STREAMLINING BRAIN TUMOR DIAGNOSIS: KNOWLEDGE DISTILLATION FOR COMPUTATIONAL EFFICIENCY

A PROJECT REPORT

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
ARTIFICIAL INTELLIGENCE

Submitted by

Jaiminkumar Moghariya (2K22/AFI/09)

Under the supervision of

Dr. Prashant Giridhar Shambharkar

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi 110042

MAY, 2024

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

ACKNOWLEDGEMENT

We wish to express our sincerest gratitude to Dr. Prashant Giridhar Shambharkar for his continuous guidance and mentorship that he provided us during the project. He showed us the path to achieve our targets by explaining all the tasks to be done and explained to us the importance of this project as well as its industrial relevance. He was always ready to help us and clear our doubts regarding any hurdles in this project. Without his constant support and motivation, this project would not have been successful.

Place: Delhi Jaiminkumar Moghariya

Date:

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Jaiminkumar Moghariya, Roll No – 2K22/AFI/09 students of M.Tech Artificial Intelligence,hereby declare that the thesis titled "Streamlining Brain Tumor Diagnosis: Knowledge Distillation for Computational Efficiency" which is submitted by me to the Department of Computer Science & Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi Jaiminkumar Moghariya

Date:

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "Streamlining Brain Tumor Diagnosis: Knowledge Distillation for Computational Efficiency" which is submitted by Jaiminkumar Moghariya, Roll No – 2K22/AFI/09, Department of Computer Science & Engineering ,Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the students under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi Dr. Prashant Giridhar Shambharkar

Date: **SUPERVISOR**

Abstract

Brain tumors are biological conditions that are complex and need proper diagnosis at the earliest in order to receive medical treatments. Two limitations of the traditional diagnostics methods such as magnetic resonance imaging (MRI) analysis are time-consuming and computational expensive. In the following regard, it is relevant to use knowledge distillation for accelerating the process of brain tumor diagnosis since it is facilitated through the transmission of the logit information from a large, very efficient model into a small, very effective model. The final distilled model is compared with the original model, and benchmark datasets are used for a competitive and quality comparison in terms of accuracy, computational cost, and inference time.

This study demonstrates an interesting potential of KD to optimize treatments for brain tumors, moving toward better and more widespread diagnosis. The database contains tumor and nontumor scans from the MRI model of the brain, which will ensure some appropriate way of value judgment. This is followed by data preprocessing, with the MRI images at a quality and significance level to be fed into the model. Initially, a high-performing baseline could be set up through training a model with Convolutional Neural Network on the preprocessed dataset. In this way, we get distillation by transferring the CNN knowledge to some compact model so that high accuracy is maintained for reducing computational demand. This approach promises to make advanced diagnostic capacity more available, especially in resource-constrained settings, with an ultimate improvement of patient outcomes realized through quicker and more efficient diagnoses.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

Brain tumors are a very serious public health issue, whereby they affect millions of persons throughout the world. In fact, in the year 2000, June 8th became recognized as World Brain Tumor Day—a day when an increasing awareness and the needs for more research sprang up in the area of brain health. Generally, the brain responds to the body's nervous system, which consists of about 100 billion nerve cells; all these perform many functions that ensure the whole body's health and well-being. Disorders from brain tumors can cause catastrophic neurological deficits and really lower the quality of life for a person. The American Brain Tumor Society had this estimated that at present about 700,000 Americans are living with brain tumors but only 36% can survive beyond five years since it was first diagnosed.

According to scientists' estimate, more than 256,000 new cases of malignant brain tumors will be detected around the world by the end of 2023. This shocking statistic really underscores the increased incidence and need for global health initiatives to be taken. Further, in this year alone, up to 84,170 brain tumors have been reported worldwide, showing a sharp rise and therefore signaling the need to continue to work on awareness and research in this area. World Brain Tumor Day reminds the world that people with brain tumors and their families face so many hardships. It calls upon public support for scientific researches and advancements in health in order to reduce the number and effect of brain tumors worldwide [3].

Timely and accurate diagnosis forms the basic foundation of optimal treatment choices and patient outcomes. With a correct diagnosis, a more focused plan of care can be developed, increasing the likelihood of an effective intervention and recovery. It is one of the most accurate and cost-effective methods available for identifying serious illnesses such as brain tumors, stomach cancer, lung cancer, and skin cancer. Medical imaging technologies, a prime example of which is MRI, CT scans, PET scans, and ultrasound, can pull out the details within the structures, which may help in early detection of any abnormality.Medical imaging plays a pivotal role in the early identification and diagnosis of various cancers. Size, location, and type details of a brain tumor, for example, can be clearly shown by highly detailed scanning that are revealed by MRI and CT— all three important factors in the choice of the right way to treat the disease. In cases of cancer of the stomach and lung malignancies, imaging techniques such as endoscopic ultrasound and low-dose CT scans may be able to trace at an early stage when it is most easily treated. In case of skin cancer, most advanced imaging tools in dermoscopy and confocal microscopy can make possible the noninvasive examination of the lesion by skin for early diagnosis so as to prevent the progression of the disease. These professionals enable better patient outcomes in tailor-made planning of treatment when advanced imaging technologies make more accurate diagnoses possible [4].

It is not surprising that this method is applied most often for the first diagnosis of brain malignancy, as MRI image analysis reveals even the smallest and appropriate images of brain tumors. This technique produces highly detailed pictures of the structure of the brain and provides the most accurate information about the size, location, and type of tumor. High detail is important for strategies of treatment either revealed in surgery or radiological, or pharmacological procedures. MRI images have great clarity and resolution, making them an indispensable tool in the initial assessment and follow-up of brain tumors.However, MRI is a complex assessment that demands high expertise and processing capabilities. It mainly includes intricate anatomy of the brain and differentiation between very subtle differences in MRI scans of normal and abnormal tissue. Specialized training is required to interpret MRI images for a diagnosis, either for radiologists or healthcare service providers. Furthermore, MRI data processing usually involves quite high computational power, which is often applied using advanced software modules and algorithms to rebuild and improve images. It becomes imperative, therefore, to provide for competent manpower and sophisticated technology to deliver a better and timely diagnosis for improved patient outcomes[5].

This is corroborated by the fact that traditional means of diagnosing a tumor in the brain are mostly subjective, labor-intensive kinds of physical examination systems that are bound to be prone to errors; the implication may be a delay in diagnosis and onset of treatment. Researchers have, as a result of those gaps, taken up ways through which artificial intelligence and machine learning can be used for an automated brain tumor diagnosis. These advanced methodologies promise to enable even broader objectiveness, efficiency, and accuracy of diagnostic tools, significantly helping health service providers to make the right clinical decision without any delay.More specifically, deep neural network models perform well in most medical imaging applications, such as brain tumor segmentation. Both kinds of models utilize very complicated algorithms to process MRI or CT images and identify areas with abnormal tissue at a very high precision rate. Quick and accurate image segmentations using deep learning are necessary to quickly and accurately locate the boundaries of tumors, which then enable the planning and monitoring of treatment to be fast and precise. It should, however, be noted that the application of deep neural networks for these purposes has a number of challenges, from requiring colossal processing power to making the results of algorithm development and implementation expertly understandable to the user. However, other more complex technologies may not easily be adopted in resource-constrained healthcare environments, meaning infrastructure and training need to be in place to yield the most from these AI-based diagnostic tools for brain tumors and other medical conditions [6][7].

In most recent years, deep learning came to form a powerful tool for enhancing the efficacy of computer-assisted healthcare diagnostics, much more in the present context of brain cancer research. It assumes much importance as it deals with the diagnosis and treatment of several serious illnesses, from brain tumors and skin cancers to neurological diseases. Deep learning, comprising convolutional neural networks, recurrent neural networks, and transformer models, promises automation of the diagnostic process and is likely to increase accuracy in detecting diseases. Most techniques that researchers utilize for adapting large-scale experience from other domains to pre-trained models, hence fine-tuning for capitalization on medical imaging datasets, include transfer learning. Through deep learning, this team of healthcare experts could expedite the process of diagnosis, respond in a timely manner for prompt intervention, and produce better patient outcomes in the challenging landscape of brain cancer and related medical conditions[8].

The classification of brain tumors and color picture identification with the help of modern deep learning methods within the framework of the Faster R-CNN algorithm was carried out. Based on the application of an advanced Faster R-CNN, three types of brain tumors were differentiated: meningiomas, gliomas, and pituitary tumors. Using this technology, detected cancer areas considerably improved in terms of accuracy of bounds and associated confidence scores. This technique not only boosted better localization of the tumors but also enabled ease in the diagnostic process; it would help the physicians to make therapeutic decisions on time to manage the patient better. This derivative methodology with deep learning methods in the Faster R-CNN is promising because it opens new ways to modernize computer-assisted diagnostics and patient care in neuro-oncology.[9].

This area of neural network research has attracted much more attention and given the most promising efficiency improvements in practice for knowledge distillation. The method involves the training of an inferior, small, low-complexity neural network which, in turn, learns to mimic the full performance and behavior of a superior, larger, and more complex model. This is often referred to as the "model student" and its superior cousin as the "teacher." This is indeed all a task of distilling knowledge out of the teacher model in a palatable manner into the more amenable student model. The method simplifies the learning process and reduces the resulting models to accurate and efficient. More specifically, knowledge distillation has become a key approach to image classification tasks, where computational resources and power consumption are taken into account when creating deployable models on low-power devices. Its potential importance was brought out in studies addressing brain tumor identification and classification, mainly with the adoption of knowledge distillation techniques to show significant promise for more accessible and effective diagnostic tools in neuro-oncology.

Chapter 2

RELATED WORK

In this section, a number of related works with respect to the Brain Tumor Diagnosis (BTD) are presented, ranging from the traditional transfer learning pre-trained models to some of the very novel pre-trained models. All these efforts are combined to yield improvements in diagnostic capabilities within neuro-oncology through the use of state-of-the-art machine learning approaches. More recently, the use of pretraining models trained on generic image datasets such as ImageNet has been repurposed to help a neural network in the domain of medial imaging tasks. In fact, very recent efforts have been aimed at producing new pretrained models tailored to diagnosing brain tumors and increasing diagnostic accuracy and effectiveness.

The following subsections enlarge on the related works in terms of the key findings, methodologies, and breakthroughs along this very line. In doing so, a view of the works accomplished in this diverse area of brain tumor diagnostics is given through pretrained models in the process of being changed.

2.1 BTD based on Pre-Trained Models and Optimization Algorithms

However, there are several challenges in the accurate and timely diagnosis of brain tumors, as follows. One major challenge is that there is a lot of similarity in the imaging characteristics for various conditions of the tumor; hence it becomes hard for practitioners to always distinguish between them. Secondly, medical expertise and advanced equipment are mostly limited, more so in resource-constrained environments. These factors delay the diagnosis and, in some cases, the initiation of treatment, eventually affecting the outcomes for patients. In an effort to overcome such roadblocks, innovative solutions directed at better diagnosing brain tumors have been one of the core areas with which machine learning and deep learning techniques are developed.

Machine and deep learning approaches have shown quite great promise in further advancing the accuracy and efficiency achieved with brain tumor diagnosis. The latter will rely on pre-trained models and optimization algorithms, developed by the researchers as part part of building a reliable diagnostic tool for accurate identification and classification of tumors in the brain based on medical imaging data. The advanced methods used will seek to address the challenges presented by the conventional screening methods in the context of automated and objective imaging scans analysis.

They can also assist in supplementing the expertise of medical professionals with decision support for healthcare professionals, especially in settings where such specialized expertise is lacking. Indeed, various works have been done on how pre-trained models and optimization algorithms can be applied to the process in diagnosing brain tumors. Not one but several highly advanced steps are achieved in the studies that prove the high accuracies of tumor detection and classification. Using machine learning and deep learning, the researchers worked to revolutionize the diagnosis of brain tumors, thereby improving the results of the patients.

