Studies in Big Data 152

M. F. Mridha
Nilanjan Dey *Editors*

Data-Driven Clinical Decision-Making Using Deep Learning in Imaging

Springer

Studies in Big Data

Volume 152

Series Editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland

The series "Studies in Big Data" (SBD) publishes new developments and advances in the various areas of Big Data- quickly and with a high quality. The intent is to cover the theory, research, development, and applications of Big Data, as embedded in the fields of engineering, computer science, physics, economics and life sciences. The books of the series refer to the analysis and understanding of large, complex, and/or distributed data sets generated from recent digital sources coming from sensors or other physical instruments as well as simulations, crowd sourcing, social networks or other internet transactions, such as emails or video click streams and other. The series contains monographs, lecture notes and edited volumes in Big Data spanning the areas of computational intelligence including neural networks, evolutionary computation, soft computing, fuzzy systems, as well as artificial intelligence, data mining, modern statistics and Operations research, as well as self-organizing systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

The books of this series are reviewed in a single blind peer review process.

Indexed by SCOPUS, EI Compendex, SCIMAGO and zbMATH.

All books published in the series are submitted for consideration in Web of Science.

M. F. Mridha · Nilanjan Dey
Editors

Data-Driven Clinical Decision-Making Using Deep Learning in Imaging

Springer

Editors
M. F. Mridha
Department of Computer Science
American International
University-Bangladesh (AIUB)
Dhaka, Bangladesh

Nilanjan Dey
Department of Computer Science
and Engineering
Techno International New Town
Kolkata, West Bengal, India

ISSN 2197-6503 ISSN 2197-6511 (electronic)
Studies in Big Data
ISBN 978-981-97-3965-3 ISBN 978-981-97-3966-0 (eBook)
https://doi.org/10.1007/978-981-97-3966-0

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Data-Driven Clinical Decision-making of healthcare is the intersection of data science and clinical decision-making has emerged as a transformative force, paving the way for unprecedented advancements in medical imaging. The book you are about to embark upon, *Data-Driven Clinical Decision-Making Using Deep Learning in Imaging*, is a comprehensive exploration of the symbiotic relationship between deep learning algorithms and the field of medical imaging.

As technology continues to reshape the healthcare paradigm, the integration of data-driven approaches has become imperative for improving diagnostic accuracy, treatment planning, quick decision, health monitoring, and overall patient intensive care. At the forefront of this revolution is the marriage of sophisticated deep learning techniques with the wealth of information embedded in medical imaging datasets. This book delves into the intricate tapestry of this union, providing clinicians, researchers, and data scientists with a profound understanding of the principles, challenges, and potential of data-driven clinical decision-making.

The book contains various topics, methodologies, and applications, providing readers with a comprehensive understanding of the field's current state and prospects. It begins with exploring domain adaptation in medical imaging and evaluating the effectiveness of transfer learning to overcome challenges associated with limited labeled data. The subsequent chapters delve into specific applications, such as improving kidney lesion classification in CT scans, elevating breast cancer research through attention-based U-Net architecture for segmentation and classifying brain tumor.

The journey of this book begins by laying a strong foundation in the fundamentals of medical imaging and deep learning, ensuring that readers from diverse backgrounds can navigate the complexities of the subject matter. We then embark on a path through the applications of deep learning in various imaging modalities, from radiology and pathology. Real-world case studies and examples illustrate the transformative impact of data-driven approaches on clinical workflows, offering insights into how these methodologies can enhance diagnostic precision and therapeutic efficacy.

Throughout the book, the emphasis is on fostering a holistic understanding of the integration of deep learning into clinical decision-making processes. Ethical considerations, regulatory frameworks, and potential pitfalls are explored, guiding practitioners in navigating the evolving landscape of healthcare technology responsibly and ethically.

As we stand on the precipice of a new era in health technology, where data-driven insights have the potential to revolutionize patient outcomes, *Data-Driven Clinical Decision-Making Using Deep Learning in Imaging* serves as a beacon for those seeking to harness the power of cutting-edge technologies for the betterment of healthcare. This book highlights the basic concept of deep learning-based medical imaging and its application in numerous scientific domains of healthcare. The book includes thirteen chapters in which *Eva* et al. in the first chapter "Domain Adaptation in Medical Imaging: Evaluating the Effectiveness of Transfer Learning" focus on the effectiveness of transfer learning, specifically within the domain adaptation framework for medical imaging, addressing the challenges posed by varying data distributions across different medical domains. *Shovon* et al. highlighted transfer learning (TL) approach fused with a squeeze-and-excitation (SE) attention mechanism to accurately diagnose brain tumors on a brain tumor MRI dataset incorporating one-hot encoding in the image preprocessing step in the second chapter "Advancing Brain Tumour Detection: Transfer Learning-Based Approach Fused with Squeeze-and-Excitation (SE) Attention Mechanism in Computer Vision". In the third chapter "A Precise Cervical Cancer Classification in the Early Stage Using Transfer Learning-Based Ensemble Method: A Deep Learning Approach", *Alam* et al. integrated the Adam optimizer into deep learning model to mitigate issues of both overfitting and underfitting to detect cervical cancer. In the fourth chapter "Unveiling Diagnostic Precision: Evaluating Machine Learning and Deep Learning Approaches for Pneumonia Recognition of COVID-19 Patients Using Chest X-Rays", *Rahman* et al. analyze the efficacy of the ML strategy, seven distinct machine learning algorithms, including K-Nearest Neighbor (KNN), Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), Naive Bayes (NB), Random Forest, and AdaBoost, are employed, followed by the best classifier being chosen using GridSearchCV. In the fifth chapter "Advanced Hybrid Deep Learning Model for Precise Multiclass Classification of Bone Marrow Cancer Cells", *Sakib* et al. discussed an automated classification method for plasma cell cancer which is pre-processed and trained with the parameterized hybrid convolutional neural network.

