving as much variance as possible. In the proposed brain cancer detection algorithm, PCA is employed to obtain a preprocessed HS cube with 200 bands. No further preprocessing is applied, as diffusion-based methods could theoretically recover the details of dark areas better through their powerful generative

TABLE 1 Patient data before and after selective class merging preprocessing. Summary of seven patient IDs and their pixel-wise distribution among each class (Normal Tissue (NT), Tumour Tissue (TT), Hypervascularised Tissue (HV), and Background (BG)).

Patient ID	Before preprocessing				After preprocessing		
	NT	TT	HV	BG	NT	TT	HV
8-01	2295	1221	1331	630	2295	1221	1331
8-02	2187	138	1000	7444	2187	138	1000
12-01	4516	855	8697	1685	4516	855	8697
12-02	6553	3139	6041	8731	6553	3139	6041
15-01	1251	2046	4089	696	1251	2046	4089
20-01	1842	3655	1513	2625	1842	3655	1513
25-02	977	1221	907	2503	977	1221	907

ability.^{41,42} A list of these patients and their pixel information for each labelled data set is shown in Table 1 both before and after data preprocessing.

3.2 | Implementation details

We used the PyTorch framework to implement and train the MedDiffHSI model. The training was done on a basic hardware setup, which consists of a POWER8NVL production-grade CPU with 128 CPU threads spread across 2 sockets for efficient processing. Additionally, four NVIDIA Tesla P100 GPUs were used for enhanced graphical computations, each offering a memory of approximately 16 GB.

The diffusion model was optimised using the Adam optimiser and trained for 35,000 epochs and was trained only once on the Patient ID 12-02 data set. The learning rate was set to 1×10^{-4} , with a batch size of 64 and a patch size of 32×32 . For classification, we extracted features from this trained model for other data sets as well. The features were combined using a multistage attention module Network. Due to hardware limitations, it is not feasible to utilise all combined features through a multistage attention module network. To evaluate the amount of spectral information preserved, PCA was employed. The number of PCA components was restricted to a maximum of $D/10$ (where D represents the diffusion features in the data set) to align with the hardware capacity. The classification model was also trained using the Adam optimiser, maintaining the same learning rate of 1×10^{-4} and a batch size of 64 for all data sets. The size of the feature patch was empirically set to 11×11 . The number of epochs for the classification model was set to 750 for all data sets. For classification, seven different samples were passed to the proposed model, and the experiment was divided into sep-

arate Train/Test sets. Each data set was divided into 5% for training and 95% for the test set.

3.3 | Performance evaluation

The classification performance was evaluated using Recall (10), Precision (11), Accuracy (OA) (12), Specificity (13), and Cohen's Kappa metrics, where TP are true positives, TN are true negatives, FN are false negatives, and FP are false positives. Cohen's Kappa was calculated as in (14), where:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (10)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (11)$$

$$\text{OA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \quad (12)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \quad (13)$$

Cohen's Kappa (κ) is a statistical measure of interrater agreement or reliability, which considers the agreement occurring by chance. It is calculated as follows:

$$\kappa = \frac{P_o - P_e}{1 - P_e}, \quad (14)$$

where P_o is the observed agreement (overall accuracy) and P_e is the expected agreement, which is calculated based on the probabilities of each class occurring by chance.

The algorithms were tested using seven HS images from five human patients with GBM tumour. To comprehensively assess the effectiveness of the proposed method, we conducted a comparative analysis with several state-of-the-art deep learning-based approaches. The methods included in this comparison are the contextual CNN (CnCNN),⁴³ deep pyramidal residual networks (DPResnet),⁴⁴ high-resolution network + object-contextual representations (HRnet + OCR),^{45,46} double-branch dual-attention mechanism network (DBDA),⁴⁷ Fusing Multiple Deep Models (FMDM),²⁰ and baseline, DiffSpectralNet.⁴ These methods have demonstrated strong performance in previous studies. The parameters for all compared methods are consistent with those reported in the respective literature, and we directly used the results from FMDM.²⁰ The classification accuracies obtained by the different deep learning-based methods are detailed in Table 2.

The proposed method achieved the highest overall accuracy, precision, and recall, underscoring its effectiveness. This is particularly significant in the context of the three-class classification problem, where the accurate identification of GBM tumour tissue is both crucial and

TABLE 2 Classification results obtained by various deep learning-based methods (in %). NT: Normal Tissue, TT: Tumour Tissue, HV: Hypervascularised Tissue. The results for CnCnn, DPRsenet, HRnet+OCR, DBDA, and FMDM are taken directly from FMDM.²⁰ The best results are highlighted in bold.

