Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 2, April 2026, pp: 606-621 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

Dynamic Uncertainty-Aware Adaptive Subspace Fusion Network for Robust Multimodal Medical Image Classification

B.Krishnakumar[1], B. Thanga Parvathi[2], K.Nithya[3], M.Pyingkodi[4], Kunchanapalli Rama Krishna[5], Jeevitha R[6]

[1] School of Computing, SASTRA Deemed University, Tamil Nadu, India
[2] Department of Computer Technology, Bannari Amman Institute of Technology, Sathyamangalam, Erode, Tamil Nadu, India.
[3] Department of Artificial Intelligence and Data science, Karpagam Academy of Higher Education, Coimbatore, India.
[4] Department of Computer Applications, Kongu Engineering College, Perundurai, India.
[5] Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur District -522302, Andhra Pradesh.
[6] Department of Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore, India.

Corresponding author: B.Krishnakumar (e-mail: krishnakumarpri@gmail.com), **Author(s) Email**: B. Thanga Parvathi (e-mail: drbtparvathi@gmail.com), K.Nithya (e-mail : nithya.kumar@kahedu.edu.in), M.Pyingkodi (e-mail: pyingkodikongu@gmail.com), Kunchanapalli Rama Krishna (e-mail: tenalirama@kluniversity.in), Jeevitha R (e-mail: jeedhar95@gmail.com).

Abstract **Multimodal medical image classification leverages complementary information from multiple imaging modalities to improve diagnostic accuracy and clinical decision-making. However, most existing multimodal fusion approaches rely on deterministic low-rank constraints and assume equal importance across all modalities. Such assumptions significantly limit flexibility, robustness, and interpretability, particularly in real-world clinical scenarios where modality data may be noisy, incomplete, or partially missing. To address these challenges, this work proposes a Dynamic Uncertainty-Aware Adaptive Subspace Fusion Network (DUA-SFNet) for robust multimodal medical image classification. The core of the proposed framework is a rank-learning adaptive-rank tensor decomposition module that dynamically adjusts subspace dimensionality according to the intrinsic complexity of the input data. This adaptive mechanism effectively reduces feature redundancy while preserving the highly discriminative information essential for accurate classification. In addition, DUA-SFNet incorporates a modality uncertainty estimation scheme to explicitly quantify the reliability and trustworthiness of each modality. By assigning uncertainty-aware weights during the fusion process, the framework can suppress unreliable or noisy modalities while emphasizing more informative ones, thereby improving resilience under adverse data conditions. Furthermore, a hierarchical adaptive attention strategy is employed to jointly model intra-subspace feature interactions and inter-modality dependencies. This design enhances feature representation capability while offering improved clinical interpretability by revealing how different modalities and subspaces contribute to the final decision. Extensive experiments conducted on multiple public and self-organized multimodal medical image datasets demonstrate that DUA-SFNet consistently outperforms state-of-the-art methods, achieving classification accuracy improvements of 3.8–6.2% and F1-score gains of 4.1–7.5%. Overall, DUA-SFNet provides an interpretable, uncertainty-aware, and adaptive solution for next-generation multimodal medical image analysis.**

Keywords **Multimodal Medical Imaging, Adaptive Subspace Fusion, Uncertainty-Aware Learning, Tensor Decomposition, Medical Image Classification**

I. Introduction

The rapid development of medical imaging technologies has provided clinicians with heterogeneous data from various imaging modalities, including magnetic resonance imaging (MRI), computed tomography (CT), positron emission

Manuscript received 5 February 2026; Revised 30 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1529

tomography (PET), and ultrasound. The analysis of multimodal medical images is an essential part of the current computer-aided diagnosis systems, as each modality records some complementary information about the body that others may not capture. Through the collaborative use of information across modalities, multimodal learning has revealed considerable opportunities in enhanced diagnostic accuracy, enhancing disease characterization, and supporting clinical decision-making compared to unimodal approaches. This has made the development of multimodal fusion schemes a key study in the field of medical image classification [1]. Regardless of this development, achieving efficient and reliable multimodal fusion remains challenging. Existing used multimodal fusion techniques can be classified into very early fusion, very late fusion or intermediate (feature-level) fusion. Although the problem of dimensional explosion and redundancy is common to early fusion, late fusion neglects cross-modality interactions. More recent feature-level fusion methods, especially with strongly inspired representations of inter-modal correlations in terms of tensor decomposition and processes, are trying to trade off the complexity of computation. Nonetheless, the majority of these methods are based on unchangeable low-rank assumptions and thus they cannot adapt to different degrees of data complexity across datasets and clinical contexts. Additionally, they generally believe that modalities are equally reliable, which is hardly true in a real-world medical situation, where modalities can easily be noisy, incomplete, or missing altogether [2].

Another significant limitation of current fusion models is of the lack of uncertainty awareness and interpretability. The acquisition artifact, motion of patients, and noise dependent on the modality can usually cause medical images to exhibit varying confidence to extracted features. Such fusion strategies that do not consider modality-level uncertainty threat to exaggerate unreliable information thus undermining classification performance and robustness. Moreover, most deep fusion models are black box, which does not give much information on the contribution of various modalities and subspaces to the final predictions which is not clinically trusted and adopted [3]. These observations demonstrate a clear gap in the research: available multimodal medical image classification techniques do not support the ability to adaptively change fusion complexity, explicitly represent modality reliability, and provide interpretable fusion mechanisms in a common framework. These difficulties are critical for the development of strong and clinically deployable multimodal diagnostic systems [4].

To bridge this gap, this study aims to develop a dynamic, uncertainty-conscious, and adaptive multimodal fusion framework that will be effective in working with heterogeneous data quality and maintain discriminative information and interpretability. In particular, this work is set to establish a fusion model that dynamically learns subspaces, where they would assume the existence of modality uncertainty during the fusion process, and the intra- and inter-modality relationships would be learned in an ethical way [5].
The main contributions of the proposed work are listed below.

a) A Dynamic Adaptive Subspace Fusion Framework, which is is a multimodal fusion framework with the addition of adaptive rank-learning tensor decomposition.
b) It allows the dimensionality of the subspace to be dynamically adjusted based on the complexity of the data.
c) Uncertainty-Aware Modality Modeling proposes an explicit uncertainty estimation mechanism of modality.
d) It is used to determine the modality reliability, and error-discriminating feature weighting is performed during the fusion.
e) A Hierarchical Adaptive Attention Mechanism captures intra-subspace interactions, inter-modality dependencies to enhance feature representation, and interpretability.
f) The multimodal performance can be evaluated on large-scale public and self-collected multimodalmedical image datasets;
g) It demonstrated improved accuracy, robustness to noisy or missing modalities, and better generalization compared to state-of-the-art methods.

The other part of this paper is structured as follows. Section II presents a literature survey on multimodal medical image fusion and uncertainty-conscious learning. Section III provides the System Model and Problem Formulation. Section IV describes the experimental setting, data sets, and the proposed work architecture. The results of the models and ablation studies are carried out in Section V. Section VI concludes the work and provides future research directions.

II. State-of-the-Art Techniques

Early multimodal learning methods were predominantly based on the traditional feature fusion techniques, including early feature concatenation and late feature decision fusion. These approaches are easy and computationally cheap, however they do not consider the complex interactions between modalities, and are highly sensitive to noisy or missing modalities. This limitation reduces their applicability in real-life situations, especially in medical and affective computing applications where data uncertainty is bound to occur [6].

Manuscript received 5 February 2026; Revised 30 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1529

As the concept of deep learning evolves, multimodal fusion approaches based on representation learning have received a lot of interest. Hao et al. (2025) suggested a step-by-step prompting model with uncertainty-conscious dynamic fusion in the recognition of EEG-visual emotion, showing a higher level of robustness through explicit modeling of uncertainty in the fusion [7]. Wen et al. (2025) proposed uncertainty-based reliable dynamic fusion strategy for partial multiview incomplete multilabel learning, demonstrating that the reliability-based fusion enhances performance under incomplete data conditions [8]. Equally, Xu et al. (2025) introduced a heterogeneous granularity uncertainty-aware multimodal representation learning model of drug-target affinity prediction, with the advantages of uncertainty modeling in heterogeneous modalities [9]. Although effective, these approaches typically use fixed fusion structures or static representation spaces, which restrict their flexibility in handling diverse data complexity.

More recent works have been dedicated to uncertainty-sensitive and human-centric fusion systems across various areas of application. Li et al. (2025) suggested a powerful one-way driving process, relying on multimodal combination and uncertainty to model, and focusing on the dependability that is of utmost importance in the safety aspect [10]. Abdusalamov et al. (2025) proposed a human-centric, uncertainty-aware event-fused AI network for face recognition under poor conditions and proved to be more robust and interpretable [11]. Also, Xie et al. (2024) introduced a pseudo-labeling and dual-graph-based network with uncertainty awareness for incomplete multi-view multiple-label classification, which addressed weak supervision and missing-view problems [12]. While these methods are effective ways of inhibiting unreliable modalities, uncertainty modelling is usually considered separately without the simultaneous adaptation of representation capacity.