In the sixth chapter "Privacy-Preserving Vision-Based Detection of Pox Diseases Using Federated Learning", *Kibriya* et al. highlighted counterfactual explanations and discussed federated learning for enhanced data analytics to introduce a privacy-preserving framework for pox disease detection. A performance analysis of utilizing transfer learning to develop an automated system capable of accurately detecting the portion of the image categorized the sample as any disease and XAI was introduced by *Chowdhury* et al. in the seventh chapter "Unveiling the Unique Dermatological Signatures of Human Pox Diseases Through Deep Transfer Learning Model Based on DenseNet and Validation with Explainable AI". In the eighth chapter "Improved Classification of Kidney Lesions in CT Scans Using CNN with Attention Layers: Achieving High Accuracy and Performance", *Afroj* et al. discussed an innovative deep learning method for precisely categorizing CT kidney images which automatically categorize the kidney disorders. In the ninth chapter "Enhancing Breast Cancer Detection Systems: Augmenting Mammogram Images Using Generative Adversarial Networks", *Rifat* et al. highlighted counterfactual explanations and discussed the efficacy of pre-existing methodologies both pre-augmentation and post-augmentation, seeking to ascertain whether an improvement in accuracy can be achieved. With the help of computer vision techniques, a robust model was proposed on the 2019 IQ-OTH/NCCD LC dataset by incorporating a multichannel convolutional neural network integrated with the residual connection by *Shovon* et al. in the tenth chapter "Incorporating Residual Connections into a Multi-channel CNN for Lung Cancer Detection in Digital Pathology".

In the eleventh chapter "Advancing Breast Cancer Diagnosis: Attention-Enhanced U-Net for Breast Cancer Segmentation", *Hasan* et al. introduce an innovative model that contributes to advancing breast cancer diagnosis and showcases promise in broader medical imaging applications, fostering a more nuanced and specialized approach within deep learning paradigms. To improve patient quality of life, reduce mortality, and enhance the privacy issues in mammography data enables personalized care, a federated learning is used for precision medicine in the twelfth chapter "Privacy Preserving Breast Cancer Prediction with Mammography Images Using Federated Learning". In the final chapter "Improving Healthcare Efficiency via Sensor-Based Remote Monitoring of Patient Health Utilizing an Enhanced AdaBoost Algorithm", *Ghosh* et al. designed a system that automatically detects and monitors the health status of patients from a remote location.

The editor expresses gratitude to the exceptional authors and referees for their valuable contributions to the book. Their hard work and cooperation have resulted in outstanding publication. Additionally, thanks are extended to the members of the Springer team for their support.

Dhaka, Bangladesh M. F. Mridha
Kolkata, India Nilanjan Dey

Contents

About the Editors

M. F. Mridha (Senior Member IEEE, Professional Member ACM) is currently working as an Associate Professor in the Department of Computer Science, American International University-Bangladesh (AIUB). He is the founder and director of Advanced Machine Intelligence Research Lab (AMIR Lab). He also worked as Associate Professor and Chairman in the Department of Computer Science and Engineering, Bangladesh University of Business and Technology (BUBT) from 2019 to 2022 and as a CSE department faculty member at the University of Asia Pacific and as a graduate head from 2012 to 2019. He received his Ph.D. in NLP in the domain of AI from Jahangirnagar University in the year 2017. He has authored/edited several books with Springer and published more than 200 papers. His research work contributed to the reputed Journal of Scientific Reports Nature, Knowledge-Based Systems, Artificial Intelligence Review, Engineering Applications of Artificial Intelligence, IEEE Access, Sensors, Cancers, Biology and Applied Sciences etc. His research interests include Artificial Intelligence (AI), Machine Learning, Deep Learning, and Natural Language Processing (NLP) etc. He has served as a program committee member in several international conferences/workshops. He served as an Editorial Board Member of several journals including PLOS ONE Journal. He has served as a reviewer of reputed journals like IEEE Transactions on Neural Networks, Artificial Intelligence Review, IEEE Access, Knowledge base System, Expert System, Bioinformatics, Springer Nature, etc.

Nilanjan Dey is Associate Professor at the Department of Computer Science and Engineering, Techno International New Town, Kolkata, India. He is Visiting Fellow at the University of Reading, UK. He also holds a position of Adjunct Professor at Ton Duc Thang University, Ho Chi Minh City, Vietnam. Previously, he held an honorary position of Visiting Scientist at Global Biomedical Technologies Inc., CA, USA (2012–2015). He was awarded his Ph.D. from Jadavpur University in 2015. He is Editor-in-Chief of the International Journal of Ambient Computing and Intelligence, IGI Global, USA. He is Series Co-editor of Springer Tracts in Nature-Inspired Computing (Springer Nature), Data-Intensive Research (Springer Nature), and Advances in Ubiquitous Sensing Applications for Healthcare (Elsevier). He is

Associate Editor of IET Image Processing and Editorial Board Member of Complex and Intelligent Systems, Springer Nature, Applied Soft Computing, Elsevier, etc. He is having 35 authored books and over 300 publications in the area of medical imaging, machine learning, computer-aided diagnosis, data mining, etc. He is Fellow of IETE and Senior Member of IEEE.

Domain Adaptation in Medical Imaging: Evaluating the Effectiveness of Transfer Learning

Arifa Akter Eva, Jamin Rahman Jim, Ashifur Rahman, Hanif Bhuiyan, and Md. Mohsin Kabir

Abstract Deep learning (DL) shows great promise in medical imaging, yet its widespread application across various medical fields encounters obstacles due to distinct data distribution variations in each domain. This research delves into the effectiveness of transfer learning, specifically within the domain adaptation framework for medical imaging, addressing the challenges posed by varying data distributions across different medical domains. This paper used two modified models, MobileNet and EfficientNet, to classify medical datasets. We studied transfer learning with metadata using two medical datasets: the MRI samples dataset and the chest X-ray samples dataset. We compared the achievement of our approach to the most advanced method. In the two models, EfficientNet B2 and MobileNet V2, total categorization accuracy was for brain tumors, 97.48 and 95.09%, and lung diseases, 97.77 and 96.67%. We created a model that could be trained on devices with low computational power, making it ideal for deployment in smaller IoT devices.

