

Automated Medical Image Analysis for Disease Detection and Diagnosis

**Thesis Submitted
In Partial Fulfilment of the Requirements for the
Degree of**

MASTER OF TECHNOLOGY

**in
Software Engineering
by**

**Madhav Bansal
(23/SWE/06)**

**Under the Supervision of
Prof. Ruchika Malhotra
(HOD, SE, DTU)**



**To the
Department of Software Engineering
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India**

May, 2025

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

CANDIDATE'S DECLARATION

I Madhav Bansal (23/SWE/06) hereby declare that the work which is being presented in the thesis entitled **Automated Medical Image Analysis for Disease Detection and Diagnosis** in partial fulfilment of the requirements for the award of the degree of Master Of Technology submitted in the Department of Software Engineering, Delhi Technological University is a bonafide record of my own work carried out during the period from August 2023 to June 2025 under the supervision of Prof. Ruchika Malhotra.

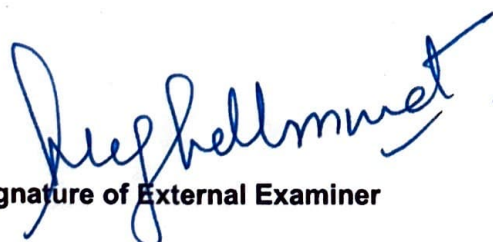
The material contained in the thesis has not been submitted by me for the award of any other degree of this or any other institute.


Date: 19 May 2025

Place: New Delhi


Madhav Bansal

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and the statement made by the candidate is correct to the best of our knowledge.


Signature of External Examiner


Signature of Supervisor



DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daultpur, Main Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR

Certified that Madhav Bansal (2K23/SWE/06) has carried out their project work presented in this thesis entitled "**Automated Medical Image Analysis for Disease Detection and Diagnosis**" for the award of **Master of Technology** from the Department of Software Engineering, Delhi Technological University, Delhi under my supervision. The thesis embodies results of original work, and studies are carried out by the student himself and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

A handwritten signature in blue ink, appearing to read "Ruchika", with a stylized flourish at the end.

Prof. Ruchika Malhotra

Head of Department
Software Engineering
DTU-Delhi, India

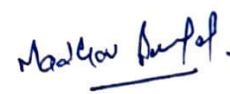
ACKNOWLEDGEMENT

I would like to express my deep appreciation to **Prof. Ruchika Malhotra**, HOD at the Department of Software Engineering, Delhi Technological University, for her invaluable guidance and unwavering encouragement throughout this research. Her vast knowledge, motivation, expertise, and insightful feedback have been instrumental in every aspect of preparing this research plan.

I am also grateful to **Prof. Ruchika Malhotra**, Head of the Department, for her valuable insights, suggestions, and meticulous evaluation of my research work. Her expertise and scholarly guidance have significantly enhanced the quality of this thesis.

My heartfelt thanks go out to the esteemed faculty members of the Department of Software Engineering at Delhi Technological University. I extend my gratitude to my colleagues and friends for their unwavering support and encouragement during this challenging journey. I have had some friends that I am thankful to be around. They made me feel truly at home. In particular, I would like to thank everyone whom I had such a great time. Last two years were great with such a lovely bunch of people around me. Their intellectual exchanges, constructive critiques, and camaraderie have enriched my research experience and made it truly fulfilling.

While it is impossible to name everyone individually, I want to acknowledge the collective efforts and contributions of all those who have been part of this journey. Their constant love, encouragement, and support have been indispensable in completing this MTech thesis.


Madhav Bansal
(2K23/SWE/06)

Automated Medical Image Analysis for Disease Detection and Diagnosis

ABSTRACT

Computerized medical image analysis has become a revolutionary technology in contemporary healthcare, making it possible to detect and diagnose diseases at a faster pace, with greater precision, and lower cost. Using cutting-edge methods like machine learning, deep learning, and computer vision, medical imaging systems can interpret sophisticated medical images like X-rays, CT scans, MRIs, and ultrasounds automatically.

These systems support clinicians by discovering patterns, outliers, and early markers of diseases like cancer, neurological disorders, cardiovascular ailments, and more. This automation not only increases diagnostic accuracy but also lightens the load of healthcare workers and minimizes the scope for human error. The integration of artificial intelligence in medical imaging opens doors to customized treatment plans and better patient outcomes.

This article examines the approaches, resources, and clinical use of computerized medical image analysis, as well as the challenges of data quality, interpretability, and ethical considerations. Recent developments in convolutional neural networks (CNNs), segmentation algorithms, and image classification methods have greatly enhanced the accuracy and consistency of computerized diagnostic systems. Furthermore, coupling with electronic health records (EHRs) and real-time data processing enables a comprehensive understanding of a patient's health status.

Keyword: Convolutional Neural Networks (CNN), Medical Imaging, Segmentation.

TABLE OF CONTENT

Title	Page No.
<i>Acknowledgment</i>	ii
<i>Candidate's Declaration</i>	iii
<i>Certificate</i>	iv
<i>Abstract</i>	v
<i>Table of Contents</i>	vi
<i>List of Table(s)</i>	viii
<i>List of Figure(s)</i>	ix
<i>List of Abbreviation(s)</i>	x
CHAPTER 1: INTRODUCTION	1
1.1 Overview	1
1.2 Problem Statement	2
CHAPTER 2: DEEP LEARNING	6
2.1 History	6
2.2 Machine Learning	7
2.3 What is Learning?	7
2.4 Convolutional Neural Network	9
2.4.1 Implementing Digital Images	10
2.4.2 Convolutional Layer	10
2.4.3 Activation Function	11
2.4.4 Pooling Layer	12
2.4.5 Dropout	12
2.4.6 Cost Function	13
2.4.7 Optimization	14
CHAPTER 3: LITERATURE REVIEW	17
3.1 Image Processing	17
3.2 Early Approaches and Classical Machine Learning Techniques	17
3.2.1 Handcrafted Feature Extraction	18
3.2.2 Classical ML Classifiers	19