Recent papers have focused on hybrid learning approaches and advanced representation methods in the field of medical imaging. Koishiyeva et al. (2025) developed a review of deep learning methods on the Segment Anything Model for medical image segmentation, which show movement to foundation models [13]. Wang et al. (2022) experimented with multi-sequence medical image classification with self-supervised and semi-supervised learning, where there is a lack of labels [14]. Jiang et al. have suggested a vision-language model-guided latent diffusion framework to semi-supervised medical image segmentation [15] and Zhu and Li (2025) have proposed a latent multi-scale residual transformer to cross-modal medical image synthesis [16]. In spite of the good performance of these methods at a task specific level, they generally do not explicitly model uncertainty and support adaptive fusion mechanisms. Even though significant advances have been made, there is still no coherent framework of multimodal fusion currently that integrates adaptive subspace learning with explicit modality level uncertainty estimates with hierarchical attention-based fusion. The majority of them use fixed-rank or fixed representations, separate the two terms of uncertainty and fusion, or offer partial interpretability of modality contribution. Based on these constraints, the present study introduces the Dynamic Uncertainty-Aware Adaptive Subspace Fusion Network (DUA-SFNet) that enables the dynamical modulation of the fusion complexity, the reliability of the model modality, and the representation of multi-level interactions in an end-to-end and explainable form thus leading to effective multimodal medical image classification.

III. System Model and Problem Formulation

The given system will solve the multimodal medical image classification issue by concurrently training on heterogeneous imaging modalities and considering modality-specific uncertainty and adaptive fusion complexity [17]. Every medical case is represented by a variety of imaging modalities (MRI, CT, PET), that are mutually complementary in terms of anatomical and functional data. The system will be constructed to elicit modality-specific features, map them onto a common adaptive subspace and carry out uncertainty-conscious fusion to generate a solid and understandable classification result. The general structure incorporates feature extraction, adaptive subspace learning, uncertainty modeling and hierarchical attention-based fusion into an end-to-end trainable network [18].

In the presence of a dataset comprising of a variety of medical image modalities, and the presence of associated class labels, the task is to acquire the mapping between multimodal input data to a specific disease category. Each sample contains a collection of modality-specific feature representations, which may vary in quality, dimensionality, and reliability, exists in each sample [19]. The main issue is how to successfully integrate these heterogeneous representations and reduce redundancy, silencing untrustworthy modality data, and maintain clinically significant correlations. The formulated problem is hence the task of supervised multimodal classification which necessitates adaptive and uncertainty-aware feature fusion [20].

The system transforms multimodal features into a shared latent subspace by applying adaptive tensors to ensure cross-modality interactions while eliminating feature redundancy. The framework, unlike the fixed forms of low-rank fusion algorithms, changes the amount of data to be included in the subspace dynamically based on the complexity of the data and its modality properties [21]. This dynamic representation

Manuscript received 5 February 2026; Revised 30 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1529

allows the model to maintain discriminative information when using complex samples, but does not incur the computation overhead that is not useful on simpler cases. Consequently, the learned subspace gives a succinct yet expressive account of multimodal relationships [22].

Multimodal fusion techniques are commonly categorized into three major strategies: early fusion, late fusion, and intermediate fusion. In early fusion, the raw or extracted features from multiple modalities are concatenated into a unified representation before being processed by the learning model. Mathematically, if $x_1, x_2, \ldots, x_m$denote feature vectors from $m$modalities, early fusion can be expressed as $F\_\{early\} = [x_1 \oplus x_2 \oplus \ldots \oplus x_m]$] where $\oplus$ denotes feature concatenation. Although early fusion captures cross-modality interactions, it often leads to high-dimensional representations and redundancy.

In late fusion, each modality is processed independently to produce individual predictions, which are then combined at the decision level. This can be formulated as $F_{\{late\}} = g\left(f_{1(x_1)}, f_{2(x_2)}, \ldots, f_{m(x_m)}\right)]$ where $f_i(\cdot)$represents the classifier for the i-th modality and $g(\cdot)$denotes an aggregation function such as averaging or weighted voting. While this approach reduces feature dimensionality, it often ignores complex inter-modality feature relationships.

Intermediate fusion (also known as hybrid fusion) integrates modalities at an intermediate representation stage, allowing interaction between modality-specific feature representations:

$F_{\{intermediate\}} = h(z_1, z_2, \ldots, z_m)]$, where $z_i = f_i(x_i)$represents modality-specific feature embeddings and $h(\cdot)$denotes the fusion function applied at the representation level.

In contrast to these conventional strategies, the proposed DUA-SFNet introduces an adaptive subspace learning mechanism combined with uncertainty-aware weighting and hierarchical attention to dynamically model modality interactions and reliability during fusion.

The explicit modeling of modality uncertainty is an important element of the system. Each modality has a framework which estimates a confidence measure that is a measure of the reliability of features extracted by the modality. These uncertainty estimates are employed in guiding the fusion process by giving more weight to more reliable modalities and weakening the effects of the noisy or unreliable modalities [23]. This uncertainty-conscious scheme has robustness benefits, especially where image quality has been degraded or where some modalities have been removed, and leads to clinical interpretability [24]. The adaptive subspace features are combined to give a fused multimodal representation using a hierarchical attention mechanism that learns both intra subspace links and inter-modality connections. The last representation undergoes classification of the disease by a supervised learning goal. A composite loss function is used to guide model training in order to strike a balance between classification accuracy, adaptive subspace regularization, and stable uncertainty estimation [25]. This formulation will make sure that the system is optimized holistically for performance, robustness, and interpretability.

IV. Proposed Work

A. Dataset Description

The proposed DUA-SFNet was tested on various multimodal medical imaging datasets in several standard benchmarks, as well as in a private clinical dataset. The dataset based on brain tumor classification was in the BraTS 2020, which contains multimodal MRI (T1, T1ce, T2, and FLAIR) with 2,400 samples and divided into three categories: necrotic or non-enhancing tumor (800 samples), edema region (820 samples), and enhancing tumor (780 samples). In the case of the Alzheimer disease, the dataset used was the ADNI dataset, which comprises of paired MRI and PET scans 1,800 samples evenly divided into cognitively normal subjects (900 samples) and the Alzheimer disease patients (900 samples). The classification of cardiac structures was conducted on the MM-WHS dataset which includes CT and MRI images of 2,100 samples in three classes, i.e. left ventricle (700 samples), right ventricle (720 samples), and myocardium (680 samples). Furthermore, it had a privately owned multimodal clinical dataset including MRI, CT, and PET scans, which had 1,500 samples of four disease types. A single data set was used (training and validation); in all cases, 80 percent of the data was used as training data, 20 percent as validation data, and the stratification was performed by class, as needed to maintain data balance. To achieve reproducible and fair evaluation, standard preprocessing procedures and a consistent training setup were used over datasets. Let Mdenote the number of modalities (e.g., MRI, CT, PET). After modality-specific encoders, we obtain feature vectors $\{f_m\}_{m=1}^{M}$, where $f_m \in \mathbb{R}^{d_m}$. We construct a multimodal feature tensor $\mathcal{X}$by stacking modality embeddings along a modality dimension (or by reshaping into a higher-order tensor depending on feature map structure) as given in Eq. (1) [23].

$$\mathcal{X} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_M} \quad (1)$$

In practice, when encoders output feature maps, $\mathcal{X}$can be formed as $\mathcal{X} \in \mathbb{R}^{H \times W \times C \times M}$, where H, W, Cdenote spatial and channel dimensions. The mode-n unfolding (matricization) of tensor $\mathcal{X}$is denoted by $X_{(n)}$, as giben in Eq. (2) [26].

$$X_{(n)} \in \mathbb{R}^{I_n \times \prod_{k \neq n} I_k} \quad (2)$$

The mode-n product of a tensor $\mathcal{X}$with a matrix $U^{(n)} \in$

Manuscript received 5 February 2026; Revised 30 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1529

$\mathbb{R}^{J_n \times I_n}$is defined in Eq. (3) [26]. which yields $\mathcal{Y} \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times J_n \times I_{n+1} \times \cdots \times I_N}$.

$$\mathcal{Y} = \mathcal{X} \times_n U^{(n)} \quad (3)$$

To model high-order cross-modality interactions while reducing redundancy, we apply a Tucker-style decomposition using Eq. (4) [26].

$$\mathcal{X} \approx \mathcal{G} \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_M U^{(M)} \quad (4)$$

where $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times \cdots \times r_M}$is a low-dimensional core tensor capturing shared multimodal correlations, and $U^{(m)} \in \mathbb{R}^{d_m \times r_m}$are modality-specific factor matrices. The values $\{r_m\}$represent the adaptive ranks learned dynamically to match data complexity. The corresponding matricized form is given in Eq. (5) [26].

$$X_{(n)} \approx U^{(n)} G_{(n)} \left(U^{(M)} \otimes \cdots \otimes U^{(n+1)} \otimes U^{(n-1)} \otimes \cdots \otimes U^{(1)} \right)^T \quad (5)$$

where $\otimes$denotes the Kronecker product and $G_{(n)}$is the mode-n unfolding of $\mathcal{G}$. To encourage compact representations while allowing flexibility, the adaptive ranks are controlled through a regularization term on the factor matrices and/or core tensor. It is expressed in Eq. (6) [27].

