An Automated Medical Image Segmentation Framework using Deep Learning and Variational Autoencoders with Conditional Neural Networks

Dustakar Surendra Rao[1], L. Koteswara Rao[2*], Bhagyaraju Vipparthi[3]
Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad, 500075, Telangana, India[1, 2]
Department of Electronics and Communication Engineering, Guru Nanak Institutions Technical Campus, Hyderabad, 501506, Telangana, India[1]
Department of Electronics and Communication Engineering, Siddhartha Institute of Engineering and Technology, Hyderabad, 501506, Telangana, India[3]

Abstract—**It is a highly difficult challenge to achieve correlation between images by reliable image authentication and this is essential for numerous therapeutic activities like combining images, creating tissue atlases and tracking the development of the tumors. The separation of healthcare data utilizing deep learning variational autoencoders and conditional neural networks is presented in this research as a paradigm. One of the essential jobs in machine vision is the partitioning of an image. Due to the requirement for low-level spatial data, this assignment is more challenging compared to other vision-related challenges. By utilizing VAEs' capacity to develop hidden representations and combining CNNs in a conditioned environment, the algorithm generates accurate and efficient results for the segmentation. Moreover, to learn the representation of latent space from labelled clinical images, the VAE is trained as part of the system that is suggested. After that, the representations that were learned and real categorizations are used to develop the conditional neural network. Furthermore, the model that has been trained is utilized to accurately separate the areas that are important in brand-new medical images during the inferential stage. Thus, the experimental findings on several healthcare imaging databases show the enhanced precision of segmentation of the structure, highlighting its ability to enhance automated diagnosis and treatment. Henceforth, the suggested Deep Learning and Variational Auto Encoders with Conditional Neural Networks (DL-VAE-CNN) are employed to solve the pixel-level problem of classification that plagues the earlier investigations.**

Keywords*—*Deep learning; variational autoencoders; CNN; medical image segmentation; automated diagnosis and treatment

I. Introduction

A harmful disease that has one of the highest five-year rates of survival is brain tumor. To determine the kind of brain tumor, neurological specialists frequently employ magnetic resonance imaging (MRI). Despite the amazing advancements in science over the past few years, certain diseases still pose a grave risk to life [1]. Of these, cancer of the brain is among the most aggressive. Brain tumors are the result of untamed, erratic polypeptide development inside and outside the brain's cells. Carcinoma brain tumors are the most dangerous kind, but benign tumors can also exist. Thus, brain cancer is the common name for the malignant variety of cerebral tumor. A tumor is deemed cancerous if it emerges from the membrane and invades nearby tissues. There are three main types of brain tumors. They are gliomas, pituitary tumors and meningiomas. The pituitary brain tumor is a malignant development of polypeptide that surrounds the pituitary gland, which is a gland that is situated at the bottom of the brain. On the outermost layer of the brain below the skull, meningioma, a benign tumor, is detected. It grows slowly. Amongst all brain tumors, glioma has the greatest fatality rate globally. Pituitary and meningioma tumors can be easily found due to the position of the different types of brain tumors; however, gliomas are challenging to find and study.

The objective of a semantic segmentation is to define the limits of important anatomical or pathological features [2]. The findings can be applied to multisensory image authorization, therapeutic targeting preparation, or operational systems for navigation. Furthermore, the field of imaging in medicine has made extensive use of its versions for segmentation. Due to the increasing adoption of Computed Tomography (CT) and Magnetic Resonance (MR) imaging for medical diagnosis, treatment planning and medical inquiries, image segmentation is necessary for radiological experts to effectively utilize technological advances to assist in helpful assessment as well as the planning of treatment. Delineating anatomical features and other areas of interest (ROI) requires the use of trustworthy methods [3]. Moreover, the utilization of three-dimensional (3D) convolutional neural networks for tissue separation on CT data is very efficient: however the process of creating labels for training is time-consuming. Therefore, the effectiveness of the algorithm that has been trained is hampered by a lack of designations, particularly when it comes to its ability to generalize to everyday clinical practice. Nevertheless, due to the complexity of person's structure and pathological situations, numerous variations and medical conditions are not included in this small training information and that may partially account for the reported discrepancy among efficiency utilizing real-world information and anticipated results [4]. Even if trained on a greater amount of data sets, there will be discrepancies among the model's probability and the real-world probability. These changes could relate to technological issues, individual traits or more uncommon pathological diseases.

Nevertheless, pathological identification is vital yet a challenging task in the computation and evaluation of clinical images. Although certain tasks such as tumor measurements or therapy surveillance refers to precise segmentation of the conditions and other applications such as assisting the physician with inspecting the image or retrieving visuals with diseases at particular points from a data base and it only require information to identify the illness and an approximate representation of its description [5]. Moreover, researchers concentrate on the identification of images with diseases as another instance of such an application. Furthermore, missing connections among images due to pathological features could result in incorrect identification that is leading to the surrounding regions. In order to estimate the regions of lacking connections, it might be good to incorporate the earlier pathological data into the process of registration. Therefore, variational autoencoders (VAEs) are broad dormant area computational simulations that have proven incredibly effective in a variety of fascinating applications in medical information technology including large-scale physiological order evaluations, incorporated multi-omics data analysis and the design of molecules and protein layout [6].

Additionally, the primary concept behind VAEs is to comprehend the data distributions in a manner that enables the generation of fresh, useful information with higher intra-class variability from the transmitted population. Moreover, deep generative algorithms have received an abundance of interest previously as a result of their numerous information creation capabilities. Therefore, one of the most well-liked methods of dynamic modelling and low-dimensional network representational learning amongst them is the variational autoencoder (VAE). Nowadays, diagnostic imaging is commonly employed to locate tumors and other anomalies in the body of a person for investigation and research. According to the illness type and the patient's body portion, several techniques for medical imaging are employed for diagnostics [7]. To diagnose an anomaly promptly and accurately, scanning needs to be done. The procedures include computer tomography (CT) that is frequently utilized to compare numerous soft tissue types including the liver, the lung cells and lipids, as well as X-ray, that is utilized for recognizing fractures in bones or to identify tissues that are hard. Positron emission tomography (PET) scans and MRI scans are both frequently employed in the recognition of anomalies in the fields of neuroscience, heart disease, malignancy and connective tissue impairment respectively.

Henceforth, a convolution's job is to find structures in the data it receives that represent characteristics that can be detected. Convolutional operations require kernels, also known as filtration systems. A convolution execution is viewed mathematically and the outcome of multiplying an element-by-element portion of the input by the filter adds the outcomes. Until the entire input area has been covered, the procedure continues by moving the location of the fragment as determined by the pace a number of instances. This creates a new result array by computing just one value for every segment [8]. Furthermore, with pooling, linguistically associated characteristics are combined into one, as opposed to convolutions. In order to decrease the size of the module and create constancy to minor changes and deformations, they are termed as the fixed processes which are not trainable. Hence, CNNs have shown they can produce images of outstanding quality with significantly fewer training time and computing demand. However, fully connected networks can also produce images. Based on vector data an image is used as input, there are two distinct kinds of picture generating algorithms. Though, improvements in technology have made it easier to acquire an image, which has resulted in the production of enormous numbers of pictures with excellent resolution at extremely low costs. The creation of biological algorithms for image processing has significantly improved as a result of this and it has made it possible to create computerized methods for gathering data through visual inspection or assessment [9].