3.2.3 Example Studies and Use Cases	19
3.2.4 Limitations of Classical Approaches	19
3.3 Emergence of Deep Learning in Medical Imaging	20
3.4 Crack Detection using CNN	20
3.4.1 Detection of Medical Images	21
3.4.2 Detection of Medical Images	23
CHAPTER 4: PROPOSED ARCHITECTURE	25
4.1 Introduction	25
4.2 Data Acquisition and Preprocessing	26
4.3 Overall Architecture	29
4.3.1 Acquisition and Preprocessing of Data	29
4.3.2 Feature Extraction via CNN Backbone	30
4.3.3 Classification and Segmentation Modules	31
4.3.4 Classification and Segmentation Modules	32
CHAPTER 5: EXPERIMENTAL EVALUATION	35
5.1 Implementation Detail	36
5.2 Training and Testing	36`
5.3 Dataset Description	38
5.4 Performance Matrices	36
5.5 Model Training and Evaluation	39
CHAPTER 6: CONCLUSION AND FUTURE SCOPE	43
REFERENCES	44

LIST OF TABLE(S)

4.1:	CNN Architecture used in Medical Imaging	31
5.1:	Comparison of Traditional vs Deep Learning Techniques	39

LIST OF FIGURE(S)

1.1	Deep Learning-Based Classification System for Automated Medical Image Analysis	4
2.1	Different ML Problem Categories	8
2.2	Concept of Generalisation and Intelligence	9
2.3	A Typical NN Containing Two Hidden Layers	10
2.4	Convolutional Operation in CNN	11
2.5	Activation Functions in DL	12
2.6	Max Pooling Mechanism	12
2.7	Net Modal Dropout Neural	13
3.1	Handcrafted Feature Extraction vs Deep Learning Flow	18
3.2	Emergence of Deep Learning in Medical Imaging	20
4.1	Data Acquisition and Preprocessing	27
4.2	Classification and Segmentation Modules for medical image analysis	32
5.1	Training and Testing Types	36
5.2	Validation and Hyperparameter Tuning in Deep Learning	40
5.3	Evaluation Metrics for medical image analysis deep learning	41

LIST OF ABBREVIATION(S)

Abbreviation	Full Form
AI	Artificial Intelligence
ANN	Artificial Neural Network
AUC	Area Under the Curve
CAD	Computer-Aided Diagnosis
CNN	Convolutional Neural Network
CT	Computed Tomography
DNN	Deep Neural Network
FN	False Negative
FP	False Positive
GPU	Graphics Processing Unit
IoU	Intersection over Union
MRI	Magnetic Resonance Imaging
NIH	National Institutes of Health
PACS	Picture Archiving and Communication System
PET	Positron Emission Tomography
ReLU	Rectified Linear Unit
ROI	Region of Interest
RNN	Recurrent Neural Network
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
U-Net	U-shaped Convolutional Neural Network
Grad-CAM	Gradient-weighted Class Activation Mapping
WHO	World Health Organization

CHAPTER 1

INTRODUCTION

1.1 Overview

In recent years, the health care sector has seen tremendous improvements with the incorporation of artificial intelligence (AI) and image processing technologies. One of the most influential uses is computerized medical image analysis, which uses computational algorithms to help interpret medical images in disease diagnosis and detection. Conventional diagnostic techniques usually rely on radiologists' subjective interpretation, which may be time-consuming, prone to errors, and variable in quality based on differential expertise. In contrast, automated approaches have the promise of high accuracy, swift analysis, and reproducibility across a wide range of clinical conditions.

Medical imaging is the keystone for diagnosing diseases such as cancer, neurologic disease, cardiovascular disease, and infections. Nevertheless, the large amount of imaging data and complexity in cases complicate the manual analysis. Automated image analysis solves these problems by employing machine learning and deep learning models to derive useful features, identify abnormalities, and assist clinicians in making more accurate decisions.

Automated medical image analysis involves a range of tasks like image classification, segmentation, feature extraction, and anomaly detection. They are applied on various kinds of medical imaging modalities such as:

- **X-rays** – widely used for bone fractures, chest infections, and lung diseases.
- **CT (Computed Tomography) scans** – useful for detecting tumors, internal injuries, and vascular diseases.
- **MRI (Magnetic Resonance Imaging)** – ideal for brain, spine, and soft tissue imaging.
- **Ultrasound** – commonly used in obstetrics and internal organ evaluation.
- **PET (Positron Emission Tomography)** – often used in oncology and metabolic disorder analysis.

Deep learning algorithms, especially Convolutional Neural Networks (CNNs), have proven to excel in image recognition and classification processes. The deep learning algorithms learn hierarchical representations of image features and are capable of recognizing refined patterns that the human eye cannot perceive. Image segmentation algorithms compound the diagnosis by isolating regions of interest (for example, tumors or lesions) for a closer look.

Automated systems are increasingly being implemented in clinical workflows to aid radiologists in screening programs (e.g., lung cancer or breast cancer screening),

triaging patients, and prioritizing emergency cases.[1] Automated systems also aid in minimizing diagnostic delays and enhancing outcomes, particularly in under-resourced healthcare facilities. Although its benefits, the area is beset with issues like the requirement for large annotated data sets, maintaining model generalization to different populations, resolving data privacy issues, and obtaining regulatory approval.[2] Ongoing efforts continue to target improving the precision, interpretability, and robustness of the systems to allow safe and ethical use in clinical settings.

Edge AI and cloud computing have also made it possible to deploy real-time diagnostic software that can be embedded into hospital networks or carried in portable devices. Telemedicine and remote diagnosis also gain a lot from these devices, particularly in pandemics or areas hit by disasters when specialists might not physically be there[3].

1.2 Problem Statement

The rising global disease burden and the escalating need for rapid and accurate diagnosis have put tremendous pressure on health care systems across the globe. Medical imaging is pivotal in the diagnosis and treatment of most diseases including cancer, cardiovascular diseases, neurological diseases, and infectious diseases. Nevertheless, image interpretation continues to rely heavily on the skills of radiologists and clinicians, thus presenting several major challenges:

Dearth of Skilled Radiologists: Most parts of the country, particularly rural or developing regions, are plagued by a critical scarcity of skilled radiologists. The shortage translates into delays in diagnosis, restricted exposure to quality healthcare, and an escalation of preventable disease advancement.

