



# An Efficient U-Net Model for Brain Tumor Segmentation in MRI Images

Shruti Patil, Dr. Mahesh Kumbhar

Department of E&TC Engineering, Kasegaon Education Society's Rajaramapu Institute of Technology, Affiliated to Shivaji University, Kolhapur, Maharashtra, MS, India

**ABSTRACT:** Brain tumor segmentation from magnetic resonance imaging (MRI) plays a crucial role in medical diagnosis and treatment planning. This study focuses on the development and evaluation of a deep learning-based approach for automated brain tumor segmentation and classification using convolutional neural network (CNNs) techniques, with particular emphasis on the U-Net architecture for segmentation. The proposed method involves pre-processing brain scans through normalization, followed by training a U-Net model to accurately delineate the tumor regions. The model was trained and validated on a publicly available dataset, and good sensitivity and precision were obtained. The experimental results demonstrate that the CNN-based U-Net model provides accurate and reliable segmentation, making it a promising tool for supporting clinical decision-making in neuroanatomy.

**KEYWORDS:** UNET; CNN; MRI Segmentation; Medical Imaging

## I. INTRODUCTION

Accurate segmentation and early detection of brain tumors are critical milestones in neuro-oncological research. As these conditions can rapidly become life threatening, precise therapeutic planning is essential to ensure patient safety. Magnetic Resonance Imaging (MRI) remains the gold standard for diagnosis because of its superior tissue contrast and detailed visualization of intracranial structures. However, manual delineation of tumor boundaries is a labor-intensive process, is susceptible to human error, and requires significant clinical expertise. To address these challenges, automated deep learning systems have emerged as vital tools for improving the segmentation accuracy and classification efficiency. Among the various architectures utilized in Medical imaging, Fully Convolutional Neural Networks (FCNNs), particularly the U-Net architecture, have demonstrated exceptional performance. U-Net's distinctive encoder-decoder structure, bolstered by skip connections, allows it to capture fine spatial details while maintaining broader contextual information. This dual capability makes it highly effective for segmenting tumors with varying shapes, sizes, and intensities. Despite their success, traditional U-Net models have limited receptive field sizes. This inherent restriction hinders the ability of the network to learn the global context, often resulting in reduced segmentation quality when dealing with structurally complex tumors or diffuse boundaries. To overcome these limitations, recent studies have focused on integrating transformer architectures into existing CNN frameworks. Transformers utilize self-attention mechanisms to model global contextual relationships, a feature that traditional CNNs struggle to achieve. Notable advancements include Trans Res UNet, which enhances feature representation for glioma segmentation, and BiTr UNet, which combines CNNs with Transformer modules to achieve superior performance over standalone models. Other hybrid approaches, such as ETUNet and Swin-UNETR, leverage efficient transformer blocks and hierarchical features to optimize 3D tumor segmentation and improve semantic accuracy. Furthermore, architectures such as HUT run U-Net and Transformer pipelines in parallel to maximize information retrieval from MRI scans. The proposed framework in this study integrates U-Net for initial segmentation with transformer modules for precise tumor-type classification. By utilizing U-Net to isolate the tumor region (ROI) and transformers to model the complex relationships between features within that region, the system can distinguish between various categories, such as Glioma, Meningioma, and Pituitary tumors, with higher diagnostic precision. The synergy between the spatial feature extraction of CNNs and the long-range dependency modeling of transformers results in a robust diagnostic system. Experimental benchmarks highlight that this integrated framework significantly outperforms traditional CNN-based methods, demonstrating its immense potential for real-world medical applications and improved patient outcomes in the field of neuro-oncology.