The key contribution of the established outline is updated as follows:

- At first, the data has been collected from the dataset.
- Next, the pre-processing step is completed by Histogram Equalization.
- After that, the segmentation of the image is done by K means clustering.
- Feature extraction process is done by VAE, this will model the data.
- Classification is done by Convolutional Neural Network.
- Finally, the effectiveness of the proposed approach has been acknowledged and compared with various methods to demonstrate its superiority in efficiency and productivity.

The other parts of this research are broken down into the following divisions as an outcome: The related works are revealed in Section II after an in-depth examination. Problem statement is covered in Section III. In Section IV, the specifics of DL-VAE-CNN are examined. The experiment's results are analyzed and reported in detail in Section V. The research's conclusion is found in Section VI.

II. Related Works

Chen et al.[10] disclosed in his paper that, the development of novel healthcare image processing techniques has attracted a lot of academic curiosity in deep learning, and this deep learning-based algorithm have achieved amazingly well across a range of healthcare scanning applications to enhance illness identification and detection. Considering their achievement, the absence of sizable and thoroughly structured data sets severely hinders the advancement of deep learning algorithms for medical image interpretation. In order to give an in-depth account of using deep learning techniques in diverse medical imaging analysis assignments, recent results are utilized in this study. Moreover, researchers focus in particular on the most recent developments and contributions made by state-of-the-art unsupervised and semi-supervised deep learning in the analysis of healthcare images that are outlined according to various application circumstances, encompassing categorization, the process of segmentation identification and registration of images. The inaccessibility problem is the research's main flaw.

Considering pixel-based deformation for evaluation, there is currently an exciting opportunity on super-resolution images with bigger up-scaling. This is due to the significant improvement in empirical precision, said by [11]. In addition to the very soften impact, the perceived resemblance is poorly grasped. Because of the development of generative adversarial networks, it is now feasible to super-resolve low-resolution images to produce distribution-sharing photo-realistic images. Generative networks, nevertheless, struggle with mode-collapse issues and unsustainable sampling creation. Owing to the probabilistic dispersion of the high-resolution images produced by the low-resolution images, researchers suggest doing Image Super Resolution via Variational Autoencoders (SR-VAE) learning. Moreover, researchers incorporate the conditioned sampling method to reduce the implicit substructure for rebuilding since Conditional Variational Autoencoders frequently produce blurry images. Thus, researchers estimate the difference among hidden vectors and the conventional Gaussian distribution using KL damage to assess the model's generalization. Finally, with the goal to negotiate the trade-off among perceptions and super-resolution deformation, researchers calculate the reconstructed image using both the revised deeper component reduction among SR and HR images as well as pixel-based reduction.

Uzunova, Ehrhardt, et al. [12] revealed in his paper that, the capacity to decode black box artificial intelligence approaches for analyzing medical images is becoming increasingly important. A system like trained neural network must be prepared to justify its choices and projections in order to be trusted by a physician. In this study, researchers take on the challenge of coming up with logical justifications for the conclusions reached by healthcare image classification algorithms, which are trained to distinguish among various illnesses and tissues that are healthy. Finding out if the results of classification alter when such image areas are removed is a logical way to figure out the fact that areas of the image have an impact on the training classification. This concept can be effectively put into practice if it is expressed as a reduction issue. This is considered as the drawback here. In order to make a difference, researchers define the omission of pathologies as the substitution of those conditions with a substitute produced by variational autoencoders that appears normal. The investigations using a classification neural network on brain lesion MRIs and OCT (Optical Coherence Tomography) images demonstrate that an important substitution of "removed" image areas has a substantial influence on the accuracy of the given interpretations. Thus, the suggested omission procedure has been demonstrated to be effective when contrasted with four other tried-and-true techniques, this approach yields superior outcomes.

Higher dimension hyperspectral images always need more computing, which renders the processing of images difficult said by [13]. Deep learning methods have excelled in many areas of processing images and they are useful for enhancing the accuracy of classification. The complete extraction of copious spectrum data, including the fusion of geographical and spectral information, still presents significant problems. This research proposes an innovative design for autonomous hyperspectral extraction of features that utilizes the spatially reviewing variation auto encoder (AE). The disadvantage in this situation is that having numerous channels that makes it hard to recognize materials and continually wastes resources. This technique's main idea is to refine the acquired spectral characteristics by collecting spatial characteristics from multiple facets using created systems. A multilayered generator collects spectral characteristics based on the data and standard errors it produces, space-time vectors are created. With the ability to modify the derived mean vectors, multilayered convolutional neural networks and long short-term memory systems are used to gather spatial information utilizing local perception and consecutive perception. Additionally, the suggested function of loss ensures the coherence of the likelihood probabilities of different implicit spatial characteristics that were acquired from the exact same neighbor area. The effectiveness of this approach is enhanced by the incorporation of spatial extraction of features algorithms and deep AE designs, which are built specifically for hyperspectral images.

Peyré and Cuturi [14] disclosed in his paper that, the generative versus discriminatory modelling is a key distinction in the arena of machine learning. In contrast to generative modelling that seeks to tackle the broader issue of developing an aggregate allocation over all the variables, discriminatory modelling strives to train an indicator provided the data available. A model that generates information imitates how data is produced in reality. Moreover, a sound architecture to develop profound latent-variable systems and related interpretation frameworks is provided by variational autoencoders. In this paper, researchers introduce variational autoencoders and discuss several significant improvements.

With multiple uses including scenario comprehension, medical imaging analysis, robotics awareness, surveillance footage, virtual reality and image compression as well, segmenting images is a fundamental subject in the fields of image processing and vision for computers said by [15]. In the fiction, a total of image separation techniques have been generated. There has been a substantial quantity of determinations subsequently focused on improving algorithms for image segmentation using models trained with deep learning as a consequence of the efficacy of deep learning algorithms in an array of visual application. Using a wide range of groundbreaking research on conceptual and instance-level classification such as entirely convolutional pixel-labeling systems, encoder-decoder designs, multi-scale and pyramid-based methods, intermittent systems, graphic focus designs and more researchers offer an in-depth examination of the research available as of the time of publication in this analysis. Moreover, researchers evaluate the similarities, advantages, and disadvantages of different deep learning simulations, look at the most popular data sets, present results, and talk about possible future study avenues in this field. The classification issue is the research's main flaw.

