Shakib Md Nazmul Hasan

BRAIN IMAGE CLASSIFICATION USING DEEP CONVOLUTIONAL NEURAL NETWORKS AND TRANSFER LEARNING

Thesis
Centria University of Applied Science
Information Technology
December 2023

ABSTRACT

Centria University of Applied Sciences	Date December 2023	Author Shakib Md Nazmul Hasan
Degree programme Information Technology		
Name of thesis BRAIN IMAGE CLASSIFICATION USING DEEP CONVOLUTIONAL NEURAL NETWORKS AND TRANSFER LEARNING		
Centria supervisor Aliasghar Khavasi		**Pages** 57+4

The goal of this thesis research is to automate brain image classification by using Deep Convolutional Neural Networks and Transfer Learning. It aims to improve classification accuracy to help in medical diagnosis and treatment. The idea is to enhance current datasets and help to diagnose neurological conditions more accurately.

The findings of the study are quite important. They enable improved brain imaging analysis, which aids in the diagnosis and treatment of neurological diseases. Furthermore, these findings have larger implications for expanding AI applications, increasing human-computer interaction, and supporting the development of Brain-Computer Interfaces, which will greatly benefit the healthcare and technology industries.

Key words
Deep Learning, Transfer Learning, Inception, ResNet, ConvNet, GoogLeNet, Data Collection, Data Preprocessing, Model Training, Evaluation Metrics, Optimization, Dataset, GPUs/TPUs, Jupiter Notebook.

Abstract

Contents

FIGURES

TABLES

1 INTRODUCTION

Magnetic resonance imaging (MRI) is a prevalent non-invasive technique utilized to visualize a wide range of abnormalities in the brain, owing to its power to contrast soft tissue and its ability to produce multispectral images. The development of computer-aided diagnosis (CAD) systems has facilitated the expeditious diagnostic process for doctors through the utilization of MRI scan data. Based on characteristics visible in medical pictures, CAD systems are capable of making diagnoses. For classifying normal/abnormal brain MR images, these systems typically use the phases of preprocessing, attribute extraction, selection, and classification. To identify abnormal brain images, several techniques using traditional machine learning algorithms have been presented in the literature. This investigation proposes a methodology for automated classification of brain images utilizing deep convolutional neural networks (CNNs) and transfer learning. (Song, Seo, Cho, Woo, Son, Kim, Cho & Kwon 2015.)

The information gained from previously trained CNNs, overcoming the limits of conventional machine-learning techniques in the analysis of clinical data. This research proposes a system for multi-classifying brain magnetic resonance images (MRI) using transfer learning. The classification of brain images is an essential endeavour in the field of medical imaging, as it has the potential to assist in the identification and management of many neurological conditions. The proper classification of brain images holds significant potential in providing valuable insights into the underlying pathophysiology of these illnesses, hence assisting doctors in making informed decisions pertaining to patient management. However, manual classification of brain images is a time-consuming and labour-intensive process that requires specialized expertise. With the advent of deep learning, automated brain image classification has become a popular research area.

Deep Convolutional Neural Networks (CNNs) have demonstrated exceptional performance in tasks related to image classification. Transfer learning has been extensively used to enhance the performance of CNNs. Transfer learning involves using a pre-trained CNN model on a large dataset, such as ImageNet, and fine-tuning it on a smaller dataset, such as brain images. This approach allows for the efficient use of computational resources and can improve the performance of CNNs on smaller datasets. Several studies have been undertaken to categorize brain images through the utilization of convolutional neural networks (CNNs) and transfer learning techniques. The aforementioned- investigations have exhibited encouraging outcomes and possess the capacity to fundamentally transform the domain of medical imaging. Nevertheless, there exist several challenges that necessitate resolution, like the insufficiency of extensive annotated datasets and the issue of interpretability pertaining to CNN models. (Kaur & Gandhi 2020.)

The objective of this thesis was to examine the efficacy of eight distinct convolutional neural network (CNN) architectures, including AlexNet, GoogleNet, ResNet-50, VGG-19, Xception, InceptionV3, DenseNet-121, Squeezenet, NASNetMobile, and MobileNetV2, in the context of automated brain picture classification through the utilization of transfer learning. The selection of these CNN designs was based on their widespread recognition and demonstrated effectiveness in diverse image classification endeavours, encompassing the realm of medical imaging as well. This thesis work examined the constraints associated with traditional machine learning approaches that necessitate the manual construction of features to carry out classification tasks. On the other hand, Deep Convolutional Neural Networks (DCNNs) carry out the task of classification by acquiring hierarchical representations of the input image via a series of convolutional and pooling operations across many layers.

Transfer learning is a machine learning methodology that entails utilizing pre-existing models as an initial foundation for training on a novel job. The pre-existing models have undergone training on extensive datasets, such as ImageNet, and have acquired the ability to extract general features that can be advantageous for other associated tasks. Transfer learning has the potential to reduce the time and computational resources needed to train deep convolutional neural networks (DCNNs) from the beginning. Additionally, it could enhance the performance of the model when applied to a new task (Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, Qing 2020). The intended methodology for this research involves data collection, preprocessing, transfer learning, model evaluation, and system development. The dataset of brain MRI images will be collected from publicly available sources or medical institutions.

The collected images will be pre-processed to remove noise, normalize the intensity, and resize images to a standard size. The brain image classification task will utilize pre-trained models, including VGG-

16, ResNet-50, and Inception-v3, as a foundational framework for training the DCNNs. The evaluation of the trained models will be conducted by applying metrics like as accuracy, precision, recall, and F1-score. The most effective model will be utilized to create an automated system for classifying brain images, which can aid radiologists in the identification and management of neurological disorders. The thesis cites several related works that have used transfer learning and DCNNs in medical imaging tasks, including brain tumour classification, Alzheimer's disease classification, and stroke diagnosis. The thesis proposes that the utilization of deep convolutional neural networks (DCNNs) and transfer learning in the creation of an automated brain image categorization system can provide valuable support to radiologists in the identification and management of neurological disorders. The findings of this study have the potential to enhance the advancement of automated medical image classification systems through the utilization of deep convolutional neural networks (DCNNs) and transfer learning techniques.

The thesis is structured in the following manner. It presents an extensive literature review on the application of convolutional neural networks (CNNs) and transfer learning in the automated classification of brain images. It delineates the methodology employed in this study, encompassing the dataset utilized, the convolutional neural network architectures implemented. The assessment criteria employed. This study includes the empirical findings obtained from the conducted experiments and provides an analysis of the performance exhibited by each convolutional neural network (CNN) architecture. The manuscript presents a comprehensive examination of the obtained outcomes and engages in a critical discourse regarding the constraints inherent in this investigation. In conclusion, it serves as the final segment of the thesis, encompassing a summary of the main findings and offering suggestions for potential avenues of future research.

2 BACKGROUND STUDY

Brain image classification is a critical task in medical imaging that can aid in the diagnosis and treatment of various neurological disorders. With the advent of deep learning, automated brain image classification has become a popular research area. Deep Convolutional Neural Networks (CNNs) have demonstrated exceptional performance in tasks related to image classification. Additionally, transfer learning has been extensively used to enhance the performance of CNNs. In recent years, there has been an abundance of studies aimed at categorizing brain images through the utilization of Convolutional Neural Networks (CNNs) and transfer learning techniques. AlexNet, an influential convolutional neural network (CNN) architecture, emerged as one of the early and widely embraced models in the field. In 2012, AlexNet—one of the first and most well-known CNN architectures—was unveiled (Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton 2012). The architecture of AlexNet comprises of a total of five convolutional layers, followed by three fully connected layers.

Additionally, the Rectified Linear Unit (ReLU) activation function is included in this network. The model demonstrated exceptional performance on the ImageNet dataset and has been widely applied in diverse medical imaging applications, such as brain picture categorization. The GoogleNet architecture, alternatively referred to as InceptionV1, was first presented in 2014 and emerged as the victor in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) during the same year (Szegedy, Christian and Liu, Wei and Jia, Yangqing and Sermanet, Pierre and Reed, Scott and Anguelov, Dragomir and Erhan, Dumitru and Vanhoucke, Vincent and Rabinovich, Andrew, 2014). It consists of 22 layers and uses a new inception module that enables optimal use of computational resources. The GoogleNet architecture has been employed in several medical imaging applications, such as the classification of brain images. The ResNet-50 architecture, which was first introduced in 2015, is a convolutional neural network (CNN) that is characterized by its increased depth. It employs residual connections to mitigate the issue of disappearing gradients. (Kaiming, Xiangyu, Shaoqing, Jian 2015.)

The architecture comprises of 50 layers and has demonstrated superior performance compared to shorter CNN architectures in diverse image classification tasks, including those related to medical imaging. The VGG-19 model, which was first proposed in 2014, is a convolutional neural network

(CNN) architecture with a total of 19 layers (Karen Simonyan, Andrew Zisserman 2014.). This architecture employs tiny filters with dimensions of 3x3 in all its convolutional layers. Research has demonstrated that it can attain a notable level of precision in jobs related to picture categorization, such as those involving medical imaging. The Xception model, which was first proposed in 2017, is a convolutional neural network (CNN) architecture that employs depth wise separable convolutions to decrease the parameter count and enhance computational performance (François Chollet 2017).

Empirical evidence has demonstrated that this approach attains exceptional performance in diverse picture categorization endeavours, encompassing the domain of medical imaging. The InceptionV3 model, which was first proposed in 2015, is a convolutional neural network (CNN) architecture that employs a blend of convolutional layers with varying filter sizes and max pooling operations to extract distinctive characteristics from input images (Christian, Vincent, Sergey, Jonathon, Zbigniew 2015). Numerous studies have demonstrated the ability of this approach to attain a notable level of precision in diverse picture classification endeavours, encompassing the field of medical imaging. The DenseNet-121 architecture, which was first introduced in 2016, is a convolutional neural network (CNN) design that leverages dense connections between layers to enhance feature reuse and minimize the parameter count (Huang, Liu, Maaten, Weinberger 2016). Numerous studies have demonstrated the ability of this method to attain a remarkable level of precision in diverse picture categorization endeavours, encompassing the realm of medical imaging.

The Squeezenet design, which was first developed in 2016, employs a blend of 1x1 and 3x3 filters to decrease the parameter count and enhance computational efficiency within convolutional neural networks (CNNs) (Forrest, Iandola, Song, Matthew, Khalid, William, Kurt 2016). Numerous studies have demonstrated the ability of this method to attain a notable level of precision in diverse picture categorization endeavors, encompassing the realm of medical imaging. The NASNetMobile, which was presented in 2018, is a convolutional neural network (CNN) architecture that was developed through the utilization of neural architecture search (NAS) methodologies (Zoph, Vasudevan, Shlens, Le 2018). Empirical evidence has demonstrated that this approach attains exceptional performance in diverse picture categorization endeavors, encompassing the realm of medical imaging. The MobileNetV2 architecture, which was launched in 2018, is a convolutional neural network (CNN) that incorporates depth wise separable convolutions and linear bottlenecks (Sandler, Howard, Zhu, Zhmoginov, Chen, 2018). These architectural features are employed to decrease the parameter count and enhance computational performance.

This study builds upon the concept of transfer learning, which is a deep learning technique aimed at enhancing performance on novel tasks by leveraging knowledge acquired from a pre-trained model. Despite the lack of labelled data for clinical data analysis, transfer learning has shown promising results. The researchers assess the existing body of literature pertaining to transfer learning in the context of medical picture categorization, highlighting the significance of transfer learning within this domain (Kaur & Gandhi, 2020) (Chelghoum, Ikhlef, Hameurlaine & Jacquir 2020.). This research centers on a comprehensive examination of transfer learning techniques employing Convolutional Neural Networks (CNNs) for the purpose of classifying brain images. The authors examine the use of transfer learning in several brain image classification tasks, encompassing the classification of brain tumours, tasks related to dementia classification, and the classification of brain functional connectomes. (Arbane, Benlamri, Brik & Djerioui 2021).

Additionally, they emphasize the advantages of transfer learning in leveraging machine learning models that have been trained to address distinct yet interconnected tasks for the specific task at hand. Most of the examined publications utilized transfer learning techniques based on CNNs, while just a few approaches clearly utilized brain MRI-specific methodology, and considered privacy issues, unobserved target domains, or unlabelled data. To predict brain tumour cells automatically, the research proposes a comparative evaluation of three transfer learning-based convolutional neural network models, namely VGG-16, ResNet-50, and Inception-v3, for the purpose of brain tumour classification. The study used a dataset of 233 magnetic resonance imaging (MRI) brain tumour images. The authors conducted a comparative analysis of the accuracy, sensitivity, specificity, and F1 score of the three models. Based on the results, it can be concluded that the VGG-16 model has superior performance compared to the other two models in terms of F1 score, sensitivity, and accuracy. (Srinivas, KS, Zakariah, Alothaibi, Shaukat, Partibane & Awal 2022.)

The study additionally illustrates the potential of transfer learning to significantly improve the performance of models. The work titled "Automated Brain Tumour Detection and Classification Using Deep Learning and Transfer Learning" presents a proposed methodology for the detection and classification of brain tumours through the utilization of deep learning and transfer learning techniques. The dataset included in the study comprises a total of 306 magnetic resonance imaging (MRI) images of the brain. The proposed methodology involves the utilization of a pre-trained deep convolutional neural network model, specifically VGG-16, to classify brain MRI images into normal and abnormal categories. This approach incorporates transfer learning techniques. The results indicate that the proposed strategy exhibited superior performance compared to the existing state-of-the-art methods, achieving an accuracy rate of 98.7%. According to the study's findings, the suggested approach can be a useful tool for the automated identification and categorization of brain tumours. In conclusion, automated brain image classification using deep CNNs, and transfer learning has become a popular research area in medical imaging. (Anantharajan & Gunasekaran 2021.)

Several CNN architectures, including AlexNet, GoogleNet, ResNet-50, VGG-19, Xception, InceptionV3, DenseNet-121, Squeezenet, NASNetMobile, and MobileNetV2, have been used in brain image classification tasks and have shown promising results (Krizhevsky, Sutskever, Hinton 2012) (Szegedy, Christian and Liu, Wei and Jia, Yangqing and Sermanet, Pierre and Reed, Scott and Anguelov, Dragomir and Erhan, Dumitru and Vanhoucke, Vincent and Rabinovich, Andrew 2014). Additional investigation is required to delve into the prospective applications of these convolutional neural network architectures inside clinical environments and to cultivate automated brain image classification systems that are both more precise and efficient.