## II. LITERATURE REVIEW

Recent research has shown that deep learning methods are becoming more effective for brain tumor segmentation in MRI. One study presented a hybrid ResNet-U-Net model that included transformer blocks. This setup allows feature extraction at different levels and the modelling of long-range dependencies. The combination of ResNet pathways and transformer attention significantly improved the consistency of spatial features during the decoding phase using skip connections and transposed convolutions. When evaluated on the BraTS 2019 dataset, it performed excellently, achieving Dice scores of 0.91, 0.89, and 0.84 for the whole, core, and enhancing tumor regions, respectively, with an overall accuracy of 98% [1]. Another study introduced BiTr-UNET, a combined CNN and Transformer framework aimed at multimodal MRI brain tumor segmentation. CNN layers extract detailed and localized features, whereas transformer modules capture broader contextual relationships. This pairing improves the segmentation precision. BiTr-UNET excelled in the BraTS 2021 challenge, achieving median Dice scores of 0.9335, 0.9304, and 0.8899, along with minimal Hausdorff distances, demonstrating its capability to outline the tumor subregions [2]. A third method combines U-Net with Transformer-based modules. It merges local feature extraction using CNNs with attention mechanisms that focus on global structural patterns. This hybrid design increases segmentation accuracy across various datasets. The model achieved Dice scores of 0.94, 0.921, 0.83, and 0.94 on the BraTS 2019–2021 datasets, confirming its effectiveness in automated brain tumor segmentation. Further developments include embedding transformer components within the U-Net framework to overcome the limitations of CNNs in capturing long-range dependencies. The new architecture features a transformer encoder, spatial-channel attention layer, and skip connections on cross-attention to improve feature representation and boundary hybrid approach not only provides high-precision localization but also supports informed clinical decision-making by accurately distinguishing between benign and malignant localization [3]. The model showed strong results on BraTS 2018 (0.854 DSC, 6.688 HD95) and BraTS 2020 (0.862 DSC, 5.455 HD95), providing high accuracy with minimal boundary errors [4]. Another noteworthy contribution is Swin UNETR, which merges blocks an inspired design. By employing shifted-window self-attention and multi-resolution feature representations, the model effectively connects global dependency while maintaining spatial details. The Swin Transformer serves as a hierarchical encoder, along with an FCNN decoder and skips connections. When tested on BraTS 2021, Swin UNETR ranked among the top models for medical-image segmentation [5]. A recent 3D U-Net, Transformer hybrid model that includes Contextual Transformer (CoT) blocks and Double Attention (DA) mechanisms. The CoT modules enhance the extraction of local features, whereas the DA improves the handling of long-range dependencies. Experimental results show that this approach surpasses existing models in terms of segmentation accuracy, offering quicker medical image processing to aid in diagnosis and treatment planning in the future. [6] Additionally, the Hybrid U-Net Transformer (HUT) model excels in both accuracy and efficiency. By running the U-Net and Transformer processes in parallel, the model benefits from the global attention and long-range feature dependencies sourced from the transformer part, whereas the U-Net decoder refines the predictions. A self-supervised training approach further enhances segmentation accuracy. HUT outperformed the SPiN model on the ATLAS dataset with a 4.84% increase in the Dice score and a 41% reduction in the Hausdorff distance. It also exceeded the performance of nnU-Net on BraTS 2020, with a 0.96% gain in the Dice score and a 4.1% decrease in HD [7-10].

## III. METHODOLOGY

### A. Dataset Preparation

The proposed method is a hybrid deep learning framework. It combines U-Net for tumor segmentation and transformers for classification. This framework efficiently processes brain MRI scans and predicts tumor subtypes. The system operates in a sequential modular pipeline. Each stage makes an important contribution to the final output of the model.

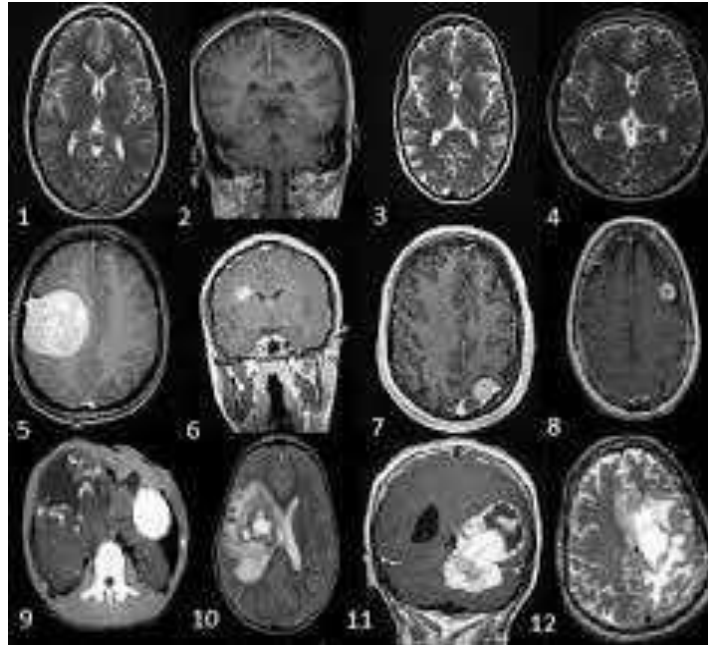


Fig.1 Dataset Preparation

## B. Dataset Acquisition

The dataset includes four main categories.