Wang et al. [16] said in his paper that, the healthcare segmentation of images has made widespread usage of deep learning, and many studies have been distributed authenticating the knowledge's presentation in the arena. Deep learning algorithms for medical image segmentation are described in a thorough topical analysis. This essay offers two distinctive contributions. Researchers categorize the most recent literary works according to a multi-level architecture from rough to satisfactory, in contrast to previous studies that explicitly split studies of deep learning on segmenting medical images into multiple categories and present studies thoroughly for each category. Secondly, while unstructured methods have been covered throughout numerous previous surveys and aren't currently in demand, this work concentrates on supervised and poorly supervised learning methods. Also, researchers examine research on supervised learning methods in three areas: the choice of backbone networks, the layout of network wedges, and the enhancement of function losses. Researchers go into the research on poorly supervised learning methods independently for data enhancement, learning by transfer, and dynamic categorization. This study organizes the literatures quite differently from previous investigations. It is also easier for users to comprehend the pertinent justification, which will help them come up with suitable deep learning-based advances for healthcare image segmentation.

III. Problem Statement

Correct interpretation in remedial and diagnosis activities depends on an accurate MRI. The precision of deep learning (DL) rebuilding, is mostly unknown due to underestimating after MRI recording and the over parameterized and opaque characteristics of DL [17]. This work intends to measure the degree of ambiguity in DL model-based image restoration. The present investigation employs an automated medical image segmentation framework using deep learning and variational autoencoders with conditional neural networks to address this uncertainty problem.

IV. Methodology

The suggested method is projected in Fig. 1. An automated medical image segmentation framework using Deep Learning and variational autoencoders with conditional neural networks (DL-VAE-CNN) is employed for this method. Regarding the purposes of training and testing, a variety of databases are accessible. To investigate the data and establish accurate labeling in this situation, the Low-Grade Gliomas (LGG) segmentation dataset is used. Furthermore, pre-processing procedures are used to increase the image precision and quality. Fig. 1 shows the proposed DL-VAE-CNN architecture.

A. Data Collection

The data set employed for the testing and training purpose is Low Grade Glioma (LGG) segmentation dataset and the data are obtained from The Cancer Imaging Archive (TCIA) brain MR images. Approximately 120 patients' brain image data is second-hand in this investigation. From that image 60% of images were designated for training process and 60% of images were designated for testing process [18]. It embraces gray scale images.

B. Pre-Processing

Histogram Equalization (HE) is used to improve the quality of images. This is done by levelling, the grey levels of the image pixels so as to continuously rearrange them within the location [19]. The image that was entered as input is turned into a final image after the histogram has been analyzed and normalized for sum computation is shown in Fig. 2.

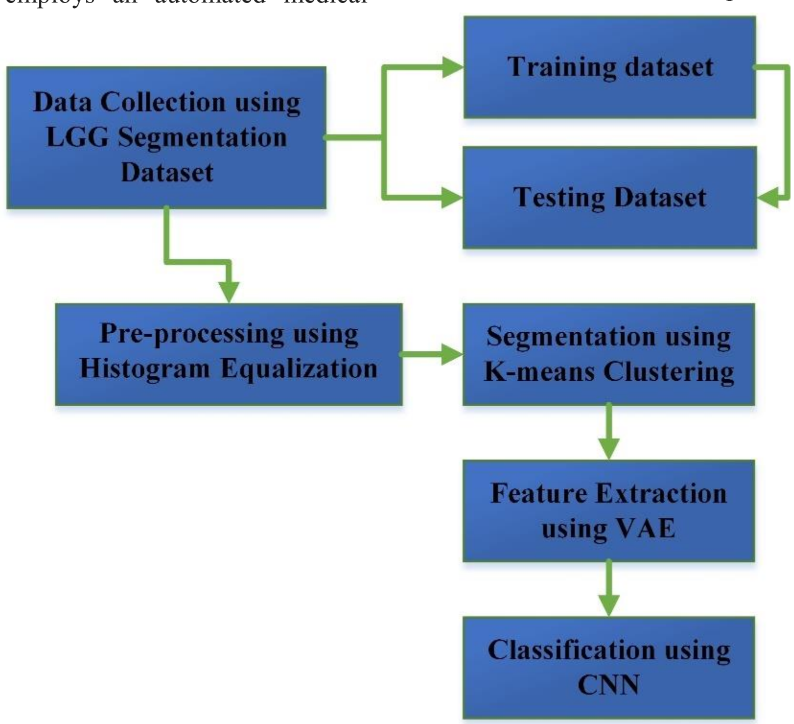

Fig. 1. Proposed DL-VAE-CNN architecture.

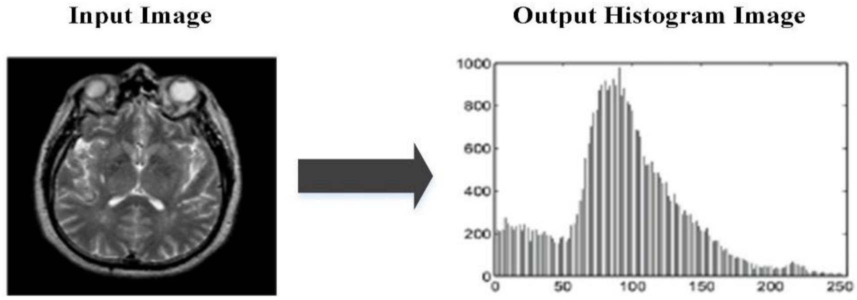

Fig. 2. Histogram equalization for healthcare image.

The method for altering the input image into output image using HE is given in the Eq. (1) below:

$$\text{Histogram Equalization image} = \text{round}\left(\frac{\text{CDF of image} - \text{CDF (min)}}{(\text{W}\times\text{H}) - \text{CDF (min)}} \times (\text{G}-1)\right) \quad (1)$$

Where, CDF is denoted as Cumulative Distribution Function; cdf (min) is denoted as the minimal non-zero value of Cumulative Distribution Function; W is denoted as the width of the image; H is denoted as the height of the image; G is denoted as the greyscale of the image.

C. Segmentation using K-Means Clustering

As a key step in many applications, segmentation has grown with the importance in the context of image processing including image recovery, classification, and recognizing objects [20]. After the pre-processing step is finished, the images are separated; then, the suggested method uses K-means clustering to find comparable areas, combine them and examine each segment properly. Moreover, the regions of interest in the grouped images are transformed to RGB layout and each predicted zone of interest is calculated by fusing attributes like color and intensity ratio.

The basic goal of the K-means clustering approach is to transform unknown data set facts into distinct group components. This will address issues related to data science or clustering. This clustering's primary significance is the fact that it will ensure resolution. It will ensure that the centroids' orientations that commence out smoothly. This centroids-calculating method iterates until it determines the right centroid. It will make the clusters of various sizes and forms simpler. The geographic coordinates of the image ought to try to find i x j first, after which the resulting images should be put together to create the k-clusters. Take into account that P_k is the function that emphasizes the set of data in pixel h (i, j). The Euclidean distance equation is given in Eq. (2):

Step 1: It is necessary to identify the center and the cluster K.