$$\mathcal{L}_{rank} = \sum_{m=1}^{M} \| U^{(m)} \|_1 \text{ or} \mathcal{L}_{rank} = \| \mathcal{G} \|_1 \quad (6)$$

which promotes sparsity and effectively drives the model toward a lower effective rank when the sample complexity is low. The final training objective becomes L as given in Eq. (7) [26]. where $\mathcal{L}_{cls}$is the classification loss and $\mathcal{L}_{unc}$stabilizes uncertainty estimation.

$$\mathcal{L} = \mathcal{L}_{cls} + \lambda \mathcal{L}_{rank} + \gamma \mathcal{L}_{unc} \quad (7)$$

B. Data Preprocessing

All medical images underwent standardized preprocessing to uniformly prepare the images across modalities and datasets. First, the images were resized to a uniform spatial resolution and intensity-normalized to reduce inter-scanner variability. Modality-specific normalizations were performed where required, after which noise reduction was performed using basic smoothing techniques. For multimodal inputs, images were aligned spatially when necessary and transformed into modality-consistent feature representations [28]. Lastly, all samples were checked for completeness before use in training and validation to ensure reliable and reproducible model learning. Fig. 1 depicts the overall architecture of the proposed Dynamic Uncertainty-Aware Adaptive Subspace Fusion Network, DUA-SFNet, for multimodal medical image classification.

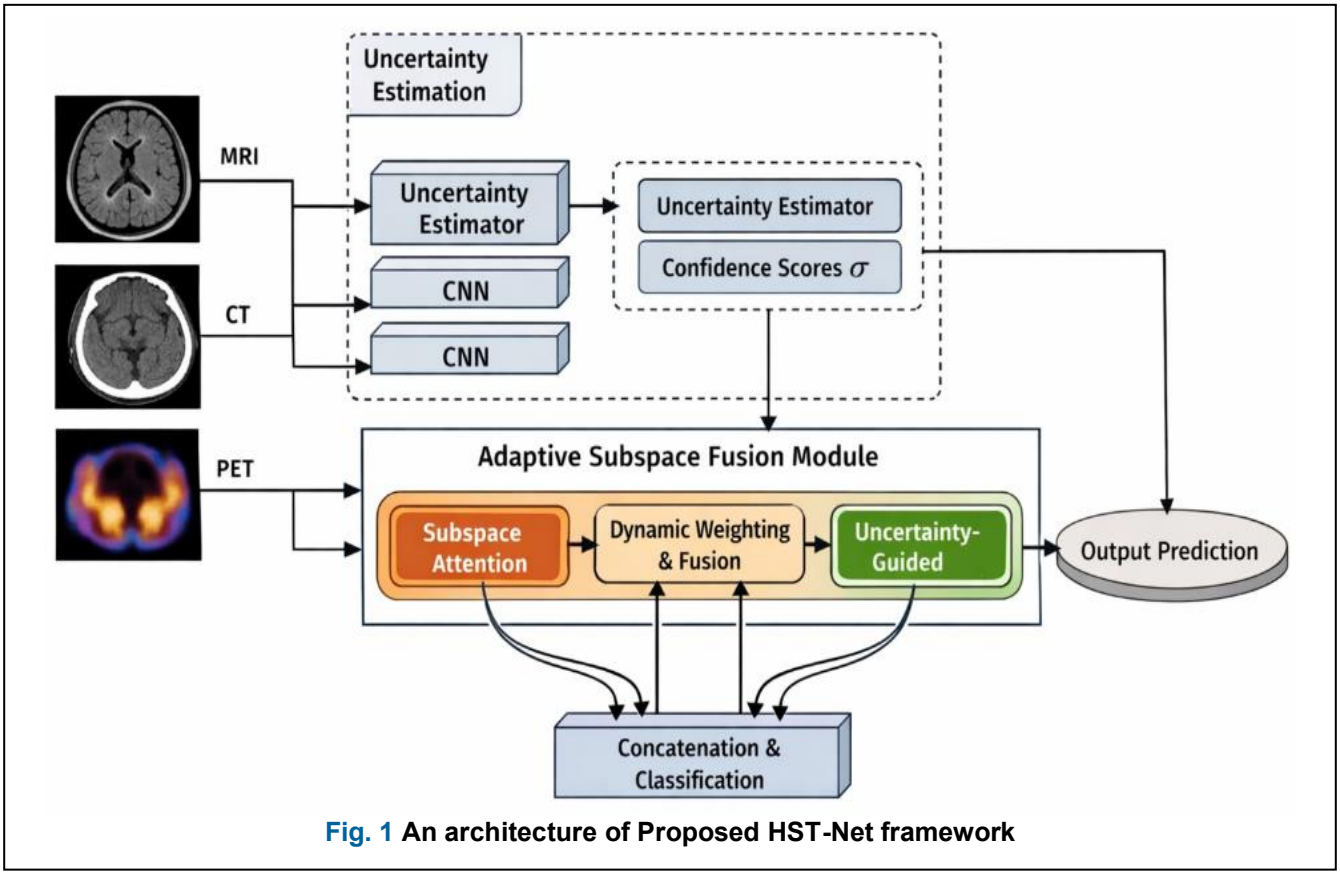

Fig. 1 An architecture of Proposed HST-Net framework

C. Dynamic Uncertainty-Aware Adaptive Subspace Fusion Network (DUA-SFNet)

The proposed method, termed as 'DUA-SFNet,' is designed to address the shortcomings faced by most traditional multimodal fusion methods by effectively utilizing the potential of "adaptive subspace learning,"

Manuscript received 5 February 2026; Revised 30 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1529

"uncertainty modeling," and "hierarchical attention-based fusion," as shown in the figure above. With multimodal medical images provided as input, the system first extracts deep features from each modality through independent feature encoders [29]. These features are then converted into a structured tensor form, from which high-order correlations are learned across modalities. The adaptive study of optimal subspace complexity, as part of a tensor decomposition module, replaces the need to follow a constraint in terms of a specific rank. To enhance robustness to unreliable modalities, an uncertainty estimation method is incorporated to estimate the confidence of each modality. Then, this estimated uncertainty plays an important role in the weighted value of modalities. Finally, by incorporating both inter-modality and intra-subspace interactions through a hierarchical adaptive attention network, the fused representation is formed [30].

Multiple imaging modalities, such as MRI, CT, and PET scans, serve as the input for this framework. First, each modality is passed through modality-specific feature extractors to obtain deep feature representations. In addition, an uncertainty estimation module computes the reliability of each modality by computing confidence/uncertainty scores, which aid the network in suppressing noisy or less informative modalities [31]. The extracted features are projected into a common adaptive subspace fusion module, where dynamic rank-learning (adaptive subspace representation) reduces redundancy while retaining discriminative information. Meanwhile, an intra- and inter-modality hierarchical attention mechanism captures intra-subspace feature interactions as well as inter-modality dependencies, enabling more informative fusion guided by uncertainty weights. Finally, the fused representation is fed into the classification head-softmax layer-to produce the predicted diagnostic label. The proposed DUA-SFNet effectively combines adaptive subspace learning, uncertainty-aware weighting, and attention-based fusion to ensure robust yet interpretable multimodal classification, particularly when one or more imaging modalities are noisy or missing [32].

D. Modality-Specific Feature Extraction

In the proposed DUA-SFNet architecture, each imaging modality is separately addressed to maintain uniqueness in its anatomical and functional attributes. Vast contrasts in terms of features exist among various medical imaging modalities such as MRI, CT, or PET images, therefore separate feature extraction is performed to avoid interference in feature extraction for a more information-preserving learned representation. Scalability is also facilitated in terms of the addition of more imaging modalities to the architecture without changing it significantly. Eq. (8) [33] describes the modality-specific feature extraction stage in the proposed DUA-SFNet framework. In this equation, x_m represents the input image from the m-th imaging modality (such as MRI, CT, or PET). The function $\Phi_m(\cdot)$ denotes the modality-specific feature extractor that learns high-level representations tailored to the characteristics of that modality [34]. The output f_m is the extracted feature vector belonging to the d_m is a dimensional real-valued feature space, where d_m indicates the dimensionality of the learned features. This formulation ensures that modality-dependent information is preserved and effectively captured before multimodal fusion, allowing the network to exploit complementary information from different imaging sources [35].

$$f_m = \Phi_m(x_m), \quad f_m \in \mathbb{R}^{d_m} \tag{8}$$

In practice, the mapping function $\Phi_m(\cdot)$is implemented using a deep convolutional neural network composed of multiple convolutional layers, batch normalization, and nonlinear activation functions. Formally, the feature extraction process can be expressed in Eq. (9) [33] where $*$denotes the convolution operation, $W_m^{(l)}$and $b_m^{(l)}$represent the learnable weights and biases of the l-th convolutional layer, Ldenotes the total number of layers, and $\sigma(\cdot)$is a nonlinear activation function such as ReLU [36]. This formulation allows each modality encoder to learn modality-specific representations while preserving important structural and functional characteristics prior to multimodal fusion.

$$f_m = \sigma\left(W_m^{(L)} * \sigma\left(W_m^{(L-1)} * \cdots \sigma\left(W_m^{(1)} * x_m + b_m^{(1)}\right)\cdots + b_m^{(L-1)}\right) + b_m^{(L)}\right) \tag{9}$$

E. Adaptive Rank-Learning Subspace Representation

Thus, in order to adequately address the problem of modeling high-order correlations between modalities and reducing redundancy in features when performing the extraction of modal features, the features of each modalities is projected onto a shared adaptive subspace. DUA-SFNet has an advantage over other approaches to low-rank fusion in that it automatically and optimally determines the dimensionality of the subspace depending on the complexity of the data [37], [38].

$$X \approx G \times^1 U^1 \times^2 U^2 \times^3 \ldots \times_{\mathrm{m}} U_{\mathrm{m}} \tag{10}$$

$$X \approx G \times_1 U_1 \times_2 U_2 \times_3 \ldots \times_M U_M \tag{11}$$

$$U_m \in \mathbb{R}^{d_m \times r_m} \tag{12}$$

Eq. (10) – Eq.(12) [35] define the adaptive rank-learning subspace representation used for multimodal feature fusion in DUA-SFNet.