Tissue	Metrics	CnCnn	DPRsenet	HRnet+OCR	DBDA	FMDM	DiffSpectralNet	MedDiffHSI
NT	Accuracy	87.59	97.19	86.83	92.92	96.61	91.11	97.37
	Precision	66.48	92.09	64.77	81.82	88.84	94.23	97.45
	Recall	95.87	98.24	94.99	85.02	95.47	96.01	98.39
TT	Accuracy	88.99	90.21	89.46	94.27	96.34	87.92	98.82
	Precision	86.78	40.30	75.02	86.16	87.76	95.41	99.12
	Recall	58.99	35.47	60.23	71.32	90.53	92.47	97.35
HV	Accuracy	94.32	90.14	96.65	91.77	95.55	90.18	97.06
	Precision	92.37	64.99	97.81	83.10	94.77	96.74	97.43
	Recall	79.67	67.78	91.69	81.89	95.53	99.28	98.50
Overall	Accuracy	90.3	92.51	90.98	92.99	96.16	89.74	98.04
	Precision	91.67	65.79	79.2	83.69	90.45	95.46	97.99
	Recall	81.87	67.16	82.30	79.41	93.84	93.73	98.08

TABLE 3 Performance metrics for different analyses.

Analysis	8-01	8-02	12-01	12-02	15-01	20-01	25-02
Specificity	97.13	99.84	99.28	97.50	98.94	98.42	99.01
k	91.14	99.10	98.22	96.28	96.83	98.43	98.44

challenging. Our method outperformed the other comparison methods in all three metrics for GBM tumour identification. The model achieved an overall accuracy of 98.04%, which is 8% higher than the accuracy achieved with the baseline DiffSpectralNet model. This improvement highlights the efficacy of the multistage attention module in enhancing and utilising the diffusion features. Additionally, the proposed model demonstrated superior accuracy in classifying tumour tissue (98.82%) compared to healthy tissue and blood vessels, which had accuracies of 97.37% and 97.06%, respectively. This further validates the robustness of our method in distinguishing between different types of tissue.

Furthermore, we evaluated the model performance using Specificity, and Kappa metrics. As shown in Table 3, the proposed model demonstrates high specificity across the data set, with values consistently above 97%. These high specificity values indicate that the model is effective in distinguishing tumour tissues from nontumour tissues. This is particularly important in a clinical setting, where accurately identifying tumour boundaries can significantly impact the success of surgical interventions. Additionally, the Kappa coefficient indicates the model consistency and reliability in classification tasks, showing strong agreement with ground truth annotations. Overall, the proposed method demonstrates state-of-the-art performance in terms of specificity, and Kappa coefficient,

highlighting its potential for clinical application in tumour identification and surgical planning.

For a comprehensive examination of the detailed settings for band selection (Table S1), train/test split (Table S2), and the performance metrics for each class across all data sets (Table S3), readers are referred to the Supplementary Material provided.

Figure 3 shows the classification maps produced by the proposed method. From top to bottom, the three figures in each row represent the synthetic RGB images, ground truth after data preprocessing, and classification map with the proposed MedDiffHSI for the testing hyperspectral images with IDs 08-01, 08-02, 12-01, 12-02, 15-01, 20-01, and 25-02, respectively. Our model shows robustness to varying illumination conditions. For example, in the testing HSI with ID 08-01, some areas are dark due to poor illumination conditions, yet the model still demonstrates satisfactory performance due to its powerful ability to extract both low and high-level features. This robustness further underscores the model potential for reliable application in diverse clinical environments.

4 | DISCUSSION

- **Model Stability across Data sets.** To thoroughly assess the robustness of our proposed framework, we extended our evaluation beyond brain tumour classification to include additional HSI data sets for breast cancer, as detailed in Table 4. By applying our model to these varied data sets, we aimed to demonstrate its versatility and effectiveness in handling diverse types of cancerous tissues. The results from our extended evaluation, as shown in Table 5, which are averages of three classes,