Step 2: For each pixel, the Euclidean distance E_d is indicated and is uttered in the Eq. (2):

$$E_d = \left||h(i,j) - Pk|\right| \quad (2)$$

Step 3: Most of the pixels are assigned to the center point using theE_d foundation.

Step 4: The work allocated to each pixel is completed, and the new center coordinates are recalculated by Eq. (3):

$$Pk = \frac{1}{k}\sum_{j\in Pk}\sum_{i\in Pk} h(i,j) \quad (3)$$

Step 5: Continue doing this until the requirement is met.

Were,

E_d is denoted as the Euclidean distance; P_k is denoted as the group of clusters; h (i, j) is denoted as the information pixel.

D. Feature Extraction using Variational Autoencoder

The procedure of choosing and displaying the most important facts or trends from unprocessed data is referred to as feature extraction [21]. It entails converting the initial information into a more condensed and useful form that may be applied to activities requiring modelling or further analysis. To improve algorithmic performance and efficiency, feature extraction is frequently used in machine learning and pattern recognition applications.

A finding in the latent space can be described probabilistically using a variational autoencoder (VAE) [22]. Thus, a distribution of probability for each underlying feature is employed to create the encoder instead of one that produces only one number to characterize every latent state feature. When using VAE, instances from the identical class wind up very near one another in the coding space, allowing for improved unsupervised representation learning. Moreover, the VAE is used in this research because it is used to detect the abnormalities in the medical images. The architectural diagram of VAE is shown in Fig. 3 and its components are explained below.

1) Input: The purpose of the application and area specificity determine the input to a Variational Autoencoder (VAE). The input for image-based VAEs, on the other hand, is frequently made up of images or patches of images.

2) Encoder: The encoder converts the input facts into the dormant space parameters associated with the distribution of probabilities. It usually consists of numerous levels of neural networks that gradually reduce the degree of dimensionality of the input data, including convolutional or fully connected layers. A collection of mean and variance vector that reflect the properties of a Gaussian distribution with multiple variables in the dormant space are the encoder's outcome.

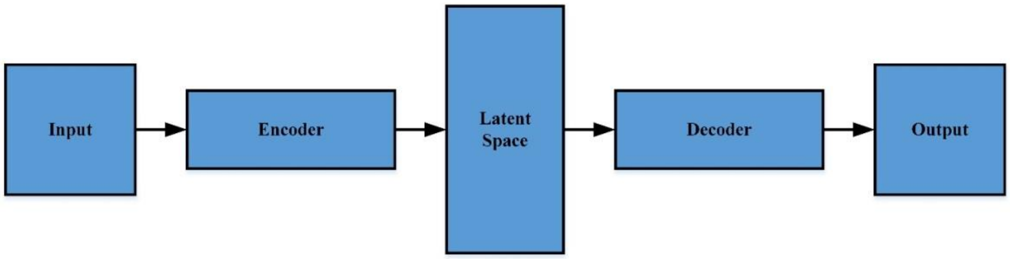

Fig. 3. Variational auto encoder architectural diagram.

3) Latent space: Every point in the latent space represents a latent code and is a reduced dimensional model of the input data. Moreover, the encoder gains the ability to create latent codes through training that precisely represent significant input data characteristics. Finding a clear format that includes the crucial data while eliminating the background noise is the goal.

4) Decoder: A sample of data from the latent space is taken by the decoder, which then transforms it to the initial input region. The decoder is prepared up of a number of strata of neural networks, similarly like the encoder and it gradually up samples the latent code to generate its result. The goal of the decoder's final product is to precisely reproduce the initial input information.

5) Output: A reconstruction of the incoming data is exactly the decoder delivers. The reconstruction loss, that gauges the variance among the input and the output, is the parameter that the VAE is trained to reduce. The VAE develops to produce useful reconstruction from latent codes by minimizing this loss of information.

The VAE possess some equations and is given under Eq. (4) and (5),

$$\mu_{pp}, \sigma_{pp} = R_{pp}(B_1, Y);\ \mu_{pv}, \sigma_{pv} = R_{pv}(B_2, Y) \quad (4)$$

$$\mu_{qq}, \sigma_{qq} = R_{qq}(B_2, Y);\ \mu_{qv}, \sigma_{qv} = R_{qv}(B_2, Y) \quad (5)$$

Where, R_{pp} and R_{pv} are determined as separators; B_1 and Y are the input and output latent variables; μ_{pp}, σ_{pp} are determined as the structural specific component; μ_{pv}, σ_{pv} are determined as the universal component.

6) Loss function: Variational Autoencoders (VAEs) train the algorithm using a loss function [23]. There were four hybrid loss functions are created. They are the classification loss (L_{cls}), the distangled cosine distance loss (L_{cos}), the reconstruct loss (L_{rec}) and the KL loss (L_{kl}). Each functions formula is given under Eq. (6), (7), (8) and (9).

$$L_{kl} = KL\left(Y_{pp}|N(0,1)\right) + KL\left(Y_{pv}|N(0,1)\right) + KL\left(Y_{qq}|N(0,1)\right) + KL\left(Y_{qv}|N(0,1)\right) \quad (6)$$

$$L_{rec} = \|B_{1'} - B_1\|2 + \|B_{2'} - B_2\|2 \quad (7)$$

$$L_{cos} = \frac{Y_{pv} \times Y_{qv}}{\|Y_{pv}\| \times \|Y_{qv}\|} \quad (8)$$

$$L_{cls} = -x \times \log(D(B_n)) \quad (9)$$

Where, x represents the one hot vector in the truth table.

E. Classification

The Convolution Neural Network (CNN) classifier identify the input images [24]. The visual representations are effectively examined by its multi-layered method, which also eliminates any necessary elements. Four layers make up the CNN classifier: input image, convolutional layer, pooling layer, fully connected layer and output. Through instruction, CNN is the instance that operates the quickest. The magnitudes of all input images' must be of identical dimensions. Fig. 4 depicts the CNN architecture.

7) Convolutional layer: After compiling a short sample of images, the convolution layer leverages all of the layers to examine the complexity of every image it gets. It has a strong relationship with the characteristics of the displayed photos. It is given in Eq. (10),

$$X_i = \sum Y_j * M_k + A_s \quad (10)$$

A_s- represents the bias term, * represents the convolutional operator, k^{th} filter convolutes by local region of Y_j called receptive field, X_i is the output of every filter.

8) Pooling layer: This layer limits the type and severity that the downstream sampling layer can use. The Pooling layer decreases the quantity of constraints, the feature map's computation quality and scale, training duration and excessive fitting. It is proceeded in Eq. (11) and Eq. (12),

$$\widetilde{x_1} = \frac{x_1 - \widetilde{m}}{\widetilde{p}} + 1 \quad (11)$$

$$\widetilde{x_2} = \frac{x_2 - \widetilde{m}}{\widetilde{p}} + 1 \quad (12)$$

X is the output layer; p is the pooling factor.