2.1 Problem Statement

To properly classify brain images (normal/abnormal or tumour types) despite their complexity and sparse labelled data, this topic focuses on automated brain image classification utilizing deep CNNs and transfer learning.

2.2 Aim

This study aims to build a strong automated brain image classification system that reliably classifies images in terms of complexity, variability, and sparse labelled data by utilizing deep CNNs and transfer learning.

2.3 Objectives

The research suggests a technique for automatically classifying brain images that makes use of deep CNNs and transfer learning. To analyse clinical data, this thesis will constraints imposed by conventional machine learning techniques. The research presents a conceptual framework for the classification of brain MRI images using transfer learning.

3 LITERATURE REVIEW

The study explores the application of transfer learning in conjunction with convolutional neural network (CNN) architectures for the purpose of classifying brain tumours based on MRI scans. The objective of this study is to examine the efficacy of transfer learning in enhancing the precision of brain tumour classification, as well as to evaluate and contrast the performance of various convolutional neural network (CNN) designs. The study commences by providing an introductory overview of brain tumours, emphasizing the significance of precise categorization to facilitate optimal treatment outcomes. Subsequently, it proceeds to offer a concise exposition of Convolutional Neural Networks (CNNs) and the concept of transfer learning. Also, it is elucidate that transfer learning encompasses the utilization of pre-trained convolutional neural network (CNN) models on extensive datasets for the purpose of extracting features from images. These extracted features are subsequently employed to train a more compact CNN model on a smaller dataset. By capitalizing on the knowledge acquired by the pre-trained model, this methodology has the potential to enhance the precision of the smaller model. (Chelghoum, Ikhlef, Hameurlaine & Jacquir 2020.)

The dataset included a total of 3064 magnetic resonance imaging (MRI) pictures of brain tumours sourced from the Brain Tumour Segmentation Challenge 2018. The paper elucidates the preprocessing procedures employed to adequately prepare the data for both training and testing purposes. Subsequently, it is delineating the convolutional neural network (CNN) architectures employed in their investigation, namely VGG16, InceptionV3, and ResNet50. It also elucidates the architectural characteristics of each model and expounds upon their adaptations specifically tailored for the purpose of brain tumour categorization. Additionally, it elucidates the employed transfer learning methodology, wherein the pre-trained layers of the models were immobilized while solely training the last few layers on the dataset pertaining to brain tumours. Subsequently, the paper proceeds to discuss the outcomes of their conducted experiments, wherein they engaged in the training and testing of each convolutional neural network (CNN) architecture, both with and without the use of transfer learning. (Chelghoum, Ikhlef, Hameurlaine & Jacquir 2020.)

The accuracy, sensitivity, specificity, and F1 score are reported for each model. Additionally, they conduct a comparative analysis of the various convolutional neural network (CNN) architectures in terms of their performance. The findings indicate that the utilization of transfer learning leads to a notable enhancement in the accuracy of all three convolutional neural network (CNN) designs. The highest performing model observed in the study was InceptionV3 with transfer learning, which attained an accuracy rate of 98.36%. It was also discovered that InceptionV3 had superior performance compared to the other two architectures in terms of accuracy, sensitivity, specificity, and F1 score. Ultimately, it engages in a comprehensive analysis of the ramifications of their findings and propose potential avenues for further scholarly inquiry. The researchers reach the conclusion that employing transfer learning using convolutional neural network (CNN) designs can yield a notable enhancement in the accuracy of brain tumour classification based on magnetic resonance imaging (MRI) pictures. (Chelghoum, Ikhlef, Hameurlaine & Jacquir 2020.)

Furthermore, they determine that the InceptionV3 architecture exhibits the most effectiveness among the considered models for this specific job. In general, the research presents a comprehensive and enlightening analysis of the application of transfer learning in convolutional neural network architectures for the purpose of classifying brain tumours based on MRI data. It provides a comprehensive account of their methodology and results, and their findings carry significant implications for advancing the development of more precise and efficient approaches in the detection and treatment of brain tumours. Using convolutional neural networks (CNNs) and transfer learning, the paper by Kaur & Gandhi (2020). suggests a deep learning method for automatic brain picture classification. It commences their discourse by elucidating the significance of automated brain picture classification in the realm of medical diagnosis and therapy planning. (Chelghoum, Ikhlef, Hameurlaine & Jacquir 2020.)

It observes that the process of manually classifying brain images is both time-consuming and susceptible to errors. They further argue that the implementation of automated classification methods can enhance the accuracy and efficiency of diagnostic procedures. (Chelghoum, Ikhlef, Hameurlaine & Jacquir, 2020.). Subsequently it defines their proposed methodology, which entails the utilization of pre-trained convolutional neural networks (CNNs) for the purpose of transfer learning. The concept of transfer learning entails the utilization of a pre-existing convolutional neural network (CNN) as an initial framework for training a novel CNN on a distinct dataset. The pre-existing convolutional neural network (CNN) has acquired the ability to identify fundamental visual characteristics in images, such as edges and textures. By employing transfer learning, the new CNN can leverage this acquired knowledge to acquire a deeper understanding of intricate features that are unique to the new dataset. (Chelghoum, Ikhlef, Hameurlaine & Jacquir 2020.)

Subsequently in this paper, it provides a detailed account of the dataset employed in their experimental study, comprising a total of 2,000 brain pictures sourced from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. It explains that the dataset underwent partitioning into training, validation, and testing sets, and that the images underwent preprocessing to eliminate noise and standardize intensity. It also explains the structural design of their convolutional neural network (CNN), which encompassed many convolutional and pooling layers, succeeded by fully connected layers. The authors elucidate their utilization of the pre-trained VGG16 Convolutional Neural Network (CNN) as an initial framework for transfer learning. This paper proceeded to refine the last layers of the network by means of fine-tuning, employing their brain picture dataset. It also proceeds to discuss the outcomes of their conducted tests, wherein it was observed that their Convolutional Neural Network (CNN) attained a noteworthy accuracy rate of 98.5% when evaluated on the testing dataset. (Chelghoum, Ikhlef, Hameurlaine & Jacquir 2020.)

A comparative analysis of their findings with those of previous studies within the same field. They observed that their methodology exhibited superior performance in comparison to a significant number of these investigations. Ultimately, it delves into the prospective clinical implementations of their methodology, specifically in terms of its utility in facilitating the identification and assessment of Alzheimer's disease and other neurological ailments. It is also acknowledging many constraints inherent in their research, including the comparatively limited scale of their dataset and the exclusive utilization of a single pre-trained convolutional neural network (CNN) as the basis for transfer learning. (Chelghoum, Ikhlef, Hameurlaine & Jacquir 2020.)

The study introduces a systematic approach for diagnosing schizophrenia (SZ) patients in comparison to healthy controls. This approach utilizes transfer learning in conjunction with deep convolutional neural networks (CNNs). The methodology employed in this study is the conversion of EEG data into visual representations by the utilization of the continuous wavelet transform (CWT) technique. These visual representations are then fed into four pre-trained convolutional neural networks (CNNs), namely

AlexNet, ResNet-18, VGG-19, and Inception-v3. The study's findings suggest that the proposed methodology has the potential to serve as a valuable tool for differentiating individuals with SZ from those without the disorder. The research work pertains to the domain of medical image analysis and deep learning, and it makes a valuable contribution towards the advancement of automated diagnostic systems for schizophrenia by utilizing EEG signals. (Shalbaf, Bagherzadeh & Maghsoudi 2020.)

It presents a comparative analysis of the performance of three transfer learning-based convolutional neural network models, namely VGG-16, ResNet-50, and Inception-v3, in the context of automated prediction of tumour cells in the brain. The study used a dataset of 233 MRI brain tumour images. It compares the accuracy, sensitivity, specificity, and F1 score of the three models. Based on the results obtained, it can be concluded that the VGG-16 model exhibits superior performance compared to the other two models in terms of F1 score, sensitivity, and accuracy. The study additionally illustrates the potential of transfer learning to greatly improve the performance of models. (Shalbaf, Bagherzadeh & Maghsoudi 2020.)

This paper proposes the utilization of pre-trained deep learning models for the classification of brain MRI data. The research compares the efficacy of pre-trained models utilizing transfer learning with the existing state-of-the-art methods in the classification of brain MRI images. In this paper, the researchers utilized a dataset of 253 magnetic resonance imaging (MRI) brain scans. The results indicate that pre-trained models utilizing transfer learning outperform the present state-of-the-art methodologies in terms of accuracy, sensitivity, and specificity. It currently investigates the potential of transfer learning in the automatic classification of cardiac cine short-axis slices. This study examines the automatic classification of cardiac short-axis slice ranges and evaluates the effectiveness of transfer learning using nine well recognized convolutional neural network architectures in both fixed feature extraction and fine-tuning scenarios. Upon analysing previously unobserved test data, it was shown that the fine-tuned VGG16 model exhibited the highest values across all considered assessment categories. Consequently, it appeared to be the most suitable choice for the classification of a cardiac cine MRI short axis slice range. (Shalbaf, Bagherzadeh & Maghsoudi 2020.)

This paper investigates the application of deep learning methodologies in the classification of Alzheimer's disease (AD) using neuroimaging data. This study assesses the efficacy of various pre-trained convolutional neural network (CNN) architectures, namely ResNet-50, VGG-19, and volumetric CNN, in classifying Alzheimer's disease (AD) based on magnetic resonance imaging (MRI) images. The study's findings indicate that deep learning techniques have the potential to serve as a valuable tool for accurately classifying Alzheimer's disease (AD) based on neuroimaging data. The research work pertains to the domain of medical image analysis and deep learning, making a valuable contribution towards the advancement of automated diagnosis systems for Alzheimer's disease (AD) through the utilization of neuroimaging data. Several further studies have been undertaken in the same field, exploring various methodologies such as the integration of convolutional neural networks (CNNs) and recurrent neural networks (RNNs), the utilization of functional magnetic resonance imaging (fMRI) and magnetic resonance imaging (MRI), as well as the application of structured deep learning techniques. (Simon, Baskar & Jayanthi 2019.)

The study offers a complete examination of transfer learning (TL) methodologies and their use in the field of medical picture classification. The significance of transfer learning (TL) in addressing the issue of limited data availability and optimizing time and hardware resource utilization is emphasized by the authors. The objective of this paper is to offer recommendations for the selection of suitable models and transfer learning methodologies for problems involving the classification of medical images. This paper examines the benefits and drawbacks of several target language (TL) approaches, while also identifying future research goals and obstacles. Deep learning (DL) methods, such as convolutional neural networks (CNNs), necessitate a substantial volume of data for training purposes. However, in the field of medical imaging, the availability of such data is generally constrained by the restricted size of medical cohorts and the expense associated with acquiring expert-annotated datasets. Transfer Learning (TL) enables the use of information acquired from extensive non-medical datasets, such as ImageNet, to address specific medical picture classification tasks. (Kim, Cosa-Linan, Santhanam, Jannesari, Maros & Gansland 2022.)

The configuration of transfer learning (TL) models in the bulk of studies within the field has been determined arbitrarily. The objective of this review is to help on the selection of suitable models and methodologies for TL. This paper advocates for the utilization of deep models, such as ResNet or Inception, by data scientists and practitioners as feature extractors. (Kim, Cosa-Linan, Santhanam, Jannesari, Maros & Gansland 2022.)

This approach is suggested as it offers the potential to reduce computing expenses and time requirements, while maintaining the predictive capabilities of the models. The review examines the various obstacles and concerns associated with the categorization of medical images.These challenges include the utilization of multiple imaging modalities, the scarcity of labelled datasets, and the necessity for interpretability and explain-ability in deep learning models (Shalbaf et al. 2020). The report outlines potential areas for future research, including the advancement of transfer learning methodologies for multi-modal and multi-task learning, the resolution of class imbalance issues, and the enhancement of interpretability in deep learning models. (Kim, Cosa-Linan, Santhanam, Jannesari, Maros & Gansland 2022.)

This research presents a study that explores the utilization of convolutional neural network (CNN) based algorithms for the purpose of medical image classification. The study specifically concentrates on a dataset of chest X-ray images, with the objective of classifying cases of pneumonia. This paper emphasizes the promise of deep neural networks, particularly convolutional neural networks (CNNs), in attaining notable performance in image classification tasks, including the categorization of medical images. The classification of medical images is of utmost importance in clinical treatment and educational endeavours. However, conventional approaches have certain drawbacks in terms of their performance and reliance on manual feature extraction and selection. (Yadav & Jadhav 2019.)

Deep neural networks, namely convolutional neural networks (CNNs), have become prominent machine learning techniques for a wide range of classification purposes, including the classification of medical images. Convolutional Neural Networks (CNNs) have demonstrated remarkable proficiency in extracting features, making them highly suitable for medical image classification tasks. Their utilization in this domain has the advantage of circumventing the intricate and costly process of feature engineering. This research presents a comprehensive literature analysis on the utilization of conventional techniques and convolutional neural network (CNN)-based transfer learning in the field of medical image classification. (Yadav & Jadhav 2019.)

Additionally, it explores the concept of capsule networks and examines their key components and limitations in the context of medical image classification. The experimental design part delineates the methodology employed in the studies, whilst the section on experimental outcomes entails the presentation of the findings obtained from these investigations, subsequently followed by a comprehensive analysis and interpretation of the data. This study adds to the expanding corpus of literature about the utilization of deep learning, particularly convolutional neural networks (CNNs), for the purpose of medical image classification. The study's primary objective, which is to classify pneumonia using a chest X-ray dataset, highlights the potential of convolutional neural network (CNN)-based algorithms in enhancing illness detection and therapy. (Yadav & Jadhav 2019.)

This publication presents a research study that explores the utilization of transfer learning (TL) and deep learning models in the context of automatically classifying brain tumours using magnetic resonance imaging (MRI) images. The primary objective of the authors is to employ transfer learning (TL) techniques to enhance the classification accuracy of brain tumours. This task presents difficulties owing to the asymmetrical shape, flexible location, and indistinct borders typically associated with such tumours. The classification of brain tumours using automated techniques is a critical undertaking within the field of medical imaging. However, conventional approaches are hindered by their limited accuracy and complex nature. (Arbane, Benlamri, Brik & Djerioui 2021.)