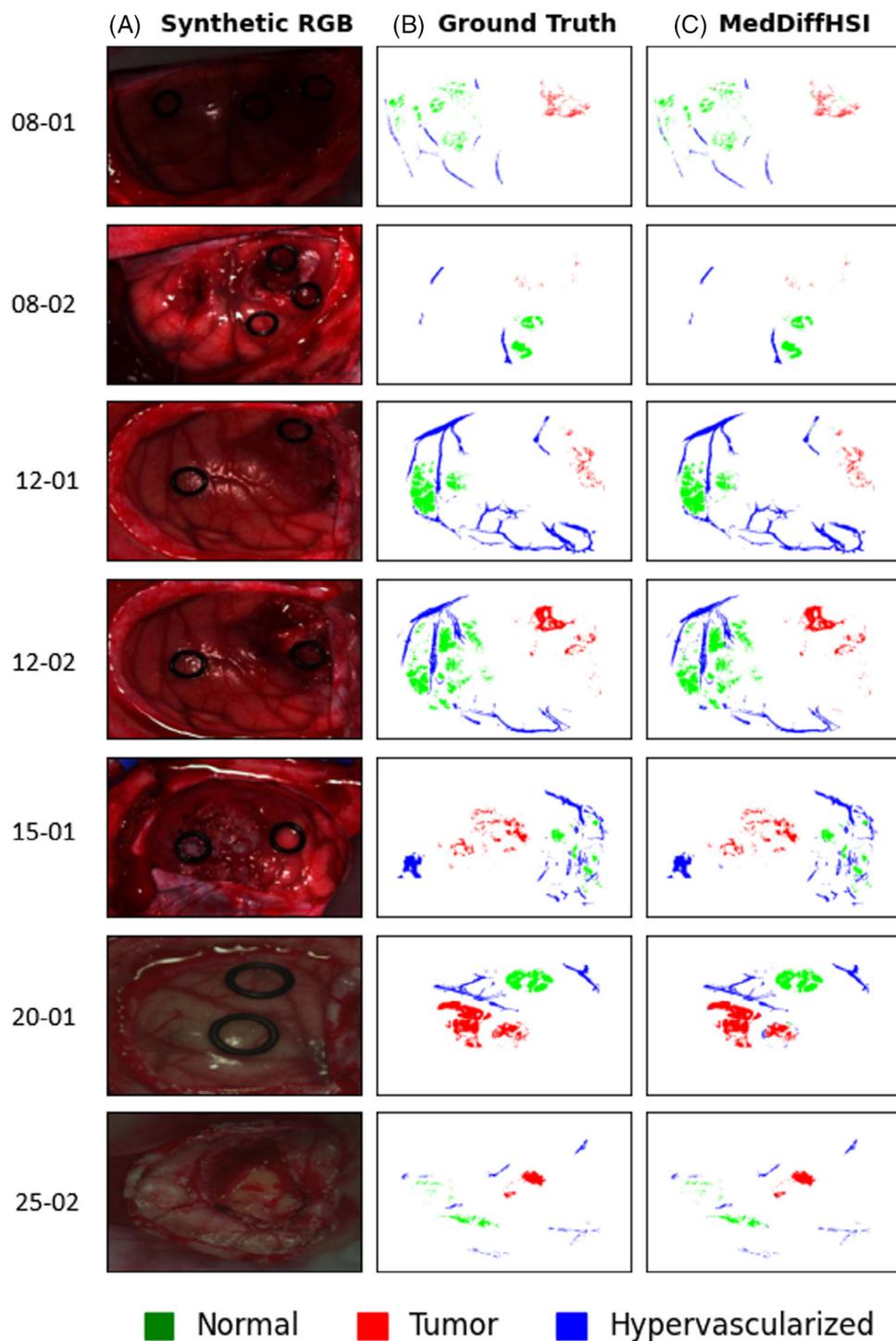


FIGURE 3 Classification maps produced by the proposed method. (A) Synthetic RGB images corresponding to test HSIs. (B) Ground truth corresponding to test HSIs after data preprocessing, with green, red, and blue colours representing normal tissue, tumour tissue, and hypervascularised tissue, respectively. (C) Classification maps obtained by MedDiffHSI.

indicate that our model maintains good performance across different analyses. The cross-data set evaluation highlights the robustness and adaptability of our proposed framework, affirming its potential for application in medical diagnostics.

- **Effects of Data Preprocessing.** The lack of dense ground truth and the absence of clinical evaluations on the tumour boundaries are significant constraints. Additionally, the nonuniform distribution in some cases may lead to high background sensitivity, resulting in a loss of