Subjectivity and Human Mistake: Interpretation of medical images by hand is subjective and is subject to inter-observer variability and fatigue-related mistakes. Even highly trained radiologists can overlook minor anomalies, particularly when reading massive numbers of images.

Increased Imaging Load: The common use of high-resolution imaging technologies like CT, MRI, and PET has led to a tremendous surge in the quantity and complexity of images. Radiologists are usually under time pressure and cannot undertake detailed reviews for all patients.

Need for Early and Correct Diagnosis: Most diseases, such as cancer and neurological disorders, need to be identified early for successful treatment. Delayed or incorrect diagnosis can lead to higher morbidity, mortality, and treatment expenses.

Inconsistency and Non standardization: Variations in imaging protocols, instrument quality, and diagnostic criteria at hospitals and nations can create inconsistencies and results in healthcare interpretations.

Scalability and Cost Limitations: There are challenges scaling diagnostic procedures manually, especially in a large population screening program, given the time, cost, and human resource implications.

To tackle these challenges, there is an increasing demand for computer-aided medical image analysis systems that are capable of accurately, efficiently, and reliably interpreting medical images to detect and diagnose disease. These systems need to tap into advanced methods in artificial intelligence (AI), specifically machine learning and deep learning, to analyze and process complicated image data, extract useful features, and deliver actionable information to clinicians.

But creating them raises a new range of technical and clinical hurdles, such as the requirement for massive, labeled, and heterogeneous datasets; the need to ensure robustness and generalizability to multiple populations and imaging scenarios; the imperative to preserve patient confidentiality; and gaining approvals from regulatory bodies to deploy clinically.

Therefore, the root challenge is in developing and deploying smart, automated image analysis technologies that are accurate, interpretable, scalable, and clinically proven to assist healthcare professionals to better and more efficiently diagnose diseases.

Despite fast-paced developments in medical imaging technologies, interpreting medical images continues to be a difficult, labor-intensive, and expertise-based activity. The increasing amounts of imaging data, coupled with worldwide shortages of trained radiologists, have resulted in diagnostic delays, increased workload for healthcare practitioners, and an elevated risk of diagnostic errors. Interpretation of images by human operators is subject to variability, fatigue, and inattention, especially in high-stress or high-throughput clinical environments. In addition, variability in imaging protocols and absence of standardization lead to diagnostic differences between institutions. The growing need for early, reliable, and high-volume diagnostic options—particularly for large-scale screening programs and emergency settings—has revealed the shortcomings of conventional methods. Parallel with this, the sophistication and level of detail in current imaging data push the boundaries of possible human examination. These are the factors emphasizing the imperative for smart, computerized medical image analysis systems capable of supporting clinicians with rapid, reliable, and reproducible diagnostic information. Yet, deploying such systems is not without its own set of challenges, such as requiring large sets of annotated data, providing model generalizability to populations, preserving transparency and interpretability, dealing with data privacy issues, and satisfying regulatory demands for clinical release.

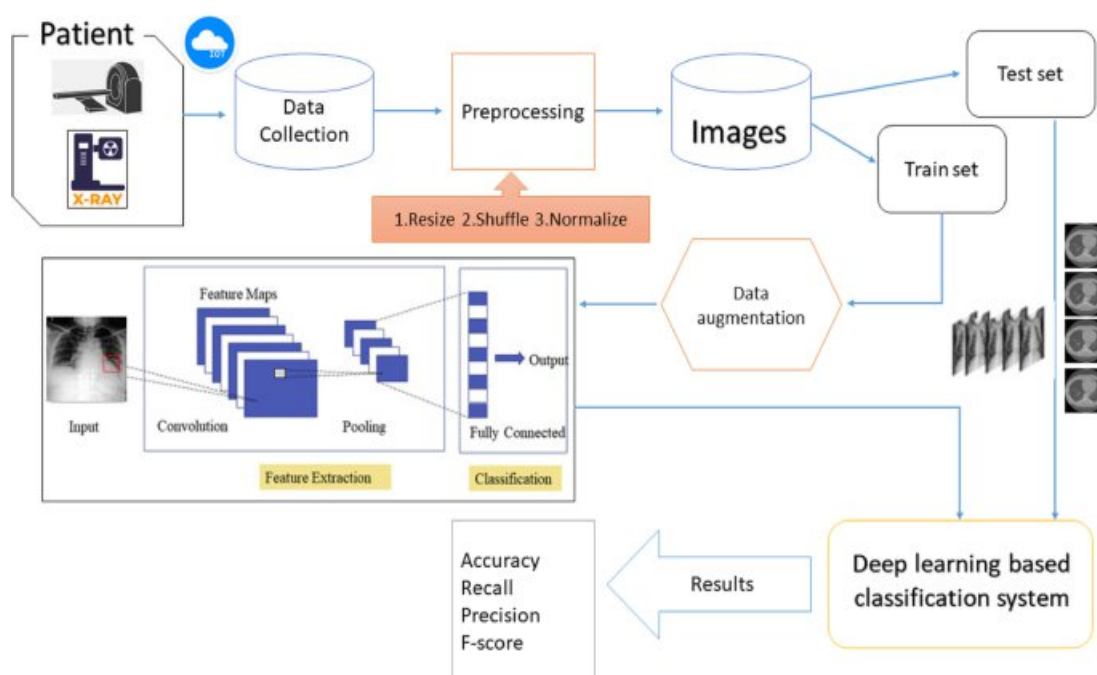


Fig 1.1. Deep Learning-Based Classification System for Automated Medical Image Analysis[4]

The figure depicts a general pipeline in applying deep learning to process medical images and classify them for disease diagnosis. It begins with data collection and goes through preprocessing, model training, testing, and ends with results.

Important Stages:

Data Collection: It begins with the collection of medical images from patients. The image particularly depicts inputs from CT scans and X-rays, meaning that the system has the potential to process various imaging modalities. A cloud symbol and a data cylinder with the label "Data Collection" indicate that patient information and their respective images are collected and stored.

Preprocessing: The images collected are subjected to preprocessing operations so that they are ready for input into the deep learning model. The diagram clearly states three important operations:

Resize: Resizing images of all inputs to a fixed size.

Shuffle: Randomizing the image order, an important step for successful training.