Keywords Transfer learning · Medical imaging · Pre-trained models · Domain adaptation · EfficientNet B2 · MobileNet V2

A. A. Eva · J. R. Jim
Advanced Machine Intelligence Research Lab (AMIR Lab), Dhaka, Bangladesh

A. Rahman
Department of CSE, Bangladesh University of Business and Technology, Dhaka, Bangladesh

H. Bhuiyan
Monash Data Futures Institute, Monash University, Melbourne, Australia
e-mail: Hanif.Bhuiyan@monash.edu

Md. M. Kabir (✉)
Superior Polytechnic School, Universitat de Girona, Montilivi, Girona, Spain
e-mail: mdmkabi@gmail.com

M. F. Mridha and N. Dey (eds.), *Data-Driven Clinical Decision-Making Using Deep Learning in Imaging*, Studies in Big Data 152,
https://doi.org/10.1007/978-981-97-3966-0_1

1 Introduction

Medical imaging is a cornerstone in modern health care, providing essential tools for accurate diagnosis, ongoing illness surveillance, and treatment planning, [1]. With the advent of DL, the landscape of medical image [2, 3] analysis has undergone a deep transformation [4], yielding remarkable breakthroughs in tasks ranging from pinpoint image classification to intricate object identification and precise segmentation [5]. However, the efficacy of these cutting-edge DL models rests heavily on the availability of expansive, high-quality tagged datasets for comprehensive training. In the realm of medical imaging, acquiring such datasets becomes an intricate challenge due to multifaceted factors such as stringent privacy concerns, the qualified scope of patient cohorts, and the inherent variability in tomography protocols across diverse healthcare institutions.

The ones who preponderating obstacles encountered in medical imaging is the phenomenon of domain shift [6]. This challenge is rooted in the inherent disparities between datasets collected from disparate sources. These variations encompass disparities in tomography hardware, patient demographics, and even the nuances of tomography procedures, collectively culminating in the discord of data distributions connecting the source and destination domains. Such shifts across domains have the potential to exert profound repercussions on the public presentation of trench learning models, particularly when these models are deployed on novel, antecedently unseen data. This phenomenon curtails the generalization capabilities of models, hampering their ability to seamlessly adapt to variable data contexts.

In response to these intricate challenges, transfer learning has emerged as an empowering paradigm that holds immense promise in the domain of medical imaging [7, 8]. Transfer learning uses knowledge gained in one area to enhance the efficiency of models in another. Of particular interest are pre-trained models, often birthed from expansive datasets of natural images, which embody a repository of generic wine features and representations [9]. The temptation lies in the potential to adapt these pre-trained models to the distinctive realm of medical imaging tasks. However, a significant wonder emerges: Can the inherent disparities between natural images and intricate medical images be effectively bridged through the utilization of these pre-trained models?

Within this context, the point direct of this chapter is the precise exploration of the transfer learning landscape painting within medical imaging. Specifically, the focus resides in the nuanced valuation of the adaptability of pre-trained models—originally tailored for the domain of cancel images—when transposed to the complex canvas of medical imaging. The overall contributions are:

- An exhaustive deliberation on two pre-trained models, providing a comprehensive exposition of their operational methodologies, accompanied by an elucidation of their inherent advantages and disadvantages.
- Evaluation of the performance of these two models is conducted across two popular medical datasets, specifically the MRI sample dataset and the Chest X-ray sample

dataset. This rigorous analysis of their relative efficacy, and accuracy when applied to pattern recognition tasks, is substantiated by the utilization of pivotal graphs.

In our chapter, we embark on a journey through various sections. Firstly, we delve into the backdrop of our work in Sect. 2, providing the necessary context. Moving on to Sect. 3, we meticulously unveil the intricate methodology behind our study, encompassing data preprocessing, model intricacies, and the systematic workflow guiding our experiments. Following this, Sect. 4 meticulously examines the datasets at hand, while Sect. 5 rigorously details the experimental phase of our study. Finally, in Sect. 6, we converge all our findings and insights to craft a conclusive discussion, paving the way for our comprehensive conclusion.

2 Background

In health care, medical imaging stands as a vital tool, offering profound insights into the human body and aiding in the diagnosis and treatment of diverse conditions. Yet, these imaging datasets often pose challenges due to their limited size and quality, hampering the effective training of deep learning models [10]. Enter transfer learning—an emerging technique that harnesses knowledge from source domains and adapts it to target domains [11]. This approach addresses data scarcity, enabling more effective utilization of resources and time, thereby making substantial strides in medical image analysis [12].

Transfer learning manifests in various forms, notably through fine-tuning and feature extraction. Fine-tuning involves retraining a pre-trained model on new data, while feature extraction utilizes the pre-trained model to derive features from input data, subsequently training a new model [13] based on these features. However, pre-trained models, initially designed for natural images, raise concerns about their efficacy when applied to medical imaging [10]. Despite this challenge, transfer learning has gained traction across different medical imaging sectors, spanning image classification, segmentation, detection, and domain adaptation [14, 15]. Innovative medical imaging algorithms with less labeled data have emerged, outperforming standard transfer learning methods in recognizing medical image challenges [16]. Recent studies delve deeper into understanding transfer learning's efficacy in medical images, emphasizing the significance of feature reuse. Contrary to prior assumptions, these studies challenge the notion that transfer learning's effectiveness solely hinges on reusing general features in the early layers of a model. This evolving research landscape seeks to unravel the nuanced factors driving transfer learning's success within the realm of medical imaging [17, 18].