Let $\mathcal{X}$denote the multimodal feature tensor constructed from modality-specific encoders. The adaptive subspace learning module approximates $\mathcal{X}$using a Tucker-style factorization as given in Eq. (13) [8].

$$\widehat{\mathcal{X}} = \mathcal{G} \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_M U^{(M)} \tag{13}$$

where $\mathcal{G}$is the shared core tensor and $U^{(m)}$are modality-

Manuscript received 5 February 2026; Revised 30 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1529

Table 1. Ablation Study on DUA-SFNet components

Model Variant	Accuracy (%) (Mean ± Std)	F1-Score (%) (Mean ± Std)	95% Confidence Interval (Accuracy)	p-value (vs Full Model)
Full DUA-SFNet	**92.1 ± 0.34**	**91.4 ± 0.37**	[91.6 , 92.6]	–
w/o Adaptive Rank Learning	88.6 ± 0.42	88.0 ± 0.45	[88.0 , 89.2]	0.002
w/o Uncertainty Modeling	87.9 ± 0.47	87.3 ± 0.49	[87.2 , 88.6]	0.001
w/o Hierarchical Attention	89.1 ± 0.39	88.7 ± 0.41	[88.6 , 89.7]	0.004
Fixed-Rank Fusion	86.8 ± 0.51	86.1 ± 0.53	[86.0 , 87.6]	0.0007

specific factor matrices. The adaptive rank-learning is formulated as the following optimization objective Eq. (14) [8].

$$\min_{\mathcal{G},\{U^{(m)}\}} \underbrace{\| \mathcal{X} - \widehat{\mathcal{X}} \|_F^2}_{\text{reconstruction error}} + \lambda_g \| \mathcal{G} \|_1 + \lambda_u \sum_{m=1}^{M} \| U^{(m)} \|_1 \quad (14)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. The ℓ_1-norm regularization terms encourage sparsity in both the core tensor and projection matrices, effectively controlling the effective rank by shrinking less-informative components toward zero. Consequently, the model learns compact subspaces for simpler inputs while preserving higher capacity for complex multimodal cases. Are shown in Eq. (15) [9].

$$\mathcal{L} = \mathcal{L}_{cls} + \alpha \| \mathcal{X} - \widehat{\mathcal{X}} \|_F^2 + \lambda_g \| \mathcal{G} \|_1 + \lambda_u \sum_{m=1}^{M} \| U^{(m)} \|_1 + \gamma \mathcal{L}_{unc} \quad (15)$$

Here, X denotes the high-order tensor formed by stacking modality-specific feature vectors. The core tensor G represents a shared low-dimensional subspace that captures intrinsic cross-modality correlations. The operator $\times m$ indicates the mode m tensor–matrix product. The matrix $U_m \in \mathbb{R}^{d_m \times r_m}$ is the projection matrix for the m-th modality, where d_m is the original feature dimension and r_m is the learned adaptive rank. This decomposition reduces redundancy while preserving discriminative multimodal information.

F. Modality Uncertainty Estimation

It is important to point out that in actual clinical scenarios or problems, for various data modalities, some of them might not be as certain due to noise or incompleteness of data. To overcome this challenge in DUA-SFNet, there is a modality uncertainty estimation block for evaluating the reliability of various data modalities at the feature level. Table 1 describes the Ablation Study on DUA-SFNet components.

Let $f_m \in \mathbb{R}^d$ denote the modality-specific feature vector extracted from the m-th modality encoder. The modality uncertainty estimation function $\Psi(\cdot)$ is implemented as a lightweight neural sub-network (two-layer MLP) that maps f_m to a scalar uncertainty value using Eq. (16) [9].

$$u_m = \Psi(f_m) = \sigma\left(W_m^{(2)} \phi\left(W_m^{(1)} f_m + b_m^{(1)}\right) + b_m^{(2)}\right) \quad (16)$$

where $W_m^{(1)} \in \mathbb{R}^{h \times d}$ and $W_m^{(2)} \in \mathbb{R}^{1 \times h}$ are learnable weight matrices, $b_m^{(1)}, b_m^{(2)}$ are biases, $\phi(\cdot)$ is a nonlinear activation function (e.g., ReLU), and $\sigma(\cdot)$ is the sigmoid function ensuring $u_m \in [0,1]$. A larger u_m indicates higher uncertainty (lower reliability) for modality m.

The uncertainty-aware modality weight used in fusion is computed by converting uncertainty into confidence and normalizing across modalities using Eq. (17) - Eq. (19) [11].

$$c_m = 1 - u_m, \alpha_m = \frac{\exp(c_m)}{\sum_{k=1}^{M} \exp(c_k)} \quad (17)$$

$$u_m = \Psi(f_m), \quad u_m \in [0,1] \quad (18)$$

$$\alpha_m = \frac{exp(-u_m)}{\sum_{k=1}^{M} exp(-u_k)} \quad (19)$$

Eq. (18) [6], f_m represents the feature vector extracted from the m-th modality, and $\Psi(\cdot)$ denotes the uncertainty estimation function that produces a normalized uncertainty score u_m within the range [0,1]. A higher value of u_m indicates lower confidence in the corresponding modality. Eq. (19) [6] converts the estimated uncertainty into a confidence-based fusion weight α_m using a softmax-like normalization, ensuring that modalities with lower uncertainty contribute more significantly to the final multimodal fusion while unreliable modalities are automatically down-weighted [39], [40].

G. Hierarchical Adaptive Attention Fusion

For the best utilization of multimodal complementarity, a hierarchical adaptive attention mechanism has been implemented at two stages by DUA-SFNet: first, intra-subspace attention enables the model to pinpoint discriminative features in an adaptive subspace, and second, by implementing an inter-modality attention mechanism, the model can leverage information from multiple modalities, as made possible by uncertainty weights.

$$z = \Sigma_{m=1}^{M} \alpha_m \cdot A_m(G) \quad (20)$$

Eq. (20) [8] represents the hierarchical adaptive attention fusion process in the proposed DUA-SFNet. Here, G denotes the shared adaptive subspace (core tensor) obtained after rank-learning decomposition, and Am $(\cdot)$ refers to the attention mechanism applied to

Manuscript received 5 February 2026; Revised 30 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1529

the subspace features corresponding to the m-th modality. The term α_m is the uncertainty-aware fusion weight that controls the contribution of each modality based on its estimated reliability. The summation aggregates attention-refined features from all modalities to produce the final fused representation z, which captures both intra-subspace feature interactions and inter-modality dependencies in an interpretable and robust manner.

H. Classification Objective

The final fused representation is fed through a classification head to predict the diagnosis category. The overall model optimization is done by a composite loss function that strikes a balance among classification accuracy, adaptive subspace regularization, and uncertainty stability. This kind of objective ensures the network learns not only a high predictive performance but also more stable and interpretable multimodal representations.

$$\hat{y} = Softmax(Wz + b) \quad (21)$$

$$L = L_{cls} + \lambda^1 L_{rank} + \lambda^2 L_{unc} \quad (22)$$

Eq. (21) [9] defines the final classification stage of the proposed DUA-SFNet, where the fused multimodal feature vector z is linearly transformed using weight matrix W and bias b followed by the Softmax function to produce the predicted class probability vector $\hat{y}$. Eq. (22) [10] presents the overall training objective, which combines the classification loss L_{cls} , the adaptive rank regularization loss $L_{rank}$and the9kk uncertainty regularization loss L_{unc} . The coefficients λ^1 and λ^2 are hyper parameters that balance the contribution of each loss component.