9) Fully connected layer: A fully connected layer has been employed to categorize the images. All of the following convolution layers are placed after the FC layers. Arranging the graphical representation amongst both the input and the output is made simpler by the FC layer. The top layers of the network are fully connected layers. The layer that is completely connected receives its input from the pooling layer's output. The algorithm for the proposed DL-VAE-CNN is assumed below and the flow chart for the planned methodology is displayed in Fig. 5.

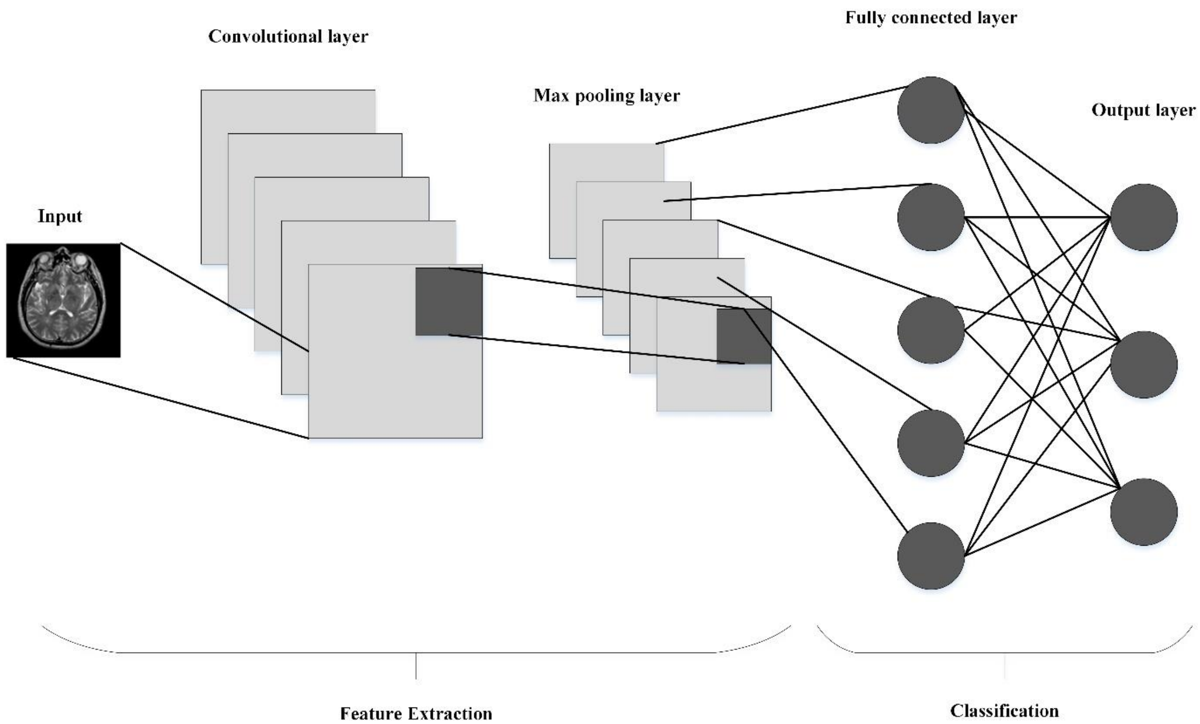

Fig. 4. Convolutional neural network architecture.

The framework's segmentation accuracy, efficiency, and resilience are greatly influenced by the choice and setup of parameters including learning rates, batch sizes, network designs, and hyper parameters. The capacity of the algorithm to handle various medical pictures, account for changes in image quality, and react to certain clinical settings may be dramatically impacted by fine-tuning these parameters. Additionally, the parameter selection may alter the amount of computer resources needed, reducing the algorithm's suitability for use in actual clinical settings. Therefore, further investigation and optimization of these parameters is necessary to guarantee the DL-VAE-CNN algorithm's optimal performance and clinical applicability for automated medical picture segmentation tasks.

Algorithm for DL-VAE-CNN

Input: *Medical image of brain*
Output: *Classification of medical image*
$Y = \{Y_1, Y_2, Y_3, ...\}$ *// LGG segmentation dataset*
Image Pre-processing
Apply Histogram Equalization to enhance image contrast
Histogram Equalization equation is given by Eqn. (1)
Image Segmentation using K-means Clustering
Detect the cluster centres using the K-means algorithm
Calculate Euclidean distance E_d as in Eqn. (2)
Assign pixels to the nearest cluster centre using E_d
Update cluster centres using the recalculated coordinates in Eqn. (3)
Repeat until convergence
Feature Extraction using VAE
Implement the VAE architecture
Encoder network with Eqn. (4) and (5) for mean and variance
Sample latent variables using the parameterization trick
Decoder network reconstructs the input image
Calculate loss functions as in Eqn. (6), (7), (8), and (9)
Train the VAE using back propagation and optimization
Classification using CNN
Implement a CNN architecture for classification
Apply convolutional layers with Eqn. (11) and (12) for feature extraction
Apply pooling layers to down sample the feature maps
Flatten the feature maps and connect to fully connected layers
Implement softmax layer for multi-class classification
Training and Evaluation
Update VAE and CNN parameters iteratively through back propagation
Monitor loss functions and classification accuracy during training
Inference
Feed an unseen medical image through the trained DL-VAE-CNN framework
The framework will automatically segment and classify the image
Output
The DL-VAE-CNN algorithm outputs a classification label for the input medical image