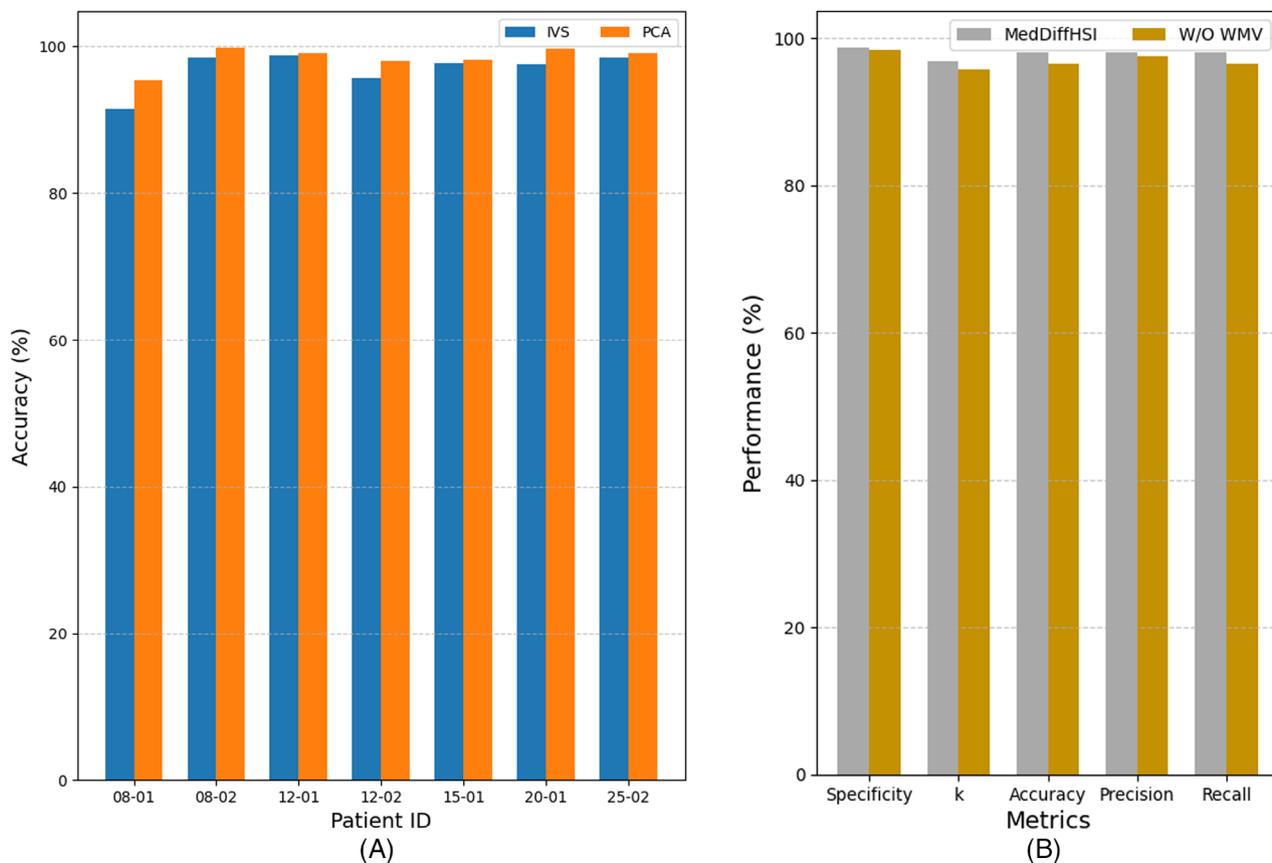


FIGURE 4 Effects of different components on the MedDiffHSI model: (A) Accuracy results on the seven HSI data sets using IVS⁴⁸ and PCA for efficient band selection. (B) Comparison of performance using different accuracy metrics for the MedDiffHSI model without the WMV component.

TABLE 4 Patient data before and after selective class merging preprocessing. Summary of 2 patient IDs and their pixel-wise distribution among each class.

Patient ID	Before preprocessing				After preprocessing		
	NT	TT	HV	BG	NT	TT	HV
21-01	3405	167	793	5330	3405	167	793
21-02	2353	31	555	2137	2353	31	555

generalisation.¹⁹ To address this, we merged the background class with unclassified pixels using a method we call selective class merging, which we explained in Section 3.1.

Moreover, effective band selection is crucial for HSI classification to retain the maximum amount of information. To achieve this, we employed two distinct meth-

ods: the approach detailed in IVS,⁴⁸ which reduces the bands to 128, and PCA, which retains only 200 bands due to hardware limitations. We compared these methods to assess their efficacy in the proposed model, MedDiffHSI, for preserving essential information for accurate classification. As illustrated in Figure 4, PCA resulted in more accurate classification outcomes. On average, PCA resulted in a 1.55% increase in classification accuracy compared to the IVS⁴⁸ method.

- **Impact of WMV.** The Figure 4B shows a comparative analysis of the MedDiffHSI model performance without WMV component. Metrics such as Specificity, Kappa (k), Accuracy, Precision, and Recall were evaluated. WMV significantly improves performance across all metrics on average over seven HSI data sets: Specificity (+0.32%), Kappa (+1.23%), Accuracy (+1.50%), Precision (+0.34%), and Recall (+1.53%). By correcting

TABLE 5 Performance metrics for different analyses.

Analysis	Specificity (%)	k	Accuracy (%)	Precision (%)	Recall (%)
21-01	98.30	94.60	97.08	97.92	94.82
21-02	98.64	94.56	98.73	93.84	94.01

TABLE 6 Average classification performance for patient IDs 8-01, 8-02, 15-01, and 20-01 (in %). NT: Normal Tissue, TT: Tumour Tissue, HV: Hypervascularised Tissue. The best results are highlighted in bold.