Convolutional neural networks (CNNs) have demonstrated considerable potential in many image classification tasks, including the categorization of medical images, with encouraging outcomes. Transfer learning (TL) is a technique that enables the application of knowledge acquired from pre-trained models to new tasks. This approach becomes advantageous in the context of medical picture classification tasks that have limited training data available. This research presents a comprehensive overview of the existing literature on transfer learning (TL) methodologies and their utilization in the classification of brain tumours using magnetic resonance imaging (MRI) images. The study emphasizes the adoption of pre-trained convolutional neural network (CNN) models, namely VGG-19, ResNet-50, DenseNet-201, MobileNet-v2, Inceptionv3, and AlexNet, for the purposes of feature extraction and classification. The experimental findings section showcases the performance of the TL-based CNN models suggested in this study. (Arbane, Benlamri, Brik & Djerioui 2021.)

These models were evaluated on a dataset consisting of 3064 input images. The results indicate a high level of accuracy and demonstrate that the proposed models outperform existing state-of-the-art methodologies. This study adds to the expanding corpus of literature concerning the utilization of transfer learning and deep learning models in the field of medical image classification. It specifically concentrates on the categorization of brain tumours using magnetic resonance imaging (MRI) pictures. The utilization of transfer learning-based convolutional neural network (CNN) models in this study, together with the exhibition of a notable level of classification accuracy, holds the potential to enhance the effectiveness and precision of brain tumour diagnosis and therapy. (Arbane, Benlamri, Brik & Djerioui 2021.)

The study introduces a novel convolutional neural network (CNN) structure designed for the purpose of classifying brain tumours into three distinct types: meningioma, glioma, and pituitary tumour (Song, Seo, Cho, Woo, Son, Kim, Cho & Kwon 2015). This paper presents a novel approach for brain tumour classification via a convolutional neural network. They employ several techniques on the dataset, including segmentation, cropping, and utilization of both cropped and uncropped tumour images (Anantharajan & Gunasekaran 2021). The performance evaluation of the proposed convolutional neural network (CNN) model is conducted by comparing it with pre-trained models, including VGG-16, ResNet-50, and Inceptionv3, through the utilization of transfer learning techniques. The investigation of brain tumour classification based on MRI images is crucial for precise diagnosis and formulation of treatment strategies. The categorization of brain tumours using conventional approaches typically necessitates physical intervention and is characterized by a time-intensive process. (Badža & Barjaktarović 2020.)

There has been a notable utilization of deep learning methodologies, namely convolutional neural networks (CNNs), in the field of medical image processing, particularly in the categorization of brain tumours. The CNN architecture presented in this research comprises a series of convolutional and pooling layers, which are subsequently followed by fully connected layers for the purpose of classification. This paper utilizes a dataset that encompasses three distinct categories of brain cancers, namely Glioma, Meningioma, and Pituitary tumours. The dataset has undergone preprocessing and augmentation techniques to enhance the performance of the Convolutional Neural Network (CNN) model. The evaluation of the proposed convolutional neural network (CNN) model encompasses the utilization of diverse metrics, including accuracy, precision, recall, and F1-score. (Badža & Barjaktarović 2020.)

The findings indicate that the CNN model suggested in this study demonstrates superior performance compared to the pre-trained models, namely VGG-16, ResNet-50, and Inceptionv3, in terms of accuracy and decrease in losses. It additionally conducts a comparative analysis of their findings with existing studies in the field, thereby showcasing the efficacy of their methodology. The utilization of Convolutional Neural Networks (CNNs) in the realm of brain tumour categorization has been extensively investigated and documented in scholarly works. Previous research studies have also suggested the utilization of hybrid models that integrate convolutional neural networks (CNNs) with support vector machines (SVMs) to classify brain MRI images. The assessment of diverse convolutional neural network (CNN) models for the purpose of brain tumour classification has garnered significant attention among researchers. They have been investigating the effectiveness of different architectural designs and methodologies to determine their performance. (Badža & Barjaktarović 2020.)

The present study provides a thorough examination of the utilization of deep learning methodologies in the identification and categorization of Alzheimer's disease (AD) through the analysis of diverse medical imaging modalities, including magnetic resonance imaging (MRI) and positron emission tomography (PET). It examines the difficulties associated with the diagnosis and classification of Alzheimer's disease (AD), explore the potential of deep learning techniques in addressing these obstacles, and outline the future possibilities for research in this domain. Alzheimer's disease (AD) is a prevalent neurodegenerative condition that impacts a significant global population. The timely and precise identification of this disorder is of utmost importance to facilitate optimal therapy and care. (Al Shehri 2022.)

Deep learning techniques, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have demonstrated considerable potential in a range of image and signal processing applications, notably in the field of medical picture analysis. This study presents a comprehensive examination of the difficulties encountered in the diagnosis and classification of Alzheimer's disease (AD). These problems include the intricate and diverse characteristics of the disease, the absence of biomarkers for timely detection, and the requirement for extensive and annotated datasets to effectively train deep learning models. It provides a comprehensive analysis of the recent progress made in the field of deep learning-based Alzheimer's disease (AD) diagnosis and classification, specifically focusing on the utilization of magnetic resonance imaging (MRI) and positron emission tomography (PET) pictures. (Al Shehri 2022.)

This paper emphasizes the significance of convolutional neural networks (CNNs) in extracting relevant features and performing accurate classification. Furthermore, it underscores the benefits of integrating multi-modal imaging data to enhance the precision of AD diagnosis and classification. The paper additionally examines the constraints of existing deep learning methodologies, including the absence of interpretability and the necessity for more resilient and comprehensible models. It proposes potential avenues for further investigation in this domain, such as the creation of personalized and adaptive deep learning models for the diagnosis and classification of Alzheimer's disease. (Al Shehri 2022.)

The primary objective of this paper is to enhance the precision and efficiency of brain tumour identification and categorization through the utilization of deep learning methodologies with magnetic resonance imaging (MRI) scans. The primary aim is to conduct a performance study of CNN models, namely VGG-16, ResNet-50, and Inception-v3, that have been pretrained using transfer learning. The objective of this research is to evaluate the models' effectiveness in automatically predicting tumour cells in the brain. Magnetic Resonance Imaging (MRI) is widely employed in the field of medical imaging due to its superior image quality and its ability to operate without the use of ionizing radiation. Deep learning, which falls under the umbrella of artificial intelligence, has demonstrated considerable potential in enhancing the accuracy of brain tumour identification through the analysis of MRI scans. (Srinivas, KS, Zakariah, Alothaibi, Shaukat, Partibane & Awal 2022.)

The researchers utilized a dataset including 233 magnetic resonance imaging (MRI) brain tumour images for their analytical investigation. The utilization of transfer learning is employed, which is a methodology enabling the utilization of pretrained models as a foundation for training on novel datasets. This approach facilitates the exploitation of the knowledge acquired from extensive image recognition tasks. The evaluation of the VGG-16, ResNet-50, and Inception-v3 models is conducted to assess their predictive capacity in detecting tumour cells within the brain. The accuracy, precision, recall, and F1-score for each model are reported by the authors, showcasing the efficacy of deep transfer learning methods in the categorization of brain tumours. The study is a valuable contribution to the advancement of effective and dependable instruments for healthcare practitioners involved in the identification and categorization of brain tumours. (Srinivas, KS, Zakariah, Alothaibi, Shaukat, Partibane & Awal 2022.)

The utilization of deep learning methodologies, particularly when employed in conjunction with transfer learning, has the potential to yield enhanced prediction accuracy and improved patient outcomes. This paper suggest a deep transfer learning model that enhances the speed of brain tumour diagnosis through the utilization of MR images. The primary objective is to examine the classification of brain tumours into multiple classes using MRI scans. In doing so, they conduct a comparative analysis of the performance of their proposed model against other contemporary models that are at the forefront of the field. The application of deep learning techniques has demonstrated potential in enhancing the accuracy of brain tumour identification through the analysis of MRI images. Nevertheless, the categorization of brain tumours with multiple classes continues to be a formidable challenge. It utilizes a dataset of 306 MRI brain tumour images for their research. It puts forth an advanced deep learning approach that integrates transfer learning with a convolutional neural network (CNN) utilizing the VGG-16 architecture. (Srinivas, KS, Zakariah, Alothaibi, Shaukat, Partibane & Awal 2022.)

In addition, data augmentation techniques are employed to augment the dataset's size and enhance the model's performance. The evaluation of the proposed model centers on its predictive capacity for three distinct categories of brain malignancies, namely glioma, meningioma, and pituitary tumours. This study present an evaluation of their model's performance in multi-class brain tumour classification, specifically focusing on accuracy, precision, recall, and F1-score metrics. The results indicate the effectiveness of their model in this task. The study makes a valuable contribution to the advancement of effective and dependable instruments utilized by healthcare practitioners in the realm of brain tumour identification and categorization. The utilization of deep transfer learning methodologies, particularly when employed alongside data augmentation procedures, has the potential to yield elevated prediction accuracies and enhanced patient outcomes. (Srinivas, KS, Zakariah, Alothaibi, Shaukat, Partibane & Awal 2022.)

This paper offers a methodology that employs deep learning techniques with transfer learning to achieve precise classification of different chickpea types. The primary objective is to enhance the efficiency and dependability of chickpea varietal classification, a critical aspect of crop management and breeding initiatives. The chickpea, a significant legume crop, exhibits a wide range of variations. Precise and effective categorization of plant varieties holds significant importance in the realms of agricultural management, breeding initiatives, and quality assurance. Convolutional neural networks (CNNs) have demonstrated considerable potential in the domain of image-based categorization problems. This paper utilizes a dataset including 1,200 photos of chickpeas, which encompasses four distinct types. It utilizes a convolutional neural network (CNN) architecture that is characterized by its depth. This architecture incorporates transfer learning, specifically by leveraging the VGG-16 model that has been pre-trained on the ImageNet dataset. (Saha & Manickavasagan 2022.)

The final layer of the VGG-16 model is substituted with a novel layer designed for the purpose of classifying chickpea varieties. The evaluation of the proposed model's performance is conducted by assessing its accuracy in categorizing different varieties of chickpea. This paper present findings indicating that their model exhibits a notable level of accuracy, thereby showcasing its efficacy in the classification of different grape varieties. Additionally, it conducts a comparative analysis between their model and other contemporary models, thereby emphasizing its superior performance. The study makes a valuable contribution to the advancement of effective and dependable instruments for crop management and breeding initiatives within the domain of chickpea cultivation. The utilization of deep convolutional neural networks (CNNs) in conjunction with transfer learning enables precise and expeditious varietal categorization, hence contributing to the enhancement of agricultural yield, quality, and sustainability. (Saha & Manickavasagan 2022.)

This paper investigates the utilization of transfer learning and deep convolutional neural networks (CNNs) in the context of automated plant identification. The objective is to enhance the efficacy and precision of plant identification, a critical aspect in multiple domains such as agriculture, ecology, and conservation. The identification of plants is of utmost importance in a range of applications, including the monitoring of biodiversity, the management of crops, and the conduct of ecological research. Deep learning approaches, particularly Convolutional Neural Networks (CNNs), have demonstrated considerable potential in the domain of image-based categorization problems. This paper utilize a dataset comprising of plant photos that depict various species for the purpose of their investigation. The utilization of transfer learning is employed, which is a technique that enables the utilization of pretrained models as a foundation for training on novel datasets. (Saha & Manickavasagan 2022.)

This approach facilitates the exploitation of the knowledge acquired from extensive image recognition tasks. This paper conduct a comparative analysis of various pretrained convolutional neural network (CNN) models, including VGG-16 and Xception, with respect to their efficacy in feature extraction.The evaluation of the suggested model's performance is conducted by assessing its accuracy in correctly identifying plant species. This paper present in their study notable levels of precision for their model, showcasing the efficacy of transfer learning and deep convolutional neural networks in the automated identification of plants. It also engages in a discussion regarding the potential obstacles and future directions within this field of study. (Saha & Manickavasagan 2022.)

This study makes a valuable contribution to the advancement of efficient and dependable methods for the identification of plants. This development has wide-ranging benefits across multiple disciplines, such as agriculture, ecology, and conservation. The application of deep learning methodologies, particularly convolutional neural networks (CNNs) integrated with transfer learning, facilitates precise and expeditious identification of plant species. This technological advancement holds potential in supporting endeavours such as biodiversity monitoring, crop management, and ecological study. (Saha & Manickavasagan 2022.)

4 METHODOLOGIES

The dataset utilized in this research comprises 3064 brain MRI scans, each consisting of raw pixel-by-pixel data and varying in size. These Joint Photographic Experts Group (JPEG) format images were taken from the Kaggle dataset and are classified into four types: healthy brain tissue (930 images), meningiomas (708 images), pituitary gland tumours (926 images), and gliomas (1426 images). The dataset is an amalgam of the Br35H, figshare, and SARTAJ datasets. The images undergo preprocessing prior to being resized to the appropriate dimensions for analysis. The dataset obtained from Kaggle combines the SARTAJ, Br35H, and figshare datasets. The collection incorporates brain MRI scan images, with a particular focus on healthy brain tissue, pituitary gland tumours, meningiomas, and gliomas. Tumours in the images from an MRI are one of the selection criteria. The dataset containing MRI brain images is presented in TABLE 1.

TABLE 1: Dataset Description

Name of Dataset	**Types**	**No. of Samples**	**Total Samples**
Brain Tumour	Glioma	1426	3064
	Meningioma	708	
	Pituitary	930	

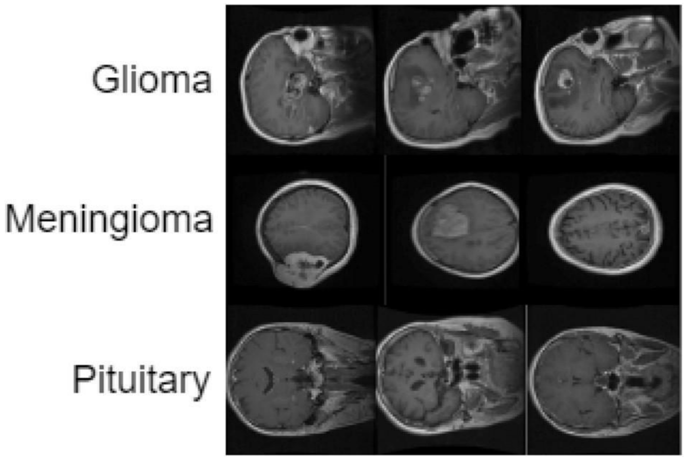


FIGURE 1: Types of Brain Tumour (adapted from Mukherkjee, Saha, Kaplun, Sinitca & Sarkar 2022).