Normalize: Rescaling pixel values into a normalized range (e.g., 0 to 1 or -1 to 1) to enhance model convergence and performance.

Images: Pre-processed images are then sorted and divided into two primary sets:

Train set: A greater fraction of the data employed to train the deep learning model to learn patterns and features corresponding to various conditions.

Test set: An independent, unseen fraction of the data employed to test the performance and the generalization power of the trained model.

The center of the system is a Convolutional Neural Network (CNN), which is an efficient and widely used architecture to analyze images. The diagram illustrates a simplified view of a CNN:

Input: A medical image (in this example, an X-ray with a highlighted area of interest).

Feature Extraction: This stage usually consists of several layers of Convolution and Pooling.

Convolutional layers learn hierarchical features of the input image using filters. The output of these layers is a series of Feature Maps, which emphasize various aspects of the image. Pooling layers reduce the size of the feature maps while keeping the essential information, hence making the model less sensitive to variations in the input. **Classification:** The extracted features are then passed through one or more Fully Connected layers after feature learning. These classify the final output into probabilities across various disease categories. **Data Augmentation (Optional but Common):** The hexagonal box titled "Data augmentation" implies that methods could be used to augment the size and variety of the training set artificially. This could include operations such as rotation, translation, scaling, and flipping of the current images, which enhances the robustness of the model and avoids overfitting.

Deep Learning Based Classification System: This is the entire trained model in a box, waiting to receive new medical images as input and predict the existence or non-existence of a particular disease.

Results: The performance of the classification system is assessed with several metrics, among which:

Accuracy: The overall accuracy in correctly classifying images.

Recall (Sensitivity): The capability of the model to identify all positive cases correctly.

Precision: The model's capacity not to misclassify negative cases as positive.

F-score (F1-score): The harmonic mean of recall and precision, which offers an average measurement for the performance of the model.

In short, this picture depicts at a high level an automated medical image diagnosis system that leverages deep learning (namely CNNs) for medical image classification towards detecting and diagnosing diseases. It identifies the principal steps required to develop and assess such a system, from data acquisition to performance assessment.

CHAPTER 2

DEEP LEARNING

Today, ML techniques have facilitated numerous parts of the contemporary civilization. More and more data is constantly produced, and it will further grow in the future. 80%-90% of the total data cannot perform most of the tasks as structured data (unstructured data). However, traditional ML techniques like logistic regression, support vector machine, decision tree and k-nearest neighbours' were constrained by the fact that they could not handle unstructured data. It has only been within the past few decades that machine learning has evolved from a methodology that required significant domain expertise and careful engineering to one where an algorithm might transform unstructured data points, such as pixel values for images, into appropriate representations that could then be used by other machine learning algorithms [7]. Representation learning is a suite of techniques that enable computers to analyse disorganised information and identify how it can be used for a given purpose without any specific instructions. Machine learning algorithms in deep learning have several layers of representation. The deep representation learning is achieved by using multiple layers of simple complex nodes, which can change the input from one form to another at a slightly higher level of abstraction. When enough of these are put together, it becomes possible to discover very sophisticated functions thereby making it easy for professionals with diverse research topics different fields take much attention. These novel technologies have been applied to tackle a difficult issue in civil engineering. Following section covers the basic idea of DL alongside the distinct constituents that must be put together without fail to develop an efficient DL model. Based on these insights and methodologies, an asphalt specific pavement-crack identifying framework will be brought forth.

2.1 History

In the early days of the AI construction, very high intelligence computing power tried very hard to resolve problems within the range of possibilities for human intellect; problems were thought of in a row of formal-mathematical rules, hence, they were simple enough for machine value. Therefore, the real aim of AI development is to handle tasks that are simple for people in such a way as they understand them "intuitively", but impossible to describe on the basis of any formal language for programming computers[8]. Solving these challenges is possible by using DL. DL aims not only to learn the mapping but as well as acquiring the most favourable data representation [8]. People have been using the terms AI and DL simultaneously ever since the first learning algorithms designed were imitative of brain functions. Essentially, the idea of artificial neural networks (ANNs) being the same as deep learning is now commonplace among practitioners in this field. About fifty years ago, Rosenblatt[9] popularized neural networks (NNs) through various types of perceptron networks. However, in 1969 Minsky and Papert considered them very limited in their function [10]. A lot of people generalized these restrictions improperly, which in turn caused a significant decrease in the popularity of neural networks. A number of deep learning techniques were developed in the 1980s and 1990s like long short term memory (LSTM)[11-12] as well as back propagation algorithm. The 1990s saw

unrealistic claims made by the artificial intelligence community that failed to meet these expectations when artificial intelligence research could not live up to it. Kernel machines and graphical models also found success in their own right; this, coupled with a drop in interest for neural networks as of 2007. This led to NNs losing their enthusiasm between 2001-04 or so [8]. In 2006 Hochreiter et al.[10], demonstrated how one might construct deep-belief-network that could be trained effectively through unsupervised layer-wise learning: while still others adopted similar techniques when dealing with different types of hierarchical architectures [13,14]. These studies have waken up AI from coma. With performance better than other techniques in multiple artificial intelligence challenges, DL is now one of effective methods among supervised, unsupervised, and reinforcement learning.

2.2 Machine Learning

Since DL falls under a wider range of other ML methods, some basic concepts in ML have to be talked about. In different fields, ML algorithms and models have been utilised hence the multiple definitions of ML. The name “machine learning” was given in the year 1959 [15] hence this relates to how mathematical models and algorithms are employed for performing specific functions using data generated by computer systems together with experience [16]. Learning from data is the process of analysing situations endowed with certain patterns that do not have a known theoretical solution. In such situations, it means that Machine Learning will always provide ways through which such patterns can be identified through which patterns can be determined. The machine learning problems generally fall under three categories: supervised, unsupervised and reinforcement learning as shown in Figure 2.1. In supervised learning a naive model can only learn a regulated data with beginners guide (The learning set). From where it gets ins and outs together Proactive Maintenance; we can travel through multiple articles including step in step Self-Instructional. For example, when it comes to detecting whether an image has a particular object, training data will involve images containing the object or images that do not have it (the in-put), with each image receiving a label depending on whether or not it contains the object [16]. Contrastingly, outputs being non-existent serves as a basis for application of unsupervised learning models. Unsupervised study focuses on how systems can find a function to reflect a latent structure from data without labels. Refereeing to reinforcement learning is a method used by machines to learn through experimentation with reward from themselves experiences and actions in an interactive setting. The agent increases its performance by automatically discovering the best way of behaving in a given situation.