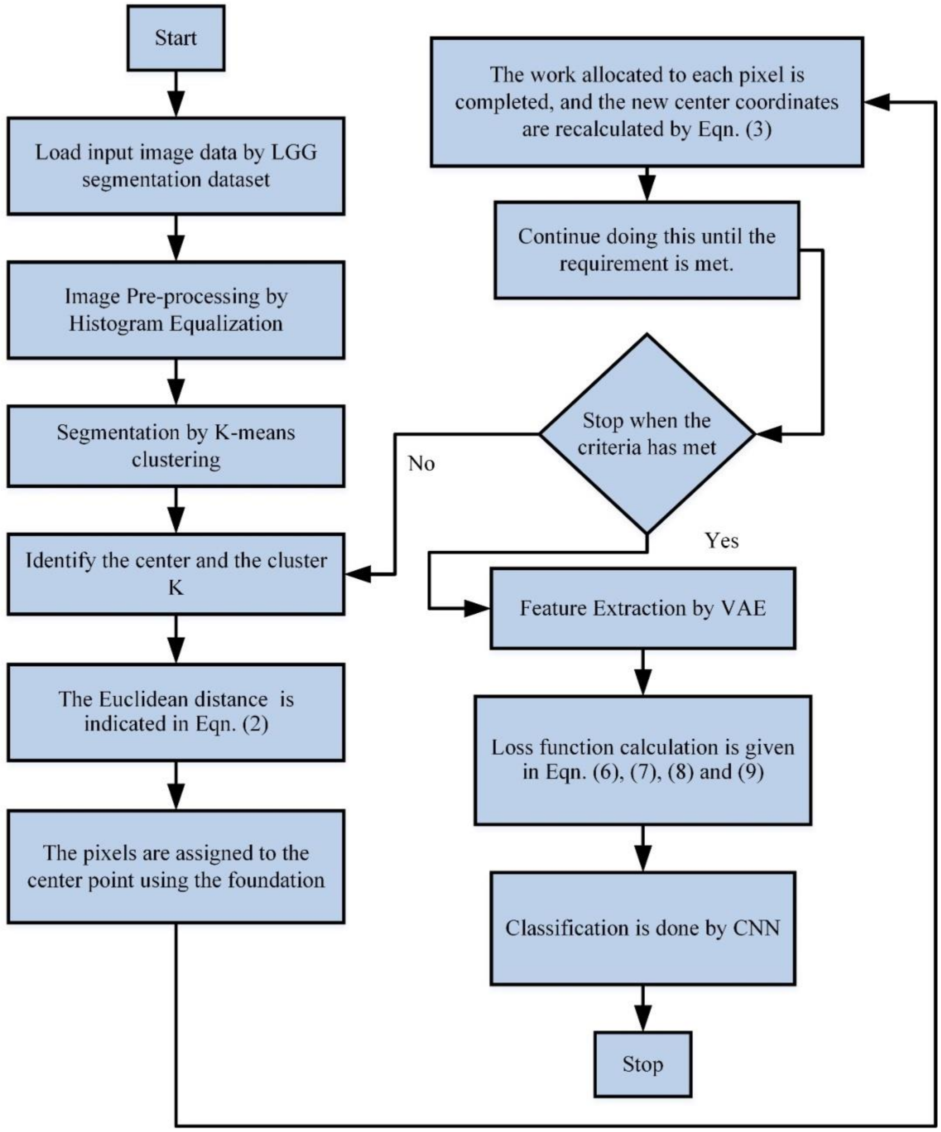

Fig. 5. Flow chart of proposed DL-VAE-CNN model.

V. Results and Discussions

Deep learning, variational autoencoders and conditional neural networks were used to attain highly precise results for segmentation in the suggested automated medical image segmentation methodology. Comparison with existing methods revealed considerable gains in segmentation clarity as indicated by measures like precision, accuracy, f1 score and recall. By efficiently employing the previously acquired hidden representations from variational autoencoders, the structure demonstrated adaptability to changes in the quality of images, especially changes in illumination, contrary and noise. By using condensed latent visualizations, the system also showed that it is capable of segmenting data quickly and effectively. The framework's generalization capabilities were especially outstanding because they allowed it to deal with a variety of anatomical characteristics and modalities for medical imaging with ease. These outcomes demonstrate that the suggested methodology for computerized medical image segmentation is efficient.

A. Performance Metrics Evaluation

1) Accuracy: The percentage of uniqueness among a calculation and its honest worth is known as accuracy. It is the ratio of precisely intended information to all observations. It is given in Eq. (13).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (13)$$

2) Precision: The level of accuracy or closeness between multiple computations is referred to as precision. The relationship between precision and accuracy reveals how reproducible the measurement is. Its formula is given in Eq. (14).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (14)$$

3) Recall: The ratio of all suitable results that were precisely organized by the method used is known as recall. The correct affirmative to true positive and false negative values proportion is used to characterize it. It is utilized in Eq. (15).

$$\text{Recall} = \frac{TP}{TP+FN} \quad (15)$$

4) F1-Score: Precision and recall are mutual to regulate the F1-score. When computing the F1 score, precision and recall levels are extremely important. Its formula is given in Eq. (16).

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision}+\text{Recall}} \quad (16)$$

Table I and Fig. 6 shows the assessment and performance evaluation of Accuracy. When comparing the proposed DL-VAE-CNN methods accuracy parameter with the subsequent three existing methods, i) CNN procedure [25], ii) VGG16-CNN [26], iii) DCNN [27], the proposed algorithm produces greater accuracy of about (99.91%).

TABLE I. COMPARISON TABLE OF PROPOSED METHOD'S ACCURACY WITH EXISTING METHOD'S ACCURACY

Method	Accuracy (%)
CNN	95.44
VGG16-CNN	93.74
DCNN	98.51
Proposed Method	99.91

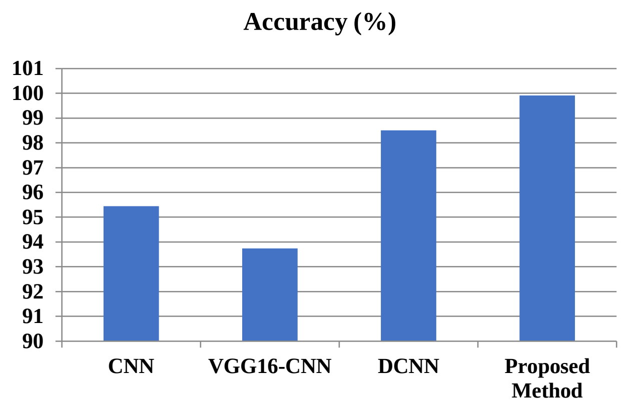

Fig. 6. Comparison graph of accuracy.

Table II and Fig. 7 shows the assessment and performance evaluation of Precision. When comparing the proposed DL-VAE-CNN methods precision parameter with the subsequent three existing methods, i) CNN procedure [25] ii) VGG16-CNN [26], iii) DCNN [27], the proposed algorithm produces greater precision of about (99.90%).

TABLE II. COMPARISON TABLE OF PROPOSED METHOD'S PRECISION WITH EXISTING METHOD'S PRECISION

Method	Precision (%)
CNN	91
VGG16-CNN	92
DCNN	99.18
Proposed Method	99.90

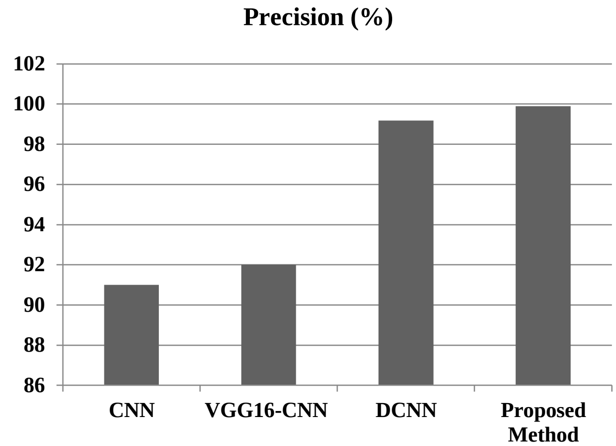

Fig. 7. Comparison graph of precision.