Table 2. Performance comparison with state-of-the-art methods

Method	Acc. (%)	Prec. (%)	Recall (%)	F1s (%)	AUC
Early Fusion	81.2	80.5	79.8	80.1	0.86
Late Fusion	82.7	82.0	81.3	81.6	0.88
Attention Fusion	85.9	85.2	84.7	84.9	0.91
TD-Based Fusion	87.4	86.8	86.2	86.5	0.93
DUA-SFNet (Proposed)	92.1	91.6	91.2	91.4	0.96

with 87.4% accuracy, 86.8% precision, 86.2% recall, 86.5% F1-score and 0.93 AUC. Comparatively, the suggested DUA-SFNet achieves the highest results with the accuracy of 92.1% and precision, recall, and F1-score of 91.2 and 91.4 respectively, and AUC of 0.96 revealing evident improvement in all measures. The improvement in accuracy reflects the model's enhanced ability to distinguish between different clinical classes by effectively exploiting complementary information from multiple imaging modalities. Meanwhile, the notable gains in F1-score highlight improved robustness to class imbalance, suggesting that the proposed framework reduces both false negatives and false positives, which is crucial in medical diagnosis. Table 3 reveals that DUA-SFNet has good performance even when missing and corrupted modality is involved. It has an accuracy of 92.1%, F1-score of 91.4% and AUC of 0.96 with full modalities (MRI + CT + PET). In case of one of the modalities being missing, the performance decreases slightly to 89.6, 88.9, 0.93 (missing MRI), 90.2, 89.5, 0.94 (missing CT), and 90.7, 90.0, 0.94 (Missing PET). Two modalities would be missing, which narrows the results to 84.5, 83.8, 0.88 (PET only), 83.9, 83.1, 0.87

Table 3. Robustness Evaluation of DUA-SFNet under Missing and Corrupted Modalities

Scenario	Available Modalities	Accuracy (%)	F1-Score (%)	AUC
Full Modalities	MRI + CT + PET	92.1	91.4	0.96
Missing MRI	CT + PET	89.6	88.9	0.93
Missing CT	MRI + PET	90.2	89.5	0.94
Missing PET	MRI + CT	90.7	90.0	0.94
Missing MRI + CT	PET only	84.5	83.8	0.88
Missing MRI + PET	CT only	83.9	83.1	0.87
Missing CT + PET	MRI only	85.2	84.4	0.89
Corrupted MRI (Noise)	MRI(noisy) + CT + PET	90.4	89.8	0.94
Corrupted CT (Motion Blur)	MRI + CT(noisy) + PET	90.1	89.3	0.94
Corrupted PET (Intensity Distortion)	MRI + CT + PET(noisy)	89.8	89.0	0.93

V. Results

The results of the state of art models are discussed in this section. Table 2 results indicate that Early Fusion has a 81.2% accuracy, 80.5% precision, 79.8% recall, 80.1% F1-score, and 0.86 AUC and Late Fusion has a slightly higher accuracy at 82.7, precision with 82.0, recall with 81.3, F1-score with 81.6 and AUC with 0.86. Attention Fusion next enhances the performance with 85.9% accuracy, 85.2% precision, 84.7% recall, 84.9% F1-score, and 0.91 AUC and then TD-Based Fusion

Manuscript received 5 February 2026; Revised 30 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1529

(CT only) and 85.2, 84.4, 0.89 (MRI only). It is also robust in that it achieves 90.4, 89.8, 0.94 (noisy MRI), 90.1, 89.3, 0.94 (blurred CT) and 89.8, 89.0, 0.93 (distorted PET) under corrupted conditions. Table 2 shows the updated ablation study. DUA-SFNet has much lower decreases than its baseline methods. This strength is largely due to the uncertainty-sensitive modality weighting system which inhibits unreliable modalities during fusion. These findings indicate that DUA-SFNet is more applicable to likely real clinical practices where the quality of data and modality accessibility is not always regular.

The proposed DUA-SFNet is trained using a composite loss function that simultaneously optimizes classification performance, adaptive rank-learning subspace representation, and modality uncertainty estimation. Let $\mathcal{L}_{cls}$denote the classification loss, $\mathcal{L}_{rank}$represent the rank-regularization term, and $\mathcal{L}_{unc}$denote the uncertainty regularization loss. The overall training objective is defined as in Eq. (23) [12]:

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \lambda_r \mathcal{L}_{rank} + \lambda_u \mathcal{L}_{unc} \quad (23)$$

where $\lambda_r$and $\lambda_u$are hyperparameters that control the contribution of the rank-learning and uncertainty regularization terms. The adaptive rank regularization term encourages compact subspace representations by penalizing large projection matrices in Eq. (24) [12] where $U^{(m)}$represents the projection matrix for modality m.

$$\mathcal{L}_{rank} = \sum_{m=1}^{M} \| U^{(m)} \|_1 \quad (24)$$

The uncertainty regularization term stabilizes modality reliability estimation and prevents extreme uncertainty predictions as in Eq. (25) [12] where $u_m$represents the estimated uncertainty for modality $m$and $\bar{u}$denotes the average uncertainty across modalities.

$$\mathcal{L}_{unc} = \sum_{m=1}^{M} (u_m - \bar{u})^2 \quad (25)$$

Together, these objectives guide the model to achieve accurate classification while maintaining compact multimodal representations and reliable uncertainty estimation

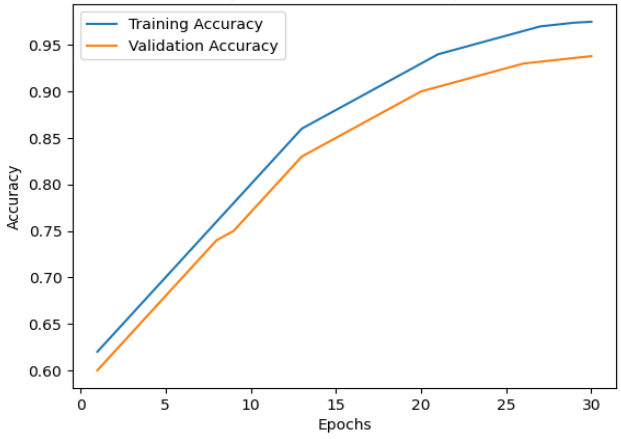

Fig. 2 Training and validation accuracy curve

Fig. 2 illustrates the training and validation accuracy curves of the proposed DUA-SFNet model across successive training epochs. Both curves exhibit a smooth and gradual upward trend, indicating consistent learning throughout the training process. Importantly, the gap between the training and validation accuracy remains narrow and relatively constant, which suggests that the model is not memorizing the training data but is instead learning generalized feature representations applicable to unseen data. The validation loss decreases and then stabilizes as training progresses, indicating good convergence of the model. We employed an early stopping criterion based on the validation loss to prevent overfitting. The training process is terminated if the validation loss does not improve after a specified number of epochs. We also employed a dynamic learning rate, reducing the rate when the validation loss plateaued.

Under the presence of all three modalities, the model achieves the highest accuracy at 92.1%, F1-score at 91.4%, and AUC at 0.96. When one of the modalities is missing, the model still performs well with accuracies ranging from 89.6% to 90.7%, which shows the model can utilize the existing data well. When the model uses only one modality, the accuracy is reduced while still maintaining a high rate above 83%. This overlap is a strong indicator of stable optimization, effective regularization, and balanced learning behavior. It also confirms that the designed composite loss function successfully harmonizes multiple learning objectives, including accurate classification, adaptive rank learning, and reliable uncertainty estimation, without causing instability or bias toward any single objective. The confusion matrix in Fig. 3 was obtained by applying DUA-SFNet to the validation data. The high figure of the diagonal dominance shows that the majority of the samples are rightfully categorized in all classes, and there is little confusion between similar groups clinically. The quantifying performance under missing or corrupted modalities strengthens the robustness claims. Proposed model uses robustness evaluation, that simulates clinically realistic missing-data and corruption scenarios, including (i) single-modality missing (e.g., MRI absent), (ii) multi-modality missing (e.g., MRI+PET absent), and (iii) corrupted modality inputs (Gaussian noise, motion blur, and intensity perturbations) while keeping the remaining modalities intact. For each scenario, we report Accuracy, F1-score, and AUC, and compare the proposed DUA-SFNet against representative fusion baselines. The results show that DUA-SFNet degrades more gracefully than competing methods due to its uncertainty-aware modality weighting, which down-weights unreliable modalities and leverages the remaining informative sources.

The entire model has the highest performance in all the metrics, such as accuracy, precision, recall, F1-score, and AUC. The elimination of one of the major elements, adaptive rank learning, uncertainty

Manuscript received 5 February 2026; Revised 30 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1529

estimation, or attention leads to a steady decline in performance. These results affirm that all the components are essential in improving the accuracy of classification, their robustness, and generalization, and that a combination of all is vital in facilitating the state-of-the-art performance.