**Glioma:** Aggressive brain tumors originating from glial cells.

**Meningioma:** tumors that arise from the meninges and are often benign.

**Pituitary:** Tumors found in the pituitary gland that can affect hormonal balance.

**No Tumor:** healthy brain scans were used for baseline comparison. MRI scans consist of multiple imaging types, including T1, T2, FLAIR, and T1 contrast-enhanced (T1c) scans. These different types provide complementary information about brain anatomy and pathology

## C. Pre Processing

Steps involved include:

### 1. Noise Reduction

Medical images often contain noise from scanner artifacts or patient movement. Techniques such as Gaussian filtering and median filtering are used to eliminate high-frequency noise while retaining essential structural details. This improves the clarity of the tumor boundaries.

### 2. Intensity Normalization.

MRI scans obtained from different machines can show different intensity levels. Normalization rescales the pixel intensity values to a standard range (e.g., 0–1). This ensured consistent inputs across the dataset and helped the model converge better.

### 3. Contrast Enhancement

Tumor areas may show low contrast compared to surrounding tissues. Techniques like Contrast Limited Adaptive Histogram Equalization (CLAHE) are used to make tumor edges and internal structures more visible.

### 4. Image Resizing and Channel Standardization

All images were resized to the same dimensions (usually 224×224 or 256×256) to fit the input requirements of the U-Net and Transformer models. If required, the images were adjusted to have three channels to work with pre-trained architectures.

## D. U-Net Based Segmentation

Segmentation is a crucial step in accurately pinpointing tumor regions before classification. The U-Net architecture was used because it can capture detailed spatial information even with little training data.

### 1. Encoder (Contracting Path)

Multiple convolutional layers extract features at different levels. Down sampling using max-pooling reduces the size of the data while increasing the richness of the features.

### 2. Decoder (Expanding Path)

Transposed convolutions up sampled the feature maps to match the original image size. Skip connections transfer features from the encoder to the decoder while maintaining the spatial context intact

### 3. Output Layer

A binary mask that highlights the tumor regions in white, while the background is black, is created.

### Advantages of U-Net:

It excels in capturing fine tumor boundaries.

It works well with relatively small datasets because it effectively reuses features.

## IV. EXPERIMENTAL EVALUATION

### Dataset Preparation

The experimental work utilized a publicly available MRI Brain Tumor Dataset from Kaggle, consisting of 7,022 images categorized into four major classes of brain tumors. The dataset was structured into separate folders, each representing a distinct tumor type. Although minor variations exist, the dataset is relatively balanced, which enables unbiased model training. Each MRI image was a 2D axial brain scan captured in grayscale or RGB format and stored in the PNG/JPG format. The scans vary in brightness, texture, and contrast owing to differences in MRI scanner settings. To ensure uniformity, all images were resized to 256×256 pixels, forming a standardized input size for both the segmentation and classification networks. The dataset was split in an 80:20 ratio, where 80% was used for training and 20% for testing. This split ensured that both segmentation and classification were evaluated on samples that had not been previously observed.

### Segmentation Methodology using u-net

Segmentation is the foundational stage of this study, responsible for extracting the tumor region from MRI images before classification. The segmentation model was implemented using the U-Net architecture, which is known for its high performance in medical-image segmentation.

### Preprocessing prior to segmentation

Before feeding the images into U-Net, the following preprocessing steps were performed.

#### a) Grayscale Conversion

If the image is RGB, it is converted into a single-channel grayscale image.

#### b) Image Resizing

All images were resized to 256 × 256 pixels using bicubic interpolation.