For example, Sultan H. Hossam et al.[7] in 2019 proposed a new network architecture and index to get better results in diagnosing brain tumors with more accuracy, substantiated by empirical facts: the high-performance accuracy reached up to 98.7%. It was able to identify the glioblastoma-related tumors and classify such into the concrete type of three most common kinds of primary tumors, including meningioma, glioma, and pituitary tumors. Moreover, it allowed for differentiation in gliomas' grades: Grade II, Grade III, and Grade IV. Therefore, it was an important tool for detailed insight, so important for real diagnosis or therapeutic decision. All these features make this comprehensive framework a great step toward improving precision and efficacy in brain tumor diagnosis and provide valuable support to healthcare professionals during clinical decision-making processes.

In a more advanced study, Ahmad Salen et al. [10] used an early 2020 brain tumor dataset to further investigate the effectiveness of pre-trained models in medical imaging tasks. For example, in an advanced study, five pre-trained models—Xception, Inception-V3, ResNet-50, VGG-16, and MobileNet—were used in evaluating the performance metrics on previously observed scans. In fact, that study showed very good F1-scores from 97.25% to 98.75% across the different pre-trained models. These results only confirm the robustness and reliability of pretained models to cope with all sophistications that may be present at medical imaging, and more specifically when diagnosing brain tumors. Such findings hint at the benefit that application of these developed models could potentially lead to in terms of improving the accuracy in diagnostic procedures and optimizing the clinical workflow at health care facilities.

In the study by S. Asif et al. [11], various optimization algorithms are queried for Nas-Net Large, Xception, DenseNet-121, and Inception-ResNetV2 against the pre-trained model. This research focused on getting effects with different kinds of optimization techniques on model performance. The authors were able to achieve very high accuracy when using the ADAM optimizer with the Xception model's architecture in solving the two-class classification problem between the Normal and No Tumor classes. This implies that very large margins of improvement in model performance can be achieved with appropriate optimization algorithms. In summary, this optimization in the training process of a model makes interior the potential full release of the pre-trained models and opens to the researcher multiple diagnostic accuracies, newness, and increased clinical outcomes regarding brain tumor detection.

Additionally, the authors emphasize the subtle in-between play between architecture of models and optimization algorithms, calling for tailored work in an ad hoc mode to maximize performance. The obtained outstanding accuracy, applying the ADAM optimizer with an Xception architecture, is demonstration power and trade-off made possible by a choice of algorithms for better results. These findings are seminal and point the way for both researchers and practitioners who are forward toward developing an optimized deep learning framework for precise and efficient brain tumor classification. Ultimately, this research contribution advances the field on how best to harness pre-trained models and strategies of optimization in real-world diagnostic applications.

In another excellent work, N. Çınar et al. [12] used a spectrum of deep learning models with their ability under the lens for medical image classifications, which vis-à-vis is being a critical area in diagnostic medicine. They evaluated an extensive set of renowned architectures, including VGG19, DenseNet169, AlexNet, InceptionV3, and ResNet101, to provide a widespread critique of various model capabilities. In fact, the ResNet-101 model surpasses them at an all-time high accuracy of 98.6% in binary classification tasks. This paper therefore presents what has been achieved using ResNet-101 but, in a general sense, presents the potential of highly advanced deep learning geometry in case one is serious with the attainment of high accuracies paramount in perfect medical image analysis. These findings drive diagnostic accuracy improvements and support the key role that deep learning methodologies have played in changing the scene in medical imaging practices.

Based on this, one can conclude that the research done on BTD using pre-trained models and optimization algorithms churned out one of the most remarkable advancements in the gamut of medical imaging. It is presented here how one could benefit from leveraging pre-trained models such as Xception, ResNet, among others, with optimization techniques like ADAM to achieve some nice accuracies in tasks associated with tumor detection and classification through meticulous research and experimentation. Finally, the subtle exploration into different deep-learning architectures underscores the fact that changes in the level of diagnostic accuracy, therefore, change clinical decision-making in neuro-oncology. The deployment of sophisticated deep learning methodologies improves the diagnostic precision for better patient outcomes, which are set to benefit from streamlined workflows in this challenging landscape of brain tumor diagnosis and treatment planning.

2.2 Advanced Architectures and Knowledge Distillation

Brain tumor diagnosis is an active field that attempts to use advanced architectures and the distillation of knowledge in better revelations. The newest studies delve into achieving development and refinement of state-of-the-art neural networking architectures, made to measure for medical imaging tasks, in improving diagnostic accuracy and efficiency. Moreover, such an application ushers in a new paradigm for the optimization of models, whereby one can come up with more compact and efficient neural networks without compromising on the quality of performance. This section critically reviews recent work related to BTD and focuses on advanced architecture developments in the latest literature and the application of methodologies of knowledge distillation to liquefy the capabilities of diagnostics in neuro-oncology.

Also, research in ultra-modern architectures concerning the diagnosis of brain tumors has so far given hopeful results, indicating that we will shortly witness great improvements in this field. For example, S. Mohsen et al. [5] worked on an experiment in March 2023, applying the stain-of-the-art model for the classification of two categories of brain tumors. In the work, the use of structures ResNext-101 32×8d and VGG-19 made it possible to achieve the extraordinary accuracy of classification of 100% and 99.98%, respectively. These results underscore the dramatically important impact that state-of-the-art neural network architectures can have on diagnostic performance and portend breakthrough potentials for neuro-oncology using the current models.

In the current research conducted by Geethanjali N. et al.[6], a significant step was taken in the domain of brain tumor diagnoses through the application of state-of-the-art neural network models. A dataset of four different classes was furthered into training and testing using ResNet-50, DenseNet, and EfficientNetB1 architectures. More importantly, the model EfficientNetB1 gives with its leadership in showing results that reach a great rate of about 99.31% on the accuracy performance count. These findings underscore potentials for the future use of sophisticated neural network architectures in revolutionizing diagnostic accuracy and efficacy in neuro-oncology. Such efforts open ways toward more precise and efficient diagnoses and add to the aid of treatment decision making that benefits patients in terms of care and outcomes.

Sohaib Asif et al.[3] have, in this regard, pursued detailed work in achieving higher classification accuracy for brain tumor diagnosis by applying multiple enhanced neural network architectures. An ensemble of Xception, DenseNet201, DenseNet121, ResNet152V2, and InceptionResNetV2 was used to combine the deep dense block with a new design for the last layer of the architecture: a softmax layer. The Xception architecture gave the following advanced results: 99.67% for a three-class dataset and 95.87% for a four-class dataset. The findings underlined that the application of very complex neural network architectures could increase accuracy in neuro-oncology diagnostics, which should lead to more precise and effective classification and treatment planning for people with brain tumors.

Table 2.1: Literature Review Summary on Brain Tumor Diagnosis Using Pre-Trained Models and Optimization Algorithms

Study	Year	Models	Key Findings	Metrics	Remarks
Sultan H. Hossam et al. [7]	2019	Custom Network Architecture	Proposed new architecture for diagnosing brain tumors with high accuracy. Differentiates between glioma grades.	Accuracy: 98.7%	Significant for providing detailed insights for clinical decision-making.
Ahmad Salen et al. [10]	2020	Xception, Inception-V3, ResNet-50, VGG-16, MobileNet	Evaluated pre-trained models on brain tumor dataset.	F1-scores: 97.25%-98.75%	Confirmed robustness and reliability of pre-trained models for medical imaging.
S. Asif et al. [11]	2020	Nas-Net Large, Xception, DenseNet-121, Inception-ResNetV2, ADAM Optimizer	Studied effects of different optimization algorithms on model performance.	High accuracy with ADAM optimizer	Demonstrated potential improvements with appropriate optimization algorithms.

Continued on next page

Table 2.1 – continued from previous page

Study	Year	Models	Key Findings	Metrics	Remarks
N. Cinar et al. [12]	2020	VGG-19, DenseNet-169, AlexNet, InceptionV3, ResNet101	Comprehensive evaluation of various architectures for medical image classification.	ResNet-101: 98.6% accuracy	Highlighted the impact of advanced architectures on diagnostic performance.
S. Mohsen et al. [5]	2023	ResNext-101 32×8d, VGG-19	Classification of two categories of brain tumors with high accuracy.	Accuracy: 100% (ResNext-101), 99.98% (VGG-19)	Showcased dramatic impact of state-of-the-art architectures on diagnostic accuracy.
Geethanjali N. et al. [6]	2023	ResNet-50, DenseNet, EfficientNetB1	Evaluated models on four-class brain tumor dataset.	Accuracy: 99.31% (EfficientNetB1)	Demonstrated potential of advanced architectures for precise diagnoses.
Sohaib Asif et al. [3]	2023	Xception, DenseNet-201, DenseNet-121, ResNet-152V2, Inception-ResNetV2	Applied ensemble of models for higher classification accuracy.	Accuracy: 99.67% (three-class), 95.87% (four-class)	Emphasized benefits of using ensemble models for improved diagnostic accuracy.

In fact, deep learning-based methods have demonstrated state-of-the-art performance on a variety of image classification tasks, and more importantly, in the field of brain diagnostics. Among the number of techniques applied in medical imaging, in practice, there is one common technique called transfer learning. This outstands capacity allows modifying the weights of the pre-existing knowledge about datasets like ImageNet or nonmedical photos for the neural networks related to the tasks of specific medical imaging [13]. This technique has almost become universal in medical imaging, from MRI through CT scans and X-rays, to provide a means of translating knowledge and expertise for the enhancement of diagnostic accuracy and efficacy.In comparison to the traditional machine learning approach, the deep learning methodologies offer a few significant advantages, mostly attributed to their ability to learn relevant features from data [14].

This automatically reduces the need for manual feature engineering, which keeps the system less complex, thus also making it easier to develop the models. Deep learning models are computationally demanding, making deployment on low-power devices non-trivial. Their application to resource-constrained contexts where computational resources are either scarce or their deployment expensive is limited as a consequence of this. To meet these challenges and promisingly apply deep learning-based approaches in the domain of brain diagnostics, researchers are moving toward novel ways. One such method is known as knowledge distillation, in which a smaller neural network, called the student model, is trained in such a way that the model behaves like some very large and more complex model of a neural network, which is called the teacher model.

This paves the way for the student model to achieve comparable performance with a teacher model but in a computationally efficient manner. It can make advanced diagnostic capabilities available in an affordable computing infrastructure and, with that, can also preserve diagnostic accuracy while deploying deep learning models on low-power devices. Should it be successful, this will bring access to advanced diagnostics tools to all and improve outcomes for patients in their care and new, medical image-related research.

Chapter 3

OVERVIEW

3.1 Brain Tumor(MRI) Detection and Classification

Groundbreaking discoveries in medical imaging are happening with the power of computer vision and deep learning models. Arguably, nowhere else are these many applications as impactful as in the case of analysis of magnetic resonance imaging (MRI) data, more specifically for brain tumor detection and classification. The very complexity and detail of MRI scans render any attempt to employ all traditional analytics methods a formidable operation and, at times, they lead to boring, manual interpretations with fairly low diagnostic accuracy.

However, this changed with novel computational approaches grounded on advanced algorithms powered by deep learning architectures and released in the processing of enormous volumes of information at high speed and much better accuracy. These use more complex networks in algorithms, which can automatically extract fine-grained features and patterns from complex MRI images to unlock a wealth of high-level clinically relevant information.