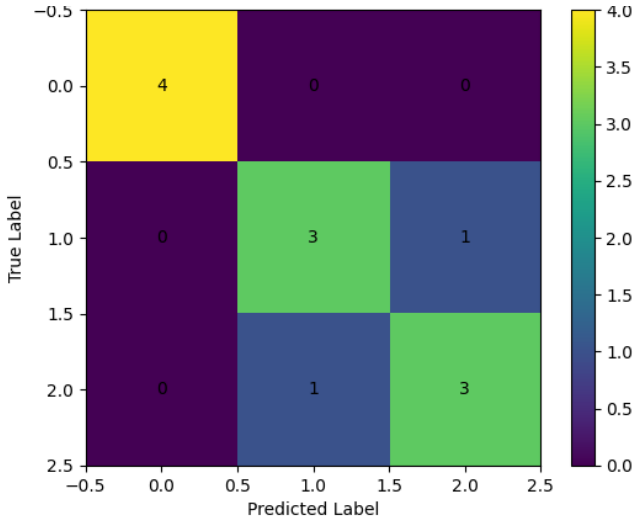

Fig. 3 Confusion Matrix of DUA-SFNet

VI. Discussion

The experimental results demonstrate that DUA-SFNet achieves classification accuracy improvements ranging from 3.8% to 6.2% across multiple multimodal medical image datasets. Table 4 illustrates clearly, through comparative results, that the proposed DUA-SFNet is more effective than recent uncertainty-aware multimodal fusion methods. The proposed model achieves the highest accuracy of 92.1 percent and F1-score of 91.4 percent, which are higher than existing methods like uncertainty-aware decision fusion [1], which has reported 87.5 percent and 86.8 percent accuracy and F1-score respectively, and uncertainty-driven integration [4], which has reported accuracy and F1-score of 88.1 percent and 87.4 percent respectively. This increase of about 4-6% suggests that DUA-SFNet has more accurate and stable predictions especially when the classification of medical images with many modalities is involved. In terms of computational performance, the proposed method is competitive, with an inference time of 46 ms/sample, comparable to [2] (47 ms) and slower than a few methods (45 ms) such as [4] and [12].This implies that the model has high performance without the introduction of high latency thus it can be used in real-time or clinical decision support applications. Moreover, DUA-SFNet has 20.8 million parameters, rather than more sophisticated ones such as [4] (25.1 M) and [12] (24.2 M), and this shows a more optimal architecture.

More so, the RAM usage of the suggested model is 2.8 GB, which is fairly average compared to more memory intensive approaches like [4] (3.4 GB) and [3] (3.1 GB). This indicate that the model can balance its performance and the use of resources. In general, the findings indicate that, in addition to increasing the classification accuracy and robustness, DUA-SFNet also preserves its computational efficiency, which proves its practical applicability and its better performance in comparison with existing uncertainty-aware fusion methods. From a clinical perspective, this improvement is substantial because even a 1–2% increase in diagnostic accuracy can lead to a meaningful reduction in misdiagnosis rates and improved patient outcomes. The observed performance gain suggests that explicit modeling of modality uncertainty enables the network to effectively identify and down-weight unreliable modalities. By suppressing noisy, corrupted, or incomplete modality inputs, the proposed model prevents the degradation in performance commonly observed in traditional multimodal fusion frameworks.

Similarly, the observed F1-score improvements of 4.1% to 7.5% indicate a balanced enhancement in both precision and recall, which is particularly important for medical diagnosis tasks characterized by severe class imbalance. In clinical settings, false negatives are especially critical because they may result in delayed or missed treatment for patients with pathological conditions. The higher F1-scores achieved by DUA-SFNet indicate that the model significantly improves sensitivity to pathological cases while maintaining robustness against false positives. This balanced performance suggests that the model effectively reduces both types of classification errors. Specifically, the uncertainty-aware modality weighting mechanism suppresses unreliable modality contributions, reducing the likelihood of false alarms, while the hierarchical attention mechanism enhances discriminative feature learning, enabling the model to better capture subtle pathological patterns present in multimodal medical images. Furthermore, the hierarchical adaptive attention mechanism strengthens the representation learning capability of the network by simultaneously capturing intra-subspace correlations and inter-modality dependencies. By modeling relationships at multiple representation levels, the network produces more discriminative and stable feature embeddings, which ultimately contributes to improved classification robustness and generalization across different datasets. This hierarchical design allows the network to selectively emphasize informative subspace features while preserving complementary information from multiple modalities.

Most prior multimodal fusion approaches rely on deterministic fusion strategies, such as feature concatenation, low-rank constraints, or fixed attention weighting schemes. These approaches implicitly assume that all modalities contribute equally and

Manuscript received 5 February 2026; Revised 30 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1529

reliably to the final prediction. However, in real-world medical imaging scenarios, modalities often suffer from noise contamination, missing data, acquisition artifacts, or modality-specific corruption, which can significantly degrade the performance of deterministic fusion models.

Unlike these conventional approaches, DUA-SFNet incorporates an explicit uncertainty estimation module that dynamically evaluates the reliability of each modality and adjusts its contribution during the fusion process. This adaptive mechanism allows the network to remain robust even when one or more modalities are partially unreliable or corrupted.

wise importance, allowing clinicians and researchers to better understand how the model reaches its predictions. This transparency is largely absent in many previously proposed black-box fusion architectures, which limits their practical adoption in clinical environments where interpretability and trustworthiness are essential.

Table 4 summarizes the statistical performance comparison between DUA-SFNet and several recent uncertainty-aware multimodal fusion methods, highlighting the superior classification accuracy, F1-score, and overall robustness achieved by the proposed framework. Despite its promising

Table 4. Statistical Performance Comparison of DUA-SFNet with Recent Uncertainty-Aware Multimodal Fusion Methods

Ref.	Method	Task / Dataset Type	Accuracy (%)	F1-Score (%)	Inference Time (ms/sample)	Model Parameters (M)	GPU Memory Usage (GB)
[1] Zhang et al. 2024	Uncertainty-aware decision fusion	Image classification	87.5	86.8	42	18.4	2.6
[2] Zhu et al. 2024	Confidence-aware dynamic fusion	Multimodal emotion recognition	86.2	85.7	47	21.3	2.9
[3] Guo et al. 2024	Uncertainty-aware dynamic fusion	Clinical prediction	85.9	84.8	51	23.5	3.1
[4] Du et al. 2024	Uncertainty-driven integration	Multi-omics tumor classification	88.1	87.4	55	25.1	3.4
[8] Wen J et al. 2024	Reliable uncertainty-driven fusion	Incomplete multiview learning	87.0	86.5	49	22.7	3.0
[12] Xie et al. 2024	Graph-based uncertainty fusion	Multi-view multi-label classification	86.8	85.9	53	24.2	3.2
Proposed	DUA-SFNet	Multimodal medical image classification	92.1	91.4	46	20.8	2.8

As a result, DUA-SFNet consistently outperforms existing multimodal fusion methods by jointly modeling uncertainty and hierarchical attention, enabling more reliable feature integration under challenging conditions. In addition to improved performance, the proposed framework offers enhanced model interpretability, which is crucial for clinical decision support systems. The hierarchical attention design enables visualization of modality-wise and subspace-

performance, DUA-SFNet has several limitations. First, the introduction of uncertainty estimation and hierarchical attention increases computational complexity, which may limit real-time deployment in resource-constrained clinical environments. Second, the current framework assumes that uncertainty can be effectively inferred from feature distributions; extreme cases of systematic modality bias may still pose challenges. Additionally, although experiments were

Manuscript received 5 February 2026; Revised 30 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1529

conducted on multiple public and self-organized datasets, broader validation on large-scale, multi-center clinical datasets is necessary to fully assess generalizability. The proposed DUA-SFNet has important implications for next-generation multimodal medical image analysis. By explicitly modeling uncertainty, the framework aligns with recent trends toward trustworthy and explainable AI in healthcare, which emphasize reliability, transparency, and robustness. Uncertainty-aware fusion can assist clinicians in understanding not only the prediction outcome but also the confidence and modality contributions behind the decision, thereby facilitating informed clinical judgment. Finally, the model has been evaluated primarily for classification tasks; its applicability to other medical imaging tasks such as segmentation or prognosis prediction remains unexplored. In the current study, the proposed DUA-SFNet model was evaluated on multiple publicly available multimodal datasets (BraTS 2020, ADNI, and MM-WHS) as well as a private clinical dataset to ensure diversity in imaging modalities and clinical conditions. While these datasets provide heterogeneous multimodal data, we acknowledge that broader validation across large-scale multi-center clinical cohorts would provide stronger evidence of real-world generalizability. Furthermore, the proposed framework can be extended to other multimodal healthcare domains, such as combining imaging with clinical records or genomic data. Supported by prior studies emphasizing uncertainty modeling and explainable attention mechanisms in medical AI, DUA-SFNet contributes a scalable and interpretable solution that advances the reliability of multimodal deep learning systems. Future research will focus on evaluating the proposed framework on larger, independent clinical datasets collected from multiple institutions to further assess robustness, scalability, and clinical applicability.

VI. Conclusion

A Dynamic Uncertainty-Aware Adaptive Subspace Fusion Network (DUA-SFNet) is proposed for multimodal medical image classification. It is designed to overcome the crucial challenges of fixed-rank fusion strategies, equal modality treatment, and limited robustness in the existing approaches. By jointly incorporating adaptive rank-learning subspace representation, modality-level uncertainty estimation, and hierarchical adaptive attention fusion, the proposed framework captures complementary information from heterogeneous imaging modalities with suppression of unreliable contributions effectively. Extensive experimental evaluations demonstrate that the proposed DUA-SFNet consistently outperforms the state-of-the-art fusion approaches, achieving 92.1% accuracy, 91.4% F1-score, and an AUC of 0.96, along with superior robustness under noisy and missing modality conditions. Component-wise ablation studies further confirm that each module contributes significantly to the overall performance and stability of the proposed framework. The proposed DUA-SFNet framework has the potential to be extended to other medical imaging tasks such as segmentation and prognosis prediction. The adaptive subspace fusion and uncertainty-aware modality weighting mechanisms can be integrated with encoder–decoder architectures commonly used for medical image segmentation, enabling pixel-level predictions while maintaining robust multimodal feature fusion. Similarly, the fused feature representations learned by the network could be incorporated into temporal prediction models for disease progression or prognosis analysis when longitudinal clinical data are available.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

Data will be available based on request.