#### c) Normalization

The pixel values were normalized as follows:

$$\frac{I}{I_{\text{normal}}} = 255$$

#### d) Contrast Enhancement using CLAHE

CLAHE improves the visibility of tumor boundaries using adaptive histogram equalization.

#### e) Pseudo mask generation

Since the dataset does not provide ground truth segmentation masks, **Otsu's thresholding** is used to create Binary pseudo-masks:

$$(x, y) = 1 \quad \text{if } (x, y) > T_{\text{Otsu}}$$
$$0 \quad \text{otherwise}$$

These synthetic masks served as supervisory labels for U-Net training.

#### f) U-net architecture description

U-Net consists of:

Encoder Path

Series of **Conv** → **ReLU** → **MaxPooling** layers Extract hierarchical features  
 Reduces spatial resolution gradually Bottleneck  
 Deepest layer  
 Captures global tumor structure Decoder Path  
 Up sampling using transpose convolution  
 Concatenation with encoder layers through skip connections Combines global and local feature information  
 Output Layer  
 A final 1×1 convolution produces a **binary mask** with the following values:  
**1** → tumor region  
**0** → background  
 The loss function used is **binary cross-entropy**, optimized using the Adam optimizer.

**a) Numerical Workflow of Segmentation Stage**

**Input Image Size: 256 × 256 × 1**

Encoding Stage: Level 1: 32 filters

Level 2: 64 filters

Level 3: 128 filters

**Bottleneck:**

256 filters

**Decoding stage:**

Up sampling from 256 → 128 → 64 → 32 feature channels.

**Output:**

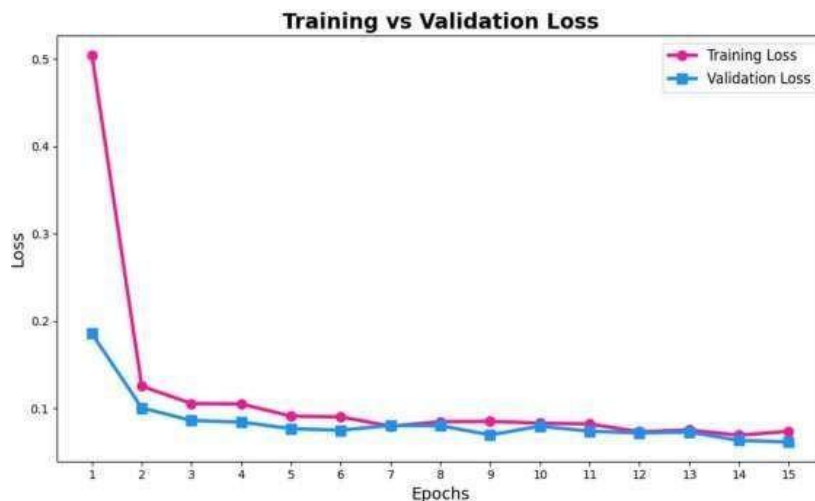
Final mask dimension: 256 × 256 × 1

**Training Setup**

Learning rate = 0.001

Batch size = 16 Epochs = 20 Validation split = 10%

**G. Evaluation of Segmentation Stage**



**Fig. 2 Evaluation of Segmentation Stage**

Once the U-Net model completes training, segmentation performance is assessed using several metrics and visual inspection.

**Quantitative Evaluation Metrics**

**A. Accuracy**

Measures pixel-wise correctness:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

### B. Loss Curve

The binary cross-entropy loss was monitored during training to observe convergence.

### C. Dice Similarity Coefficient (DSC)

Although pseudo masks were used, the Dice score was still computed as follows:

$$\text{DSE} = \frac{2|A \cap B|}{|A| + |B|}$$

### D. Visual Evaluation

- the tumor boundary is captured,
- unwanted noise is suppressed
- the mask accurately isolates tumor regions
- Experimental Evaluation

### H. Segmentation Performance Evaluation

The model achieved the following scores:

Mean IoU: 0.9383

Mean Dice: 0.9670

These values indicate excellent segmentation performance

- Bottleneck: 256 filters
- Decoding stage:  
Up sampling from 256 → 128 → 64 → 32 feature channels.
- Output:  
Final mask dimension: 256 × 256 × 1
- Training Setup  
Learning rate = 0.001