The preprocessing methods for Kaggle's brain image dataset prior to the use of pre-trained CNN models are outlined as the necessary packages should be imported. Import the two data folders labelled as "Yes" and "No". The photos should be read and subsequently transformed into labelled images, where the label "Tumour=Yes" indicates the presence of a tumour, and the label "Tumour=No" indicates the absence of a tumour. The MRI images that have been appropriately labelled should be stored within the data frames. The photos should be resized to a dimension of 256 × 256. The MRI images present in the dataset underwent preprocessing procedures as illustrated in Figure 2.

FIGURE 2: Dataset steps (adapted from Mukherkjee, Saha, Kaplun, Sinitca & Sarkar 2022).

The dataset encompasses several characteristics, such as the classification of brain cancers into three types: glioma, meningioma, and pituitary (FIGURE 1). It has a total of 3064 samples, with specific sample counts for each tumour type: 708 meningiomas, 930 pituitary tumours, and 1426 gliomas. Each image possesses varying dimensions and is encoded in the JPEG file format. The utilization of datasets is crucial in the application of deep learning and various machine learning methodologies for the purposes of training, testing, and validating brain cancer research. Intensity normalization, which addresses potential differences in pixel intensity owing to variable scanning periods and equipment restrictions, is a crucial step in standardizing pixel values across multiple photographs. To comply with the specifications of CNN pre-trained models, MRI pictures are resized to a precise dimension of 224x224x3. (FIGURE 2.)

Data augmentation techniques, such as scaling, cropping, resizing, flipping, rotating, and viewpoint manipulation, are used to get beyond the dataset's limits. To generate synthetic data for CNN model training, Affine image transformation and pixel-level picture transformation are employed.
Initially, open-source computer vision (CV) is employed to extract the brain region from MRI brain images by trimming the input MRI photos. Numerous researchers have developed computer-aided diagnostic (CAD) models on publicly available small-scale datasets (Song, Seo, Cho, Woo, Son, Kim, Cho & Kwon 2015.). Nevertheless, it is acknowledged that contemporary deep transfer learning models necessitate a substantial amount of data to achieve enhanced classification accuracy.

Consequently, a substantial quantity of data is required for training the Convolutional Neural Network (CNN) model to mitigate the potential occurrence of overfitting problems. The present study employed data augmentation, a preprocessing technique in transfer learning, to overcome the limitations imposed by the restricted datasets utilized in the study. Various techniques, such as scaling, cropping, resizing, flipping, rotating, and perspective alteration, are employed based on the specific requirements.
The utilization of this specific methodology is to generate artificial visual data from the primary dataset with the purpose of facilitating training. The proposed diagnostic model can potentially achieve improved performance by the utilization of artificial data augmentation techniques. These techniques generate new and unique data examples to be used for training the model.

The empirical results demonstrate that the deep learning-based computer-aided design (CAD) model, which has been trained using artificially augmented data, exhibits superior performance and produces outputs that are more precise when compared to the authentic picture dataset. The generation of additional training examples for the model involves the utilization of data augmentation techniques, specifically the affine image transformation approach and pixel-level picture alteration methodologies.

5 MODEL DESCRIPTION

In the next subsections, a full description of the modelling technique, including its theoretical background and practical execution has been provided.

5.1CNN

Convolutional neural networks (CNNs), which belong to the category of deep learning models, are commonly employed in the field of image processing for various tasks such as segmentation, object recognition, and classification. Furthermore, these techniques have been utilized in several domains, such as natural language processing and medical imaging. Since they excel in identifying spatial hierarchies and patterns in data, CNNs are well suited for applications that use grids, such as image processing. (Arbane, Benlamri, Brik & Djerioui 2021.)

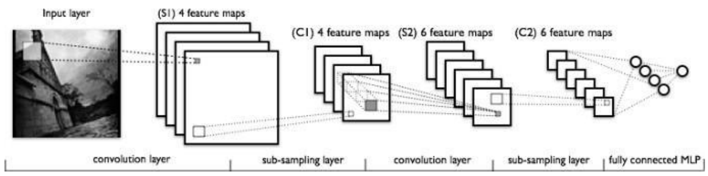


FIGURE 3: CNN Architecture (adapted from Arbane, Benlamri, Brik & Djerioui 2021.)

5.2 Mathematical Representation

CNN is made of several layers, each of which serves a particular purpose. Mathematically, it analyses the fundamental components and explains how each of them functions. Input Layer (I) in an image classification task, the input is a 3D tensor representing an image. Suppose the input image has dimensions width (W), height (H), and channels (C, typically 3 for RGB images). The input layer can be represented as-

$$I \in R \wedge (W \times H \times C)$$

Where ‘I ’stands for Input Layer, R is Real numbers, W stands for Width, H for Height and C for Channels.

The convolution operation is the heart of CNN. The process entails the application of a collection of trainable filters, sometimes known as kernels, to the input image. Each filter possesses a limited receptive field and moves over the input to generate feature maps. Mathematically, the convolution operation can be expressed as follows. (FIGURE 3.)

$$y_{i,j} = b + \sum_{m=0}^{F-1} \sum_{n=0}^{F-1} x_{i+m,\, j+n} W_{m,n}$$

where the output value at position (i,j) is denoted as yi,,j. The bias term is represented by b. The input value at position (i+m,j+n) is denoted as xi + m, j+n. The filter value at position (m,n) is represented by wm,n. F represents the filter size. This equation assumes that the input and the filter have the same number of channels and that the stride and padding are both equal to 1.

The output value at position (i,j) is denoted as $y_{i,,j}$. The bias term is represented by b. The input value at position (i+m,j+n) is denoted as x_i + m, j+$_n$. The filter value at position (m,n) is represented by $w_{m,n}$. F represents the filter size. This equation assumes that the input and the filter have the same number of channels and that the stride and padding are both equal to 1.

In Activation Function (e.g., ReLU), a convolutional layer or a fully linked layer's output is transformed into a non-linear form using an activation function. Typically, it is applied to each value in the output feature map element-by-element. The neural network's performance, capacity for learning, and the kinds of predictions it can make may all be influenced by the activation function. Many other activation function types exist, including sigmoid, tanh, ReLU, Leaky ReLU, SoftMax. The job and the data determine which activation function should be used, and each activation function has advantages and disadvantages of its own. One such instance involves the utilization of Rectified Linear Unit (ReLU), which is employed in an element-wise manner to induce non-linearity.

$$A_{(i,j)} = \max(0, F_{(i,j)})$$

where i stands for row, j for column, A for Output and F is for the value of the original feature map.

A pooling layer is a layer within a convolutional neural network (CNN) that serves to decrease the spatial dimensions of the input feature maps while maintaining the depth which refers to the number of channels. The pooling layer operates by partitioning the input feature map into a collection of nonoverlapping regions, referred to as pooling regions, and subsequently performing a pooling operation on each individual region. The pooling operation can be categorized into two types: max pooling and average pooling. These operations summarize the maximum or average value of the elements inside each zone, respectively. The pooling layer generates a novel feature map that possesses reduced dimensions while preserving the salient characteristics from the original input. Pooling layers are employed in convolutional neural networks (CNNs) to enhance their performance and efficiency. This is achieved by decreasing the number of parameters and computations required. Additionally, pooling layers introduce a degree of translation invariance to the extracted features.

$$P_{(i,j)} = \max(A_{(2i,2j)}, A_{(2i,2j+1)}, A_{(2i+1,2j)}, A_{(2i+1,2j+1)})$$

where P stands for Result, A for input feature map value, (i, j) for the position.

A Fully Connected (FC) layer within a convolutional neural network (CNN) is characterized by its ability to establish connections between each unit present in the input feature map and every unit present in the output feature map. The fully connected layer is responsible for executing a linear transformation, which is subsequently followed by an optional activation function. The utilization of a fully connected layer is versatile, since it can serve multiple objectives including classification, regression, and dimensionality reduction. This layer can be regarded as a unique instance of a convolutional layer, in which the dimensions of the filter are identical to those of the input, and the quantity of filters corresponds to the dimensions of the output. This implies that a convolutional layer with identical parameters can serve as a substitute for a fully linked layer. Nevertheless, it is worth noting that a fully connected layer typically possesses a greater number of parameters and computational requirements compared to a convolutional layer. This is mostly because a fully connected layer fails to leverage the spatial structure inherent in the input data. (Thomas Wiatowski, Helmut Bölcskei 2017.)

Following the application of many convolutional and pooling layers, predictions are made using fully connected layers. The aforementioned layers can be characterized as tightly connected neural networks.

$$O = \sigma (W_x + b)$$

In the equation, the output vector is denoted as O, the weight matrix as W, the input vector as x, the bias vector as b, and the activation function as σ.

In the given context, the output vector is denoted as O, the weight matrix as W, the input vector as x, the bias vector as b, and the activation function as σ.

The primary purpose of the output layer of a Convolutional Neural Network (CNN) is to generate the ultimate output of the network by utilizing the features that have been extracted by the preceding levels (Wiatowski, Bölcskei 2017). Typically, the output layer is a fully connected layer that establishes connections between each unit in the input feature map and each unit in the output feature map. The composition of units in the output layer might vary in terms of both quantity and kind, contingent upon the specific task at hand and the characteristics of the dataset. In the context of image classification, it is common practice to configure the output layer with a few units equal to the total number of classes. This allows for the utilization of a SoftMax activation function, which facilitates the generation of a probability distribution over the various classes.

6 TRANSFER LEARNING

Transfer learning is a machine learning technique that enables the utilization of a pre-existing model produced for one specific task as a foundation for constructing a model for another distinct activity (Sinno Jialin Pan, Qiang Yang 2010.). When the second job is identical to the first or has little data, this approach may be successful. By employing the learned knowledge characteristics from the first assignment as a starting point for the second challenge, the model may learn more rapidly and efficiently. Because the model has already learned broad traits that will be useful in the second assignment, this can also help prevent overfitting. One approach for illustrating the concept of transfer learning involves the utilization of a diagram that visually represents the hierarchical structure of the neural network and its modifications or adaptations for the subsequent task. The illustration below, in Figure 4, provides an example of a transfer learning scenario.

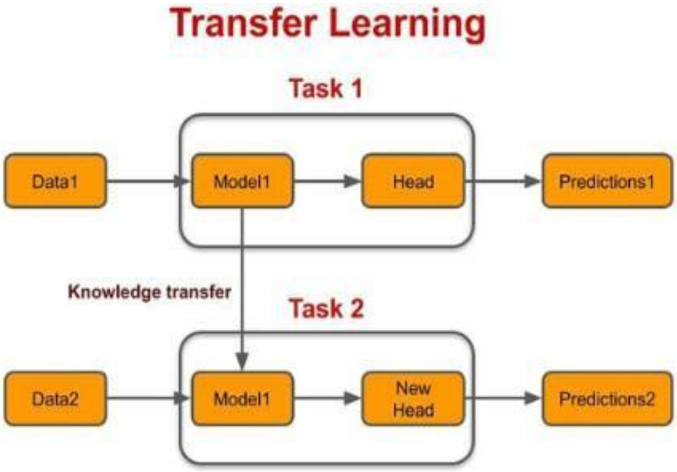


FIGURE 4: Transfer Learning

The transfer learning process consists of two distinct steps, namely feature extraction and fine-tuning. The basic model in feature extraction is frozen, which prevents weight updates during training. The result of the base model is used as input for the new classification layer while it is being trained from scratch on the flower dataset. As a result, the newly added layer can benefit from the basic model's general and advantageous picture recognition features. With fine-tuning, some, or all the layers in the basic model are unfrozen, allowing their weights to fluctuate throughout training. Using a lower learning rate than in feature extraction, the new classification layer and the basic model's unfrozen layers are simultaneously trained on the flower dataset. This allows the base model to adjust its features to better suit the new task while preserving its general knowledge. (FIGURE 4.)

TABLE 2: Model Overview

No	Model	Core Model	Architecture	Input Matrix	Year
1	AlexNet	CNN	5-Conv, 3-FC, 3 Max-pool Layers	(256, 256, 3)	2012
2	GoogleNet	CNN+Inception	59-Conv, 9-Inception, 1-FC, 5 Max-pool Layers, Avg-pool Layers	(224, 224, 3)	2014
3	ResNet-50	CNN+Residual Connection	48-Conv, 1-Max-pool Layer, 1-Avg-pool Layers	(224, 224, 3)	2015
4	VGG-19	CNN	19-Conv, 3-FC, 5 Max-pool Layer	(224, 224, 3)	2014
5	InceptionV3	CNN+Inception	86-Conv, 11-Inception, 3-FC, Max-pool Layer, Global Avg-pool Layers	(256, 256, 3)	2015
6	DenseNet-121	Dense CNN	103-Conv, 4-Transition layers, 1-Global Avg-pool Layers	(224, 224, 3)	2016
7	SqueezeNet	Compact CNN	8-Fire Module, 2-Conv Layers	(224, 224, 3)	2016
8	MobileNetV2	Region-based CNN	86-Conv, Residual Blocks, Global Avg-pool Layer	(224, 224, 3)	2018

6.1 AlexNet

AlexNet is a convolutional neural network (CNN) architecture developed by Alex Krizhevsky and colleagues in 2012. It won the ImageNet Large Scale Visual Recognition Challenge with a top-5 error of 15.3%, far lower than the previous best-in-class models. It comprises eight layers, five of which are convolutional and three of which are fully connected. To decrease memory use and speed up training, the network is divided into two parallel streams, each executing on a single GPU. The convolutional layer in AlexNet is a layer that convolutions input data, which is a linear combination of the input and a filter (or kernel) matrix. A convolutional layer may extract characteristics like as edges, shapes, textures, and so on from input data. AlexNet features has five convolutional layers, each with its own set of filters, padding, strides, and filter sizes. (FIGURE 5.) The first convolutional layer consists of 96 11x11 filters with a stride of 4 and no padding, resulting. (TABLE 2.) (Krizhevsky, Sutskever, Hinton 2012.)