Author Contribution

B. Krishnakumar contributed to the conceptualization of the study, development of the methodology, and drafting of the initial manuscript. B. Thanga Parvathi was responsible for data collection, preprocessing, and implementation of the experimental work. K. Nithya contributed to model development, validation, and performance analysis. M. Pyingkodi assisted in data interpretation, result analysis, and manuscript editing. Kunchanapalli Rama Krishna provided overall supervision, critical review of the manuscript, and guidance throughout the research. Jeevitha R contributed significantly to methodology refinement, experimental design, detailed analysis, and preparation of the manuscript, playing a major role in shaping the final work. All authors read and approved the final version of the manuscript.

Declarations

Ethical Approval

Not Applicable.

Consent for Publication Participants.

Consent for publication was given by all participants

Competing Interests

The authors declare no competing interests.

References

[1] Zhang, X., Xie, Z., Yu, H., Wang, Q., Wang, P., & Wang, W. (2024, October). Enhancing Adaptive Deep Networks for Image Classification via Uncertainty-aware Decision Fusion. *In*

Manuscript received 5 February 2026; Revised 30 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1529

Proceedings of the 32nd ACM International Conference on Multimedia (pp. 8595-8603). https://doi.org/10.1145/3664647.3681368

[2] Zhu, Q., Zheng, C., Zhang, Z., Shao, W., & Zhang, D. (2023). Dynamic confidence-aware multi-modal emotion recognition. IEEE Transactions on Affective Computing, 15(3), 1358-1370. https://doi.org/10.1109/TAFFC.2023.3340924

[3] Guo, J., Cheng, Y., He, W., Zhang, Y., Feng, R., & Zhang, X. (2025, April). Uncertainty-Aware Dynamic Fusion for Multimodal Clinical Prediction Tasks. *In ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICASSP49660.2025.10889595

[4] Du, L., Liu, C., Wei, R., & Chen, J. (2023). Uncertainty-aware dynamic integration for multi-omics classification of tumors. *Journal of Cancer Research and Clinical Oncology,* 149(7), 3301-3312. https://doi.org/10.1007/s00432-022-04219-3

[5] Shao, Z., Wang, H., Cai, Y., Chen, L., & Li, Y. (2025). UA-Fusion: Uncertainty-Aware Multimodal Data Fusion Framework for 3D Object Detection of Autonomous Vehicles. *IEEE Transactions on Instrumentation and Measurement.* https://doi.org/10.1109/TIM.2025.3548184

[6] E. Munari, A. Scarpa, L. Cima, M. Pozzi, F. Pagni, F. Vasuri, S. Marletta, A.P. Dei Tos, A. Eccher, Cutting-edge technology and automation in the pathology laboratory, *Virchows Arch.* 484 (4) (2024) 555–566. https://doi.org/10.1007/s00428-023-03637-z

[7] Hao, D., Meng, M., Gao, Y., Lou, X., & Kong, W. (2025). Step-wise Prompting Meets Uncertainty-Aware Dynamic Fusion for Robust EEG-Visual Emotion Recognition. *IEEE Transactions on Affective Computing*. https://doi.org/10.1109/TAFFC.2025.3632304

[8] Wen, J., Long, J., Lu, X., Liu, C., Fang, X., & Xu, Y. (2025). Partial multiview incomplete multilabel learning via uncertainty-driven reliable dynamic fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* https://doi.org/10.1109/TPAMI.2025.3603677

[9] Xu, W., Liu, X., Wang, J., Zhang, F., Hu, D., & Zong, L. (2025). UAMRL: multi-granularity uncertainty-aware multimodal representation learning for drug-target affinity prediction. *Bioinformatics,* 41(10), btaf512. https://doi.org/10.1093/bioinformatics/btaf512

[10] Li, F. J., Zhang, C. Y., & Chen, C. P. (2025). Robust Decision-Making Method for Autonomous Driving Based on Multi-Modal Fusion and Uncertainty Modeling. *IEEE Transactions on Vehicular Technology.* https://doi.org/10.1109/TVT.2025.3613839.

[11] Abdusalomov, A., Umirzakova, S., Boymatov, E., Zaripova, D., Kamalov, S., Temirov, Z., ... & Whangbo, T. K. (2025). A Human-Centric, Uncertainty-Aware Event-Fused AI Network for Robust Face Recognition in Adverse Conditions. *Applied Sciences,* 15(13), 7381. https://doi.org/10.3390/app15137381

[12] Xie, W., Lu, X., Liu, Y., Long, J., Zhang, B., Zhao, S., & Wen, J. (2024, October). Uncertainty-aware pseudo-labeling and dual graph driven network for incomplete multi-view multi-label classification. *In Proceedings of the 32nd ACM international conference on multimedia* (pp. 6656-6665). https://doi.org/10.1145/3664647.3680932

[13] Koishiyeva, D., Mukhammejanova, D., Kang, J. W., & Mukasheva, A. (2025). A Review of Deep Learning Approaches Based on Segment Anything Model for Medical Image Segmentation. *Bioengineering,* 12(12), 1312. https://doi.org/10.3390/bioengineering12121312

[14] Wang, Y., Song, D., Wang, W., Rao, S., Wang, X., & Wang, M. (2022). Self-supervised learning and semi-supervised learning for multi-sequence medical image classification. *Neurocomputing,* 513, 383-394. https://doi.org/10.1016/j.neucom.2022.09.097

[15] Jiang, J., Zhou, Q., Chen, N., He, H., Zhang, J., & He, C. Davlmf-Seg: Vision-Language Model Guided Latent Frequency-Aware Diffusion for Semi-Supervised Medical Image Segmentation. Available at SSRN 5387264. http://dx.doi.org/10.2139/ssrn.5387264

[16] S. Iqbal, A.N. Qureshi, M. Alhussein, K. Aurangzeb, M. Zubair, A. Hussain, A novel reciprocal domain adaptation neural network for enhanced diagnosis of chronic kidney disease, *Expert Syst.* 42 (2) (2025) e13825. https://doi.org/10.1111/exsy.13825.

[17] A. Shabbir, M. Zubair, Interpretable deep learning classifier using explainable AI for non-small cell lung cancer, in: *2024 Horizons of Information Technology and Engineering, HITE, IEEE,* 2024, pp. 1–6. https://doi.org/10.1109/HITE63532.2024.10777248

[18] E. Bercovich, M.C. Javitt, Medical imaging: from roentgen to the digital revolution, and beyond, *Rambam Maimonides Med. J.* 9 (4) (2018). https://doi.org/10.5041/RMMJ.10355

[19] Lu, Y., Zhao, Y., Chen, X., & Guo, X. (2022). A Novel U-Net Based Deep Learning Method for 3D Cardiovascular MRI Segmentation. *Computational Intelligence and Neuroscience,* 2022(1), 4103524. https://doi.org/10.1155/2022/4103524

Manuscript received 5 February 2026; Revised 30 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1529

[20] Suganyadevi, S., Pershiya, A. S., Balasamy, K., et al. "Deep learning based alzheimer disease diagnosis: A comprehensive review". *SN Computer Science*, Vol.5 no.4, pp.391, 2024, https://doi.org/10.1007/s42979-024-02743-2.

[21] Balasamy, K., Krishnaraj, N., & Vijayalakshmi, K. "An adaptive neuro-fuzzy based region selection and authenticating medical image through watermarking for secure communication", *Wireless Personal Communications*, Vol.122, no.3, pp. 2817–2837, 2021, https://doi.org/10.1007/s11277-021-09031-9.

[22] Suganyadevi, S., & Seethalakshmi, V. "CVD-HNet: Classifying Pneumonia and COVID-19 in Chest X-ray Images Using Deep Network". *Wireless Personal Communications*, Vol.126, no. 4, pp.3279–3303, 2022, https://doi.org/10.1007/s11277-022-09864-y

[23] Balasamy, K., & Suganyadevi, S. "Multi-dimensional fuzzy based diabetic retinopathy detection in retinal images through deep CNN method". *Multimedia Tools and Applications*, Vol 83, no. 5, pp.1–23. 2024, https://doi.org/10.1007/s11042-024-19798-1

[24] Balasamy, K., Seethalakshmi, V. & Suganyadevi, S. Medical Image Analysis Through Deep Learning Techniques: A Comprehensive Survey. *Wireless Pers Commun*, 137, 1685–1714 (2024). https://doi.org/10.1007/s11277-024-11428-1

[25] Suganyadevi, S., Seethalakshmi, V. Deep recurrent learning based qualified sequence segment analytical model (QS2AM) for infectious disease detection using CT images. *Evolving Systems*, 15, 505–521 (2024). https://doi.org/10.1007/s12530-023-09554-5.