Table III and Fig. 8 shows the assessment and performance evaluation of Recall. When comparing the proposed DL-VAE-CNN methods recall parameter with the subsequent three existing methods, i) CNN procedure [25] ii) VGG16-CNN [26], iii) DCNN [27], the proposed algorithm produces greater recall of about (98.99%).

TABLE III. COMPARISON TABLE OF PROPOSED METHOD'S RECALL WITH EXISTING METHOD'S RECALL

Method	Recall (%)
CNN	95
VGG16-CNN	92.1
DCNN	97.90
Proposed Method	98.99

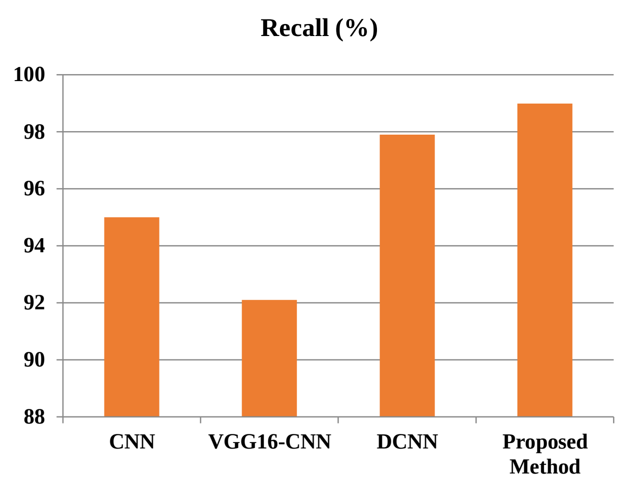

Fig. 8. Comparison graph of recall.

Table IV and Fig. 9 shows the assessment and performance evaluation of F1-Score. When comparing the proposed DL-VAE-CNN methods f1-score parameter with the subsequent three existing methods, i) CNN procedure [25] ii) VGG16-CNN [26], iii) DCNN [27], the proposed algorithm produces greater f1-score of about (99.99%).

TABLE IV. COMPARISON TABLE OF PROPOSED METHOD'S F1-SCORE WITH EXISTING METHOD'S F1-SCORE

Method	F1-Score (%)
CNN	93
VGG16-CNN	67.08
DCNN	98.53
Proposed Method	99.99

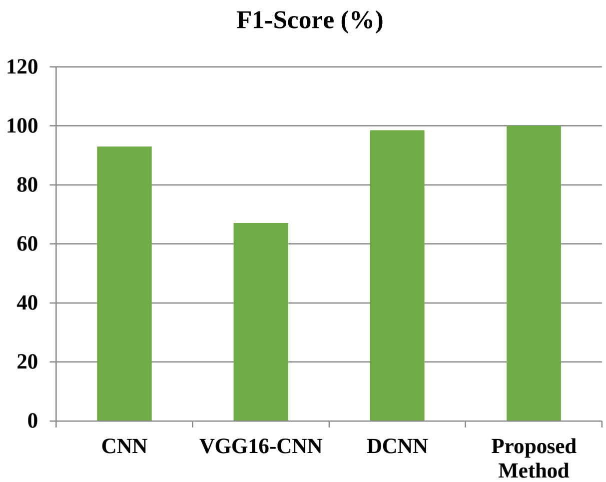

Fig. 9. Comparison graph of F1-score.

B. *Discussions*

An automated medical image segmentation framework using Deep Learning and Variational Auto Encoder with Convolutional Neural Network is employed in this paper. This section explains about the things that have explained above in a short way. Initially, the introduction, related works, problem statements are completed. Next, it steps on to the methodology part. Here, it discusses about the planned method that is used in this paper. Fig. 1 shows the architectural illustration of the planned DL-VAE-CNN technique. Fig. 2 shows the Histogram Equalization of the healthcare image. Next, the segmentation process takes place and it is done by K-means clustering. It is followed by that the feature extraction has done. Fig. 3 shows the architectural diagram of VAE and followed by that its components are explained. Next, is the classification process. It is done by CNN. Fig. 4 shows the architectural diagram of CNN and followed by that the various layers of CNN has been explained. Next, is the algorithm of DL-VAE-CNN. Fig. (5) shows the flow chart of DL-VAE-CNN. Next, is the results and discussion's part. Here, the various performance metrics like Precision, Accuracy, F1-score and Recall are associated with existing methods to show the planned method will produce better results. Table I gives the comparison table of accuracy and Fig. 6 shows the evaluation graph of accuracy. Table II gives the comparison table of precision and Fig. 7 shows the evaluation graph of precision. Table III gives the comparison table of recall and Fig. 8 shows the evaluation graph of recall. Table IV gives the comparison table of f1-score and Fig. 9 shows the evaluation graph of f1-score. Finally, the discussion part and followed by that conclusion and the future work of the paper is provided.

VI. CONCLUSION AND FUTURE WORK

In the field of medical imaging, establishing trustworthy image authentication to establish correlations between images is a difficult and critical task that is addressed in this study. The generation of tissue atlases, tumor development tracking, and image fusion are just a few therapeutic applications where this association is crucial. The paper offers a fresh method for addressing this issue, using CNNs and VAEs for healthcare data separation as deep learning approaches. Due to the requirement for precise image division at a low-level spatial data level, the fundamental machine vision task of image segmentation is extremely difficult. The suggested approach uses CNNs and VAEs in a conditioned environment to integrate latent representations that VAEs can produce using. When the trained model delineates pertinent regions in recently collected medical images during the inferential step, it successfully demonstrates its efficacy. The experimental findings drawn from several medical imaging databases highlight the increased accuracy in segmenting anatomical entities. This accuracy highlights the substantial contribution of this method to the field of medical imaging and expands the possibilities of automated diagnosis and therapy. The complex pixel-level classification issue, which has presented difficulties in earlier studies, is effectively addressed by the suggested Deep Learning and DL-VAE-CNN approach. This research provides a strong framework that has the potential to revolutionize image identification and segmentation in the context of medical imaging by using the capabilities of deep learning, latent space representation, and conditional neural networks. Deep learning frameworks sometimes require a lot of processing power, both during inference and training. Deploying it in clinical settings with limited resources may thus be difficult.

REFERENCES

[1] B. Ahmad, J. Sun, Q. You, V. Palade, and Z. Mao, "Brain Tumor Classification Using a Combination of Variational Autoencoders and Generative Adversarial Networks," Biomedicines, vol. 10, no. 2, p. 223, Jan. 2022, doi: 10.3390/biomedicines10020223.

[2] W. T. Le, F. Maleki, F. P. Romero, R. Forghani, and S. Kadoury, "Overview of Machine Learning: Part 2," Neuroimaging Clin. N. Am., vol. 30, no. 4, pp. 417–431, Nov. 2020, doi: 10.1016/j.nic.2020.06.003.