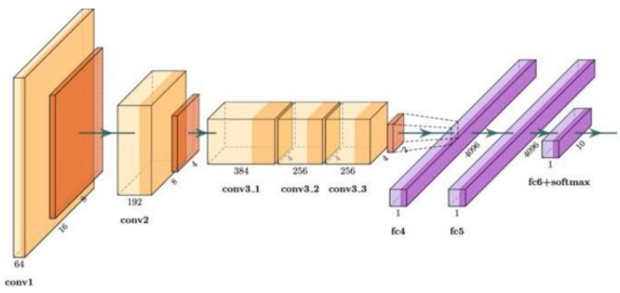


FIGURE 5: Architecture of AlexNet (adapted from Krizhevsky, Sutskever, Hinton, 2012.)

In a feature map with dimensions of 55x55x96. The second convolutional layer contains 256 5x5 filters with stride 1 and padding 2, yielding a 27x27x256 output feature map. (Krizhevsky, Sutskever, Hinton 2012.)

The third convolutional layer contains 384 3x3 filters with stride of 1 and padding of 1, yielding a 13x13x384 output feature map. The fourth convolutional layer contains 256 3x3 filters with stride of 1 and padding of 1, yielding a 13x13x256 output feature map. The fifth convolutional layer contains 256 3x3 filters with stride of 1 and padding of 1, yielding a 13x13x256 output feature map. (TABLE 2.) (Krizhevsky, Sutskever, Hinton 2012.)

$$m = f_a (W_m \times n + bv)$$

where, n stands for input, m stands for output, W_m stands for weight matrix, bv stands for bias vector, f_a stands for activation function.

AlexNet has three pooling layers, each with its own filter size and stride. The first pooling layer is applied after the first convolutional layer, with a filter size of 3x3 and a stride of 2, creating an output feature map with dimensions of 27x27x96. Following the second convolutional layer, the second pooling layer is applied, which employs max pooling with the same parameters to provide an output feature map with dimensions of 13x13x256. Following the fifth convolutional layer, the third pooling layer is utilized to generate an output feature map with dimensions 6x6x2563 utilizing max pooling, a filter size of 3x3, and a stride of 2. (Krizhevsky, Sutskever, Hinton 2012.)

AlexNet normalization layers are layers that normalize the output of a convolutional layer, which is a feature map with several channels. The goal of normalization layers is to reduce internal covariate shift, which is the change in the distribution of layer activations caused by changes in network parameters during training. Normalization layers can also improve the network's generalization ability, avoid overfitting, and accelerate convergence. AlexNet uses local response normalization (LRN) layers, which are a type of divisive normalization. LRN layers normalize the activations across nearby channels at the same spatial location. (Krizhevsky, Sutskever, Hinton 2012.)

LRN layers apply the following formula to each activation -

$$y_{i,\,j,\,k} = \frac{x_{i,j,k}}{k+\alpha \sum_{l=max(0,j-n2)}^{min=(N-1,j+n2)} x_{i,l,k}^{2}}$$

In $x_{i,j,k}$, i stands for batch index, j stands for channel index, k stands for spatial index.
In formula, N stands for number of channels, n stands for size of local region, k stands for constant, α stands for scaling parameter, β stands for exponent parameter.

AlexNet features two LRN layers, one after each of the first and second convolutional layers. The filter size of the LRN layers is 5, the scaling parameter is 0.0001, the exponent parameter is 0.75, and the constant is 2. In Fully Connected Layer, the architecture of AlexNet has three completely connected layers, wherein each layer exhibits distinct neuron counts. The initial layer of the neural network consists of 4096 neurons, followed by a second layer also including 4096 neurons. The subsequent layer comprises 1000 neurons, which aligns with the total number of classes included in the ImageNet dataset. The final prediction of the network is obtained by passing the output of the last completely connected layer through a SoftMax function. (Krizhevsky, Sutskever, Hinton 2012.)

Data augmentation is a methodology employed to artificially enhance the magnitude and heterogeneity of a dataset through the application of diverse modifications to the original images. These transformations may include but are not limited to flipping, cropping, rotating, scaling, altering colours, introducing noise, and other similar techniques. Data augmentation can help improve the generalization ability of a neural network by reducing overfitting and increasing the robustness of different inputs. Some examples of the data augmentation techniques used by AlexNet are randomly cropping 224x224 patches from the 256x256 images, resulting in a 2048-fold increase in the number of training samples. Randomly flipping the images horizontally. The intensities of the RGB channels are modified by using Principal Component Analysis (PCA) to the collection of RGB pixel values across the training dataset. (Krizhevsky, Sutskever, Hinton 2012.). These data augmentation techniques helped AlexNet to learn more invariant and discriminative features from the images and reduced the risk of overfitting to the limited training data. The AlexNet model underwent training on a GTX 580 GPU, which possessed a limited memory capacity of 3 GB, rendering it unable to accommodate the whole network structure. The network was partitioned into two separate processing units, specifically two GPUs, where an equal distribution of neurons (also referred to as feature maps) was allocated to each GPU. The presence of a division inside the architecture diagram can be attributed to this factor. The technique of model ensembling was employed to achieve optimal outcomes.

6.2 GoogleNet

The aforementioned architectural design achieved the top position in the 2014 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) pertaining to the task of image categorization. The current model has demonstrated a significant decrease in error rate in comparison to previous winners of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) such as AlexNet (winner in 2012), ZF-Net (winner in 2013), and VGG (runner-up in 2014). Within the realm of architectural design, this design employs methodologies such as global average pooling and 1-1 convolutions. In features of GoogleNet, GoogleNet uses a variety of approaches, including global average pooling and 1X1 convolution, to create deeper architecture. (Szegedy, Christian, Liu, Wei, Jia, Yangqing, Sermanet, Pierre, Reed, Scott, Anguelov, Dragomir, Erhan, Dumitru, Vanhoucke, Vincent, Rabinovich & Andrew 2014.)

For example, the use of one-to-one convolution is a notable characteristic inside the inception architecture. The number of weights and biases in the design was decreased through the utilization of these convolutions. Decreasing the parameters additionally enhances the complexity of the system. Presented here is an illustration depicting a 1:1 convolution, to get a 6x6 convolution with 54 filters without utilizing 1x1 convolution as an intermediary step, the following approach can be applied here. Total operations performed- (14 x 14 x 48) x (5 x 5 x 480) = 112.9

Applying a 1x1 convolution- (1 × 1 × 480) × (14 × 14 × 16) +(14 × 14 × 48) × (5 × 5 × 16) = 1.5M + 3.8M = 5.3M, which is less than 112.9M.

In Global Average Pooling, Prior designs, like AlexNet used entirely connected layers at the network's edge. Most of the parameters in many designs, which increase computation costs, are in these entirely connected layers. At the edge of the network, a technique known as global average pooling is applied in the GooLeNet architecture. In this layer, a 7x7 feature map is averaged down to a 1x1 size. Additionally, this reduces the number of trainable parameters to 0 and increases the accuracy of the top-1 by 0.6. (Szegedy, Christian, Liu, Wei, Jia, Yangqing, Sermanet, Pierre, Reed, Scott, Anguelov, Dragomir, Erhan, Dumitru, Vanhoucke, Vincent, Rabinovich & Andrew 2014.). In Inception Module, unlike ZF-Net and AlexNet, which are older systems, the inception module is unique. Each layer in this design has a convolution size assigned to it. The final output is created in the Inception module by stacking the outputs of the parallel 1x1, 3x3, 5x5, and 3x3 max pooling operations that were performed at the input.

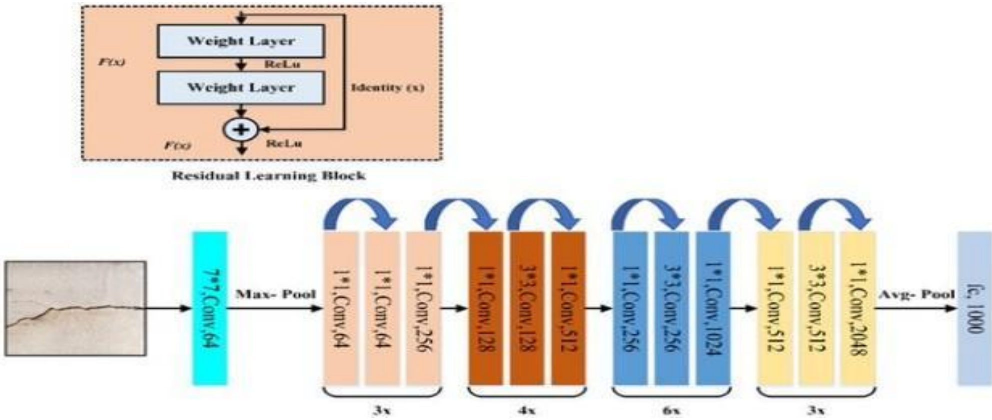
Theoretically, convolution filters of different sizes will be better able to handle objects of different scales. In Auxiliary Classifier for Training, a few intermediate classifier branches scattered throughout the Inception architecture are only used during training. These branches comprise a SoftMax classification layer, two layers comprising a layer with 55 average pooling and a stride of 3, a layer with 11 convolutions, 128 filters, and 1024 outputs each with full connections between them., and two layers with 1000 outputs each. The loss generated by these layers weighed in at 0.3 of the total loss. These layers assist in addressing the gradient vanishing issue in addition to regularization. (FIGURE 6.) (Szegedy, Christian, Liu, Wei, Jia, Yangqing, Sermanet, Pierre, Reed, Scott, Anguelov, Dragomir, Erhan, Dumitru, Vanhoucke, Vincent, Rabinovich & Andrew 2014.)

The layers of GoogleNet, the overall architecture consists of 22 levels. The architecture was developed with computing efficiency in consideration. The architecture is made to function on solitary devices with constrained computing power. The architecture links two additional classifier layers to the outputs of the Inception (4a) and Inception (4d) layers. The auxiliary classifiers' architectural specifications are a standard pooling layer with three strides and a 5x5 filter size. ReLU activation and dimension reduction using a 1x1 convolution with 128 filters. (c) A completely linked layer that activates ReLU and has 1025 outputs. Dropout Regularization using a 0.7 dropout ratio. A 1000-class softmax classifier that achieves performance levels that are like those of the main classifier. (FIGURE 7.)

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

FIGURE 6: GoogleNet Model (adapted from Szegedy, Christian, Liu, Wei, Jia, Yangqing, Sermanet,

Pierre, Reed, Scott, Anguelov, Dragomir, Erhan, Dumitru, Vanhoucke, Vincent, Rabinovich & Andrew 2014).



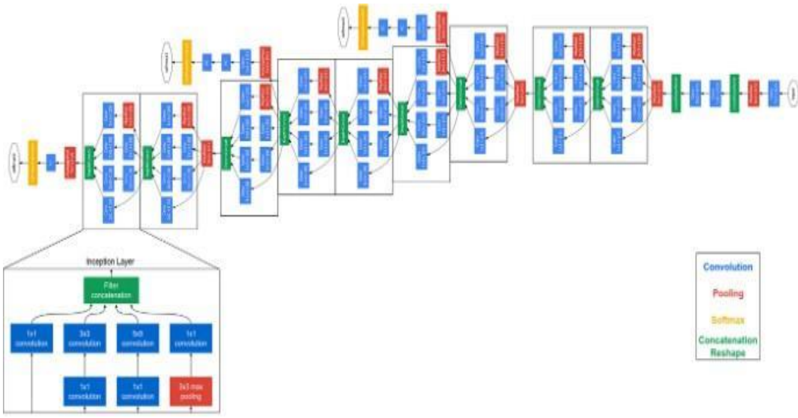


FIGURE 7: Architecture of GoogleNet (adapted from Szegedy, Christian, Liu, Wei, Jia, Yangqing, Sermanet, Pierre, Reed, Scott, Anguelov, Dragomir, Erhan, Dumitru, Vanhoucke, Vincent, Rabinovich & Andrew 2014).

6.3 ResNet-50

The deep convolutional neural network architecture ResNet-50, also referred to as "Residual Network-50," was first presented. ResNet-50 is a version of the original ResNet architecture that is commonly used in several computer vision applications, most notably image categorization. (He, Zhang, Ren, Sun, 2015.)

FIGURE 8: Architecture of ResNet-50

The primary innovation in the ResNet architecture is the use of residual connections, often known as skip connections or shortcut connections. These connections enable the formation of incredibly deep neural networks by solving the vanishing gradient problem, which can make it challenging to train deep networks without residual connections. The ResNet-50 architecture's main characteristics and elements are depth of ResNet-50 has 50 layers and is a deep network. (He, Zhang, Ren, Sun, 2015.) Each of these blocks or stages, which make up these layers, has a distinct number of residual units. In a nutshell, the architecture is as the first layer of convolution is the maxpooling layer. There are four phases, each with several residual units. A pooling layer with a global average layer. A categorization layer that is completely linked. Residual Units of ResNet-50 has numerous residual units for each level. Two or three convolutional layers, batch normalization, and ReLU activation functions constitute a residual unit. The important concept is that residual units learn the residual (difference) between the input and the desired output rather than the desired mapping directly. Before passing through the subsequent unit, the output of one residual unit is added to the input (shortcut connection). (FIGURE 8.)

Bottleneck Architecture of ResNet-50's residual units have a bottleneck design, which helps to decrease computational costs while retaining excellent performance. (He, Zhang, Ren, Sun, 2015.)

A typical ResNet-50 residual unit has the following structure: 1x1 convolution, 3x3 convolution, and another 1x1 convolution. Global Average Pooling of ResNet-50 leverages global average pooling rather than fully connected layers with a high number of parameters for final classification. This method computes the average of each feature map's overall spatial dimensions, providing a tiny fixed-size tensor that may be used directly for classification.

TABLE 3: ResNet-50 Model (adapted from He, Zhang, Ren, Sun, 2015.)