[26] T. Gopalakrishnan, S. Ramakrishnan, K. Balasamy and A. S. Muthananda Murugavel, "Semi fragile watermarking using Gaussian mixture model for malicious image attacks," *2011 World Congress on Information and Communication Technologies, Mumbai, India,* 2011, pp. 120-125, https://doi.org/10.1109/WICT.2011.6141229.

[27] Renuka Devi, K., Suganyadevi, S and Balasamy, K. "Healthcare Data Analysis U sing Deep Learning Paradigm". *Deep Learning for Cognitive Computing Systems: Technological Advancements and Applications*, edited by M.G. Sumithra, Rajesh Kumar Dhanaraj, Celestine Iwendi and Anto Merline Manoharan, Berlin, Boston:De Gruyter, 2023, pp. 129–148. https://doi.org/10.1515/9783110750584-008.

[28] Shamia, D., Balasamy, K., and Suganyadevi, S. "A secure framework for medical image by integrating watermarking and encryption through fuzzy based roi selection", *Journal of Intelligent & Fuzzy systems*, 2023, Vol. 44, no.5, pp.7449-7457, https://doi.org/10.3233/JIFS-222618.

[29] E. Elyan, P. Vuttipittayamongkol, P. Johnston, K. Martin, K. McPherson, C.F. Moreno-García, C. Jayne, M.M.K. Sarker, Computer vision and machine learning for medical image analysis: recent advances, challenges, and way forward, *Artif. Intell. Surg.* 2 (1) (2022) 24–45. https://doi.org/10.20517/ais.2021.15

[30] O. Ayo-Farai, B.A. Olaide, C.P. Maduka, C.C. Okongwu, Engineering innovations in healthcare: a review of developments in the USA, *Eng. Sci. Technol. J.* 4 (6) (2023) 381–400. https://doi.org/10.51594/estj.v4i6.638

[31] S. Yin, C. Fu, S. Zhao, K. Li, X. Sun, T. Xu, and E. Chen, "A survey on multimodal large language models," arXiv preprint arXiv:2306.13549, 2023. https://doi.org/10.1093/nsr/nwae403

[32] D. Hao, M. Meng, Y. Gao, X. Lou and W. Kong, "Step-Wise Prompting Meets Uncertainty-Aware Dynamic Fusion for Robust EEG-Visual Emotion Recognition," in IEEE Transactions on Affective Computing, vol. 17, no. 1, pp. 694-707, Jan.-March 2026, https://doi.org/10.1109/TAFFC.2025.3632304

[33] Z. Lin, X. Hu, Y. Zhang, Z. Chen, Z. Fang, X. Chen, A. Li, P. Vepakomma, and Y. Gao, "SplitLoRA: A Split Parameter-Efficient Fine-Tuning Framework for Large Language Models," arXiv preprint arXiv:2407.00952, 2024. https://doi.org/10.48550/arXiv.2407.00952

[34] S. Hu, Z. Fang, Z. Fang, Y. Deng, X. Chen, Y. Fang, and S. T. W. Kwong, "Agentscomerge: Large language model empowered collaborative decision making for ramp merging," IEEE Transactions on Mobile Computing, 2025. https://doi.org/10.1109/TMC.2025.3564163

[35] S. Hu, Z. Fang, Z. Fang, Y. Deng, X. Chen, and Y. Fang, "Agentscodriver: Large language model empowered collaborative driving with lifelong learning," arXiv preprint arXiv:2404.06345, 2024. https://doi.org/10.48550/arXiv.2404.06345

[36] K. Wu, B. Jiang, Z. Jiang, Q. He, D. Luo, S. Wang, Q. Liu, and C. Wang, "Noiseboost: Alleviating hallucination with noise perturbation for multimodal large language models," arXiv preprint arXiv:2405.20081, 2024. https://doi.org/10.48550/arXiv.2405.20081

[37] Lin, Qinghua, Guang-Hai Liu, Zuoyong Li, Yang Li, Yuting Jiang, and Xiang Wu. "Multimodal Medical Image Classification via Synergistic Learning Pre-training." arXiv preprint arXiv:2509.17492 (2025). https://doi.org/10.48550/arXiv.2509.17492

[38] Q. Wang, L. Zhan, P. Thompson, and J. Zhou, "Multimodal learning with incomplete modalities

Manuscript received 5 February 2026; Revised 30 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1529

by knowledge distillation," in Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 2020, pp. 1828–1838. https://doi.org/10.1145/3394486.3403234

[39] J. Zhao, R. Li, and Q. Jin, "Missing modality imagination network for emotion recognition with uncertain missing modalities," in Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021, pp. 2608–2618. https://doi.org/10.18653/v1/2021.acl-long.203

[40] K. Zhou, J. Li, Y. Xiao, J. Yang, J. Cheng, W. Liu, W. Luo, J. Liu, and S. Gao, "Memorizing structure-texture correspondence for image anomaly detection," IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 6, pp. 2335–2349, 2021. https://doi.org/10.1109/TNNLS.2021.3101403

Author Biography

Dr. B. Krishnakumar is working as an Assistant Professor in the School of Computing at SASTRA Deemed University, Tamil Nadu, India. He holds a B.Tech degree in Information Technology and an M.E. degree in Computer Science and Engineering from Anna University. He received his Ph.D. from Anna University with a specialization in Deep Learning. He has more than 18 years of teaching experience and has handled a wide range of undergraduate and postgraduate courses in Computer Science. He has published over 12 research articles in international journals, including reputed SCI-indexed journals, and more than 17 papers in international and national conferences. He has also authored two book chapters with internationally reputed publishers such as IGI Global and Wiley.

Dr. Thanga Parvathi B is a distinguished Professor at Bannari Amman Institute of Technology, Sathyamangalam, with more than seven years of dedicated experience in engineering education. She holds a Ph.D. in Information and Communication Engineering from Anna University, Chennai, demonstrating her strong research orientation and academic excellence. She completed her Bachelor's and Master's degrees in Computer Science from institutions affiliated with Manonmaniam Sundaranar University, Tirunelveli. Her teaching and research interests focus on advanced computing technologies, information systems, and innovative pedagogical practices. Dr. Thanga Parvathi is committed to mentoring students, promoting academic integrity, and contributing actively to institutional and professional development.

Mrs. K. Nithya is currently serving as an Assistant Professor in the Department of Artificial Intelligence and Data Science at Karpagam Academy of Higher Education, Coimbatore, India. She earned her Master's degree from Vivekanandha College of Engineering, Tiruchengode, and brings over one year of academic experience at Karpagam Academy of Higher Education. Her primary research interests lie in Deep Learning, machine learning, and data-driven intelligent systems. She has actively participated in numerous seminars, workshops, and faculty development programs to enhance her technical and teaching skills. Mrs. Nithya has also contributed to academic research through the publication of several papers in reputed international conferences and journals.

Dr. M. Pyingkodi is an Associate Professor in the Department of Computer Applications at Kongu Engineering College, Tamil Nadu, with 19 years of professional experience, including 17 years in teaching and 2 years as a Software Developer at Colizeeum Technology, Chennai. She holds a Ph.D. in Computer Applications from Anna University, Chennai (2020), and her academic background includes a Bachelor's in Computer Science (2003) and a Master's in Computer Applications (2006). Her areas of specialization and current research interests include Machine Learning, Deep Learning, IoT, Cloud Computing, and Bioinformatics. She has published around 71 research articles in reputed journals and delivered numerous guest lectures. She has successfully organized several sponsored events, including the AICTE-sponsored international conference RCHI 2019, FDP on Deep Learning (2021), and a CSIR-sponsored seminar on Precision Agriculture (2024). She received the Best Faculty Award (2019–2020) and was honored

Manuscript received 5 February 2026; Revised 30 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1529

by the Scientific International Publishing House in April 2024.

Dr. Kunchanapalli Rama Krishna is a Professor in the CSIT Department at KL University with over 33 years of academic and research experience in Computer Science & Engineering. He holds a Ph.D. in CSE and has served at reputed institutions including Galgotias University, C-DAC, and IIMT College of Engineering. As an AICTE Margdarshak and NAAC e-Assessor, he has guided institutions toward NBA accreditation and academic excellence. He has published over 30 research papers in reputed national and international journals and conferences. Dr. Rama Krishna has also led funded projects and national conferences on cryptography and modernizing education. He actively participates in FDPs and holds certifications from IITs, IIMs, and NPTEL/SWAYAM. His guidance has shaped numerous UG, PG, and Ph.D. scholars, contributing to quality education and innovation.

Jeevitha R is an Assistant Professor in the Department of Computer Science and Engineering at KPR Institute of Engineering and Technology, Coimbatore. She completed her Bachelor of Engineering in Computer Science and Engineering in 2016 and her Master of Engineering in the same discipline in 2018, both from KPR Institute of Engineering and Technology. With strong academic credentials and a passion for research, she has published around ten papers in reputed international and national journals and conferences. Her primary areas of interest include Machine Learning, Image Processing, and Blockchain Technologies. She is actively involved in teaching, mentoring students, and contributing to institutional academic and research activities.

.

Manuscript received 5 February 2026; Revised 30 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1529