[3] N. Sharma et al., "Automated medical image segmentation techniques," J. Med. Phys., vol. 35, no. 1, p. 3, 2010, doi: 10.4103/0971-6203.58777.

[4] V. Sandfort, K. Yan, P. M. Graffy, P. J. Pickhardt, and R. M. Summers, "Use of Variational Autoencoders with Unsupervised Learning to Detect Incorrect Organ Segmentations at CT," Radiol. Artif. Intell., vol. 3, no. 4, p. e200218, Jul. 2021, doi: 10.1148/ryai.2021200218.

[5] H. Uzunova, S. Schultz, H. Handels, and J. Ehrhardt, "Unsupervised pathology detection in medical images using conditional variational autoencoders," Int. J. Comput. Assist. Radiol. Surg., vol. 14, no. 3, pp. 451–461, Mar. 2019, doi: 10.1007/s11548-018-1898-0.

[6] R. Wei and A. Mahmood, "Recent Advances in Variational Autoencoders With Representation Learning for Biomedical Informatics: A Survey," IEEE Access, vol. 9, pp. 4939–4956, 2021, doi: 10.1109/ACCESS.2020.3048309.

[7] P. Naga Srinivasu, T. B. Krishna, S. Ahmed, N. Almusallam, F. Khaled Alarfaj, and N. Allheeib, "Variational Autoencoders-BasedSelf-Learning Model for Tumor Identification and Impact Analysis from 2-D MRI Images," J. Healthc. Eng., vol. 2023, pp. 1–17, Jan. 2023, doi: 10.1155/2023/1566123.

[8] P. Celard, E. L. Iglesias, J. M. Sorribes-Fdez, R. Romero, A. S. Vieira, and L. Borrajo, "A survey on deep learning applied to medical images:

from simple artificial neural networks to generative models," Neural Comput. Appl., vol. 35, no. 3, pp. 2291–2323, Jan. 2023, doi: 10.1007/s00521-022-07953-4.

[9] I. Rizwan I Haque and J. Neubert, "Deep learning approaches to biomedical image segmentation," Inform. Med. Unlocked, vol. 18, p. 100297, 2020, doi: 10.1016/j.imu.2020.100297.

[10] X. Chen et al., "Recent advances and clinical applications of deep learning in medical image analysis," Med. Image Anal., vol. 79, p. 102444, Jul. 2022, doi: 10.1016/j.media.2022.102444.

[11] Z.-S. Liu, W.-C. Siu, and Y.-L. Chan, "Photo-Realistic Image Super-Resolution via Variational Autoencoders," IEEE Trans. Circuits Syst. Video Technol., vol. 31, no. 4, pp. 1351–1365, Apr. 2021, doi: 10.1109/TCSVT.2020.3003832.

[12] H. Uzunova, J. Ehrhardt, T. Kepp, and H. Handels, "Interpretable explanations of black box classifiers applied on medical images by meaningful perturbations using variational autoencoders," in Medical Imaging 2019: Image Processing, E. D. Angelini and B. A. Landman, Eds., San Diego, United States: SPIE, Mar. 2019, p. 36. doi: 10.1117/12.2511964.

[13] W. Yu, M. Zhang, and Y. Shen, "Spatial Revising Variational Autoencoder-Based Feature Extraction Method for Hyperspectral Images," IEEE Trans. Geosci. Remote Sens., vol. 59, no. 2, pp. 1410–1423, Feb. 2021, doi: 10.1109/TGRS.2020.2997835.

[14] G. Peyré and M. Cuturi, Computational Optimal Transport. now Publishers Inc, 2019. doi: 10.1561/9781680835519.

[15] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey." arXiv, Nov. 14, 2020. Accessed: Jul. 06, 2023. [Online]. Available: http://arxiv.org/abs/2001.05566

[16] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi, "Medical image segmentation using deep learning: A survey," IET Image Process., vol. 16, no. 5, pp. 1243–1267, Apr. 2022, doi: 10.1049/ipr2.12419.

[17] V. Edupuganti, M. Mardani, S. Vasanawala, and J. Pauly, "Uncertainty Quantification in Deep MRI Reconstruction," IEEE Trans. Med. Imaging, vol. 40, no. 1, pp. 239–250, Jan. 2021, doi: 10.1109/TMI.2020.3025065.

[18] T. Le-Tien, T.-N. To, and G. Vo, "Graph-based signal processing to convolutional neural networks for medical image segmentation," SEATUC J. Sci. Eng., vol. 3, no. 1, pp. 9–15, 2022.

[19] R. J. S. Raj, S. J. Shobana, I. V. Pustokhina, D. A. Pustokhin, D. Gupta, and K. Shankar, "Optimal Feature Selection-Based Medical Image Classification Using Deep Learning Model in Internet of Medical Things," IEEE Access, vol. 8, pp. 58006–58017, 2020, doi: 10.1109/ACCESS.2020.2981337.

[20] Dr. S. Manoharan, "Performance Analysis of Clustering Based Image Segmentation Techniques," J. Innov. Image Process., vol. 2, no. 1, pp. 14–24, Mar. 2020, doi: 10.36548/jiip.2020.1.002.

[21] M. Imani and H. Ghassemian, "An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges," Inf. Fusion, vol. 59, pp. 59–83, Jul. 2020, doi: 10.1016/j.inffus.2020.01.007.

[22] K.-L. Lim, X. Jiang, and C. Yi, "Deep Clustering With Variational Autoencoder," IEEE Signal Process. Lett., vol. 27, pp. 231–235, 2020, doi: 10.1109/LSP.2020.2965328.

[23] Q. Zuo, Y. Zhu, L. Lu, Z. Yang, Y. Li, and N. Zhang, "Fusing Structural and Functional Connectivities using Disentangled VAE for Detecting MCI." arXiv, Jun. 16, 2023. Accessed: Jul. 07, 2023. [Online]. Available: http://arxiv.org/abs/2306.09629

[24] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges," J. Digit. Imaging, vol. 32, no. 4, pp. 582–596, Aug. 2019, doi: 10.1007/s10278-019-00227-x.

[25] M. A. Mahjoubi, S. Hamida, O. E. Gannour, B. Cherradi, A. E. Abbassi, and A. Raihani, "Improved Multiclass Brain Tumor Detection using Convolutional Neural Networks and Magnetic Resonance Imaging," Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 3, 2023, doi: 10.14569/IJACSA.2023.0140346.

[26] A. Alshammari, "Construction of VGG16 Convolution Neural Network (VGG16_CNN) Classifier with NestNet-Based Segmentation Paradigm for Brain Metastasis Classification," Sensors, vol. 22, no. 20, p. 8076, Oct. 2022, doi: 10.3390/s22208076.

[27] L. Alzubaidi et al., "Novel Transfer Learning Approach for Medical Imaging with Limited Labeled Data," Cancers, vol. 13, no. 7, p. 1590, Mar. 2021, doi: 10.3390/cancers13071590.