Fig. 2 illustrates the model exhibits low false positives (307k) and low false negatives (137k), leading to an overall high accuracy of ~98.8%, precision of ~95.5%, and recall of ~97.9%. These results confirm the robustness and reliability of the proposed tumor segmentation and classification. The model exhibited excellent training behavior. Loss decreases smoothly validation loss mirrors training loss No over fitting. Final losses are low → high accuracy and strong predictive performance this is the ideal type of loss curve for a deep learning model

Interpretation:

IOU ranges from 0 to 1

Values above 0.80 → very good

Values above 0.90 → high-quality segmentation IOU = 0.9383

This indicates that the predicted tumor regions by the model overlapped 93.8% with the actual tumor regions.

This is extremely strong performance, showing accurate boundaries and minimal leakage/missing regions

## V. CLASSIFICATION USING CNN

Once the tumor region is isolated through segmentation, the next step is to identify the tumor type. A Transformer-based classification model achieves this by using self-attention mechanisms to analyze complex visual structures. Originally designed for processing language, transformers have become highly effective in image classification because they can model long-range dependencies within an image more efficiently than traditional convolutional networks. Before entering the transformer, the segmented tumor image was divided into small patches, treating each patch as an individual token. This patch-based representation allows the model to study the local regions of the tumor while also understanding how these regions connect across the entire image. Each patch was then transformed into a numerical representation that captured its texture, shape, and structural features. Positional encodings were added to maintain the original arrangement of these patches. This helps the model understand the relative placement of each patch, which is crucial because tumors often show spatial patterns and directional growth that aid in classification. The heart of the Transformer is the Multi-Head Self-Attention

mechanism. Here, every patch is connected to every other patch, allowing the model to recognize global correlations. For example, a benign tumor may exhibit smooth boundaries and consistent intensities, whereas malignant tumors often have irregular shapes or mixed textures. Self-attention enables the classifier to analyze these relationships as a whole, leading to better performance than CNNs, which only consider local regions at a time. After several layers of attention and feed-forward processing, the model produces a high-dimensional feature representation of the tumor. This representation is then sent through a classification head that assigns the tumor to one of four categories. The Transformer model is particularly effective because it can detect subtle differences in tumor morphology, making it more reliable than traditional models such as VGG16.

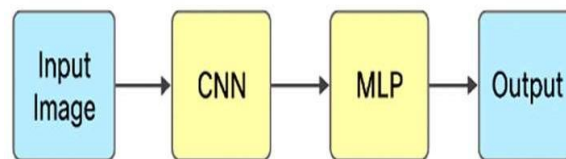


Fig. 3 Flow of Classification Architecture

Before segmentation and classification, the system preprocessed the data to ensure consistent input quality. Preprocessing usually involves converting the MRI scans to grayscale, resizing them to the same dimensions, normalizing pixel intensity values, and applying contrast improvement techniques, such as CLAHE. These steps helped standardise the differences across MRI devices, reduce noise, and highlight the tumor boundaries. After preprocessing, the dataset was split into training and testing subsets. The training set was used to teach both the U-Net and Transformer models, whereas the testing set was used to evaluate the performance. This ensured that the system was tested on images that it had never encountered, providing a realistic measure of accuracy. Once both models were trained, the segmentation mask from U-Net was applied to the original input image to extract a clean ROI. This ROI was then sent to the transformer classifier, which made the final tumor type prediction. The system compiles various evaluation metrics, such as accuracy, precision, recall, F1-scores, and confusion matrices, to show performance. Additionally, it generates visual outputs that include the original MRI images, predicted masks, extracted ROIs, and classification results. At the end of the workflow, both models were saved for future use or deployment. This allows the entire system to be integrated into a real-time diagnostic tool or used in a clinical setting. Saving the model also removes the need to retrain it from scratch each time