2.3 What is learning?

The traditional frameworks are used to explain the aspects of learning algorithms and for learning to be considered as feasible, provide mathematical proof of this fact—Shai Shalev-Shwartz, Shai Ben-David [18] presented examples that could help in understanding how basic learning process work alongside what have been identified as principal challenges within machine learning (ML). Rats learn how

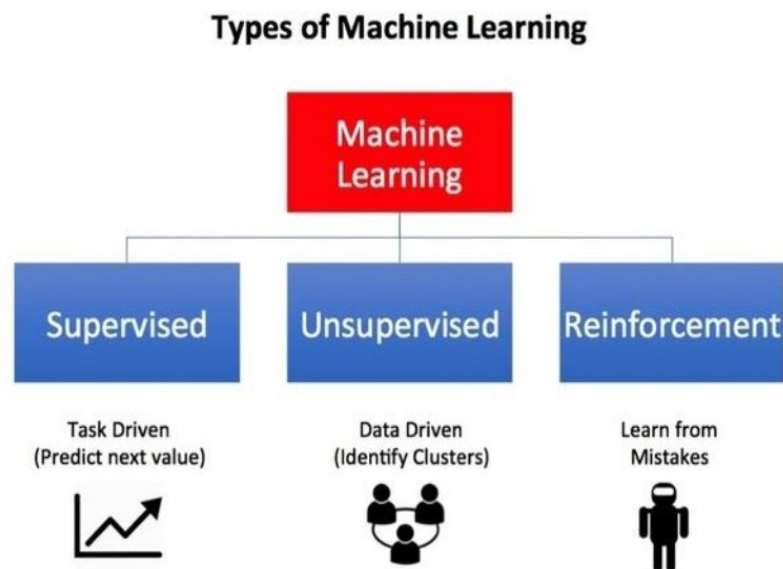


Figure 2.1 . Different ML problem categories [17]

to avoid poisoned food starting from their childhood. Rats usually take a small amount of new food first and are careful to investigate the physical consequences. If the food causes sickness, they never eat it forever. The experiment involved an animal in search of a harmless meal. In this case, the animal would expect that if it experienced a negative label then it would also develop negatively. Assume we are attempting to write a spam detector program. For instance, one straightforward way is to remember every email determined to be spam by a user. When an incoming email is received, it is verified against the spam set. If it is found in the spam set, then it is marked as a spam message; else, it is saved in the inbox folder. Memorization is occasionally helpful, but it does not have much in common with learning because it cannot be generalized. An intelligent learner who truly understood should be able to extract wider generalizations from diverse instances. It therefore means that generalizing constitutes the ultimate definition of intelligence. When compared with other creatures, man's special gift is his ability to think and understand concepts widely, putting us one step ahead. For instance, given a realistic picture of an elephant, a child might be able to recognize a drawn elephant that looks very different (Figure 2.2). Another problem is when the learner comes to a wrong conclusion. In explaining this notion, Skinner's superstition experiments are the most useful example. To be precise, Skinner put some hungry pigeons in a box that came with an automatic device meant to supply food for the hen occasionally with no consideration given to its actions. He found that pigeons would exhibit behaviours signalling expectancy only during feeding time and for more or less two minutes after that. While waiting for food, a particular bird spun round and round in a counter clockwise direction before making one or two turns in the opposite direction before it was rewarded. But there were sometimes when it was fed by Andy and would peck continuously at the upper edge of its basin." "A bird thrust its head out and swung it sharply rightwards from leftwards then back again with some slowness so as to make it like a pendulum while another bird began shaping up like it was making quotations (this means they stuck their heads beneath an unseen pole raised them up multiple times'[19].

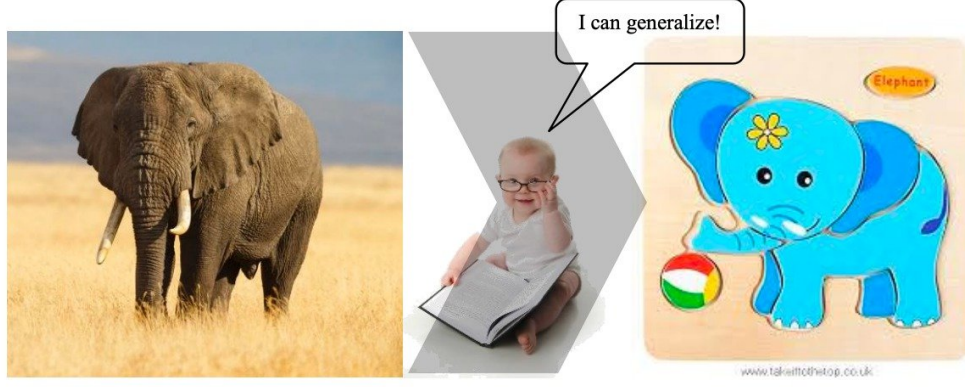


Figure 2.2. Concept of generalisation and intelligence[11]

When humans learn, they use their common sense and ignore random patterns or conclusions from learning that are meaningless, but machines do not. A machine requires well defined principles to steer it out of arriving at irrelevant conclusions. In simpler terms, the algorithm should be able to discern a pattern in the data but not in the noise.

2.4 Convolutional Neural Network (CNN)

In this section, we will introduce the basic notions of NNs and discuss various parts of CNNs before explaining why each architecture is worth considering. There is a standard NN architecture shown in Figure 2.3 with input i given as a single feature vector, x^k . The input is passed through successive hidden layers, to estimate an output \hat{y} . All the layer consists of neurons (nodes), each of which is completely linked to all nodes in the previous layer and the following layer. You can do this at arbitrary patches because each layer has no connections to the others. With respect to this particular patch, the output of the one that came before it $a_k^{[l-1]}$ is modified by the weight $\omega_{jk}^{[l]}$ and added to a bias term $b_j^{[l]}$.