The implications of these breakthroughs are rather profound and may lead to a possible landmark in having better and accurate identification of brain tumors on time. For the first time, researchers and clinicians may actually use the computational power of deep learning to wade through a complex sea of MRI data much more efficiently and effectively than ever before. Further, the chance of classifying tumors according to unique characteristics offers huge potential for actually best-tailored individualized treatment techniques that would be translated into greater patient outcomes and quality of life. In fact, the synergy between computer vision, deep learning, and medical imaging has brought in a new era of hope and possibility for people suffering from brain tumors. As we continue to push the borders of technical innovation, the concept of the use of advanced techniques for unlocking the total diagnostic potential of MRI data steadily approaches the tipping point to promise a future in which early detection and classification are not just aspired to, but real moments in the fight against this formidable disease.

The recurrent neural networks and the convolutional neural networks have been important deep learning models. Among the RNNs, special attention has been paid to the analysis of sequential data, which makes it very good for learning temporal dependencies inherent in imaging data, such as tumor feature formation in images. At the same time, CNNs are really powerful for recognizing and finding spatial hierarchies hidden in visual data.

Accordingly, stacked architectures of RNNs and CNNs are trained with respect to brain tumor detection and classification via the backpropagation process. This is an iterative procedure in such a way that the models automatically enhance the most important features related to brain tumors for the purpose of their diagnosis and in planning treatment. These features may consist of characteristics like shape, size, and texture. They develop the capability to meaningfully extract large and complex MRI images, which will be used with a combined effort of RNNs and CNNs to aid clinical decisions quickly and precisely [15].

Moreover, the proposed method also included transfer learning, a crucial approach in which fine-tuned pre-trained models are applied to large datasets adapted for various tasks relevant to medical imaging. Transfer learning has played an important role in improving performance. In particular, models using ResNet, EfficientNet, and DenseNet have shown advanced brain tumor classification tasks' remarkable effectiveness for this special application [8][16]. Indeed, further improved methodologies—for instance, the merging of ResNext101 with VGG19—have greatly enhanced the generalizability and robustness of these models in detection and classification for brain tumors [5]. In fact, the better task transfer learning methodologies in complex medical imaging can only be realized in parallel hybrid model architectures that include other original citations.

In this regard, deep learning is likely to be an indispensable tool in classifying malignant brain tumors, as the capability required to handle this task is unparalleled. This is due to the diseases imaged being incredibly complex and diverse; therefore, the ability to navigate through a plethora of such images is unlike any other task or setting. Brain tumors are not easy to diagnose clinically because of their huge morphological and textural variations. Deep learning algorithms have a unique potential to discriminate subtle details and intricate patterns within medical imaging data. This algorithm can detect these very minute characteristics in the images with high precision, pointing toward the specific subtypes of cancer or the grades of malignancy. Such precision guides healthcare professionals on properly informed decisions about patient care and strategies for treatment in such a way that interventions can be tailored according to individual needs, maximizing therapeutic benefits. In this context, deep learning has been an enabler that is going to push forward the transition in modern medicine, which will allow neurosurgeons to categorize brain tumors with more accuracy and effectiveness, directly leading to better patient care and prognosis[17].

The capacity of deep learning to generalize across vast training datasets bestows on it robustness and accuracy crucial for testing data sets. If exposed with a diverse dataset containing tumors of most sizes, shapes, locations, and imaged under different conditions, the deep learning model will adapt to ensure versatility within real-world applications. The depth of these models allows for the identification of even the subtlest and the rarest variations of tumors—something beyond the reach of the traditional methods. This is especially important in medical imaging because tumors are often very variable in the way they present, from one patient to another, and when different imaging techniques are used. Deep learning provides a nuanced understanding of complex imaging data, opening great promises for breakthroughs in brain tumor classification to clinicians, thus affording insight into pathology regarding guided, tailored treatment strategies for improved patient outcomes[17].

Moreover, other advanced techniques added by single-image super-resolution and knowledge distillation have formed an integral step in the effort to make MRI images more resolute and clearer on the whole, thereby increasing the effectiveness of tumor detection and classification[5][18]. Application of these modern techniques would not only result in a higher level of scientific diagnostic accuracy but would also directly impact patient decisions, demonstrating the possibility of earlier and more precise detection of brain tumors, leading to quicker intervention with greater positive outcomes.

On the other hand, this paradigm shift in brain tumor detection and classification landscapes not only permits harnessing the synergistic potential between RNNs and CNNs but is versatile enough to allow researchers, particularly clinicians, to reach new highs in precision and efficiency of medical imaging [15]. It will undoubtedly revolutionize diagnostic methodologies and open ways for the most tailor-made treatment modalities, heralding an era of personalized medicine and optimized patient care. These innovations and collaborations will continue to fuel this field.

3.2 Knowledge Distillation

Knowledge distillation is a machine learning learning technique that enables the transfer of ideas embodied in a bulky teacher model to a more straightforward student model, conceived by Hinton et al.[18] in 2015, and today it is applied into practice with virtually all major current applications as a significant breakthrough in deployment and compression of models.

One of the training techniques is knowledge distillation, where a student model learns the possibility of the tractor model class labels and their output probabilities. In other words, soft targets increase data that the model can use for further learning about the underlying distribution of the data.

Apart from the fundamental concepts and algorithms, there are some training strategies in knowledge distillation which have been developed for boosting performance. Basically, these methods are critical for the enhancement of student model performance and also in reduction of the entire process used in transferring information involved.

3.2.1 Knowledge

Probably the most common thing to call knowledge in a neural network is what was learned, weights and biases. But for a large deep neural network, the sources of knowledge go far beyond that. In classical knowledge distillation, the source of teacher knowledge is the raw output of the neural network in the form of logits, i.e., without applying the activation functions. Some explore the weights or the activations of layers in between. Important related knowledge involves how kinds of activations of the teacher model and/or its neurons or parameters relate to one another.

These different forms of knowledge are further broken down into categories such as Response-based knowledge, Feature-based knowledge all play a very critical role in the distillation process. These knowledge sources are discussed below:.

Types of Knowledge

Response based Knowledge

Such knowledge is based on the output layer of the teacher model. The student model is made to learn from the response of the teacher model. This loss, which might capture a difference between the logits of the student and teacher models, can be expressed as a loss function and is termed distillation loss. The training will thus minimize the loss and will improve the predictions of the student model in this direction of training to achieve those predicted by the teacher Figure 3.1.

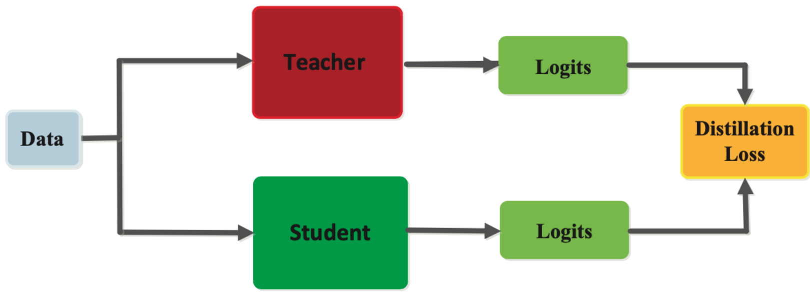

Figure 3.1: Response-based Knowledge[1]

Soft targets are usually response-based knowledge distillation in most computer vision tasks, like picture categorization. They are typically predicted by the use of a softmax function. The soft targets yield the probability distribution over the output classes, while the temperature parameter controls the contribution of every soft target to the information content. When it comes to the general setting of the response-based knowledge distillation in supervised learning, this is the place in which we usually work with soft objectives. [19].

Feature based Knowledge

In the case of deep neural networks, a well-trained teacher model in place will already capture valuable information with regard to the data in question within its intermediate layers. These intermediate layers can then be used to train a student model by transferring the prowess of the teacher model into recognizing specific features in data. The general objective is to transfer such an ability to the student model with a distillation loss function that will drive the student to learn exactly similar feature activations to those of the well-trained teacher model.

Relation based Knowledge

A model for the student could also be trained with information inlaying the innateness among feature maps, together with the output and intermediate layers. Such relation-based knowledge is given to account for the relations between feature maps, graphs, similarity matrices, feature embeddings, or the probabilistic distributions derived from the features representations. This helps in effective capture of interactions and larger context in the data, relevant for a better student model [19].

3.2.2 Training Methods

There are mainly three ways in which the training of student and teacher models goes: offline, online, and self-distillation. These methods are categorized in a manner dependent on whether they are done with a modified teacher model simultaneously with the student model. Common training techniques include:

Offline Distillation

In the conventional methods for knowledge distillation, there is offline distillation. The teacher network is trained first under fixed conditions. Because the teacher model underwent training, the student model can observe the behavior of the teacher model without him being corrected frequently by the latter. The learning process of the student is enriched in such a case. This was a common method, and most of the original work on knowledge distillation follows this paradigm, including the seminal paper by Hinton et al[18].

Other research works have also focused on building architectures that would enhance the transmission in offline distillation, which is in contrast to the architecture guiding the teacher network. This makes knowledge transfer possible between well-trained and high-performance teacher models and student models to enhance better performance.

Online Distillation

One of the key advantages of online distillation over offline distillation is the ability to update the teacher models and student models in real time during the training process, potentially leading to an enhancement in learning efficiency and performance for student models. Another name for this process is continual or continuous distillation because the source model gradually flows into the target model [19].

In the knowledge distillation scenario, online processing is applied for updating both teacher and student models at training. While new data is available on-line for updating, through one of the schemes, the student model gets continuously updated by each of the teacher updates.

The teacher model is continuously updated with new data, and its output is then used as one of the inputs to the development of a student model. We wished that the student model learn online from all the updates that are performed on the teacher model in real-time and do the same for the student model in presenting a method that allows the continuous improvement of the latter's performance.

In a more general formulation, the output of a teacher model updates the student model and the output of the student model is fed back as feedback to the teacher. This error signal, the second of the two sources generating feedback, primarily takes an indication form; more precisely, it is a measure of how well the student model is performing in relation to the teacher model. It then adjusts its parameters and makes fresh predictions on the basis of this learning signal. Other major advantages: these methods work well with non-stationary or streaming data, which does really change in NLP applications pretty much all the time. Online knowledge distillation is a method of training a student model that can handle further feedback from the teacher.

Self-Distillation

Self-distillation makes use of the same model as both the teacher and student by mostly copying knowledge from deeper layers to shallower layers in the very network. By learning its internal representations during training, it does not necessarily require the existence of an external teacher model, which can refine its own knowledge. Both two mainstream knowledge distillation methods, namely Offline Distillation and Online Distillation, suffer from two deficiencies. First, there is high dependency on the teacher's choice by student accuracy; the best in attaining high teacher accuracy often is not the best one for distillation. Many times student models do not show quite the same level of accuracy that their teachers achieve, thereby degrading and yielding lower accuracy during inference [19].

The Self Distillation approach assumes a form of likeness in topology between the teacher and student in addressing them. The model has been applied to the shallow classifiers based on attention in relation to the various depths of the intermediate layers of the neural network. More specifically, while training, the deeper classifiers play the role of teacher models that condition the training of the student classifiers with a loss based on the divergence metric between the outputs and a loss base of L2 of the feature maps. These extra classifiers are then used to infer, just as much.

3.2.3 Distillation Knowledge Algorithms

Herein, we describe a few of the leading algorithms applied in the area of knowledge distillation, which is one of the methodologies for transferring knowledge from a large model to a smaller one. More specifically, the three principal algorithms we detail consider adversarial distillation, multi-teacher distillation, and cross-modal distillation.

Adversarial Distillation

Adversarial distillation compresses the model and augments by adversarially generating synthetic data, providing hard examples, respectively: These processes involve the following two steps:

- Teacher Model Training : Initially, the teacher model is trained on a dataset to obtain ground truth labels.
- Student Model Training : This is trained over the original training data and data synthesized from an adversarial network.

On the other hand, the adversarial networks generate samples that are hard for the teacher model to classify, just due to slightly added small perturbations to the sample. The student learns from the classification of this new bothersome synthetic sample and, consequently, leads to better generalization and robustness against the adversarial attack.