Model	NT	TT	HV	Overall accuracy
EAGC	99.62	90.8	98.97	96.46
MedDiffHSI	98.1	98.18	98.03	98.10

pixels using neighbouring information and eliminating misclassification like noisy scattered points, WMV enhances the overall performance.

- Performance Analysis.** Additional performance analysis was conducted using the available results for Patient IDs 8-01, 8-02, 15-01, and 20-01. For comparison, the inpatient approach (which provided the best results) using Xtreme Gradient Booster (XGB), as recommended by the authors, from the EAGC⁴⁹ method was utilised. Table 6 presents the comparison with the EAGC⁴⁹ model, showing the average performance across the four (8-01, 8-02, 15-01, and 20-01) HS images. The MedDiffHSI model outperformed the EAGC method overall and showed consistent results across all classes, while the EAGC method was less effective in the tumour tissue. Additionally, it is worth mentioning that the EAGC method used 70% of the data for training, while the MedDiffHSI model used only 5%.

HSI is a noninvasive, nonionising technique already employed for brain tumour detection.^{19,20} This study presents new experiments conducted on an in vivo hyperspectral brain database.¹³ Models were trained and tested using data from GBM patients to classify tumour, healthy, and hypervascularised tissue. The proposed MedDiffHSI method shows promising results, achieving superior multiclass classification performance compared to state-of-the-art deep learning approaches. Specifically, we achieved an accuracy of 98.82%, a precision of 99.12%, and a recall of 97.35% for tumour tissue classification. These results are promising for identifying brain tumour margins to aid surgeons during resection and show improvements over the current state-of-the-art for multiclass classification of in vivo human-brain HS data sets. Furthermore, the classification model demonstrates very high specificity, consistently above 97%, as shown in Table 3, averaging across all classes for each data set. Therefore, we can conclude that the proposed technique excels in correctly classifying cases as disease-free, indicating high confidence in 'ruling in' cases of disease.⁵⁰ In brain cancer resection, an intraoperative guidance system must have very high specificity to ensure that resected areas are not normal brain tissue, which is crucial for better patient outcomes.

This study also has several limitations. Firstly, we directly used results from FMDM²⁰ for comparison with most deep learning approaches. FMDM employs four-class data, while we utilised three-class HSI data. We attempted a one-to-one comparison for each class; however, the differences in class configurations may affect the validity of these comparisons. This discrepancy should be considered when interpreting the results, as it might lead to variations in the outcomes. Another limitation is that most hyperparameters of the proposed method in classification are selected empirically. Future work will focus on developing automatic hyperparameter selection techniques. We plan to train and test the model on the entire data set rather than using a limited subset. Future experiments should use a much larger data set to enhance the model generalisation. This study has several limitations. Firstly, we used results from FMDM²⁰ for comparison with most deep learning approaches. FMDM utilises four-class data, whereas we utilised three-class HSI data. We attempted to make direct comparisons for each class; however, differences in class configurations may affect the validity of these comparisons. This discrepancy should be considered when interpreting the results, as it might lead to variations in the outcomes. Additionally, most hyperparameters in our proposed classification method were selected empirically. There is a need to develop automatic hyperparameter selection techniques. It is also necessary to train and test the model on the entire data set rather than a limited subset. Future experiments should incorporate a much larger data set to enhance the model generalisation and robustness.

5 | CONCLUSION

This study introduces MedDiffHSI, a novel method for brain tissue classification in GBM patients. Unlike traditional CNN models, our approach uses a spectral-spatial diffusion process to model complex relationships and extract both high-level and low-level features efficiently. By employing the diffusion model, we leverage the spatial-spectral neighbourhood structure of hyperspectral data, extracting deep features more effectively. We also utilise a spectral-spatial attention module and a transformer-based model with cross-layer skip connections to prevent information loss during propagation. Our method processes data in patches rather than on a pixel-by-pixel basis, improving detail capture for accurate classification. MedDiffHSI achieved state-of-the-art results in HSI classification in our quantitative trials. Future studies will focus on validating and enhancing this model with additional hyperspectral data sets, including mineralised tissues such as teeth and bones. This approach shows great potential

for generalising HSI classification by capturing complex inter-band relationships.

AUTHOR CONTRIBUTIONS

G.T. and T.N. initialised concepts and directions. N.S. conducted experiments and analysed results. T.N. and G.T. provided critical updates and suggestions that significantly enhanced the scope and direction of the research. N.S. wrote the paper with important input from G.T. and T.N. All authors, N.S., G.T., T.N. reviewed and approved the final manuscript.