Lately, domain adaptation has gained substantial traction within machine learning-based medical image analysis [19], captivating the interest of researchers as a significant research avenue. This technique, a facet of transfer learning, specifically aims to adapt a model applied on a source domain to better suit a destination domain [20], effectively mitigating the domain shift effect. In the medical imaging domain,

[21], domain adaptation proves instrumental in surmounting challenges arising from limited labeled data and domain discrepancies when deploying deep learning models across diverse healthcare tasks [22]. Recent advancements propose domain adaptation methodologies that transfer knowledge gleaned from an initial domain to effectively execute tasks within a destination domain. These approaches strive to enhance model performance across varied datasets and bolster the resilience of machine learning models against domain shifts [23]. Transductive transfer learning techniques suited for domain adaptation in brain MRI segmentation have been introduced in studies in this domain [19], effectively reducing the impact of domain shifts. Additionally, there are benchmarks proposed specifically for T1-weighted brain MRI segmentation, aimed at evaluating domain adaptation techniques [24]. Overall, domain adaptation exhibits promise in mitigating domain shift challenges in medical imaging. Researchers continue to explore innovative methodologies, striving to further enhance their effectiveness and applicability in this crucial domain [7].

3 Methodology

3.1 Transfer Learning

Transfer learning is a type of machine learning (ML) method in which a trained model is converted for another related job. The first model has already learned to extract important properties from images, such as edges, forms, and colors, which is why transfer learning works. These properties can be employed in the second model, which can then focus on distinguishing between flowers and other things [25]. To learn data representation, neural networks often, a huge amount of training data is required. When working with limited training data, there are numerous techniques to aid neural network models in exploring data representation. Among these approaches, transfer learning stands out as a powerful tool for improving the performance of machine learning models, particularly in situations where the new task has minimal data for training [26]. It can also be used to accelerate training. It transfers data from one to another and this method is called the self-learning method or transfer learning [27].

Transfer learning uses pre-trained models as a foundation for a new task or area. Because of their ability to adapt to diverse tasks, many pre-trained models are especially intended to assist transfer learning. An architecture that has been trained is an ML model that has been developed on a large dataset for a specific task before being made accessible for use or further fine-tuning to developers, academics, or practitioners. Some prominent pre-trained models for transfer learning are widely used for medical imaging. They are ResNet [28], MobileNet [29], VGG [30], inception [31], and generative pre-trained transform [32]. Some key objects about pre-trained models are:

- Large Dataset Training: pre-trained models are frequently trained on large and diverse datasets. Models in computer vision, for example, may be trained on millions of labeled images (e.g., the ImageNet dataset), whereas models in natural language processing may be trained on massive volumes of text data.

- Reusability and flexibility: These models, once trained, can be utilized for other related tasks. Their learned representations can be used as an initial point for new tasks, domains, or datasets (by transfer learning), allowing for faster convergence and potentially greater performance on these tasks.
- Time and resource efficiency: Training models from scratch may be computationally and time-consuming, particularly for complicated models such as deep neural networks. Pre-trained models act as the foundation for new tasks, significantly reducing training time and computational resources required by using previous expertise.
- Improved Performance: Because they are trained on large datasets, pre-trained models frequently perform well in a variety of domains. They may generalize well and perform competitively when applied to related tasks, even with minimum fine-tuning.
- Accessibility: Most pre-trained models are publicly available and freely accessible via libraries such as TensorFlow Hub, Hugging Face Transformers, or PyTorch Hub. This openness encourages their reuse across projects, allowing for speedier machine learning development and testing.

3.2 MobileNet

MobileNet represents a CNN uniquely designed to cater to the demands of mobile and embedded vision applications. MobileNet is a specialized CNN engineered to operate efficiently on devices constrained by limited CPU power and memory capacities. Within MobileNet's architecture, convolution is segmented into depthwise and pointwise convolutions. These components of MobileNet leverage batch normalization (BN) and rectified-linear units (ReLU) to optimize the process of both depthwise and pointwise convolutions [33].

Indeed, the main contrast between MobileNet and traditional CNN architectures lies in their focus: CNNs typically emphasize accuracy through the utilization of more parameters and layers. In contrast, MobileNet is built for efficiency with depthwise separable convolutions, making it suitable for mobile devices are shown in Figs. 1 and 2 [34].

Figure 2 illustrates the architecture on the left side, demonstrating a standard convolutional layer integrated with batch normalization and ReLU. Additionally, it showcases the concept of Depthwise Separable convolutions, incorporating Depthwise and Pointwise layers along with batch normalization and ReLU. The MobileNet model encompasses two versions: MobileNet V1 [35] and MobileNet V2 [36]. Particularly, MobileNet V2 [37] is optimized for simpler mobile and on-board vision applications. It's noteworthy that deep learning techniques have expanded their influence beyond computer vision, permeating domains such as robotics, the Internet of