Multi Teacher Distillation

Multi teacher distillation deploys most of the teacher models for one single student model, thereby assisting the student to learn a more general feature set and gain all-round performance. This is done in two steps:

- Teacher Model Training: First, the teacher models are trained on the training data and their outputs are obtained.
- Student Model Training: The very next step is training the student model on the same pool of data but with the targets generated by the concatenated output of all the teacher models, so it learns from a multi-perspective view.

Cross-Modal Distillation

The main take-home benefit of cross-modal distillation is the ability to transfer learning from one modality to another so long as data are available for one and not the other. For instance, textual description of images that has reinforced image recognition model.

It takes place in two stages:

- Teacher Model Training: It is done over the source modality itself, such as text data, and this model is made to produce the outputs.
- Student Model Training: Train the student model on the target modality (e.g., image data) with the teacher model's outputs acting as target values.

This is to enable the student model to better learn knowledge from the source modality as presented by the teacher model with better performance on the target modality.

A distillation-based offline technique is newly employed in the proposed methodology to train the student model with optimized knowledge transfer and streamlined model training. The approach with the insights and knowledge distilled from the teacher model would enable compact and efficient model development, preferably reaching performance levels close to those of larger models. In so doing, the rate of adoption towards practical applicability using distillation in real-world challenges will be driven by novel applications aimed at solution providers for markedly improved algorithms.

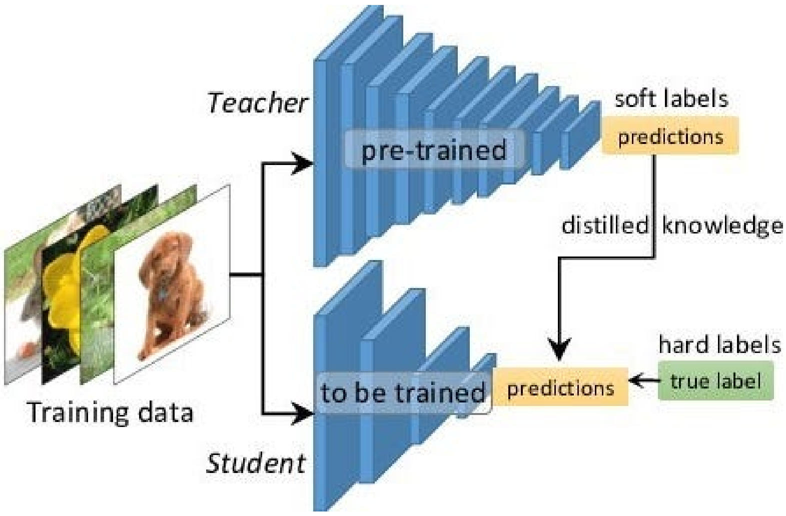

Figure 3.2: Knowledge Distilation [2]

The training process in knowledge distillation unfolds in two key phases shown in Fig 3.2. Firstly, the teacher model undergoes training on a large dataset using standard loss functions like cross-entropy. Subsequently, the student model is trained using the soft targets generated by the teacher model. Here, the loss function integrates the conventional loss term with the Kullback-Leibler (KL) divergence between the teacher and student model outputs. By quantifying the disparity between the output distributions of the two models, the KL divergence enables the student model to effectively assimilate the knowledge distilled from the teacher model.

Using knowledge distillation, this brain tumor classification framework converts the capacity of a large deep learning model (teacher) into a faster, smaller, and low-power device-friendly model (student). The teacher model first gains information from a sizable dataset of MRI pictures. Later, its understanding is refined by strategies including attention transfer and teacher student training, which enable the student to attain high accuracy with fewer resources. This makes it possible to diagnose brain tumors on even low-powered devices, potentially increasing access to vital medical care.

Chapter 4

RESEARCH METHODOLOGY

The proposed framework for brain tumor detection and classification using knowledge distillation consists of several steps, as described below:

4.1 Dataset Preparation and Preprocessing

The dataset used in the project was part of Masoud Nicparvar's Brain Tumor MRI dataset [20], a wide compilation sourced from three available datasets on Kaggle: Br35H, SARTAJ, and figshare. Such an aggregation of datasets will ensure strong diversity and strength towards the development of a very effective brain tumor detection and classification framework using brain MRI images.

4.1.1 Dataset Overview

The final dataset consists of 7023 images of human brains' MRIs, which are well separated into the following four distinct classes:

- **Pituitary Tumor:** The brain MRI will show a focal mass lesion centered at the pituitary gland—an abnormal growth that goes by the name of pituitary adenoma. This gland is significant in the production of hormones that regulate a number of functions within the body, among them being growth, metabolism, and reproduction. Some tumors might be functioning in their behavior (characterized by heightened hormone production) or non-functioning (expressed with symptoms caused by pressure). Complaints include hormone-related issues, headache, problems with vision, and hypopituitarism. MRI is the diagnostic tool of choice. It will show the characteristic findings of the tumors in respect of size, shape, and enhancement after contrast. Management includes medication, surgery, sometimes radiation, and close follow-ups for recurrence. High-quality MRI of pituitary tumors greatly improves training deep learning models for early detection and effective patient management.

- **Glioma:** The MRI image clearly shows gliomas, which are tumors that result from the uncontrolled multiplication of glial cells in the brain and spinal cord. Under normal conditions, these cells play a supporting and a protective role for the neurons. The various types of glial cells involved in the formation of these tumors include the astrocytes, oligodendrocytes, and the ependymal cells. Depending on the location and size of the tumor, different symptoms may become evident, such as headache, seizure, cognitive or motor impairments, and personality change. MRI is usually considered as the method of choice for imaging due to revelation of size, shape, and contrast enhancement patterns of gliomas. Treatment usually comprises either surgery, radiation therapy, or chemotherapy according to the type and grade of a tumor. Hence, accurate MRI images are required to train our model for an early detection and classification of glioma for an improved treatment planning process intended to bring better patient results.

- **Meningioma:** MRI can detect meningiomas, which are usually benign tumors originating in the thin layers of tissue enveloping the brain and spinal cord just inside the skull. These tumors arise from the meninges and are typically slow-growing. The symptoms will depend on the location of the tumor and how large it is; these tumors can result in headaches, seizures, vision problems, or neurological deficits. MRI is the imaging modality of choice. Meningiomas are enhanced, well-circumscribed masses that often seem to be dura-based. Treatment options for meningioma include surgery, radiation therapy, or active surveillance only for patients with small, asymptomatic tumors such as meningioma. Better and accurate detection, classification, and an improved approach using the training of models of deep neural networks in meningiomas can save lots of precious time for responsive patient care.

- **No Tumor:** MRI scans without any evidence of tumor represent scans of a normal and healthy brain. These become very important reference values for comparison in diagnosing and evaluating patients with suspected brain tumors. It shows the general anatomy and features of a normal, healthy brain in relation to structures, tissues, and blood vessels. Normal MRI images without tumors can help set a baseline for comparison. Radiologists and clinicians use this information to correctly identify and describe abnormal findings in other scans.

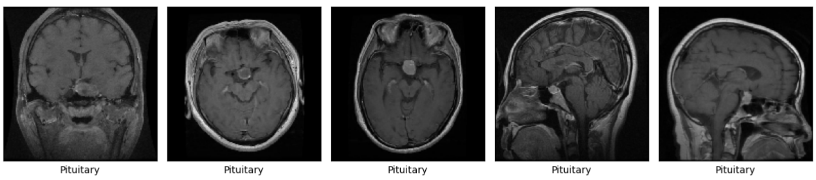

Figure 4.1: Pituitary Dataset

Two subsets of the dataset comprise a test set of 1311 photos and a training set of 5712 images. Every image in the collection has a multiple classification label applied to it, which indicates if the brain MRI shows a pituitary, glioma, meningioma, or tumor-free state shown above in Fig 4.1, Fig 4.2, Fig 4.3, Fig 4.4. The images differ in terms of size and color, and some can have extraneous details like noise.

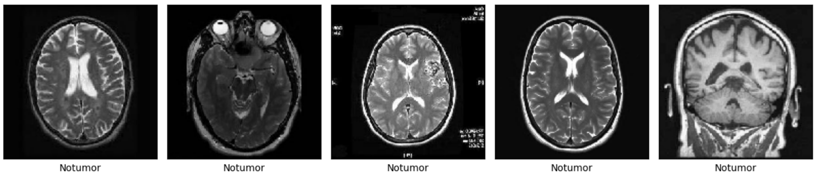

Figure 4.2: No Tumor Dataset

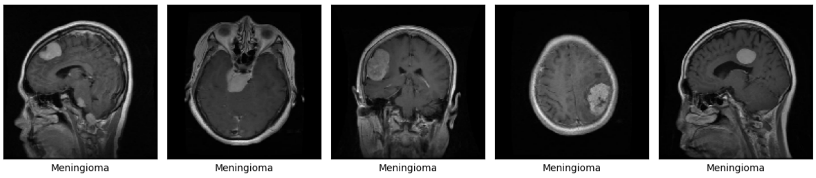

Figure 4.3: Meningioma Dataset

4.1.2 Splitting the Dataset

Dataset Splitting Dataset splitting is one of the most important steps taking place to prepare the dataset for insulation into deep learning models. Splitting the dataset in a procedural manner, this was ensured by sorting downloaded images with classification labels in different folders. To meet the commonest convention, an 80:20 split ratio was done in such a way that 80% would come to the training part and the 20

This will be required to strictly guard the integrity of the model training and evaluation exercise. Training on a large portion of the data enables models to learn complex patterns and features possible for effective representation of the underlying data distribution. Therefore, the test set will be used as an entirely new dataset for evaluation of the performance of trained models so that judgment about generalization on unseen data is unbiased.

The ability of the model can be evaluated by evaluating 80% of the data and testing it on 20% of data instances. As a result, the 80:20 rule gives a good balance to not only the volume of data in the training set but also the emphasis placed on evaluation in the training surface. The training surface is extremely important in building up sufficiently robust models that can capture characteristics associated with pictures of brain tumors. Meanwhile, a big test set allows comprehensive evaluation to find possible overfitting and get an estimation for how well our model does with respect to different data sub-samples.

Training Set

All this constitutes 5,712 images in the training set, which is the main data source for training models. Being very high in number allows the models to learn very complex patterns and features pertaining to different tumor types and healthy brain scans.

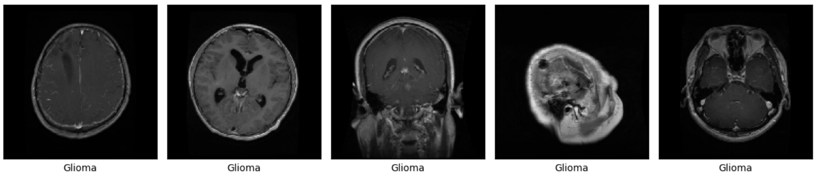

Figure 4.4: Glioma Dataset

Test Set

The total number of images in the test set is 1311; this dataset works hand in hand with proving grounds to get an assessment of the performance of the trained model independently. This withholding of a part of the data by the test set assures generalization ability of the model on unseen data and at the same time proves the reliability of the predictions made by this model in real applications.

4.1.3 Data Augmentation

We embed extremely extensive data augmentation techniques for increasing the size and diversity of the training set. In the process, perhaps numerous kinds of data augmentations are one of the major pre-processing pipelines in an attempt to breed a more varied and representative training dataset out of already existing images. In that sense, it enhances models' learning capacity to generalize and protects against overfitting—the most common problem during learning deep models with small datasets.

Data augmentation was conducted by applying various transformation techniques to the original images, commonly involving:

- **Random Rotations:** Random rotations of an image within a defined range. This helps the model train itself to detect a tumor in any orientation of the original brain scan by simulating natural variability in the test data.