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		[3×3, 64 3×3, 64] ×2	[3×3, 64 3×3, 64] ×3	[1×1, 64 3×3, 64 1×1, 256] ×3	[1×1, 64 3×3, 64 1×1, 256] ×3	[1×1, 64 3×3, 64 1×1, 256] ×3
conv3_x	28×28	[3×3, 128 3×3, 128] ×2	[3×3, 128 3×3, 128] ×4	[1×1, 128 3×3, 128 1×1, 512] ×4	[1×1, 128 3×3, 128 1×1, 512] ×4	[1×1, 128 3×3, 128 1×1, 512] ×8
conv4_x	14×14	[3×3, 256 3×3, 256] ×2	[3×3, 256 3×3, 256] ×6	[1×1, 256 3×3, 256 1×1, 1024] ×6	[1×1, 256 3×3, 256 1×1, 1024] ×23	[1×1, 256 3×3, 256 1×1, 1024] ×36
conv5_x	7×7	[3×3, 512 3×3, 512] ×2	[3×3, 512 3×3, 512] ×3	[1×1, 512 3×3, 512 1×1, 2048] ×3	[1×1, 512 3×3, 512 1×1, 2048] ×3	[1×1, 512 3×3, 512 1×1, 2048] ×3
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

In Output Layer, the number of classes in the classification task is correlated with the number of neurons in the last layer, which is fully connected. A SoftMax activation function is used to generate class probabilities. The ImageNet large scale visual recognition challenge is one of the image classification benchmarks where ResNet-50 has demonstrated its remarkable performance. Object identification, semantic segmentation, and other computer vision applications have shown widespread adoption and use of this methodology. Because of their efficacy and efficiency, pre-trained ResNet-50 models are frequently used as a beginning point for a variety of computer vision applications by researchers and professionals. (He, Zhang, Ren, Sun, 2015.) (TABLE 3.)

6.4 VGG-19

The VGG-19 model was developed by Simonyan and Zisserman of the University of Oxford. It has 19 layers (16 linked, 3 fully linked), 22 max-pooling layers with stride 2, rigorously applies 33 filters with stride and pad of 1. The VGG-19 is a CNN that is more intricate and has more layers than AlexNet. It makes the best use of its modest 33 filters and 7.3% error rate over all convolutional layers to decrease the number of parameters in such deep networks. Even though the VGG-19 model did not take home the top prize at ILSVRC 2014, the VGG Net article remains one of the most important ones since it solidified the idea that CNNs must have a sizable number of layers for this hierarchical representation of visual input to function. The VGG-19 model, with a total of 138 million parameters, came in first for localization and second for classification at the ILSVRC 2014. (FIGURE 9.) (Simonyan & Zisserman, 2014.)

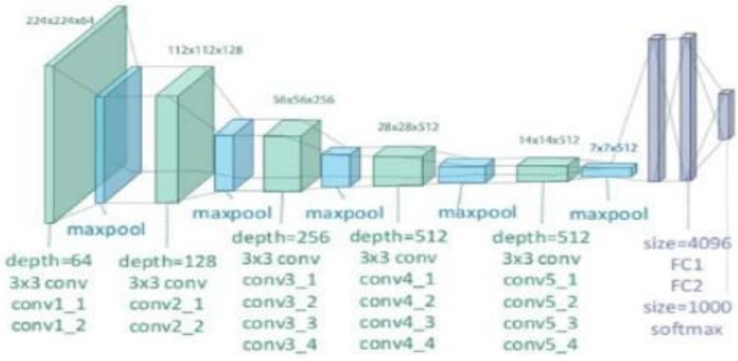


FIGURE 9: Architecture of VGG-19 (adapted from Simonyan & Zisserman, 2014.)

The ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) uses a portion of the ImageNet database, which is utilized to train this model. The VGG-19 can categorize photos in 1,000 categories, including categories for keyboards, mice, pencils, and various animals. For training, about a million pictures were used. (Simonyan & Zisserman, 2014.). As a result, the model has produced detailed feature representations for a variety of pictures. The deep neural network architecture VGG-19 is renowned for being straightforward and simple to build. VGG's ability to learn fine-grained features is largely due to its usage of small 3x3 convolutional filters with a stride of 1. To decrease spatial dimensions and expand the network's receptive field, max-pooling layers are employed. The network's completely linked layers function as a classifier.

The ImageNet dataset, which has 1.2 million training images and 1000 categories, was used in the original research to train VGG-19 for image classification. The term "VGG-19" refers to the total weight layers of 19 weight layers (16 convolutional layers and 3 fully linked layers). VGG-19 is regarded as relatively deep for its time and training it from scratch on big datasets can be computationally expensive, even though it did well on a variety of image classification tasks. Pre-trained VGG-19 models are frequently used for transfer learning, in which the weights of the network are adjusted for tasks or datasets. VGG-19 and its derivatives are helpful for practical computer vision applications as well as providing as a foundation for more complex models, which has greatly influenced the development of deep learning architectures.

6.5 InceptionV3

A well-liked deep learning architecture is to address picture categorization challenges, InceptionV3 was developed. Google unveiled it in 2015 as an extension of the original Inception architecture (Szegedy, Vanhoucke, Ioffe, Shlens, Wojna 2015). To better depict complicated patterns in images, InceptionV3 combines a variety of different-sized filters to collect data at various sizes. It was designed for image classification tasks which is an extension of the original Inception architecture. InceptionV3 applies a range of different-sized filters to collect data at various sizes to better describe intricate patterns in images. The principal adjustments made to the Inception V3 model include Factorization into Smaller Convolutions such as Spatial Factorization into Asymmetric Convolutions, Utility of Auxiliary Classifiers and Efficient Grid Size Reduction.

In Mathematical Representation of Inception V3, the input image to the InceptionV3 architecture can be denoted as X, with dimensions H × W × C. Here, H indicates the height, W represents the width, and C represents the number of channels. The architectural design comprises a sequence of modules, wherein each module utilizes filters of different sizes to extract distinctive properties from the input. Basic Convolution Block of The InceptionV3 architecture commences with a foundational convolutional block that employs a sequence of convolutional layers with diverse filter sizes on the input image X. Following each convolutional layer, batch normalization and a non-linear activation function, commonly Rectified Linear Unit (ReLU), are employed. The outputs of these layers are subsequently merged to produce a unified output. (Szegedy, Vanhoucke, Ioffe, Shlens, Wojna 2015.)

The Inception module functions as the fundamental building block of the InceptionV3 architecture. To gather features at different scales, multiple concurrent convolutional operations are employed, utilizing a range of filter sizes. In 1x1 Convolution, operation applies 1x1 filters to reduce the dimensionality of the input. It helps to reduce computational complexity. 3x3 Convolution, this operation applies 3x3 filters to capture more spatial information in the input. 5x5 Convolution, this operation applies 5x5 filters to capture features on a larger receptive field. This operation applies max pooling with a stride of 1 to capture the most important features. The ultimate outcome of the Inception module is produced by combining the outcomes of multiple concurrent approaches. The inception V3 model consists of a total of 42 layers, representing a little increase compared to the preceding inception V1 and V2 models. (FIGURE 10.) (Szegedy, Vanhoucke, Ioffe, Shlens, Wojna 2015.)

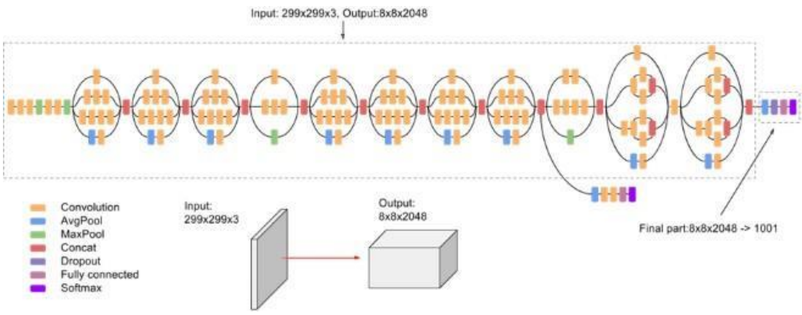


FIGURE 10: Architecture of InceptionV3 (adapted from Szegedy, Vanhoucke, Ioffe, Shlens, Wojna 2015).

6.6 DenseNet-121

In a typical feed-forward convolutional neural network (CNN), the initial convolutional layer is the sole layer that directly receives the output from the preceding convolutional. The convolutional layer generates an output feature map, which is subsequently forwarded to the following convolutional layer. Consequently, each layer is characterized by "L" direct connections, linking it to the subsequent layer. The issue of the "vanishing gradient" becomes increasingly apparent as the number and depth of layers in the convolutional neural network (CNN) increase. This suggests that when the pathway for transmitting information from the input layer to the output layer becomes longer, it can result in the loss or disappearance of certain information. This can impede the network's ability to train effectively. The problem at hand is effectively addressed by DenseNet by the modification of the conventional CNN architecture and the optimization of the inter-layer connectivity. (Huang, Liu, Maaten & Weinberger 2016.)

The nomenclature "Densely Connected Convolutional Network" is derived from the architectural characteristic wherein each layer inside a DenseNet configuration exhibits intimate interconnections with all other layers. The number of direct connections for layer 'L' can be expressed as L(L+1)/2. The connectivity of the feature maps derived from the prior layers are concatenated and employed as inputs in each subsequent layer, as opposed to being averaged. Dense-Nets exhibit a reduction in parameter count compared to regular CNNs due to the elimination of redundant feature mappings, hence facilitating the reuse of features. The feature maps obtained from the preceding layers, denoted as x0, x1, …, xl-1, serve as the input for the l0th layer. The concatenation of these feature maps, represented as [x0, x1, …, xl-1], is used as the input for the current layer. To streamline the implementation process, the several inputs of H1 are consolidated into a unified tensor. (Huang, Liu, Maaten, Weinberger, 2016).

$$X_l = H_l([x_0, x_1, x_2,\ldots\ldots, x_{l-1}])$$

Here, H1 are consolidated into a unified tensor and [x0, x1, …, xl-1] is used as the input for the current layer.

The utilization of the concatenation method becomes impractical when there is variation in the size of feature maps. Down sampling of layers is an essential component of Convolutional Neural Networks (CNNs). It involves reducing the size of feature-maps by dimensionality reduction, resulting in faster calculation rates. To do this, DenseNets are partitioned into DenseBlocks, where the dimensions of the feature maps inside a block remain consistent while the number of filters between them fluctuates. Transition layers refer to the intermediate layers situated between blocks, which effectively reduce the number of channels by half in comparison to the present number of channels being utilized. Growth Rate of the features can be seen as the comprehensive state of the network. (Huang, Liu, Maaten, Weinberger, 2016).

As the input data traverses each successive layer, the dimensions of the feature map expand due to the addition of 'K' features in each layer, which are built upon the existing global state features. The rate of growth of the network, represented by the parameter "K," governs the extent to which information is incorporated into each layer of the network. If each function Hl generates k feature maps, then the lth layer has k feature maps -

$$k_l = k_0 + k \times (l - 1)$$

where, the input feature maps consist of k0 channels k0 is the number of channels in the input layer. In contrast to prevailing network topologies, DenseNets has the capability to incorporate very narrow layers.

The input feature maps consist of k0 channels, where k0 is the number of channels in the input layer. In contrast to prevailing network topologies, DenseNets has the capability to incorporate very narrow layers. (Huang, Liu, Maaten, Weinberger, 2016.)

TABLE 4: DenseNet-121(adapted from Huang, Liu, Maaten, Weinberger, 2016).

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112 × 112	7 × 7 conv, stride 2			
Pooling	56 × 56	3 × 3 max pool, stride 2			
Dense Block (1)	56 × 56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56 × 56	1 × 1 conv			
	28 × 28	2 × 2 average pool, stride 2			
Dense Block (2)	28 × 28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28 × 28	1 × 1 conv			
	14 × 14	2 × 2 average pool, stride 2			
Dense Block (3)	14 × 14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14 × 14	1 × 1 conv			
	7 × 7	2 × 2 average pool, stride 2			
Dense Block (4)	7 × 7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1 × 1	7 × 7 global average pool			
		1000D fully-connected, softmax			

In Bottleneck Layers, despite the fact that each layer only generates k output feature maps, there can be a significant amount of input, especially for subsequent levels. To increase computing efficiency and speed, a 1x1 convolution layer might be added as a bottleneck layer before each 3x3 convolution. The table presented above offers a concise overview of the many architectural approaches employed in the construction of the ImageNet database. The stride refers to the number of pixels that are shifted across the input matrix. When the stride parameter is set to 'n' (with a default value of 1), the filters are shifted by 'n' pixels at each step. Analysing the table with the DenseNet-121 architecture, it becomes apparent that each dense block consists of many layers (repetitions), each comprising two convolutions. These convolutions consist of a bottleneck layer with a kernel size of 1x1 and a convolution layer with a kernel size of 3x3. Additionally, it should be noted that each transition layer in the model architecture consists of a 1x1 convolutional layer and a 2x2 average pooling layer, where the pooling layer has a stride of 2.

Therefore, the layers that are present can be a simple convolution layer with 64 7X7 filters and a 2 stride. A basic pooling layer with a stride of two and 3x3 maxpooling. Dense Block 1, which repeats twice through two convolutions. Transition layer 1 (1 Conv + 1 AvgPool) V. Dense Block 2 repeating twice with two convolutions. Layer 2 Transition (1 Conv + 1 AvgPool). Dense Block 3 with 24 repetitions of 2 convolutions. Layer 3 of transition (1 Conv + 1 AvgPool). Dense Block 4 with 16 repetitions of two convolutions. In Global Average Pooling Layer, the layer uses all the network's feature maps to carry out classification. DenseNet-121 has four average pooling pools and 120 convolutions. Because all levels, including those in the same dense block and transition layers, distribute their weights over numerous inputs, deeper layers can use characteristics that were acquired earlier. Because they generate a large amount of duplicated data, the layers in the second and third dense blocks give the transition layers' output the lowest weights. Furthermore, because trials seemed to place more focus on final feature maps, it's likely that higher-level features are formed later into the model, even if the final layers rely on the weights of the entire dense block. (TABLE 4.)

6.7 SqueezeNet

SqueezeNet is a compact deep learning network optimized for resource-constrained inference. DeepScale and UC Berkeley academics introduced it in 2016. SqueezeNet uses fire modules, which are created to decrease the number of parameters while maintaining representational power, to strike a compromise between model size reduction and accuracy. Certain techniques are used by the SqueezeNet model to reduce the bulk of parameters. Which are replacing the 3x3 filters with 1x1 filters. Using 3x3 filters only on the input channels. Later network down sampling. Switching out the 3x3 filters for 1x1 filters, the model used 1x1 filters instead of the more common 3x3 filters because of financial limitations. (Iandola, Han, Moskewicz, Ashraf, Dally, Keutzer, 2016.).

Hence, compared to a conventional filter, the model contains nine times less parameters. Reducing the number of input channels to 3x3 filters, to lower the filter size to 1x1, fewer input channels must be used. The squeeze layer, which will be covered later, is used for this. It can obtain larger activation maps for the convolution layers by down sampling later in the network. Every convolution layer in a convolutional network generates an activation map output with a spatial resolution of at least 1x1 and frequently much larger than 1x1. Two factors determine the height and width of these activation maps- The size of the input data (256x256 pictures) and The CNN architecture's choice of layers for down sampling.