ORCID

Tuan T. Nguyen  <https://orcid.org/0000-0003-0055-8218>

REFERENCES

- World Health Organization. (2024). *Global cancer burden growing amidst mounting need for services*. World Health Organization News.
- Halicek, M. T., Fabelo, H., Ortega, S., Callicó, G. M., & Fei, B. (2019). In-vivo and ex-vivo tissue analysis through hyperspectral imaging techniques: Revealing the invisible features of cancer. *Cancers, 11*.
- Walter, S., Stocker, S., Wagner, S., Stepp, H., Fritsch, C., Goetz, C., Goetz, A. E., Kieffmann, R., & Reulen, H. J. (1998). Intra-operative detection of malignant gliomas by 5-aminolevulinic acid-induced porphyrin fluorescence. *Neurosurgery, 42*(3), 518–25; discussion 525–6.
- Sigger, N., Vien, Q.-T., Nguyen, S. V., Tozzi, G., & Nguyen, T. T. (2024). Unveiling the potential of diffusion model-based framework with transformer for hyperspectral image classification. *Scientific Reports, 14*, 8438. <https://doi.org/10.1038/s41598-024-58125-4>
- Hong, D., Han, Z., Yao, J., Gao, L., Zhang, B., Plaza, A., & Chanussot, J. (2022). Spectralformer: Rethinking hyperspectral image classification with transformers. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1–15. <https://doi.org/10.1109/TGRS.2021.3130716>
- Shah, S. T. H., Qureshi, S. A., Rehman, A. U., Shah, S. A. H., & Hussain, J. (2021). Classification and segmentation models for hyperspectral imaging – An overview. In S. Y. Yayilgan, I. S. Bajwa, & F. Sanfilippo (Eds.), *Intelligent technologies and applications* (pp. 3–16). Cham: Springer International Publishing.
- Audebert, N., Le Saux, B., & Lefevre, S. (2019). Deep learning for classification of hyperspectral data: A comparative review. *IEEE Geoscience and Remote Sensing Magazine, 7*, 159–173. <https://doi.org/10.1109/MGRS.2019.2912563>
- Stuart, M. B., McGonigle, A. J. S., & Willmott, J. R. (2019). Hyperspectral imaging in environmental monitoring: A review of recent developments and technological advances in compact field deployable systems. *Sensors (Basel, Switzerland), 19*.
- Deepthi, Devassy, B. M., Sony, G., Nussbaum, P., & Thomas, T. (2021). Classification of forensic hyperspectral paper data using hybrid spectral similarity algorithms. *Journal of Chemometrics, 36*.
- de Cássia Mariotti, K., Ortiz, R. S., & Ferrão, M. F. (2023). Hyperspectral imaging in forensic science: An overview of major application areas. *Science & Justice, 63*, 387–395. <https://doi.org/10.1016/j.scijus.2023.04.003>
- Shimoni, M., Haelterman, R., & Perneel, C. (2019). Hypersectoral imaging for military and security applications: Combining myriad processing and sensing techniques. *IEEE Geoscience and Remote Sensing Magazine, 7*, 101–117. <https://doi.org/10.1109/MGRS.2019.2902525>
- Lu, G., & Fei, B. (2014). Medical hyperspectral imaging: A review. *Journal of Biomedical Optics, 19*, 010901. <https://doi.org/10.1117/1.JBO.19.1.010901>
- Fabelo, H., Ortega, S., Szolna, A., Bulters, D., Pineiro, J. F., Kabwama, S., J-O'Shanahan, A., Bulstrode, H., Bisshopp, S., Kiran, B. R., Ravi, D., Lazcano, R., Madroñal, D., Sosa, C., Espino, C., Marquez, M., de La Luz Plaza, M., Galán, R. C., Carrera, D., ... Sarmiento, R. (2019). In-vivo hyperspectral human brain image database for brain cancer detection. *IEEE Access, 7*, 39098–39116. <https://doi.org/10.1109/ACCESS.2019.2904788>
- Collins, T., Maktabi, M., Barberio, M., Bencteux, V., Jansen-Winkel, B., Chalopin, C., Marescaux, J., Hostettler, A., Diana, M., & Gockel, I. (2021). Automatic recognition of colon and esophagogastric cancer with machine learning and hyperspectral imaging. *Diagnostics, 11*. <https://doi.org/10.3390/diagnostics11101810>
- Liu, N., Guo, Y., Jiang, H., & Yi, W. (2020). Gastric cancer diagnosis using hyperspectral imaging with principal component analysis and spectral angle mapper. *Journal of Biomedical Optics, 25*, 066005. <https://doi.org/10.1117/1.JBO.25.6.066005>
- Ortega, S., Halicek, M., Fabelo, H., Guerra, R., Lopez, C., Lejaune, M., Godtliebsen, F., Callico, G. M., & Fei, B. (2020). Hyperspectral imaging and deep learning for the detection of breast cancer cells in digitized histological images. *Proceedings of SPIE – The International Society for Optical Engineering*, 11320.
- Jong, L.-J. S., Post, A. L., Veluponnar, D., Geldof, F., Sterenborg, H. J. C. M., Ruers, T. J. M., & Dashtbozorg, B. (2023). Tissue classification of breast cancer by hyperspectral unmixing. *Cancers, 15*. <https://doi.org/10.3390/cancers15102679>
- Halicek, M. T., Lu, G., Little, J. V., Wang, X., Patel, M., Griffith, C. C., El-Deiry, M. W., Chen, A. Y., & Fei, B. (2017). Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging. *Journal of Biomedical Optics, 22*.
- Fabelo, H., Halicek, M., Ortega, S., Shahedi, M., Szolna, A., Piñeiro, J. F., Sosa, C., O'Shanahan, A. J., Bisshopp, S., Espino, C., Márquez, M., Hernández, M., Carrera, D., Morera, J., Callico, G. M., Sarmiento, R., & Fei, B. (2019). Deep learning-based framework for in vivo identification of glioblastoma tumor using hyperspectral images of human brain. *Sensors, 19*. <https://doi.org/10.3390/s19040920>
- Hao, Q., Pei, Y., Zhou, R., Sun, B., Sun, J., Li, S., & Kang, X. (2021). Fusing multiple deep models for in vivo human brain hyperspectral image classification to identify glioblastoma tumor. *IEEE Transactions on Instrumentation and Measurement, 70*, 1–14. <https://doi.org/10.1109/TIM.2021.3117634>
- Fabelo, H., Ortega, S., Ravi, D., Kiran, B. R., Sosa, C., Bulters, D., Callicó, G. M., Bulstrode, H., Szolna, A., Piñeiro, J. F., Kabwama, S., Madroñal, D., Lazcano, R., J-O'Shanahan, A., Bisshopp, S., Hernández, M., Báez, A., Yang, G.-Z., Stanculescu, B., ... Sarmiento, R. (2018). Spatio-spectral classification of hyperspectral images for brain cancer detection during