- **Flips:** Horizontal and vertical flips were performed on the images. This augmentation procedure helps the model be invariant to the directionality of the image—important for detecting and classifying tumors correctly at different orientations.

- **Zoom** Random zoom was applied to the images by zooming within and between intervals randomly. It helps the model to look at different parts in an image, so it can detect a larger variety of tumor sizes and locations.

- **Shift and Translation:** Random shifts and translations were applied to images in the x and y axes. This makes sure the model is capable of highly confident prediction for tumor detection and, even if they are not perfectly centered within the image, it will be immune from these robust predictions.

- **Brightness and Contrast Adjustments:** Brightness and contrast adjustments are done to the images to bring them to a form similar to those taken under different conditions of imaging. This augmentation helps the model be robust against variations in imagery.

These augmentation techniques are applied in such a way that they help us effectively to increase the size of our training dataset. In this way, we could practically expand the nature of the original dataset containing about 7000 images, since every single image was passed through various transformations. This also not only increased the total number of samples of each class but diversified it under new image orientations and conditions, thus enlarging the training data.Data augmentation considerably helps to fight overfitting. The invariance is brought by data augmentation to the model: exposure to different data while training will enable the model to adjust in the presence of new data during testing.

Also, the model has to handle variations in image orientations and conditions since end-users of this system may not upload MRI images in a standardized format. The augmentation process will help the model be able to infer much better from varied conditions of the input images that were not seen at training times. In conclusion, data augmentation does play a critical role in improving the generalization power of our deep learning models. We, therefore, brought much diversity in terms of transformations, making our models robust, which could detect and classify brain tumors under very varying imaging conditions and orientations.

This is because, in this step, datasets are prepared for training while ensuring that the developed models can be used to perform very well in live clinical situations and, therefore, help in early detection and treatment of brain tumors [21].

4.1.4 Data Preprocessing

Data preprocessing is one of the most important steps in preparing the data set for training a deep learning model. Good preprocessing assures that data is in a consistent and optimal format, which is very important to obtain high performance on the model and to reduce computational overhead.All images were first resized to a common size of 224x224 pixels.

This resizing step ensured that the image size was uniform before feeding the data into the Convolutional Neural Network. The standardization of image dimensions removes variability and therefore improves both the training process and performance of the model. The 224×224 size was chosen based on the common practice in the field; ideally, this size keeps at a minimum the trade-off between keeping the detail of the image and reducing the load size on the computer. The images were then converted into grayscale. In MRI images, intensity values contain the structural information required for detecting and classifying tumors. Using grayscale simplifies the data to reduce three color channels—RGB to one, which reduces computational requirements and memory loads but maintains all necessary information for the tasks at hand.

Grayscale images find special utility in the field of medical imaging because color information is far less important than structural and textural information.

Subsequently, the images are resized and converted into grayscale, followed by normalization to achieve a zero mean and unit variance. Normalization is the process of adjusting pixel values to fall within a small interval, often [0, 1] or [-1, 1].

The steps of preprocessing in which resizing, grayscaling and normalization are done can be visibly summarized through Fig. 4.5 in tandem, prepare the dataset for training of deep learning models that are very robust and efficient.

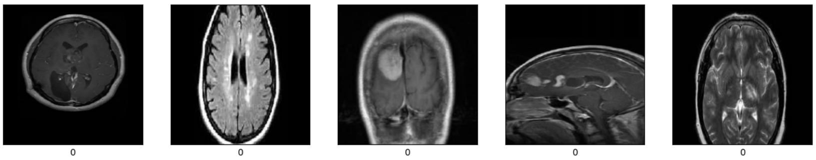

Figure 4.5: Preprocessed Images

All things considered, the dataset is a useful tool for developing and testing deep learning models for the detection of brain tumor. It is appropriate for creating reliable and accurate models due to its big image collection and high-quality labeling, and its capacity to generalize is enhanced by the incorporation of various tumor kinds and image artifacts.

4.2 Teacher Model Training

To train the student models in the proposed framework of knowledge distillation-based brain tumor detection and classification, we first trained a teacher model that would act as the reference model. Using the DenseNet121 architecture [16] - which has been pre-trained on the ImageNet dataset - transfer learning was used to create the teacher model. Using the training data, the model was refined on the brain tumor MRI dataset and assessed on the validation set[22].

4.2.1 DenseNet121

DenseNet121 represents one of the few CNN structures that implements a dense connectivity pattern. In traditional CNNs, the layers are connected one after another. In a DenseNet, dense connections are implemented wherein each layer participates in all subsequent layers with a direct connection. Such a hallmark has a significant implication for feature reuse across the network and can help to alleviate effectively the vanishing gradient problems faced by deep networks.

Specifically, DenseNet121 is 121-layered; such dense connections allow the features to be propagated deep into the network and hence learn very complex patterns very effectively. DenseNet121 reuses its characteristics down the layers, therefore making it computationally efficient and at the same time very high performance due to this design, which dramatically decreases the number of parametric counts.

These characteristics make the model particularly applicable to image classification, where complicated features and spatial relations between pixels are of paramount importance in order to derive good predictions. Dense connectivity helps ensure free flow of information through the network, which in turn makes the learning robust and effectively represents the complex data.

4.2.2 Transfer Learning

Transfer learning applies to a machine learning procedure where a model could first be trained on one task and therefore then used on a second related task. In this project, we used off-the-shelf DenseNet121 with transfer learning by taking pre-trained weights on ImageNet data. It contains over one million labeled images in the dataset. DenseNet121 on ImageNet is already pre-trained by supervised learning, which has rich feature representation through a large dataset in the process. Various patterns, shapes, and textures, characteristic of natural images, have been captured by these features—which set the foundation that DenseNet121 will later become an empowered feature extractor.

Here, we fine-tuned DenseNet121 pre-trained on ImageNet with our dataset of brain tumor MRI. We fine-tune the network by learning better parameters over the pre-trained model features according to the new data characteristics. In this process, the model learns the new features it should extract according to the new task but still retains the prior knowledge attained from the ImageNet.

Transfer learning with DenseNet121 enables the learning through the generalization power of a model pre-trained on a large and diverse dataset like ImageNet, which will speed up this training process. In addition, it will prevent the issues faced while training deep neural networks right from the beginning of their training processes, especially when using scarce labeled data, which most tasks concerning medical imaging usually comprise.

With pre-trained weights from ImageNet, a DenseNet-121 backbone made up the teacher model. In addition, we included a fully linked layer with 512 hidden units and ReLU activation, as well as a global average pooling layer. To avoid overfitting, a 0.3 dropout rate was used to the fully linked layer. For the purpose of multiclass classifying MRI images into no tumor, pituitary, glioma and meningioma categories, the output layer was composed of four units with softmax activation and is characterized by its substantial size, comprising approximately 11 million parameters and occupying 44.87 MB of memory[23], architecture shown in Fig 4.6.

With a sparse categorical cross - entropy loss and a learning rate of 0.001, the model was built using the Adam optimizer.

Adam Optimizer: Adam is short for Adaptive Moment Estimation. It is an optimization algorithm that computes learning rates adaptively with respect to parameters. The other two adaptations were AdaGrad and RMSProp, which are also versions of stochastic gradient descent algorithms. Being flexible with sparse gradients and noisy data sets, Adam is a very hot optimizer for deep learning models.

It used early stopping in the training process with a patience of three epochs. In case the validation loss did not improve for three epochs, then the training was

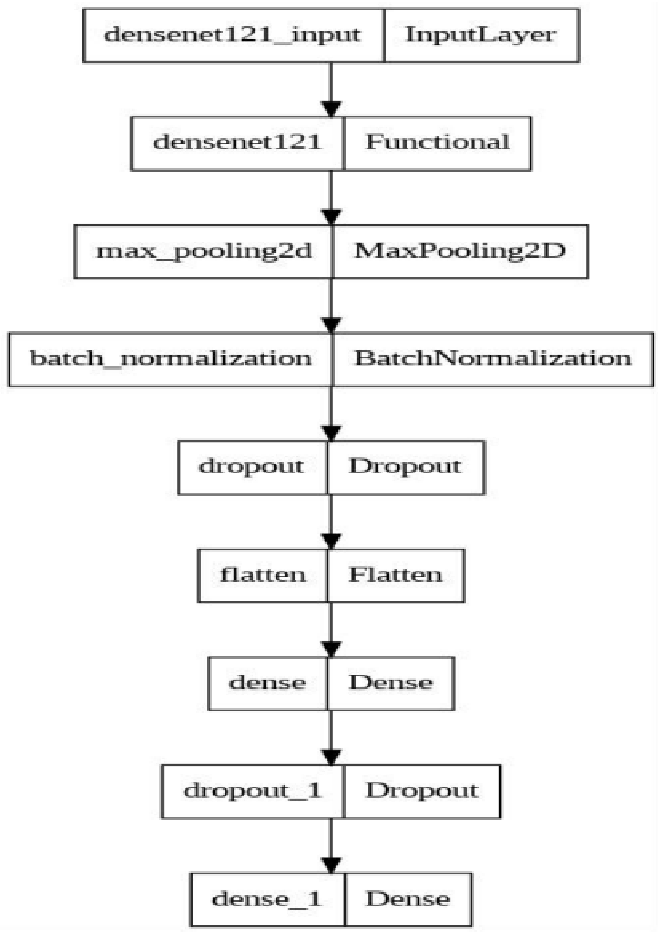

Figure 4.6: Teacher Model Architecture

terminated. It also reduces the learning rate by 0.3 times when the validation loss does not show any improvement using the callback ReduceLROnPlateau. The validation accuracy when training the model at any point in time was recorded.

Early stopping: Early stopping a technique for regularizing. It terminates the training by the number of epochs if the model does not show improved performance on a validation set compared to previous iterations, possibly saving time and computational expenditure in the process.

ReduceLROnPlateau: This is a callback that reduces the learning rate once a metric stops improving. It allows the user to fine-tune a learning process that has been already computed and further attains small adjustments in the learning dynamics if the learning curve stagnates.

Using the training set, the teacher model was trained for 20 epochs before being assessed using the validation set. When the validation loss did not improve after one epoch, the best model was saved using the ModelCheckpoint function, and the learning rate was decreased using the ReduceLROnPlateau callback. The teacher model's training history was recorded and utilized as a benchmark to evaluate how well the student models performed.

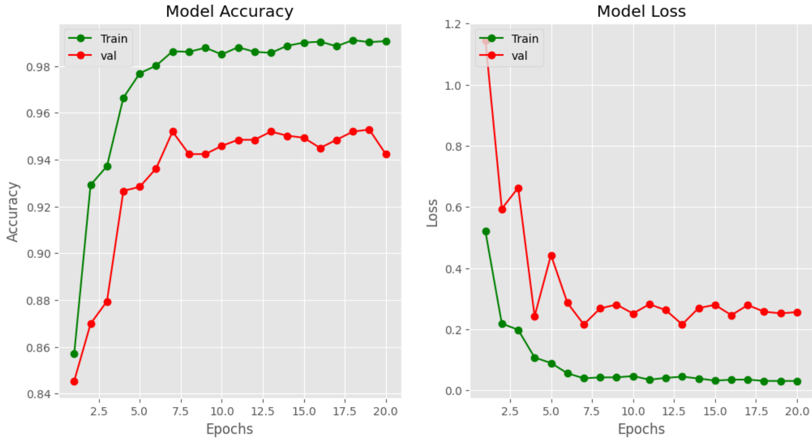

Figure 4.7: Accuracy,loss for training and validation

Using the Keras save function, the teacher model was saved to a file once it had been trained. After 20 iterations of training, the model reached a loss of 0.0307 and an accuracy of 99.06% on the training set. The model obtained an accuracy of 94.25& and a loss of 0.2559 on the validation data shown in Fig 4.7. The model's performance on test datasets, or unknown data, demonstrated its good generalization ability with a loss of 0.1399 and an accuracy of 97.25% using DenseNet121 as its base model trained on ImageNet.