CNN designs incorporate down sampling by setting the (stride > 1) in some of the convolution or pooling layers. If the layer’s nearest to the input layer makes significant progress, the majority of the network's layers will have small activation maps. Conversely, if most of the layers have a stride of 1, and the strides greater than 1 are grouped around the classifier at the end of the network, then many of the layers will have high activation maps. The primary goal of strategies 1 and 2 is to decrease the number of parameters. The main goal of Strategy 3 is to maximize accuracy while using a limited number of parameters.

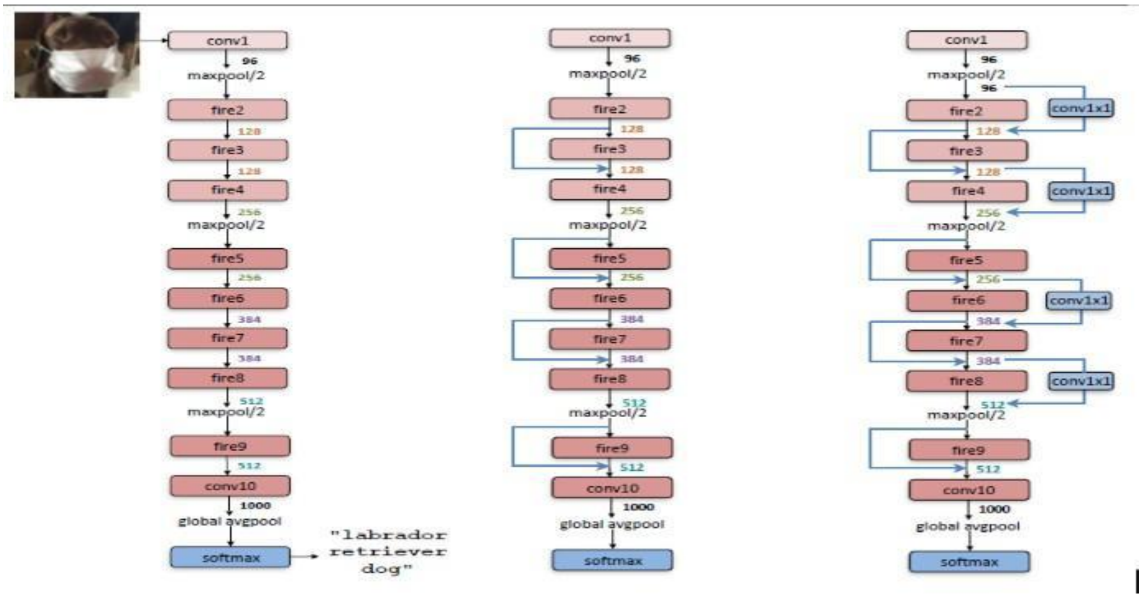


FIGURE 11: Architecture of SqueezeNet (adapted from Iandola, Han, Moskewicz, Ashraf, Dally & Keutzer, 2016).

In the SqueezeNet architecture, an isolated convolution layer (conv1) is followed by eight Fire modules (fire2-9) before a final convolution layer (conv10). Every fire module has a constant number of filters from the beginning to the end of the network. Following layers conv1, fire4, fire8, and conv10, with a stride of 2, SqueezeNet places pooling somewhat late, in accordance with Strategy 3. A summary of the entire SqueezeNet architecture may be found below. The first three screens show SqueezeNet, SqueezeNet with a simple bypass, and SqueezeNet with an advanced bypass. Squeezenet is a CNN architecture that retains accuracy comparable to AlexNet but having 50 times less parameters. (FIGURE 11.)

6.8 MobileNetV2

A convolutional neural network (CNN) architecture called MobileNetV2 was created for mobile and resource- constrained devices. In 2018, Google researchers announced it as an upgrade from the original MobileNet design. MobileNetV2 strives for improved accuracy while keeping efficient processing and model size. MobileNetV2 is a mobile-friendly convolutional neural network (CNN) architecture for a range of computer vision applications, including object identification, image categorization, and more. It replaces the original MobileNet architecture and aims to increase performance while preserving deployment effectiveness for mobile and embedded devices. (Sandler, Howard, Zhu, Zhmoginov & Chen 2018.)

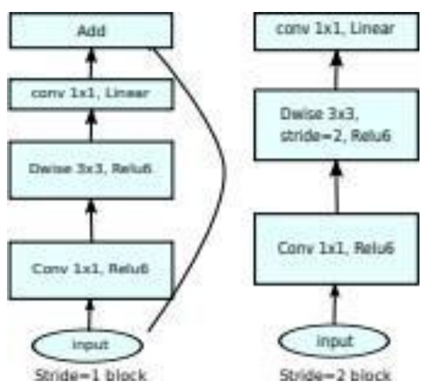


FIGURE 12: Architecture of MobileNetV2 (adapted from Sandler, Howard, Zhu, Zhmoginov & Chen 2018.)

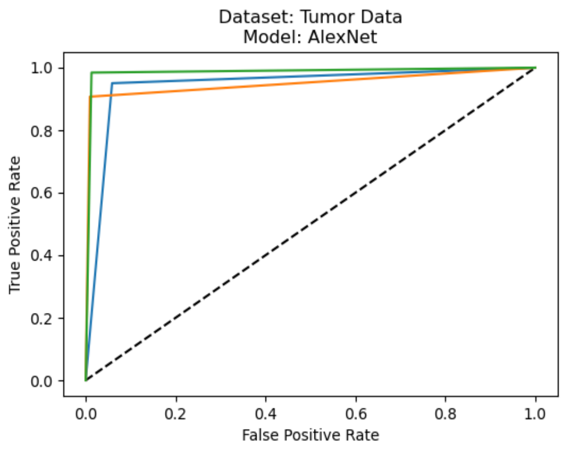
In MobileNetV2 architecture, MobileNetV2 typically takes an input image with a size of 224x224 pixels, which is a common size for many image classification tasks. Initial Convolution is the input image is passed through an initial convolutional layer, which is a standard 3x3 convolution with batch normalization and ReLU activation. The input image's first features are extracted by this layer. The core building blocks of MobileNetV2 are called inverted residual blocks. Each inverted residual block consists of three key components. A 1x1 depthwise convolution (pointwise convolution with 1x1 kernel) to reduce the number of input channels. Batch normalization and ReLU activation. A depth wise separable convolution, which includes a depth wise convolution (3x3) and pointwise convolution (1x1). Batch normalization and ReLU activation are applied after each convolution. Depth wise convolution reduces computational cost by applying a separate 3x3 convolution to each input channel. (FIGURE 12.)

To increase the number of channels back to the original width, perform one more 1x1 pointwise convolution. Batch normalization but no ReLU activation. These blocks are designed to reduce computation while capturing important features. Multiple inverted residual blocks are stacked together, forming sequences of these blocks. The number of blocks and their specific configurations may vary depending on the desired model size and task.In feature Pyramid, MobileNetV2 often includes additional layers for feature pyramid construction, which is useful for object detection tasks. These layers capture features at multiple scales within the network. Global Average Pooling (GAP), The feature maps are globally average-pooled to produce a fixed-size feature vector. This feature vector is used for final classification or regression tasks. In some versions of MobileNetV2, a fully connected layer may be added for classification tasks. Typically followed by a SoftMax activation for classification or a linear activation for regression.

MobileNetV2 is well-known for its ability to balance model size, computational cost, and performance. It has been frequently used in applications for mobile and edge devices where resource limitations are an issue. Users can select the best MobileNetV2 variations for their particular use cases by adjusting the width multipliers and input resolutions to meet their individual accuracy and computational needs.

7 RESULT ANALYSIS

The study investigates eight distinct deep pre-trained models, namely AlexNet, GoogleNet, ResNet, SqueezeNet, VGG19, and three others, for tumour classification based on medical imaging data. Three common designs were used to analyse the performance metrics, which included average accuracy, weighted precision, recall, and F1-score. The performance of eight pre-trained models on the tumour dataset reveals intriguing patterns and considerations for tumour classification. Starting with AlexNet, it consistently demonstrates robust accuracy, maintaining an average of 94.09% across different folds. (TABLE 5). Its effectiveness and stability are supported by the F1-score, weighted precision, and recall. The ROC curve analysis underlines its discriminative power, portraying a favourable performance in tumor classification (FIGURE 13). On the contrary, DenseNet exhibits a comparatively lower average accuracy of approximately 71.8% (TABLE 5). While its weighted precision, recall, and F1-score show moderate values, the ROC curve suggests a balanced trade-off between sensitivity and specificity, positioning DenseNet as a model with potential in certain scenarios (FIGURE 14).



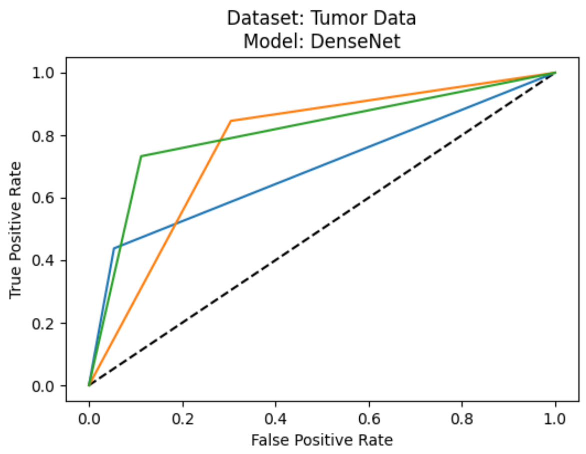


FIGURE 13: ROC curve of AlexNet

FIGURE 14: ROC curve of DenseNet

TABLE 5: Model Results

Model	Accuracy	Weighted Precision	Weighted Recall	Weighted F1-Score
AlexNet	94.09%	94.73%	94.09%	94.21%
DenseNet	71.80%	71.63%	71.80%	70.22%
GoogleNet	96.84%	96.79%	96.84%	96.79%
InceptionV3	64.52%	50.52%	64.52%	55.84%
MobileNet	97.72%	97.77%	97.72%	97.73%
ResNet	73.34%	73.57%	73.34%	72.46%
SqueezeNet	90.18%	90.18%	90.18%	90.11%
VGG-19	94.06%	94.96%	94.06%	94.22%

GoogleNet stands out prominently, boasting an impressive average accuracy of 96.8% and high-weighted precision, recall, and F1-score values (TABLE 5). The ROC curve analysis reinforces its efficacy in tumour classification, making it a strong contender for such tasks (FIGURE 15). InceptionV3, however, presents a moderate average accuracy of about 64.5% (TABLE 5). The weighted precision, recall, and F1-score indicate a moderate performance level, and the ROC curve analysis further highlights potential limitations in specific aspects of tumour classification (FIGURE 16).

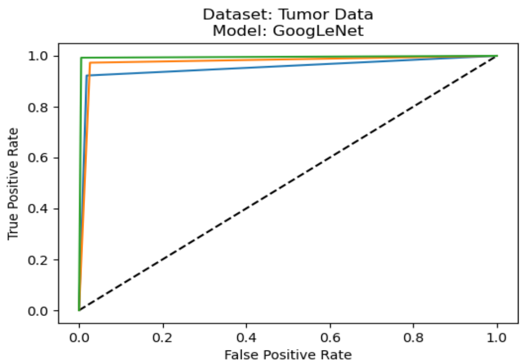


FIGURE 15: ROC curve of GoogLeNet

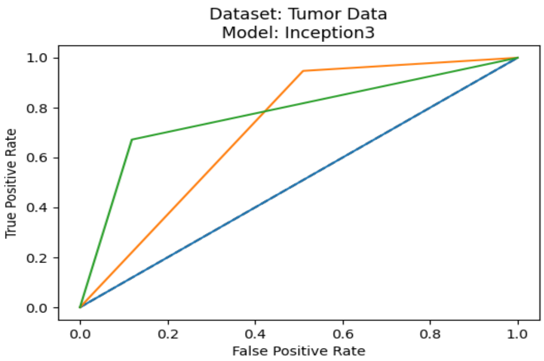


FIGURE 16: ROC curve of Inception3

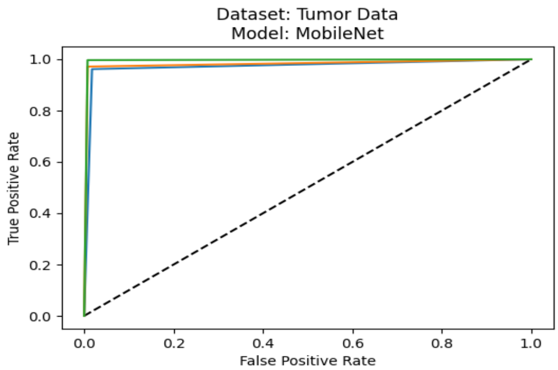


FIGURE 17: ROC curve of MobileNet

MobileNet emerges as a top-performing model, achieving a remarkable average accuracy of 97.7% (TABLE 5). Its high weighted precision, recall, and F1-score values underline its effectiveness in tumor detection. The ROC curve analysis affirms its discriminative capability, solidifying its position as a standout performer (FIGURE 17). MobileNet's superior performance on the tumor dataset can be attributed to its unique architecture, specifically designed for efficiency and speed. MobileNet employs depth wise separable convolutions, which, in comparison to conventional solutions, drastically cut down on the number of parameters and computations. Because of this feature, MobileNet works effectively in contexts with restricted resources, including those involving mobile devices or scarce computational capabilities.