After this happens, it passes through an activation function $g^{[l]}$ which decides what will be outputted from the node $a_j^{[l]}$.

The result of each node is generally formulated as

$$a_j^{[1]} = g^{[1]}(\sum_k \omega_{jk}^{[1]} a_k^{[1-1]} + b_j^{[1]}) \quad (1)$$

The input vector is denoted by $a^{[0]}$. The final fully-connected layer $a^{[3]}$ is given the name “output layer” in this example, while in classification tasks it shows the likelihoods of classes. It should be noted that the weights $\omega_{jk}^{[l]}$ as well as the biases $b_j^{[l]}$ are actually calculated while training the model.

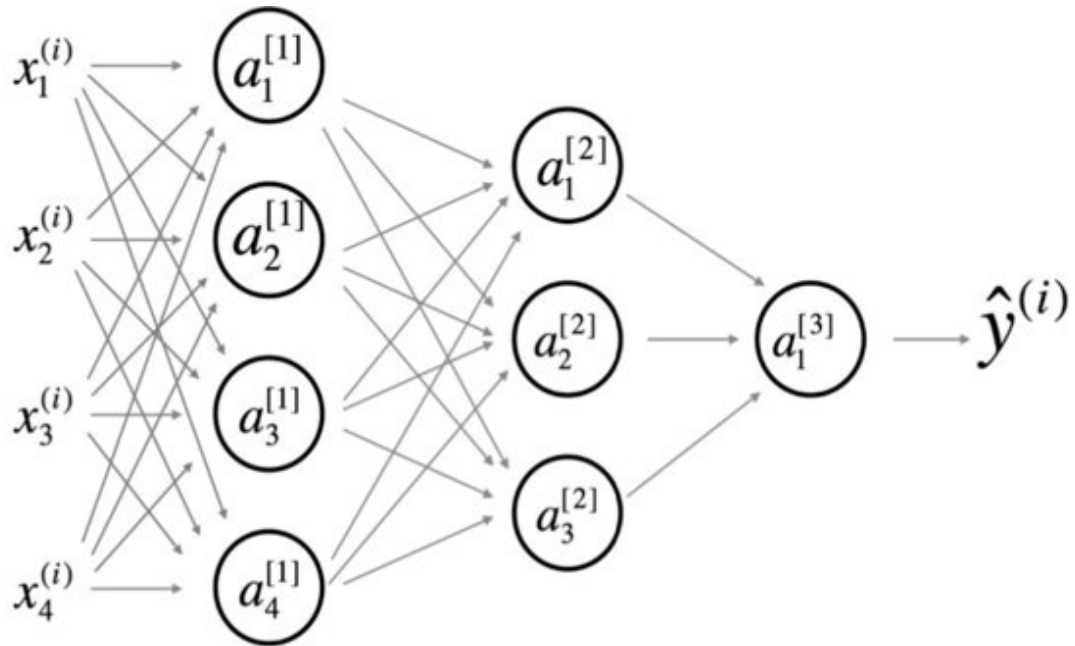


Figure 2.3. A typical NN containing two hidden layers[23]

2.4.1 Implementing Digital Images

We first convert a tensor with three channels (an order of 3) into one with a smaller order, meaning one (a vector) when you intend on using an ordinary network to handle a digital image. For example, consider an image of 100×100 pixels resolution stored in RGB format with 3 channels hence appearing as a vector of 30,000 elements while each element represents a single input feature. Building an NN model requires thirty thousand weight parameters for one node located at layer one. Therefore, it implies that if you want to employ larger images or insert additional nodes into the first layer then you will have increased number of parameters. This approach does not really work for image NN development and it is cumbersome. Convolutional neural networks take better advantage of the forms of input data to set an architecture using weights more effectively. CNNs capitalizes on two vital ideas to enhance network performance: handy interactions and shared parameters. In an ordinary neural system, each output node $a_j^{[l]}$ interacts with every input neuron $a_k^{[l-1]}$ whereas CNNs are sparse in terms of connections usually. This can be achieved by making use of filters having less size compared to the initial data. For example an input image may contain many pixels while filters consisting merely tens or hundreds of pixels can identify minor yet important characteristics like contours. Other techniques like a dropout layer can be used to improve performance and avoid over-fitting of data. This paper describes how each of these layers works as well as their configurations within the CNN system.

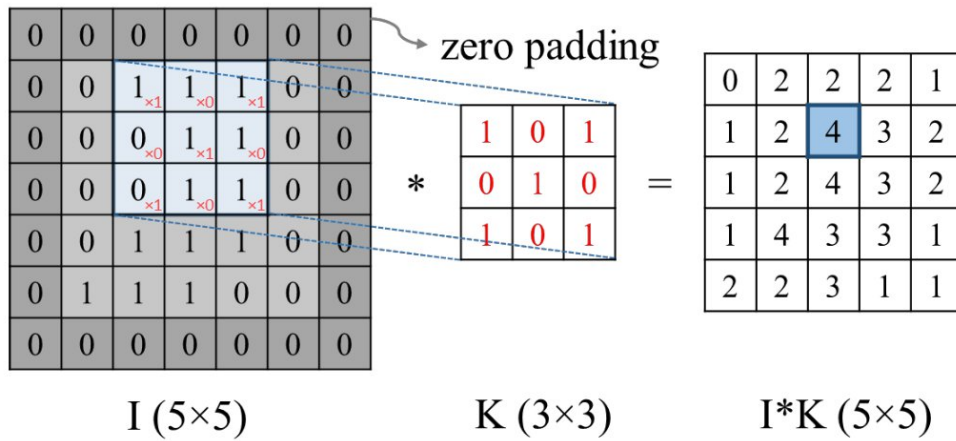
2.4.2 Convolution Layer

Convolution layers are the main computational elements of CNNs. A series of filters with learnable weights is included in each block. These filters are convolved with input from the previous layer to look for important patterns across the whole picture. An error function is minimized by designing filters in a certain manner for each network. In a CNN, a convolution operation is identical to a cross-correlation operation in two-dimensional signal processing. In Figure 39, we see the 2D image I of size 5×5