4.3 Student Model Training and Distillation

4.3.1 Student Model Architecture

The student model is a customized convolution neural network (CNN) model to extract the features from the input images. The convolutional layers are stacked in much of the model's structure and followed by max-pooling layers, which will help extract features and downsample them. Fully connected layers with dropout regularization are included after the convolution layers to avoid overfitting [24].

The Keras input function is used to define the input layer. A max-pooling layer comes after the two layers of convolution with eight filters each that make up the first convolutional block. There is a max - pooling layer after every convolutional block, with the second and third blocks having 16 and 32 filters, respectively. A max-pooling layer is positioned after 64 filters in the fourth convolutional block. The max-pooling layer comes after a two layers of convolutional with 64 filters apiece in the fifth convolutional block.

There are two fully connected layers of 256 and 4 units, respectively, after the convolutional layers. Regression modeling is evident in the last layer, which contains linear activation. To minimize overfitting, a 0.2 dropout rate is given to the second fully linked layer. The student model, characterized by its reduced parameter count, comprises approximately 2.96 million parameters, occupying a mere 1.13 MB of memory.

4.3.2 Initial Performance Metrics

The student model, without knowledge distillation, showed us some raw predictive capabilities that guide the performance of the model. This distilled knowledge runs in an autonomous manner and sets the starting accuracy for the training set at 33.84%, with a loss of 4.9847. These metrics set a starter benchmark for the application of evaluation procedures before being optimized with enhancement techniques. This is important to gain an understanding of how the model is currently performing in this raw state for assessing how effective such ensuing strategies for refinement—like, say, knowledge distillation—are going to be in increasing predictive accuracy and decreasing loss.

4.3.3 Model Distillation: Enhancing Student Model Performance

Model distillation is one of the very important steps in developing the student model with the rich knowledge residing in the predictions from the teacher model. The training regimen, therefore, is focused on not only forcing the student model to make output label predictions correctly but also learning the probability of classes the teacher model would have modeled. The distillation loss function is useful due to the fact that it drives optimization and enables a simpler transfer of knowledge from teacher to student with the minimization of difference between both the models. During model distillation, various parameters and techniques are employed to optimize the transfer of knowledge like:

- **Temperature Parameter:** The parameter temperature is, in fact, one of the central parts of the distillation model. It is responsible for managing the softening of probability distributions when transferring the knowledge from the teacher to the student. The model in this way can balance iteration between exploitation from the rich knowledge in the prediction of the teacher and exploration of new solutions by changing temperatures comfortably. However, a small value of this parameter will give more importance to matchings of the probabilities assigned by the teacher model at a cost from a broad, more exploratory alternative.

- **Weight Coefficients:** Further, weight coefficients such as alpha play a significant role in setting up the distillation process so that they set relative importance between the distillation loss and the student loss through an optimization. Such further allows the fine-tuning of learning dynamics to allow the model to either copy the predictions of the teacher or optimize for his performance. Adapting weight coefficients allows one to optimize the compromise between the fidelity towards the teacher and the learning of the student.

- **Loss Functions:** The reason why loss functions are one of the major constituents of the distillation model is that they quantify deviations in the teacher's prediction and student's output. Two loss functions that are common to many model distillation tasks are Sparse Categorical Crossentropy and KLDivergence. Sparse Categorical Crossentropy quantifies the discrepancy between predicted and true labels, and KL Divergence quantifies divergence between probability distributions. These aim at optimizing the process of better learning from the teacher's prediction by the model.

- **Optimizer and Hyperparameters:** The biggest issue with this process is the choice of optimizer and its hyperparameters, as they greatly influence the efficiency and effectiveness of distillation. An optimizer like Adam, and specific hyperparameters such as epsilon, beta parameters, and learning rate, all help to reach an optimal solution during convergence of the model. In addition, schemes like dynamic learning rate adjustment using callbacks ensure adaptive optimization based on validation performance. Fine-tuning these hyperparameters will result in obtaining optimal performance while distilling the model.

4.3.4 Implementation and Training Procedure

The user can set up several parameters for the distillation process, including temperature to soften probability distributions, weight coefficients, and loss functions, by initializing the Distiller class with the student and instructor models. The student model's output in this implementation is represented by the Sparse Categorical Crossentropy loss function, while the distillation loss is represented by the KLDivergence loss function. The weight coefficient alpha is set to 0.5 and the temperature parameter to 2[25] [26].

The Adam optimizer is used to train the distillation process, with an epsilon of 1e-08, beta parameters of 0.9 and 0.999, and a learning rate of 1e-4. With a batch size of 32, the training procedure is conducted across 30 epochs utilizing the supplied validation and training data. Using the ModelCheckpoint callback, the best-performing model gets saved during training, and the ReduceLROnPlateau callback lowers the learning rate when the validation student loss approaches a plateau.

4.3.5 Performance Evaluation

In the last epoch of the model distillation process, the training set accuracy was 75.16%, with a distillation loss of 0.0201 and a student loss of 0.7166. The validation set accuracy was 71.38%, and the validation student loss was 0.8060. These results suggest that while there has been some development in the student model's imitation of the behaviour of the instructor model, there is still room for improvement in validation performance. Overall, it has been demonstrated that knowledge transfer may successfully enhance model performance by condensing a bigger, more complicated instructor model into a smaller one.

Chapter 5

RESULTS AND DISCUSSION

This section presents our method's findings while taking into account the consequences for computing efficiency while utilising a student model that has a lot less parameters than the instructor model.

Table 5.1: Comparison of Model Performance

Author	Models	Accuracy
Geethanjali N. et al.[6]	DenseNet	90.24
Sohaib Asif et al.[3]	Xception + DDB	95.87
Sohaib Asif et al.[3]	DenseNet121 + DDB	94.49
Proposed Teacher Model	**DenseNet121**	**97.25**

Note: The table provides a comparison of model accuracy scores for different pre-trained models with less resource intensive.

Teacher Model Performance Another reason a teacher model is a good foundational framework is that it has been trained with a huge number of eleven million parameters, hence being able to capture the varieties of patterns and subtleties in data very well. The parameterization provides the model with the ability to learn and generalize well. Further, the model works efficiently with a relatively modest memory footprint of 44.87 MB RAM for such a complex and capable model. Such efficiency would be very helpful in environments where computational resources are scarce.

The performance metrics of the teacher model are on the better side compared to either the outcome or the training and testing datasets. Achieves a remarkable testing set accuracy at 97.25%, meaning it is very robust and reliable once exposed to new, unseen data. This testing accuracy means that the model learned the underlying data distribution very well, and hence accurate prediction follows. The perfect accuracy on the training set is 99.06%. This near-perfect performance on the training data proves that the model can capture and learn from training examples.

It is, therefore judged that this teacher model was very well-capable of the given task of knowledge extraction and transfer learning by virtue of these attributes. This makes it a prime candidate with high performance relative to baselines for gaining further insights or transfer learning, where learned representations are passed on to other models or tasks. Its high number of parameters and high RAM efficiency, along with its exceptional accuracy on both training and testing data, underpin the value of this model as being strong and reliable for a vast number of applications involving very different fields. Hence, it is the best candidate for knowledge acquisition and to be carried out with further research in model development. These findings give a good foundation for comparison as shown in Table 5.

Student Model Performance It has the potential for great computational efficiency while being greatly reduced in size, such as containing only 2.96 million parameters and a compact memory footprint of just 1.13 MB. This reduced model size in terms of parameters and memory use makes the student model particularly attractive for deployment into setups that are resource-lean in terms of computation and memory, for example, into mobile or embedded systems.

The training of a student model independently, without using knowledge distillation for this task, presents very poor performance. Accuracies are 33.84% on the training set and 28.12% on the testing set, hinting at the fact that the student model might not correctly learn and generalize from raw data. A model with rather low accuracy thus means that it does not manage to catch the underlying patterns properly from within a set of data and is thus poor in prediction.

The dramatic change of scenery is eminently related to knowledge distillation. Knowledge distillation is a technique where the student model learns from the outputs of a more complex, well-trained teacher model, effectively transferring knowledge from the teacher to the student.

Hereby, the steep upward slope of student model performance curves can enable it. The training shows an accuracy of 75.16%, and on the testing set, it clocks to 70.27%. Therefore, a dramatic increase is provided as evidence that knowledge distillation effectively boosts the learning capability of the student model.

The huge improvement in the performance of a student model through knowledge distillation points toward the influence of the task in model training. This allows the student model to draw from rich knowledge that resides within the teacher model and affords generalization capabilities between training data and novel data. In fact, it is a way to bridge the performance gap between a much smaller student model and its substantially larger teacher counterpart. Concretely, the student model distillation that results in a better accuracy allows its use as a practical model for applications that require a trade-off between model accuracy and computational efficiency. The process of knowledge distillation by a student model enables it to obtain the capability of performing to demand from various real-world applications while benefiting from a suite of properties usually associated with low-sized models and less use of memory. This balance makes the student model a very good choice when saving computational resources and high performance are both critical.

Summary of the model performance as shown in Table 5 illustrates the transforming effect that knowledge distillation has on the student model. It really draws a contrast between the performance of the student model before and after the process of distillation. It is evidently illustrating the addition to its worth by way of the technique. By and large, the student model, when put in an up-boosted position using knowledge distillation, looks very strong and powerful, providing an alternative to bigger models high in accuracy yet not in much need of computational resources.

Table 5.2: Summary of Model Performance

Model	**Params (Millions)**	**Memory Size (MB)**	**Training Accuracy (%)**	**Testing Accuracy (%)**
Teacher (DenseNet121)	11	44.87	99.06	97.25
Student (Pre-distillation)	2.96	1.13	33.84	28.12
Student (Post-distillation)	2.96	1.13	75.16	70.27

Note: The table provides a comparison of model parameters, memory size, and accuracy scores for different models before and after distillation where student model is **97.49% compressed** model than teacher.

After knowledge distillation, both the parameter count and the size of a student model are significantly reduced. This enables high processing efficiency within deep learning applications. Therefore, through knowledge distillation, the student model can be small yet still reach performance metrics comparable to its large teacher, or even exceed it. This improvement in performance makes it easily deployable on low-resource devices like mobile phones or edge computing platforms. The less onerous computing requirements allow for a more real-time use of applications, which immensely increases user experience in various disciplines by lowering inference times and memory use.

Moreover, the student model is low on parameters and memory as compared to the teacher model in size, hence making training more resource-sensible. This thus means a more efficient training process, hence fewer computational resources during training, energy savings, and cutting down costs in operation during deployment. Reduces the model production cycle and speeds up the time to market for new solutions.

Techniques of knowledge distillation effectively transfer the expertise of the teacher model to the student model. This way, it provides efficient solutions for deep learning through high performance with minimum usage of resources. These are gains in efficiency and scalability from such techniques that basically play a paramount role in making successful deployment of deep learning models into real-world scenarios where computational constraints and resource availability matter a lot.

Enabling efficient and scalable deep learning models is important in building an open and innovative environment for artificial intelligence. Knowledge distillation has been one of the key enablers in the development of high-performance, deployable models—both on leading-edge server environments and on tiny resource devices. This means more disruptive AI technologies can be adopted and used, hence advancing research in most application domains.

In essence, knowledge distillation provides effective student models with several advantages related to efficiency, cost, and scalability—the basic motivations to apply deep learning models in real-world applications to achieve satisfactory performance under conditions of resource constraints. The resulting models are much more effective and efficient, hence the high and broad use of AI technologies in various applications.