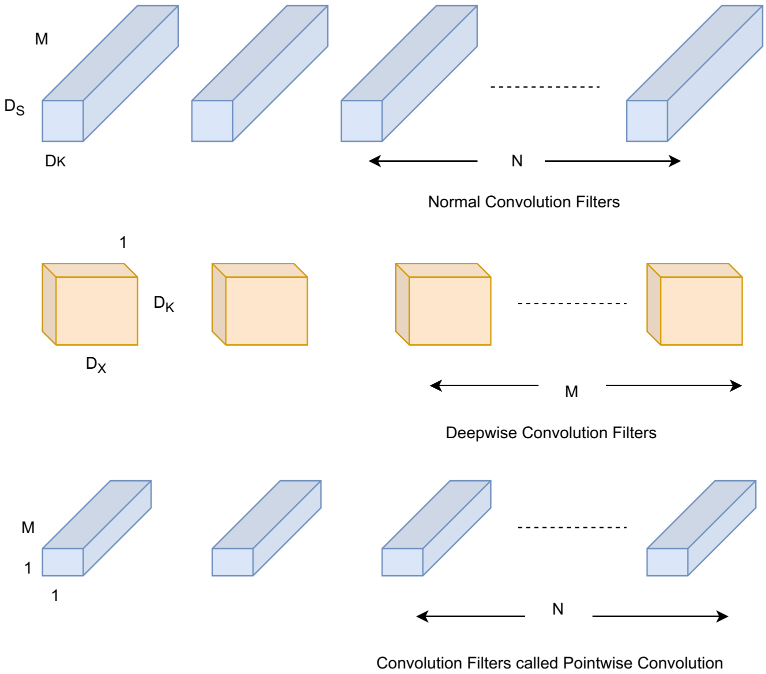


Fig. 1 Convolution standard

Fig. 2 Convolution layer

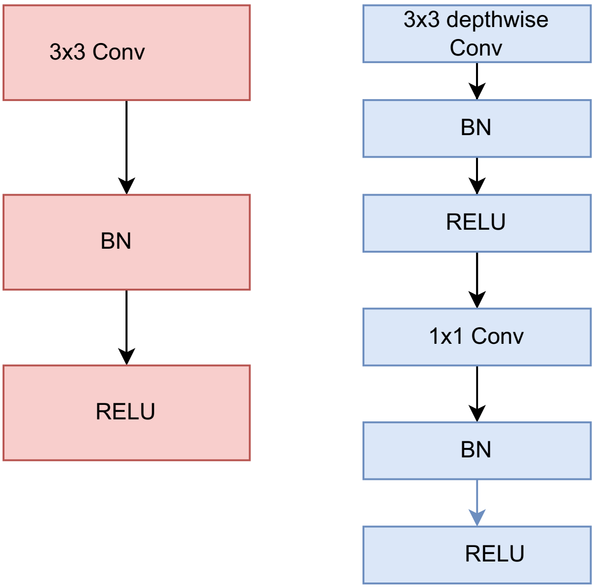


Table 1 MobileNet V2 architecture

Input	Operator	t	e	n	s
$224^2 \times 3$	Conv2d	–	32	1	2
$112^2 \times 32$	Bottleneck	6	16	1	1
$112^2 \times 16$	Bottleneck	6	24	2	2
$56^2 \times 24$	Bottleneck	6	32	3	2
$28^2 \times 32$	Bottleneck	6	64	4	2
$14^2 \times 64$	Bottleneck	6	96	3	1
$14^2 \times 96$	Bottleneck	6	160	3	2
$7^2 \times 160$	Bottleneck	6	320	1	1
$7^2 \times 320$	Conv2d 1×1	–	1280	1	1
$7^2 \times 1280$	Avg pool 7×7	–	–	1	–
$1 \times 1 \times 1280$	Conv2d 1×1	–	k	–	

Things (IoT), natural language processing (NLP), and medical image processing applications.

Indeed, in MobileNet V2, the pointwise convolution operates inversely by decreasing the number of channels. This specific layer is commonly termed the projection layer because its primary function is to convert high-dimensional data into a tensor while concurrently reducing its dimensionality. The bottleneck layer is a tiny convolution layer that decreases how many feature map channels there are [38]. The expansion layer then restores the amount of available channels to their real value. The Table 1 illustrates the architectural layout of MobileNet V2.

Within MobileNet, a bottleneck exists between the model's inputs and outputs. This bottleneck is situated within an inner layer that encompasses the model's capacity to change input data from lower-level concepts into more abstract, higher-level descriptors (Fig. 4). Additionally, similar to residual connections in conventional CNNs, these bypasses established between bottlenecks contribute to accelerated training and enhanced accuracy.

MobileNet V2 introduces the extension layer as a pioneering element. This layer, termed the expansion layer, leverages one-to-one convolutions to amplify the channel count in image data before delving into depthwise convolution [39]. Unlike the projection layer, this expansion layer consistently generates a greater number of output channels than input channels. Another innovative addition is the residual connection, showcased in Fig. 4, mirroring ResNet's [38] functionality and facilitating seamless gradient flow throughout the network. A parameter termed the expansion factor governs the manipulation of feature channels. During testing, MobileNet V2 was employed using 0.5 and a channel-1 multiplier, accommodating an input size of 224×224.

MobileNet V2 places a strong emphasis on optimizing latency while accommodating tiny networks to efficiently handle inputs of varying sizes. This method