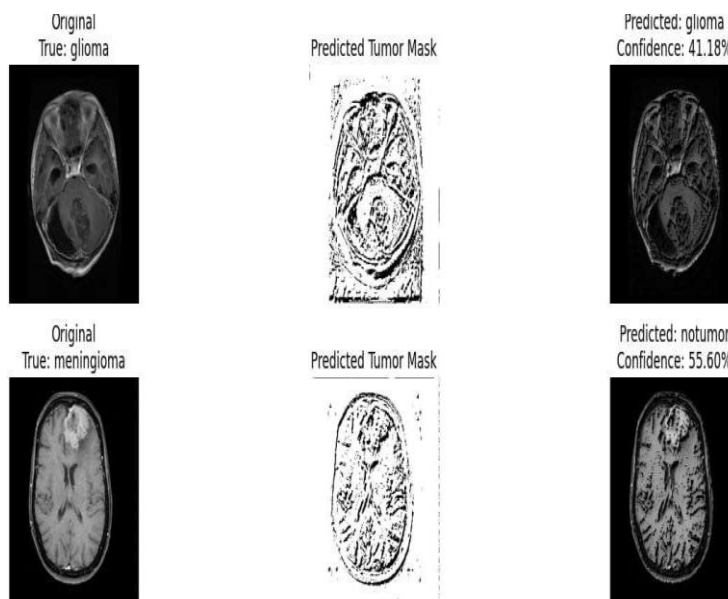


Fig. 4 Predicted output using Classification (CNN)

## VI. CONCLUSION

The proposed U-Net + CNN-based system exhibited outstanding performance in both segmentation and classification tasks. The consistent decrease in training and validation losses suggests stable learning without evidence of over fitting. The model achieved a high Mean Intersection over Union (IoU) of 0.938, indicating a substantial overlap between the predicted and actual tumor regions. Furthermore, a high Dice score of 0.967 confirmed excellent mask similarity and reliable extraction of tumor boundaries. The confusion matrix revealed a high number of true positives and true negatives, demonstrating accurate detection of both tumor and non-tumor regions. The low rates of false positives and false negatives indicate that the model infrequently generated false alarms or missed actual tumors. Performance metrics, including precision (95.5%), recall (97.9%), and accuracy (98.8%), underscored the model's robust behavior. Overall, the system provided reliable, consistent, and clinically meaningful segmentation. The results validate the effectiveness of the proposed hybrid deep learning pipeline for medical image analysis. The model is suitable for research applications and can be extended for real-time medical imaging support.

## REFERENCES

- [1] Rasool N, Bhat JI, Wani NA, Ahmad N, Alshara M. TransResUNET: Revolutionizing glioma brain tumor segmentation through transformer-enhanced residual UNET. 2024 IEEE Access.
- [2] Jia Q, Shu H. Bitr-UNET: A CNN-transformer combined network for MRI brain tumor segmentation. In International MICCAI Brainlesion Workshop 2021:3-14. Cham: Springer International Publishing.
- [3] Shedbalkar J, Prabhushetty K, Havaladar RH. UNET and transformer- based model for multi-modality brain tumor segmentation. SSRG International Journal of Electronics and Communication Engineering, 2023; 10: 22–35.
- [4] Zhang W, Chen S, Ma Y, Liu Y, Sri Varshini E, Cao X. ETUNET: Exploring efficient transformer-enhanced UNET for 3D brain tumor segmentation. Computers in Biology and Medicine, 2024; 171: 108005.
- [5] Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D. Swin UNETr: Swin transformers for semantic segmentation of brain tumors in MRI images. In International MICCAI Brainlesion Workshop. 2021:272–284. Cham: Springer International Publishing.
- [6] Nguyen-Tat TB, Nguyen TQT, Nguyen HN, Ngo VM. Enhancing brain tumor segmentation in MRI images: A hybrid approach using UNET, attention mechanisms, and transformers. Egyptian Informatics Journal, 2024;27:100528.
- [7] Soh WK, Yuen HY, Rajapakse JC. HUT: Hybrid UNET transformer for brain lesion and tumour segmentation. Heliyon, 2023; 9(12): e22412.
- [8] Liang J, Yang C, Zeng L. 3D PSwinBTS: An efficient transformer-based UNET using 3D parallel shifted windows for brain tumor segmentation. Digital Signal Processing, 131:2022:103784.
- [9] Tiwary PK, Johri P, Katiyar A, Chhipa MK. Deep learning-based MRI brain tumor segmentation with EfficientNet-enhanced UNET. 2025 IEEE Access



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details