Chapter 6

CONCLUSION

In the field of HAR, the pursuit of greater accuracy and efficiency has led the recognition community from considerations about everything from traditional pre-trained models to more advanced architectures like ConvNeXt. Here, this paper deals with the deep evaluation and comparison of their performances on a dataset comprising 12,000 images in 15 different activity classes. The results of the experiment produced an interesting story that definitively highlights the superiority of ConvNeXt architectures applied to HAR over all others, with ConvNeXtLarge leading this group. Even though promising performance initially appeared using traditional models pretrained as VGG16, VGG19, EfficientNetV2S, and Xception, their performance was largely poorer than for newly formulated ConvNeXt ones.

Of all the trained models, the leading model proved to be ConvNeXtLarge, which touched on an accuracy of training and validation better than the others. This major success has therefore proven the power of ConvNeXt architectures in capturing complicated patterns and features that are generally displayed in human activities. By using a mixed convolutional layer architecture in ways never before done, ConvNeXtLarge managed to state-of-the-art accuracy and robustness of activity recognition. It can pick out fine flavor and adapt to mixed datasets, thus putting it at the core of where HAR is moving. Moreover, success of model ConvNeXtLarge is of more general importance for the machine learning/artificial intelligence community in that this attests to the fact that new design in architecture might solve major practical tasks which claim high precision/reliability, in particular HAR.

This research also depicted how important continuous innovation and exploration are to the development of models. With the ever-evolving technologies and the increasing complexity of the datasets, new methodologies and architectures are coming forward which will now start to come into reality. The great performance of ConvNeXt Large will not only consolidate itself further among models for HAR but also open new possibilities and potential for deep learning.

References

[1] N. AI, "Knowledge distillation illustration," 2022. Accessed: 2024-05-22.

[2] P. Ganesh, "Knowledge distillation simplified," 2019. Accessed: 2024-05-29.

[3] S. Asif, M. Zhao, F. Tang, *et al.*, "An enhanced deep learning method for multi-class brain tumor classification using deep transfer learning," *Multimed Tools Appl*, vol. 82, pp. 31709–31736, 2023.

[4] A. Rehman, S. Naz, M. I. Razzak, F. Akram, and M. Imran, "A deep learning-based framework for automatic brain tumors classification using transfer learning," *Circ Syst Signal Process*, vol. 39, no. 2, pp. 757–775, 2020.

[5] S. Mohsen, A. M. Ali, E. S. M. El-Rabaie, A. ElKaseer, S. G. Scholz, and A. M. A. Hassan, "Brain tumor classification using hybrid single image super-resolution technique with resnext101_32× 8d and vgg19 pre-trained models," *IEEE Access*, vol. 11, pp. 55582–55595, 2023.

[6] G. N, V. Pushpalatha, R. C, S. L, and S. S, "Brain tumor detection and classification using deep learning," in *2023 Winter Summit on Smart Computing and Networks (WiSSCoN)*, (Chennai, India), pp. 1–6, 2023.

[7] H. H. Sultan, N. M. Salem, and W. Al-Atabany, "Multi-classification of brain tumor images using deep neural network," *IEEE Access*, vol. 7, pp. 69215–69225, 2019.

[8] H. A. Shah, F. Saeed, S. Yun, J. H. Park, A. Paul, and J. M. Kang, "A robust approach for brain tumor detection in magnetic resonance images using finetuned efficientnet," *IEEE Access*, vol. 10, pp. 65426–65438, 2022.

[9] Y. Bhanothu, A. Kamalakannan, and G. Rajamanickam, "Detection and classification of brain tumor in mri images using deep convolutional network," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, (Coimbatore, India), pp. 248–252, 2020.

[10] A. Saleh, R. Sukaik, and S. S. Abu-Naser, "Brain tumor classification using deep learning," in *2020 International Conference on Assistive and Rehabilitation Technologies (iCareTech)*, (Gaza, Palestine), pp. 131–136, 2020.

[11] S. Asif, W. Yi, Q. U. Ain, J. Hou, T. Yi, and J. Si, "Improving effectiveness of different deep transfer learning-based models for detecting brain tumors from mr images," *IEEE Access*, vol. 10, pp. 34716–34730, 2022.

[12] N. Çınar, B. Kaya, and M. Kaya, "Comparison of deep learning models for brain tumor classification using mri images," in *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, (Chiangrai, Thailand), pp. 1382–1385, 2022.

[13] K. Muhammad, S. Khan, J. D. Ser, and V. H. C. d. Albuquerque, "Deep learning for multigrade brain tumor classification in smart healthcare systems: A prospective survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, pp. 507–522, Feb 2021.

[14] S. Solanki, U. P. Singh, S. S. Chouhan, and S. Jain, "Brain tumor detection and classification using intelligence techniques: An overview," *IEEE Access*, vol. 11, pp. 12870–12886, 2023.

[15] K. Neamah *et al.*, "Brain tumor classification and detection based dl models: A systematic review," *IEEE Access*, vol. 12, pp. 2517–2542, 2024.

[16] N. Zebari, R. Marqas, M. Salih, and A. Alkurdi, "Enhancing brain tumor classification with data augmentation and densenet121," *Academic Journal of Nawroz University*, vol. 12, pp. 323–334, 2023.

[17] N. Abiwinanda, M. Hanif, S. T. Hesaputra, A. Handayani, and T. R. Mengko, "Brain tumor classification using convolutional neural network," in *Proc. World Congr. Med. Phys. Biomed. Eng.*, (Cham, Switzerland), pp. 183–189, 2019.

[18] G. E. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *ArXiv*, vol. abs/1503.02531, 2015.

[19] Neptune.ai, "Knowledge distillation: A simple guide." https://neptune.ai/blog/knowledge-distillation?ref=blog.paperspace.com.

[20] M. Nickparvar, "Brain tumor mri dataset," 2024.

[21] R. M. Prakash and R. S. S. Kumari, "Classification of mr brain images for detection of tumor with transfer learning from pre-trained cnn models," in *2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, (Chennai, India), pp. 508–511, 2019.

[22] R. Agarwal, S. D. Pande, S. N. Mohanty, and S. K. Panda, "A novel hybrid system of detecting brain tumors in mri," *IEEE Access*, vol. 11, pp. 118372–118385, 2023.

[23] B. Ojha, R. Maharjan, and T. Acharya, "Brain tumor detection using convolutional neural networks: A comparative study," *International Journal on Engineering Technology*, vol. 1, pp. 11–26, 2023.

[24] A. Garg, A. Chauhan, A. Sahgal, and P. G. Shambharkar, "Cnn based deep learning approach for brain tumor detection in mri images," in *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, (Salem, India), pp. 834–841, 2023.

[25] J. Si, W. Jia, and H. Jiang, "Breast cancer histopathology images classification through multi-view augmented contrastive learning and pre-learning knowledge distillation," *IEEE Access*, vol. 12, pp. 25359–25371, 2024.

[26] K. Deng, J. Lei, X. Li, S. Shuai, M. Lin, and S. Li, "An improved lightweight segmentation neural network for dermoscopic lesion images based on knowledge distillation," in *2023 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)*, (Chengdu, China), pp. 268–271, 2023.

List of Publications

The following publications have resulted from the research work presented in this thesis:

1. **Paper 1**:

   Paper Title: "Streamlining Brain Tumor Diagnosis: Knowledge Distillation for Computational Efficiency"

   The following is the proof of acceptance letter and the registration fee payments.

27/05/2024, 12:00 Gmail - Acceptance : ICOTET 2024

Gmail Jaimin Moghariya <jaimin.moghariya369@gmail.com>

Acceptance : ICOTET 2024
1 message

ICOTET2024 <icotetdgi2024@gmail.com> Fri, May 10, 2024 at 4:05 PM
To: jaimin.moghariya369@gmail.com
Cc: icotet@gnindia.dronacharya.info

Greetings from ICOTET 2024!

Dear Author (s)

We are pleased to inform you that **Paper ID 2476** entitled " STREAMLINING BRAIN TUMOR DIAGNOSIS: KNOWLEDGE DISTILLATION FOR COMPUTATIONAL EFFICIENCY" submitted by you has been accepted by the 2nd International Conference on Optimization Techniques in Engineering and Technology Engineering (ICOTET 2024).

You are advised to register for the conference by 16th of May, 2024! Payment details for registration can be found at the bottom of this email.

You are requested to fill out the following Google form for the registration and payment information etc.:

https://forms.gle/mqyRFhx45hqkJ6cx7

All the registered and presented papers for the 2nd ICOTET 2024 will be published in the AIP Conference Proceedings (Scopus Index) and Springer Nature Conference Proceedings (Scopus Index). Please note that the plagiarism level of the paper should not exceed 15%.

For further details, please visit the official website: https://www.icotet.in/registration

Thanks & Regards

Organizing Committee

ICOTET 2024.

Figure 6.1: Paper 1: Acceptance letter

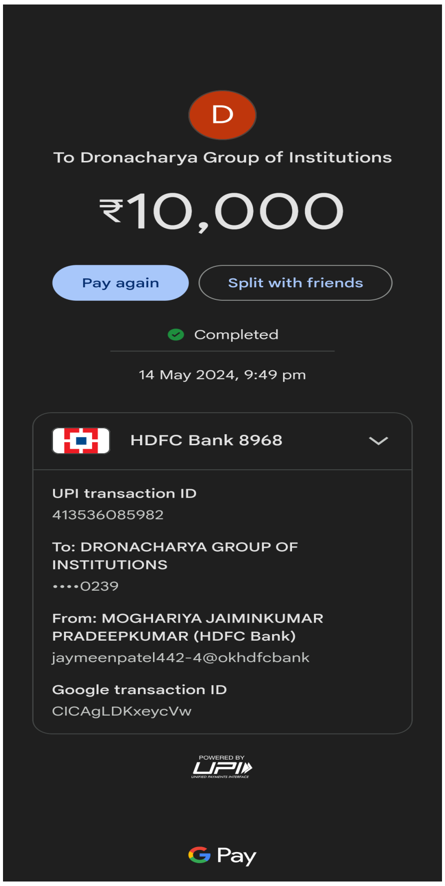

Figure 6.2: Paper 1: Fee Reciept

2. **Paper 2**:

   J. Moghariya and P. G. Shambharkar, "Blockchain-Enabled IoT (B-IoT): Overview, Security, Scalability Challenges," 2023 Second International Conference on Trends in Electrical, Electronics, and Computer Engineering (TEECCON), Bangalore, India, 2023, pp. 210-217, doi: 10.1109/TEECCON59234.2023.10335786.

   .

30/05/2024, 00:07 Gmail - Paper ID [7]: ACCEPTANCE Notification of the paper, IEEE TEECCON-2023, REVA University and Sub...

Gmail Jaimin Moghariya <jaimin.moghariya369@gmail.com>

Paper ID [7]: ACCEPTANCE Notification of the paper, IEEE TEECCON-2023, REVA University and Submit Registration Details by July 5th, 2023

1 message

Microsoft CMT <email@msr-cmt.org> Tue, Jun 27, 2023 at 11:45 AM
Reply-To: Jyotheeswara Reddy K <jyotheeswarareddy.k@reva.edu.in>
To: Jaiminkumar P Moghariya <jaimin.moghariya369@gmail.com>

Dear Jaiminkumar P Moghariya,

CONGRATULATIONS!!!!!

Greetings from the IEEE TEECCON-2023, REVA UNIVERSITY!

Hearty Congratulations, based on the recommendation of Expert reviewers and the Conference chair, we are pleased to inform you that your paper (7) has been ACCEPTED for the presentation (online) at the "IEEE 2022 Trends in Electrical, Electronics, Computer Engineering Conference (TEECCON-2022)" (IEEE Conference Record# 54414) to be held on 26th & 27th May 2022, organized by School of EEE, in association with IEEE Student Branch REVA University. The conference is Technical Sponsored by IEEE Bangalore Section and IEEE Power & Energy Society Bangalore chapter.