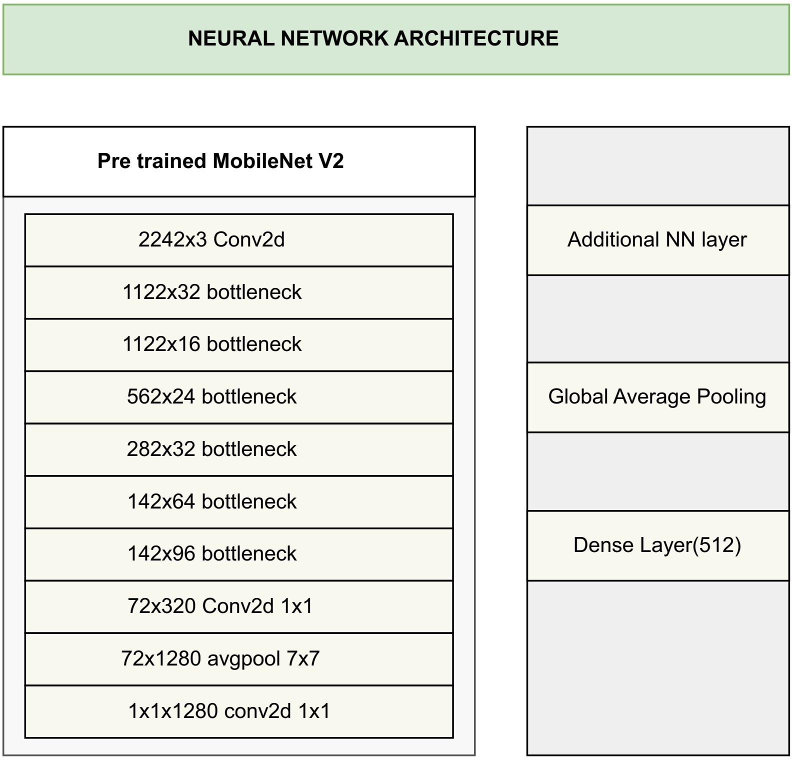
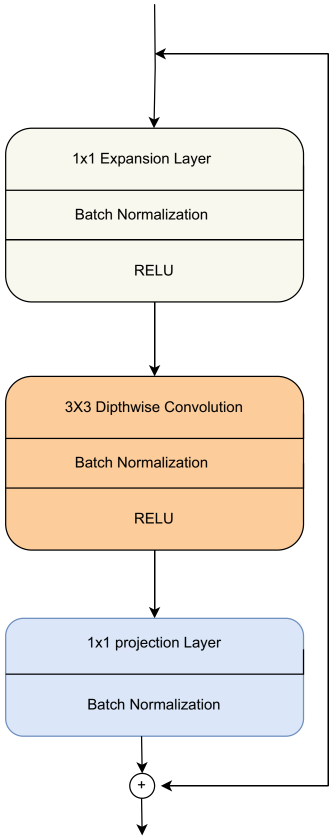


Fig. 3 The CNN pre-trained architecture is made up of an ImageNet pre-trained MobileNet V2 with extra avg pooling and one dense layer

delivers enhanced performance by employing ReLU6 as the activation function in every layer, coupled with batch normalization. Notably, performance assessments on the ImageNet [40] categorization task indicate that MobileNet V2 exceeds both MobileNet and ShuffleNet [41] (width multipliers 1.5) with similar design sizes and computational costs. Furthermore, MobileNet V2 exhibits faster inference times compared to ShuffleNet and NASNet [41], especially notable with a width multiplier of 1.5. In the domain of object detection [42] tasks like MS COCO, MobileNet V2 + SSDLite [43] demonstrate a 20% increase in efficiency and a 10% reduction in size compared to YOLOv2. This version stands as an advanced, real-time object identification system on the COCO dataset [44]. The comprehensive architecture of MobileNet V2, illustrated in the accompanying Fig. 3, comprises 17 of these building blocks sequentially. It concludes with a regular 1×1 convolution, a global mean pooling layer, and a classification layer. The initial block was slightly different as it employed a standard 3×3 convolution with 32 channels instead of incorporating an expansion layer.

Fig. 4 Inside, there are three blocks, the first of which is a new feature of the MobileNet V2 design



3.3 EfficientNet

EfficientNet is a CNN family that is intended to be both accurate and efficient. They are based on a compound scaling algorithm that scales a CNN's width, depth, and resolution consistently [45]. This enables EfficientNets to achieve cutting-edge accuracy on a wide range of image classification applications while being much more efficient than regular CNNs. The EfficientNet family includes several models, each with its own set of scaling factors. EfficientNet B0, EfficientNet B1, EfficientNet B2, EfficientNet B3, EfficientNet B4, and EfficientNet B7 are the most widely used models. The number following "B" shows the model's size, with larger numbers referring to larger models.

The model's input layer incorporates both normal and lung disease chest X-rays at the outset. To ensure uniformity, each pixel values of the chest X-rays are standardized within the range of 0–1. The hidden layer plays a pivotal role in optimal feature extraction, diligently discerning the most relevant features essential for precise lung

Table 2 Properties of the efficient model

Model	Size	Input dimension	Parameters
EfficientNet B0	75	240×240	4,050,845
EfficientNet B1	31	260×260	6,576,513
EfficientNet B2	36	300×300	7,769,971

illness classification. Within this hidden layer [46], an array of EfficientNet multichannel models, including variants and others, operate. A comprehensive Table 2 outlines the distinctive properties of these various models. Notably, the training of EfficientNet models is conducted leveraging ImageNet datasets, harnessing their rich and diverse data for model refinement and enhancement.

ImageNet databases predominantly comprise primary images, yet fine-tuning the model's weights on medical image classification yields superior performance. In this study, the pre-existing EfficientNet models were harnessed as transfer learning tools, aimed at enhancing the accuracy of lung disease categorization using chest X-ray (CXR) images. This method of instruction significantly reduces training duration, accelerates convergence, and yields optimal outcomes in distinguishing between chest CXR patient data samples as either indicative of lung disease or normal. The flexibility to adjust the weight of the hidden layer during backpropagation remains a notable feature of this work. Furthermore, the employed methodology involves utilizing the dual cross-entropy loss function, which proves effective in guiding the model's learning process toward achieving precise lung disease classification based on CXR images.