- surgical operations. *PLoS ONE*, 13, e0193721. <https://doi.org/10.1371/journal.pone.0193721>
22. Manni, F., van der Sommen, F., Fabelo, H., Zinger, S., Shan, C., Edström, E., Elmi-Terander, A., Ortega, S., Callicó, G. M., & de Peter, H. N. (2020). Hyperspectral imaging for glioblastoma surgery: Improving tumor identification using a deep spectral-spatial approach. *Sensors*, 20. <https://doi.org/10.3390/s20236955>
 23. Kavitha, M., Gayathri, R., Polat, K., Alhudhaif, A., & Alenezi, F. (2022). Performance evaluation of deep E-CNN with integrated spatial-spectral features in hyperspectral image classification. *Measurement*, 191, 110760. <https://doi.org/10.1016/j.measurement.2022.110760>
 24. Wu, X., Hong, D., & Chanussot, J. (2022). Convolutional neural networks for multimodal remote sensing data classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–10. <https://doi.org/10.1109/TGRS.2021.3124913>
 25. Hong, D., Hong, D., Yao, J., Li, C., Meng, D., Yokoya, N., & Chanussot, J. (2023). Decoupled-and-coupled networks: Self-supervised hyperspectral image super-resolution with subpixel fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1–12. <https://doi.org/10.1109/TGRS.2023.3324497>
 26. Hong, D., Zhang, B., Li, H., Li, Y., Yao, J., Li, C., Werner, M., Chanussot, J., Zipf, A., & Zhu, X. X. (2023). Cross-city matters: A multimodal remote sensing benchmark dataset for cross-city semantic segmentation using high-resolution domain adaptation networks. *Remote Sensing of Environment*, 299, 113856.
 27. Hong, D., Zhang, B., Li, X., Li, Y., Li, C., Yao, J., Yokoya, N., Li, H., Ghamisi, P., Jia, X., Plaza, A., Gamba, P., Benediktsson, J. A., & Chanussot, J. (2024). Spectralgpt: Spectral remote sensing foundation model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46, 5227–5244. <https://doi.org/10.1109/tpami.2024.3362475>
 28. Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. In *NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems*. Curran Associates Inc.
 29. Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D. J., & Norouzi, M. (2021). Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4), 4713–4726.
 30. Liu, Y., Zhu, J., Feng, J., & Mu, C. (2023). A feature embedding network with multiscale attention for hyperspectral image classification. *Remote Sensing*, 15. <https://doi.org/10.3390/rs15133338>
 31. Chen, J., Wei, Y., & Xie, Y. (2023). Combining attention mechanism and feature selection module for real-time semantic segmentation. In *2023 4th International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, 334–337. <https://doi.org/10.1109/CVIDL58838.2023.10167275>
 32. Sigger, N., Al-Jawed, N., & Nguyen, T. (2022). Spatial-temporal autoencoder with attention network for video compression. In *International Conference on Image Analysis and Processing* (pp. 290–300). Springer.
 33. Hong, D., Han, Z., Yao, J., Gao, L., Zhang, B., Plaza, A., & Chanussot, J. (2022). Spectralformer: Rethinking hyperspectral image classification with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–15. <https://doi.org/10.1109/tgrs.2021.3130716>
 34. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In: N. Navab, J. Hornegger, W. Wells, & A. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science*, vol 9351. Cham: Springer.
 35. He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). Las Vegas, NV, USA.
 36. Littlestone, N., & Warmuth, M. (1994). The weighted majority algorithm. *Information and Computation*, 108, 212–261. <https://doi.org/10.1006/inco.1994.1009>
 37. Devarajan, G., Aatre, V., & Sridhar, C. (1990). Analysis of median filter. In *ACE '90. Proceedings of XVI Annual Convention and Exhibition of the IEEE* (pp. 274–276). India. <https://doi.org/10.1109/ACE.1990.762694>
 38. Lafferty, J. D., McCallum, A., & Pereira, F. C. N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01*, 282–289. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
 39. Vaddi, R., Kumar, B. P., Manoharan, P., Agilandeswari, L., & Sangeetha, V. (2024). Strategies for dimensionality reduction in hyperspectral remote sensing: A comprehensive overview. *The Egyptian Journal of Remote Sensing and Space Sciences*, 27, 82–92. <https://doi.org/10.1016/j.ejrs.2024.01.005>
 40. Goodenberger, M., & Jenkins, R. (2012). Genetics of adult glioma. *Cancer Genetics*, 205. <https://doi.org/10.1016/j.cancergen.2012.10.009>
 41. Nguyen, C. M., Chan, E. R., Bergman, A. W., & Wetzstein, G. (2023). Diffusion in the dark: A diffusion model for low-light text recognition. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (pp. 4134–4145). Waikoloa, HI, USA.
 42. Yi, X., Xu, H., Zhang, H., Tang, L., & Ma, J. (2023). Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 12268–12277). Paris, France.
 43. Lee, H., & Kwon, H. (2017). Going deeper with contextual CNN for hyperspectral image classification. *IEEE Transactions on Image Processing*, 26, 4843–4855. <https://doi.org/10.1109/TIP.2017.2725580>
 44. Paoletti, M. E., Haut, J. M., Fernandez-Beltran, R., Plaza, J., Plaza, A. J., & Pla, F. (2019). Deep pyramidal residual networks for spectral-spatial hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57, 740–754. <https://doi.org/10.1109/TGRS.2018.2860125>
 45. Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., & Xiao, B. (2021). Deep high-resolution representation learning for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43, 3349–3364. <https://doi.org/10.1109/TPAMI.2020.2983686>
 46. Yuan, Y., Chen, X., Chen, X., & Wang, J. (2021). Segmentation transformer: Object-contextual representations for semantic segmentation. In A. Vedaldi, H. Bischof, T. Brox, & J. M. Frahm (Eds.), *Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science*, vol 12351. Cham: Springer.

