Introduction

Brain tumors are among the most severe and deadly tumors that affect both adults and children. According to the World Health Organization (WHO) classification standards, The malignancy of brain tumors, also known as central nervous system tumors, is further classified into grades I (benign) to grade IV (high malignancy) [1]. In 2015, brain tumors were discovered in about 23,000 patients in the United States, according to cancer data [2]. Statistics show that after two years, this tumor is thought to be the primary cause of cancer-related deaths in the world for both adults and children [2].

Early brain tumor detection is crucial for treating patients appropriately and gauging how well they will respond to the selected course of treatment. Nowadays, the use of digital images in the medical field is increasingly common for diagnosis [13]. Specialized radiologists employ imaging methods to identify tumors. Of the several methods accessible for the non-invasive detection of brain tumor, Magnetic Resonance Imaging (MRI) is usually preferred [13]. However, the categorization process of brain tumors classification is time-consuming, relies on radiologist expertise, and gets more complicated with the increase in the number of patients [2]. The impact is that the daily volume of data that requires analysis is growing, resulting in high costs and the risk of errors in visual interpretation. Furthermore, categorizing brain tumors that include multiple pathological types is more challenging than simple binary classification. Incorrect diagnoses of brain tumors can lead to significant issues and drastically diminish a patient's survival chances [2]. To mitigate this downside, it is crucial to focus on developing an automated image processing system [2].

Brain tumor detection is conducted using Machine Learning (ML) and Artificial Intelligence (AI) approaches, each of which has different features [3]. Deep Learning (DL) is a branch of machine learning that eliminates the necessity for handcrafted features. It has been shown to be effective in bridging the gap between human and computer vision in pattern recognition, offering superior performance over traditional methods [2]. As one of the primary computational intelligence methods for medical imaging, deep learning is extensively utilized in medical image classification, which is a crucial aspect of pattern recognition. This approach, based on deep learning, is commonly employed in automated disease diagnosis systems [4].

The most successful deep learning architecture for image analysis to date is the Convolutional Neural Network (CNN) [5]. This model includes convolutional layers for automated segmentation and feature extraction, followed by traditional neural networks for classification tasks [6]. CNNs are predominantly utilized for classifying images, segmenting images, detecting objects, processing videos, natural language processing, and recognizing speech [7]. In computer vision, CNNs are the primary approach and a widely favored architecture in medical image recognition [8]. Numerous studies have applied CNNs to predict brain tumors from medical images, one of which was conducted by Azizy et al. in 2023 [9].

In 2014, Hemanth et al. conducted a study aimed at enhancing the performance of non-iterative artificial neural networks for the classification of abnormal brain images from MRI scans. They introduced two modified neural network approaches, revealing that the altered Kohonen Neural Network reached a sensitivity of 0.90, a specificity of 0.96, and an accuracy of 95%. In contrast, the modified Counter Propagation Neural Network achieved a sensitivity of 0.95, a specificity of 0.98, and an accuracy of 98% [4]. In 2018, Afshar et al. used the Capsule Network, a sophisticated CNN architecture to categorize brain tumors and surrounding tissues from MRI images. According to their research, this model achieved the highest success rate of 86.56% in brain tumor segmentation, while its accuracy for raw brain images was 72.13% [10]. In a 2023 study by Zhongxiao Li et al., a Vision Transformer model with weak supervision was employed to analyze primary brain tumor histopathology images from MRI scans. The results indicated that the ViT-WSI model achieved patient-level AUC scores of 0.960, 0.874, and 0.845, respectively [11]. Another 2023 study by Azizy et al. used a CNN model optimized with the Cuckoo Search Algorithm to predict brain tumors from MRI images. The findings showed that the developed model achieved a success rate of approximately 92.6% for brain tumor detection [9].

CNNs have demonstrated superior performance on extensive, labelled datasets, such as ImageNet, which comprises over a million images. However, utilizing deep CNNs in the medical domain presents several challenges. One major challenge is the limited size of medical datasets, necessitating expert radiologists to manually inspect and label the images; a process that is time-consuming, exhausting, and costly. Furthermore, training deep CNNs on small datasets is difficult due to overfitting and convergence issues. Expertise is needed to iteratively refine models and adjust learning parameters to enhance performance [2]. Alexey Dosovitskiy et al. showed that reliance on CNNs is not mandatory and that a pure transformer, when applied directly to sequences of image patches, can achieve outstanding performance on image classification tasks [12]. Vision Transformers (ViTs) have emerged as an AI technique for medical image classification. The Vision Transformer (ViT) method has garnered significant interest from researchers in computer vision because of its superior performance across numerous computer vision tasks [8]. Employing heavy training or altering the ViT architecture has become a prevalent strategy for applying

ViTs to computer vision tasks [8]. Once pre-trained on large datasets and transferred to various medium or small-sized image recognition tasks, Vision Transformers deliver excellent results compared to state-of-the-art convolutional networks and require significantly fewer computational resources to train [12].

This study aims to detect brain tumors in MRI scan data using the Vision Transformer. This deep learning model is chosen for its use of the ViT technique, which offers unique advantages for brain tumor detection. ViT processes images as sequences of patches, capturing global contextual information more effectively. This is particularly beneficial for medical imaging, enhancing diagnostic accuracy. These strengths make the Vision Transformer an excellent choice for accurately identifying brain tumors in MRI scans. This study is conducted to measure the accuracy achieved in detecting brain tumors using the proposed method. The contribution of this study is to emphasize the potential of ViT in image classification tasks.