5 pixels and filter K of size 3 x 3 pixels being subjected to convolution operation. Applying the filter to one pixel at a time means moving one pixel at a time on the input image; this stride (i.e., one) makes the output image smaller than the input image. This is solved by placing zero pixels on the edge of the input image during convolution while using a filter (see Figure 2.4). The convolutional layer's output is calculated using the addition operation on the result of the convolution operation and a bias "b," which is then passed through an activation function "a". A formula; $\text{Conv}(I, K)_{xy}$ of a pixel's convolutional layer in (x, y) coordinate is:

$$\text{Conv}(I, K)_{xy} = a \left(b + \sum_{i=1}^h \sum_{j=1}^w \sum_{k=1}^d K_{ijk} * I_{x+i-1, y+j-1, k} \right) \quad (2)$$

Figure 2.4. Convolution operation in CNN.[17]

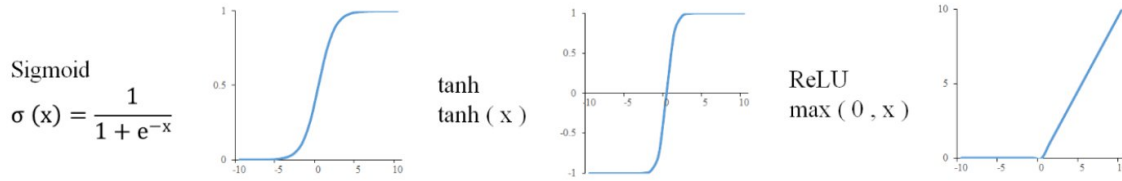


2.4.3 Activation Function

To introduce nonlinearity, it is important for you to include a nonlinear activation function in the network. Check the following diagram where the three activation functions are commonly seen in DL. In the early days of DL, many people loved the sigmoid function. Nevertheless, nowadays it is well known that tanh function outperforms it [20]. One issue with these functions is that their gradients vanish at the end points, making them stagnate. As a result, learning becomes drastically slow when a gradient-based optimizer is employed. In recent times, the Rectified Linear Unit (ReLU), which is non-saturating, has gained popularity as an activation function [21,22]. The use of this activation function has been found to increase network performance. In this study, we are using ReLU activation functions for all activation functions except the final layer of the network; it will consist of a softmax activation function to help classify input data. Softmax function $\text{si}(\vec{x})$ of class i defines probabilities of input points belonging to each class, defined as:

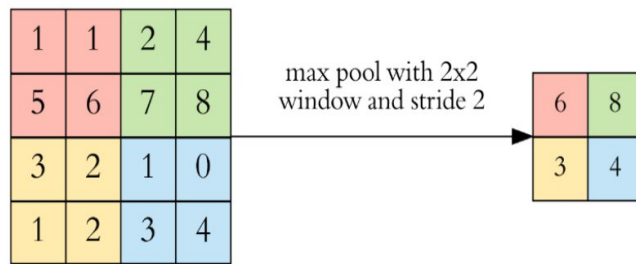
$$S_i(\vec{X}) = \frac{e^{x_i}}{\sum_{j=1}^2 e^{x_j}} \quad (3)$$

Figure 2.5. Activation functions in DL.[18]



2.4.4 Pooling Layer

Pooling layers are mostly used by CNNs for reducing the size of the input layers so that computation is accelerated while at the same time increasing detection robustness. The most commonly used types of pooling are max-pooling and average pooling in DL. For image-like data, max-pooling has been shown to be far much better [23]. Every pooling layer in this study is a max-pooling layer unless otherwise indicated. In Figure 2.6, notice that with a 2×2 window and a stride of 2 the max-pooling mechanism is illustrated. As it goes through the input data, the highest value in the 2×2 window is selected. With each two-pixel shift of the 2×2 window, the whole input will be operated upon in this way. In this way, size of input data is reduced (in this example, the output data is half the size of the input data).



2.4.5 Dropout

Figure 2.6. Max pooling mechanism[24].

Dropout could make neural net more flexible by applying diverse architectures or avoiding from overfitting-many various nets can combine into one net [25]. Actually, it means to randomly remove neurons in a NN(“dropout”). Figure 2.7 illustrates the disappearing of temporarily their input-output links together with output layers themselves during dropout applied on neural networks (see fig. forty-two). As part of

this analysis, the dropout method will be used on all the layers with 0.5 as the threshold probability rate [25].

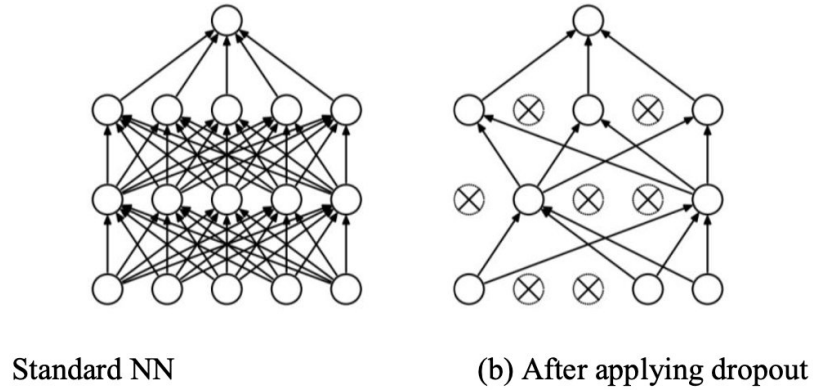


Figure 2.7. Net Modal Dropout Neural [4]

2.4.6 Cost Function

Training a CNN is essential for finding a group of weights and bias that minimizes mistake in prediction and actual. To numerically measure error, we would have to define loss functions. Categorical cross entropy (Equation 4) is the designated loss function L_i that is used for estimating the difference between the true class y from the probability distribution over the predicted class \hat{y} of a single image. Through the use of the softmax function, it becomes easy to calculate the probability distribution of the anticipated class.

$$L_i(\hat{y}_i, y_i) = \sum_{i=1}^k -y_i \ln \hat{y}_i \quad (4)$$

When it comes to incorporating image labels into neural networks, modelers use a one-hot encoding scheme. Two classes- one and two are represented in binary classification by (0, 1) as well as (1, 0). Consequently, the network model output is represented in the form probabilities for every class which are denoted by (\hat{y}_1, \hat{y}_2) . This means that in our example case, if we have output vector (0.3, 0.7) then it implies that there is a 30% possibility that it belongs to class one while there is a 70% chance that this same vector belongs to class two.