The reviewers' comments on your paper are attached at the end of this email.

REGISTRATION DETAILS:

Kindly fill in the Registration Details using the following link on or before July 5, 2023, to enable the submission of Camera-ready paper, Copyright form submission, Registration payment process, and further proceedings.

Fill Registration details, Link: https://forms.gle/bu84kZkVQPZ9V7t46

Figure 6.3: Paper 2: Acceptance Letter

REVA UNIVERSITY
Bengaluru, India

IEEE BANGALORE SECTION

TEECCON

CERTIFICATE OF PRESENTATION

School of Electrical and Electronics Engineering

This certificate is proudly presented to Dr/Mr/Ms Jaiminkumar Moghariya who have/has participated and successfully presentedtheir/his/her presentation at **Trends in Electrical, Electronics & Computer Engineering(TEECCON-2023) Conference,** (IEEE Record #59234), held on 23rd and 24th August 2023.

Director, School of EEE
REVA University

Vice Chancellor
REVA University

Figure 6.4: Paper 2: Certificate of Presentation

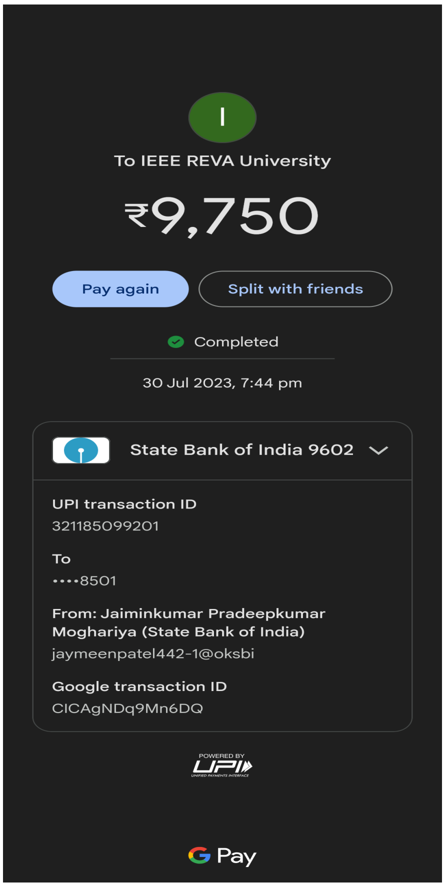

Figure 6.5: Paper 2: Fee Reciept

ANNEXURE-IV

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-42

PLAGIARISM VERIFICATION

Title of the Thesis Streamlining Brain Tumor Diagnosis: Knowledge Distillation for Computational Efficiency

Total Pages 41 Name of the Scholar Jaiminkumar Moghariya Supervisor (s)

(1) Dr. Prashant Giridhar Shambharkar

(2) ____

(3) ____

Department Computer Science & Engineering

This is to report that the above thesis was scanned for similarity detection. Process and outcome is given below:

Software used: Turnitin Similarity Index: 10%, Total Word Count: 13630

Date: ____________

Candidate's Signature **Signature of Supervisor(s)**

PAPER NAME

Thesis_v1 (1)-10-50.pdf

WORD COUNT

13630 Words

CHARACTER COUNT

75995 Characters

PAGE COUNT

41 Pages

FILE SIZE

1.8MB

SUBMISSION DATE

May 30, 2024 12:16 AM GMT+5:30

REPORT DATE

May 30, 2024 12:17 AM GMT+5:30

10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 5% Internet database
- 5% Publications database
- Crossref database
- Crossref Posted Content database
- 7% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Small Matches (Less then 8 words)

10% Overall Similarity

Top sources found in the following databases:

- 5% Internet database
- 5% Publications database
- Crossref database
- Crossref Posted Content database
- 7% Submitted Works database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

#	Source	Similarity
1	blog.roboflow.com Internet	1%
2	export.arxiv.org Internet	<1%
3	ijraset.com Internet	<1%
4	mdpi.com Internet	<1%
5	Stylianos Tsanakas, Aroosa Hameed, John Violos, Aris Leivadeas. "A li... Crossref	<1%
6	Indian Institute of Information Technology, Design and Manufacturing -... Submitted works	<1%
7	iq.opengenus.org Internet	<1%
8	ijiemr.org Internet	<1%

9	Liverpool John Moores University on 2023-10-19 Submitted works	<1%
10	Liverpool John Moores University on 2021-12-08 Submitted works	<1%
11	Sanjukta Chakraborty, Dilip Kumar Banerjee. "A review of Brain Cancer ... Crossref	<1%
12	Southampton Solent University on 2024-01-08 Submitted works	<1%
13	"ECAI 2020", IOS Press, 2020 Crossref	<1%
14	V P Nithya, N Mohanasundaram, R. Santhosh. "An Early Detection and ... Crossref	<1%
15	Wright State University on 2023-12-08 Submitted works	<1%
16	deci.ai Internet	<1%
17	trepo.tuni.fi Internet	<1%
18	ijisae.org Internet	<1%
19	Georgia Institute of Technology Main Campus on 2023-12-08 Submitted works	<1%
20	Hosseini, Mahshid. "Deep Learning Methods for Low-Resource Emotio... Publication	<1%

21	hindawi.com Internet	<1%
22	Ahmeed Suliman Farhan, Muhammad Khalid, Umar Manzoor. "PRCnet: ... Crossref posted content	<1%
23	Karrar Neamah, Farhan Mohamed, Myasar Mundher Adnan, Tanzila Sa... Crossref	<1%
24	University of Cincinnati on 2023-04-22 Submitted works	<1%
25	link.springer.com Internet	<1%
26	Hangtao Feng, Lu Zhang, Xu Yang, Zhiyong Liu. "Enhancing class-incre... Crossref	<1%
27	fdocuments.us Internet	<1%
28	Bipin Ch. Mohanty, P. K. Subudhi, Ratnakar Dash, Bidyadhar Mohanty. "... Crossref	<1%
29	Ebrahim Mohammed Senan, Mukti E. Jadhav, Taha H. Rassem, Abdula... Crossref	<1%
30	My-Diem Nguyen Pham, Chinh Tran-To Su, Thanh-Nhan Nguyen, Hoai-... Crossref posted content	<1%
31	Asia Pacific Instutute of Information Technology on 2022-02-11 Submitted works	<1%
32	Neng Dong, Shuanglin Yan, Hao Tang, Jinhui Tang, Liyan Zhang. "Multi... Crossref	<1%

33	North South University on 2024-01-11 Submitted works	<1%
34	Wen-Ting Li, Shang-Bing Gao, Jun-Qiang Zhang, Shu-Xing Guo. "Trainin... Crossref	<1%
35	digitalcommons.fiu.edu Internet	<1%
36	etd.aau.edu.et Internet	<1%
37	publikationen.bibliothek.kit.edu Internet	<1%
38	Jianping Gou, Baosheng Yu, Stephen J. Maybank, Dacheng Tao. "Know... Crossref	<1%
39	Napier University on 2023-05-18 Submitted works	<1%
40	University College London on 2021-12-17 Submitted works	<1%
41	"International Conference on Innovative Computing and Communicatio... Crossref	<1%
42	123dok.net Internet	<1%
43	Amena Mahmoud, Nancy Awadallah Awad, Najah Alsubaie, Syed Imma... Crossref	<1%
44	Baolong Liu, Tianyi Zheng, Peng Zheng, Daizong Liu, Xiaoye Qu, Junyu ... Crossref	<1%

45	Brooke Lampe, Weizhi Meng. " Intrusion Detection in the Automotive D... Crossref	<1%
46	Curtin University of Technology on 2021-09-26 Submitted works	<1%
47	Ebin P M, Roja Thomas, Safad Ismail, Kavitha Nair R. "A Novel Approac... Crossref	<1%
48	Jialong Si, Wei Jia, Haifeng Jiang. "Breast Cancer Histopathology Imag... Crossref	<1%
49	Liverpool John Moores University on 2021-06-19 Submitted works	<1%
50	Liverpool John Moores University on 2023-09-01 Submitted works	<1%
51	Shaojie Li, Mingbao Lin, Yan Wang, Yongjian Wu, Yonghong Tian, Ling ... Crossref	<1%
52	Sohaib Asif, Ming Zhao, Fengxiao Tang, Yusen Zhu. "DCDS-Net: Deep tr... Crossref	<1%
53	Tarun Agrawal, Prakash Choudhary, Vijay Kumar, Prabhishek Singh, M... Crossref	<1%
54	The Hong Kong Polytechnic University on 2024-04-11 Submitted works	<1%
55	Universiti Teknologi Petronas on 2024-05-02 Submitted works	<1%
56	University of East London on 2021-09-08 Submitted works	<1%

57	University of London External System on 2022-04-04 Submitted works	<1%
58	University of Queensland on 2023-05-15 Submitted works	<1%
59	University of Surrey on 2022-09-02 Submitted works	<1%
60	University of Sydney on 2021-06-04 Submitted works	<1%
61	Wang Wenjian, Xiao Qian, Xue Jun, Hu Zhikun. "DynamicSleepNet: a m... Crossref	<1%
62	arxiv.org Internet	<1%
63	"Cognitive Computing and Cyber Physical Systems", Springer Science ... Crossref	<1%
64	"Soft Computing for Security Applications", Springer Science and Busin... Crossref	<1%
65	Aston University on 2023-04-27 Submitted works	<1%
66	B.V. B College of Engineering and Technology, Hubli on 2024-01-12 Submitted works	<1%
67	Coventry University on 2023-08-08 Submitted works	<1%
68	Kaplan Kaplan, Yılmaz Kaya, Melih Kuncan, H. Metin Ertunç. "Brain tum... Crossref	<1%

69	Liverpool John Moores University on 2024-03-10 Submitted works	<1%
70	Mansoura University on 2023-07-07 Submitted works	<1%
71	Mepco Schlenk Engineering college on 2024-04-14 Submitted works	<1%
72	Monash University on 2024-05-19 Submitted works	<1%
73	NIT Imphal on 2019-10-04 Submitted works	<1%
74	Ngangbam Herojit Singh, N. R. Gladiss Merlin, R. Thandaiah Prabu, Dee... Crossref	<1%
75	Sarah Ali Abdelaziz Ismael, Ammar Mohammed, Hesham Hefny. "An e... Crossref	<1%
76	Sikkim Manipal University on 2024-04-08 Submitted works	<1%
77	Sohaib Asif, Wenhui Yi, Qurrat ul Ain, Jin Hou, Tao Yi, Jinhai Si. "Improv... Crossref	<1%
78	The Robert Gordon University on 2021-05-04 Submitted works	<1%
79	The University of Manchester on 2020-09-04 Submitted works	<1%
80	Tilburg University on 2023-12-04 Submitted works	<1%

81	University of Anbar on 2023-04-16 Submitted works	<1%
82	University of Edinburgh on 2022-01-27 Submitted works	<1%
83	University of Sunderland on 2024-04-05 Submitted works	<1%
84	University of Western Ontario on 2023-04-11 Submitted works	<1%
85	University of Wollongong on 2023-05-29 Submitted works	<1%
86	Virginia International University on 2021-01-28 Submitted works	<1%
87	Virginia Polytechnic Institute and State University on 2024-01-12 Submitted works	<1%
88	Zar Nawab Khan Swati, Qinghua Zhao, Muhammad Kabir, Farman Ali, Z... Crossref	<1%
89	dergipark.org.tr Internet	<1%
90	dokumen.pub Internet	<1%
91	dr.ntu.edu.sg Internet	<1%
92	ijeecs.iaescore.com Internet	<1%

93	vdoc.pub Internet	<1%
94	matec-conferences.org Internet	<1%