A multiobjective neural architecture search was employed to craft the EfficientNet architecture, visualized in Fig. 5. This approach aimed to enhance both precision and reduce floating-point operations. The inception of EfficientNet stemmed from the foundation laid by EfficientNet B0, which paved the way for a family of models spanning from B1 to B7. This lineage achieved top-notch accuracy, ranking within the top 1% on the challenging ImageNet dataset [47]. The evaluation showcased the performance of eight distinct scaled CNN architectures on the ImageNet dataset. For instance, the baseline design of EfficientNet B0 comprised 5.3 million parameters and utilized 224×224 images as input, while the more complex EfficientNet B7

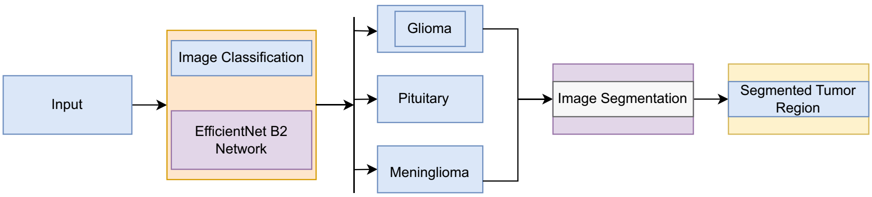


Fig. 5 Overview of the proposed architecture for brain tumors

model contained 66 million parameters and employed 600×600 images as input. Leveraging the substance scaling method, this work aims to yield precise findings. During testing, EfficientNet B2 was employed with a 0-channel multiplier and a channel-1 multiplier, utilizing a maximum input size of 300×300, later resized to 224×224 for analysis. This strategic resizing ensured compatibility with the model's specifications for comprehensive evaluation.

4 Dataset

4.1 Brain Tumor

Brain tumor detection and classification is a key challenge in computer-aided diagnosis (CAD) for use in medicine. The complexity of various tumor types necessitates a system that effectively distinguishes between normal brain activity and pathological conditions [48]. While specialist expertise remains crucial in diagnosing specific tumor types, recent strides in deep learning-based networks have ushered in groundbreaking advancements in brain tumor classification. Elevating the accuracy and efficiency of brain tumor categorization [49] could significantly influence diagnosis and treatment outcomes, presenting a promising avenue by harnessing the capabilities of deep learning methodologies.

Exploring the transformative impact of artificial intelligence (AI) on health care stands as a crucial endeavor aimed at advancing patient outcomes. This research endeavors to devise a preprocessed classifier capable of detecting three distinct types of brain tumors: meningioma, glioma, and pituitary [50], as illustrated in Fig. 6. Below is a comprehensive rundown of the detailed information required to operationalize this deep model within the presented algorithm.

4.1.1 Dataset—Brain Tumor

The strategy's effectiveness was evaluated using a widely recognized brain tumor dataset. Our study utilized a specialized brain tumor dataset consisting of 3064 T1-weighted contrast-enhanced images collected from 233 patients [51]. This dataset encompassed three distinct types of tumors:

- Meningioma: 708 images obtained from 82 patients.
- Glioma: 1426 images sourced from 89 patients.
- Pituitary: 930 images were gathered across 62 patients.

For rigorous evaluation, the dataset was divided into three subsets:

- 80% for training the model.
- 10% for validation purposes.
- 10% specifically allocated for testing the model's performance.

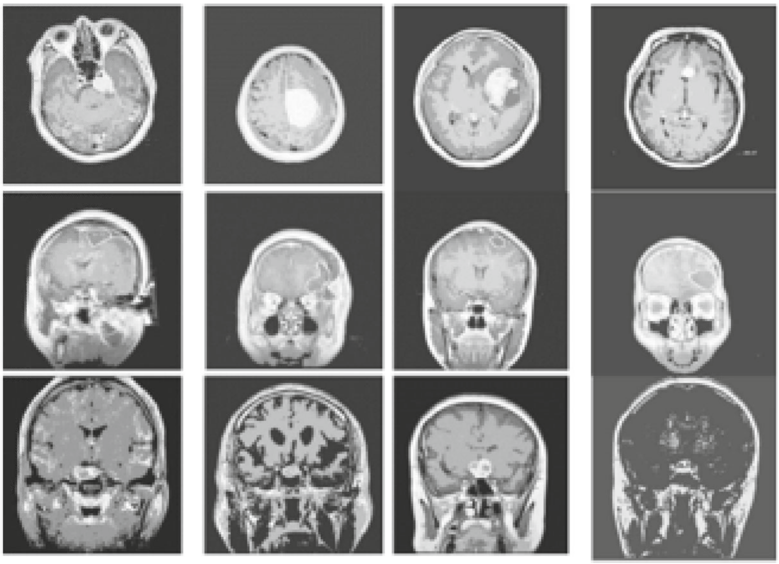


Fig. 6 Example of MRI images of a brain tumor. First row: Meningloma, Second row: Glioma Third row: Pituitary glide

Table 3 List of tumor datasets

Class	Total	Training	Validation	Testing
Glioma	826	676	75	75
Meninglioma	822	674	74	74
Pituitary	827	677	75	75

The images, obtained from the T1-CE MRI modality, encompass coronal, sagittal, and axial views, each with pixel sizes of 49 mm $\times$ 49 mm across the three planes. This freely accessible dataset online includes crucial information for each image, such as patient ID, tumor mask, tumor boundary, and class label. Particularly significant among these is the lesion mask, utilized for cropping the tumor's region of interest (ROI). The dataset provides a diverse range of brain tumor picture samples, exemplified in Table 3. To optimize model training by expediting the learning process and addressing memory constraints, preprocessing stages were integrated. The initial preprocessing step involved normalizing the intensity values of MRI images via a min-max normalization algorithm, effectively scaling the intensity values within the range of [0, 1].

$$I`(x, y) = \frac{I(x, y)}{Z_{\max} - Z_{\min}} \tag{1}$$

where I and I‘ are the images of an original and healthy brain, respectively, and x, and y are the locations of an MRI image. $Z_{\max}$ represents the greatest value of pixel, and $Z_{\min}$ represents the smallest value of pixel. Resizing the images to dimensions 256×256 was the initial step, followed by rescaling them to match the input layer size of the deep model, intended for class prediction with the updated input data. Given that MRI images inherently contain grayscale structures, a transformation was applied to generate three channels. This process involved replicating the grayscale values thrice, resulting in an image size of $256 \times 256 \times 3$. This adjustment to the suitable dimensions ensures compatibility between the input images and the DNN architecture.