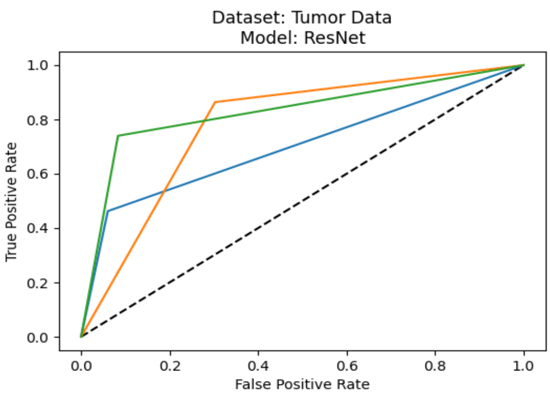


FIGURE 18: ROC curve of ResNet

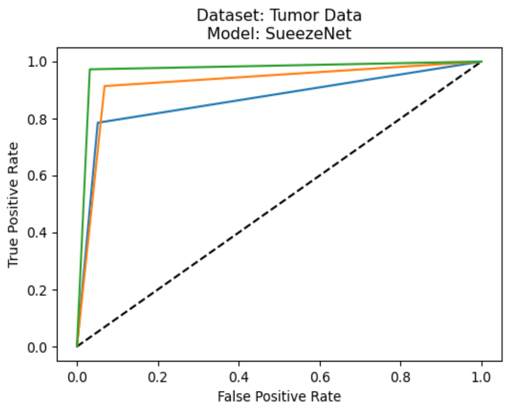


FIGURE 19: ROC curve of SqueezeNet

In our dataset, MobileNet allows a faster inference without compromising accuracy. MobileNet captures our dataset features efficiently and contributes to its high accuracy. Additionally, MobileNet's architecture includes a series of lightweight depth wise separable convolutions that enable it to learn and represent complex features while maintaining computational efficiency. ResNet demonstrates respectable performance, with an average accuracy of 73.3% (TABLE 5). The balanced weighted precision, recall, and F1-score values and a robust ROC curve suggest reliability in tumour classification tasks (FIGURE 18). SqueezeNet produces competitive weighted precision, recall, and F1-score values together with a strong average accuracy of 90.2% (TABLE 5). It is more successful in differentiating between tumour and non-tumour classes when combined with the ROC curve analysis (FIGURE 19).

With its impressive, weighted precision, re-call, and F1-score values, VGG19 stands out with an impressive average accuracy of 94.1% (TABLE 5). The ROC curve study highlights how well-suited and extraordinarily discriminative it is for tumour detection (FIGURE 20).

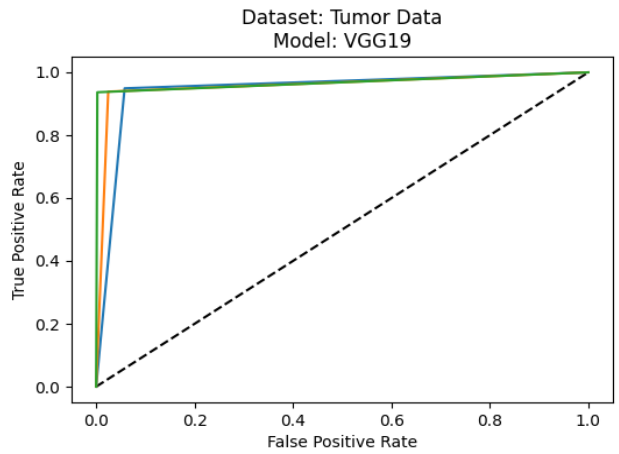


FIGURE 20: ROC curve of VGG19

In summary, MobileNet, GoogleNet, and VGG19 emerge as top-performing models, excelling in accuracy and discriminative ability. These models are well-suited for classification tasks on the used dataset in this task. While DenseNet and InceptionV3 exhibit moderate performance, the ROC curve analysis offers nuanced insights, guiding model selection based on the desired balance between sensitivity and specificity in tumour detection applications.

8 RESULT DISCUSSION AND CONCLUSION

To gain important insights into the performance and applicability of the eight pre-trained models for tumour classification by examining their outcomes on the tumour dataset, AlexNet consistently demonstrated robust accuracy, achieving an average of 94.1% across different folds. Its effectiveness is further supported by high weighted precision, recall, and F1-score values, making it a reliable choice for tumor detection tasks. DenseNet, while exhibiting a lower average accuracy of approximately 71.8%, showcases a balanced trade-off between sensitivity and specificity, suggesting potential use in specific scenarios. GoogleNet stands out with an impressive average accuracy of 96.8%, indicating its strong capability in tumor classification. The model is well-suited for these kinds of tasks because of its excellent precision, recall, and F1-score values. In contrast, InceptionV3 presents a moderate average accuracy of about 64.5%, highlighting potential limitations in specific aspects of tumour classification. MobileNet emerges as a top performer, achieving a remarkable average accuracy of 97.7% and demonstrating high precision, recall, and F1-score values. Its discriminative capability is further affirmed by the ROC curve analysis.

ResNet demonstrates respectable performance, with an average accuracy of 73.3% and balanced weighted precision, recall, and F1-score values. SqueezeNet delivers a solid average accuracy of 90.2%, competitive with other models. VGG19 impresses with high accuracy of 94.1%, superior weighted precision, recall, and F1-score values, and exceptional discriminative ability as indicated by the ROC curve analysis. Challenges such as computational costs, especially notable in VGG19, and considerations like hyperparameter tuning and class imbalances were not explicitly discussed in the provided descriptions. These factors can significantly impact model performance and are essential considerations for real-world applications. According to the study, the requirements of the work should be taken into consideration when selecting a tumour classification model.

While ResNet and SqueezeNet offer computational efficiency, VGG19 excels in detailed feature extraction. Consideration of computational resources, dataset characteristics, and the task's intricacies should guide the selection of an appropriate model. Limitations in model performance may arise from dataset characteristics, such as imbalances in class distribution or variations in tumour types. Future work could involve fine-tuning the models to better adapt to specific tumour subtypes and exploring ensemble methods for improved robustness. Additionally, addressing interpretability challenges and incorporating domain-specific knowledge could enhance the models' clinical utility. Overall, this work offers a thorough comprehension of the advantages and disadvantages of each model, opening the door for additional improvements and developments in the field of tumour classification utilizing deep learning models that have already been trained. All things considered, subsequent study in this area might examine other deep learning model designs and variations, look at the effects of various transfer learning techniques, and extend the analysis to bigger and more varied brain imaging datasets. These advancements will contribute to further enhancing the accuracy and reliability of brain image classification in clinical applications.

REFERENCES

Al Shehri, W. 2022. Alzheimer's disease diagnosis and classification using deep learning techniques. *PeerJ Computer Science* 8 (2022) e1177. Available at: https://www.researchgate.net/publication/366444515_Alzheimer's_disease_diagnosis_and_classification_using_deep_learning_techniques Accessed 1 November 2023.

Anantharajan, S. & Gunasekaran, S. 2021. Automated brain tumor detection and classification using weighted fuzzy clustering algorithm, deep auto encoder with barnacle mating algorithm and random forest classifier techniques. *International Journal of Imaging Systems and Technology* 31 (4) (2021) 1970–1988. Available at: https://www.researchgate.net/publication/351086233_Automated_brain_tumor_detection_and_classification_using_weighted_fuzzy_clustering_algorithm_deep_auto_encoder_with_barnacle_mating_algorithm_and_random_forest_classifier_techniques Accessed 12 October 2023.

Arbane, M., Benlamri, R., Brik, Y. & Djerioui M. 2021. *Transfer learning for automatic brain tumor classification using MRI images.* Available at: https://www.researchgate.net/publication/350197770_Transfer_Learning_for_Automatic_Brain_Tumor_Classification_Using_MRI_Images Accessed: 28 October 2023.

Asif, S., Zhao, M., Tang, F. & Zhu Y. 2023. *An enhanced deep learning method for multi-class brain tumor classification using deep transfer learning, Multimedia Tools and Applications.* Available at: https://www.researchgate.net/publication/368843315_An_enhanced_deep_learning_method_for_multi-class_brain_tumor_classification_using_deep_transfer_learning. Accessed 1 November 2023.

Badža, M. M. & Barjaktarović, M. Č. 2020. *Classification of brain tumors from mri images using a convolutional neural network.* Available at: https://www.researchgate.net/publication/339994574_Classification_of_Brain_Tumors_from_MRI_Images_Using_a_Convolutional_Neural_Network Accessed 28 October 2023.

Chelghoum, R., Ikhlef, A., Hameurlaine, A. & Jacquir, S. 2020. *Transfer learning using convolutional neural network architectures for brain tumor classification from mri images.* Available at: https://www.researchgate.net/publication/341728934_Transfer_Learning_Using_Convolutional_Neural_Network_Architectures_for_Brain_Tumor_Classification_from_MRI_Images Accessed 18 October 18 2023.

Chollet F. 2017. *Xception: Deep Learning with Depthwise Separable Convolutions.* Available at: https://doi.org/10.48550/arXiv.1610.02357 Accessed 16 October 2023.

He, K., Zhang, X., Ren, S. & Sun J. 2015. *Deep Residual Learning for Image Recognition.* Available at: https://doi.org/10.48550/arXiv.1512.03385 Accessed 5 November 2023.

Huang, G., Liu, Z., Maaten L. & Weinberger Q. 2016. *Densely Connected Convolutional Networks*. Available at: https://doi.org/10.48550/arXiv.1608.06993 Accessed 4 November 2023.

Iandola N., Han, S., Moskewicz, W., Ashraf, K., Dally J. & Keutzer K. 2016. *SqueezeNet: AlexNet level accuracy with 50x fewer parameters and <0.5MB model size.* Available at: https://doi.org/10.48550/arXiv.1602.07360 Accessed 12 November 2023**.**

Kaur, T. & Gandhi, T. K. 2020. *Deep convolutional neural networks with transfer learning for automated brain image classification, Machine vision and applications.* Available at: https://www.researchgate.net/publication/340225334_Deep_convolutional_neural_networks_with_transfer_learning_for_automated_brain_image_classification Accessed 9 October 2023.

Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E. & Gansland, T. 2022. *Transfer learning for medical image classification: a literature review.* Available at: https://www.researchgate.net/publication/359935888_Transfer_learning_for_medical_image_classification_a_literature_review Accessed 26 October 2023.

Krizhevsky, A., Sutskever I. & Hinton, E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems.* 25, 1097-1105. Available at:

https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html Accessed 18 October 2023.

Karen Simonyan & K., Zisserman A. 2015. *Very Deep Convolutional Networks for Large-Scale Image Recognition.* Available at:
https://arxiv.org/abs/1409.1556 Accessed 16 October 2023.

Liu, W., Han, H. & Han, G. 2022. *Transfer learning with deep convolutional neural network for automated plant identification.* Available at: https://ieeexplore.ieee.org/document/9886149. Accessed 2 November 2023.

Mukherkjee, D., Saha, P., Kaplun, D. *et al.* 2022. *Brain tumor image generation using an aggregation of GAN models with style transfer*. *Sci Rep* 12, 9141 (2022).
Available at: https://doi.org/10.1038/s41598-022-12646-y Accessed 1 November 2023.

Pan, S. & Yang Q. 2010. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering.* 22,1345-1359. Available at: https://www.semanticscholar.org/paper/A-Survey-on-Transfer-Learning-Pan-Yang/a25fbcbbae1e8f79c4360d26aa11a3abf1a11972 Accessed 16 October 2023.

Saha, D. & Manickavasagan, A. 2022. *Chickpea varietal classification using deep convolutional neural networks with transfer learning.* Available at:
https://www.researchgate.net/publication/358247234_Chickpea_varietal_classification_using_deep_convolutional_neural_networks_with_transfer_learning. Accessed 2 November 2023.

Shalbaf, A., Bagherzadeh, S. & Maghsoudi, A. 2020. *Transfer learning with deep convolutional neural network for automated detection of schizophrenia from eeg signals, Physical and Engineering Sciences in Medicine.* Available at: https://link.springer.com/article/10.1007/s13246-020-00925-9 Accessed 18 October 2023.

Simon, B. C, Baskar, D. & Jayanthi V. 2019. *Alzheimer's disease classification using deep convolutional neural network.* Available at: https://www.researchgate.net/publication/339174590_Alzheimer's_Disease_Classification_Using_Deep_Convolutional_Neural_Network Accessed 24 October 2023.

Srinivas, C., KS, N. P., Zakariah, M., Alothaibi Y. A., Shaukat, K., Partibane B. & Awal, H. et al. 2022. *Deep transfer learning approaches in performance analysis of brain tumor classification using mri images.* Available at: https://www.hindawi.com/journals/jhe/2022/3264367/ Accessed 12 October 2023.

Szegedy, C., Liu, W., Jia, Y., Sermanet P., Reed, S., Anguelov D., Erhan, D., Vanhoucke, V. & Rabinovich, A. 2015. Going Deeper with Convolutions. *CoRR Journal.* Available at: https://doi.org/10.48550/arXiv.1409.4842 Accessed 18 October 2023.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen L. 2019. *MobileNetV2: Inverted Residuals and Linear Bottlenecks.* Available at: https://doi.org/10.48550/arXiv.1801.04381 Accessed 8 November 2023.

Szegedy, C., Vanhoucke V., Ioffe, S., Shlens, J. & Wojna, Z 2015. *Rethinking the Inception Architecture for Computer Vision.* Available at: https://doi.org/10.48550/arXiv.1512.00567 Accessed 4 October 2023.

Song, S. E., Seo, B. K., Cho, K. R., Woo, O. H., Son, G. S., Kim, C., Cho, S. B. & Kwon, S.-S. 2015. *Computer-aided detection (cad) system for breast mri in assessment of local tumor extent, nodal status, and multifocality of invasive breast cancers: preliminary study, Cancer Imaging.* Available at: https://www.researchgate.net/publication/272377201_Computer-aided_detection_CAD_system_for_breast_MRI_in_assessment_of_local_tumor_extent_nodal_status_and_multifocality_of_invasive_breast_cancers_Preliminary_study. Accessed 6 October 2023.

Wiatowski, T. & Bölcskei H. 2017. *A Mathematical Theory of Deep Convolutional Neural Networks for Feature Extraction*. Available at: https://doi.org/10.48550/arXiv.1512.06293 Accessed 1 October 2023.

Yadav, S. S. & Jadhav, S. M. 2019. *Deep convolutional neural network based medical image classification for disease diagnosis.* Available at: https://www.researchgate.net/publication/338001332_Deep_convolutional_neural_network_based_medical_image_classification_for_disease_diagnosis Accessed 27 October, 2023

Zhuang, F., Qi Z., Duan, K., Xi D., Zhu, Y., Zhu, H., Xiong, H., He, Q. (2020), *Comprehensive Survey on Transfer Learning.* Available at:

https://www.researchwithrutgers.com/en/publications/a-comprehensive-survey-on-transfer-learning Accessed 24 October 2023.

Zoph, B., Vasudevan, V., Shlens J. & Le V. 2018. *Learning Transferable Architectures for Scalable Image Recognition.* Available at: https://ieeexplore.ieee.org/document/8579005 Accessed 8 November 2023.