47. Shu, Z., Wang, Y., & Yu, Z. (2024). Dual attention transformer network for hyperspectral image classification. *Engineering Applications of Artificial Intelligence*, 127. <https://doi.org/10.1016/j.engappai.2023.107351>
48. Fabelo, H., Ortega, S., Lazcano, R., Madroñal, D., Callicó, G. M., Juárez, E., Salvador, R., Bulters, D., Bulstrode, H., Szolna, A., Piñeiro, J. F., Sosa, C., O'Shanahan, A. J., Bisshopp, S., Hernández, M., Morera, J., Ravi, D., Kiran, B. R., Vega, A., ... Sarmiento, R. (2018). An intraoperative visualization system using hyperspectral imaging to aid in brain tumor delineation. *Sensors*, 18. <https://doi.org/10.3390/s18020430>
49. Chavarría, J. N. M., Cruz-Guerrero, I. A., Fabelo, H., Ortega, S., Callico, G. M., & Delgado, D. U. C. (2023). Ensemble of artificial intelligence classifiers for in-vivo identification of glioblastoma tumours using hyperspectral images. In *2023 IEEE EMBS R9 Conference* (pp. 1–5). <https://doi.org/10.1109/IEEECONF60929.2023.10525309>
50. Parikh, R. S., Mathai, A., Parikh, S., Sekhar, G. C., & Thomas, R. (2008). Understanding and using sensitivity, specificity and

predictive values. *Indian Journal of Ophthalmology*, 56, 45–50.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Sigger, N., Nguyen, T. T., & Tozzi, G. (2024). Brain tissue classification in hyperspectral images using multistage diffusion features and transformer. *Journal of Microscopy*, 1–15. <https://doi.org/10.1111/jmi.13372>