4.2 Lung Diseases

Lung disease encompasses a large range of medical issues that impact the function and structure of the lungs. These disorders can be moderate to severe, and they can be acute or chronic. Among the most frequent lung diseases are: COPD (chronic obstructive pulmonary disease), asthma, lung interstitial disease, cancer of the lung, pneumonia, and thyroid hypertension. In numerous aspects, artificial intelligence (AI) can be extremely useful in the diagnosis, care, and study of lung diseases: diagnosis and detection at an early stage, patient education and engagement, treatment penalization, monitoring and predictive analytics, drug discovery and research, telemedicine, and remote monitoring.

4.2.1 Dataset—Lung Diseases

Creating an artificial neural network model for classifying X-ray reports into COVID-19, healthy cases and pneumonia using deep learning is a valuable endeavor [52]. Additionally, developing an algorithm to detect lung disorders and visualize affected areas on X-ray images through Computer Vision techniques represents a significant step in aiding medical diagnosis.

The Table 4 displays the data description and the number of data points utilized in our process. Initially, we segregated the information into two distinct categories: train data and test data. In the initial phase, we built the artificial training model using supervised machine learning techniques on labeled photos [53]. Following that, in the subsequent phase, we employed this newly developed model to assess fresh data, predict final outputs, and classify X-ray images (Shown in Figs. 7 and 8) into three categories: COVID-19, pneumonia, and healthy instances. Our data collection included images from Kaggle, comprising 856 pneumonia images. We further subdivided this data into two subsets, allocating 80% of the first subset used in train and validate purposes and the remaining 20% of the second subset for testing the model’s performance.

Table 4 List of lung datasets

Class	Total	Training	Validation	Testing
COVID-19	1000	600	200	200
Pneumonia	856	514	171	171
Healthy	1000	600	200	200

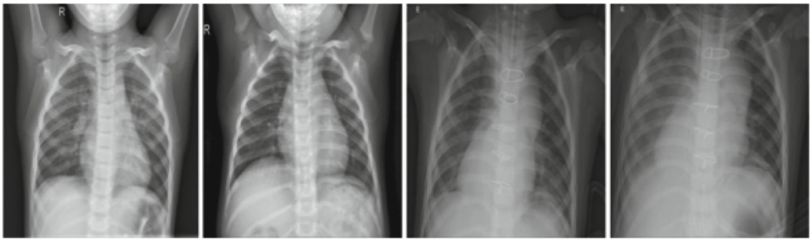


Fig. 7 Chest X-rays of normal (From left to right, the first two images are for normal chest X-rays) and Pneumonia (the next two are for Pneumonia Chest X-rays)

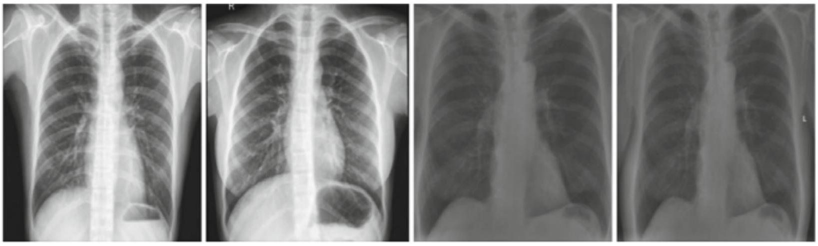


Fig. 8 Chest X-rays of healthy (From left to right, the first two images are for normal X-rays) and COVID-19 (the next two are for COVID-19 X-rays)

Medical X-rays typically exist as grayscale images where each pixel carries an integer value denoting the image's color intensity, ranging from 0 to 255. In this scale, 0 represents complete black, while 255 signifies pure white, with intermediate values representing varying shades of gray. Before diving into model development, enhancing image features stands as a crucial initial step. Each deep learning model operates within specific standard dimensions, necessitating resizing the images to fit these specifications. Furthermore, optimizing the input image involves various transformations like resizing, adjustments in brightness, random rotation, positional changes, horizontal or vertical flipping, and rescaling of the image data. When executed effectively, this preprocessing method significantly reduces the model's processing time while enhancing the accuracy and efficiency of image categorization

5 Experiment

This research was conducted using Google Collaboratory, leveraging multiple libraries such as TensorFlow, Matplotlib, Numpy, and Sklearn. Transfer learning played a pivotal role, incorporating the latest pre-trained models—specifically, MobileNet and EfficientNet. The model of the MobileNet body is outlined in Table 1. To expedite the learning process, fine-tuning techniques were applied to these models. Employing a batch size of 32, we utilized the MobileNet v2 224 × 224 and EfficientNet B2 224 × 224 models. The study spanned 30 training epochs, employing the learning rate of 0.02, and convolution layers with a kernel size of 3 × 3 were employed.

The datasets underwent a meticulous division, with 80% allocated to training data, 10% to validation data, and the remaining 10% designated for testing. In the instance of the tumor dataset, the training dataset consisted of 2027 images, while the validate and test datasets both comprised 224 images each. For the lung dataset, the training dataset included 1714 images, During the verification and test datasets contained 571 images each. This careful data distribution strategy aimed to ensure a comprehensive and representative set of samples for train, validate, and test purposes. The data distribution is shown in Tables 3 and 4.

5.1 Results and Discussion

In evaluation, parameters must first be understood before performing a brief analysis of the achievement of an image categorization model. A classifier is evaluated using the following factors.

Precision

$$\text{precision} = \frac{TP}{TP + FP} \tag{2}$$

Recall

$$\text{Recall} = \frac{TP}{TP + Fn} \tag{3}$$

F1-score

$$F1\,\text{score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + Fn + TN + FP} \tag{5}$$

The elementary objective of the study was to thoroughly analyze the efficacy of MobileNet and EfficientNet concepts within the context of a medical image