In equation 4, the lost value is 0.36 given that the true class is one. To illustrate, $(-0 * \ln(0.3) - 1 * \ln(0.7))$. is what the context is? For instance, poor prediction for the same example shall attract loss of 0.92 (0.6, 0.4) while good predictions like (0.05, 0.95) have such small losses as 0.05 The cost function, C , is merely a summation over the loss function L that has been applied to all images divided by their number, N .

$$Cost = \frac{1}{N} \sum_i^N L_i(\hat{y}_i, y_i) \quad (5)$$

In order to add regularization into the model, it is necessary to insert the L2 regularization formula. This formula is defined as the sum of the squares of the weights on features and then is put into cost function together with its parameter.

2.4.7 Optimization

Broadly speaking, the question of learning is about optimisation. It follows therefore, that this optimisation aims at determining the most appropriate parameters that help to reduce the cost function, that is, weights and biases. In cases where neural networks are large, closed form solutions to their optimisation are not available so these are determined via gradient descent among other methods while using iterative algorithms. Common neural networks have non-convex search spaces, therefore it is logical to consider using a modified stochastic gradient descent algorithm. The cost function has millions of parameters in the proposed CNNs which should be fine-tuned. In this research work, we use Adam (adaptive moment estimation) to minimise the cost function. Adam is a first-order gradient-based optimisation algorithm for stochastic objective functions.

In the training phase, this optimisation algorithm will be administered as the optimisation algorithm [26, 28] since Adam optimiser is computationally efficient, with small amounts of memory required, invariants to gradient rescaling along diagonals, and suitable for huge data and/or parameter non-convex optimisation problems in ML [26,27]. An effective method of finding gradients of parameters through backwards and forwards application of chain rule on a computational graph is back propagation. When every forward pass is done, the expense function calculates; thus, depending on the output from such activity, besides inputs used as ground truths we can compute value derivative with respect to learning parameters by performing back propagation. Furthermore, this information feeds into Adam optimiser which modifies learning rates according to them.

The computational parallelism gets quicker-due to vectorisation that is increased on Graphics Processing Unit (GPU) processors. Nonetheless, the computation will move slower since with a larger data set there is need for large memory to implement vectorisation. As a remedy to this problem, the training data is broken into smaller mini-batches [29]. Despite enlarger mini-batches offers more computational parallelism, smaller mini-batch training, nevertheless, tends to give better generalisation performance, as well as having a much smaller memory footprint that can be leveraged to increase the speed of machines used for this purpose [30].

According to Masters & Luschi, mini-batches with fewer samples lead to gradients being calculated closer to their current value thus giving rise for both stable learning procedures with less noise in them or simply put improving reliability of such systems [30]. Thus this survey is going to employ mini-batches of 32 images (N in Equation (5) will be 32 instead of the total number of images).

CHAPTER 3

LITERATURE REVIEW

3.1 Image Processing

A 2D function $f(X, Y)$ is employed to describe an image, with X and Y acting as spatial coordinates indicating a point's position within the given image and with the "f" value representing how much light a pixel contains at this specific point. Both pixels along with their intensity levels are discrete and of a finite number. Digital Image Processing, as defined by Gonzalez[31], is the use of computer to process digital images which is also termed as "The field of Digital Image Processing". One of the controversial issues that is confronting researchers with regard to the interface within image processing and other fields like computer vision and image analysis. It is difficult to distinguish between computer vision and image processing. But it must be realized that these computerized processes can further be classified into low-level, mid-level, and high-level processes. Medical imaging has transformed the healthcare sector in such a way that non-invasive visualization of the internal structures of the human body is now possible. Due to the increasing need for fast and accurate diagnosis, medical image analysis has become a much-needed tool for clinical decision-making. Historically, image interpretation was done manually by radiologists and experts. With growing image volume and complexity, manual diagnosis is time consuming and prone to errors. Recent breakthroughs in the field of artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), have resulted in automated medical image analysis systems that can support clinicians with disease detection and diagnosis. This review of the literature examines significant research trends, methods, pitfalls, and prospects in this fast-growing area.

1. Low-level processing is the provision of such as receipt or provision of images that comprise operations such as noise elimination, difference enlargement as well as image sharpness.
2. The images are taken as inputs by middle-level processes and then they produce image features such as contours and edges.
3. High-level processing like image analysis, includes sophisticated operations such as recognition and object detection.

3.2 Early Approaches and Classical Machine Learning Techniques

Early CAD used conventional ML methodologies like Support Vector Machines (SVM), Decision Trees, k-Nearest Neighbors (k-NN), and Random Forests. These used mostly handcrafted feature extraction, where one would look for relevant image features such as texture, shape, and intensity. For example, S. Aylward and J. Jomier (2003) wrote about classical ML segmentation and feature extraction in CT and MRI scans. Doi (2007) presented a comprehensive review of CAD systems, especially in the field of mammography, and illustrated how early machine learning methods could

be used to help identify breast cancer. Although these systems were promising, their efficacy was hindered by the quality of manually engineered features and the lack of ability to generalize across various datasets. In conventional maintenance systems, technicians or experts usually locate and assess medical images under expert guidance, a process that requires a great deal of time and effort. It is therefore anticipated that automated or semi-automated processes using image analysis will facilitate the process, resulting in timeliness and enhanced performance in crack index and condition assessment. There are several studies on automated detection of medical images with the use of CNN. [32, 33, 34, 35–38].

3.2.1 Handcrafted Feature Extraction

Handcrafted features were obtained employing mathematical and statistical methods to represent visual attributes of the image. Typical features are: Texture Features: Like Gray Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP), and Gabor filters. Shape Features: Boundary descriptors, moments, and geometric measures (e.g., circularity, area, compactness). Intensity Features: Histogram-based descriptors and pixel statistics.

Edge Features: Sobel or Canny edge detectors to detect boundaries. The features were used as input to machine learning algorithms. Yet, their performance was frequently hampered by how well the feature engineer could represent clinically relevant patterns, and they tended to be sensitive to noise, scale, orientation, and illumination.

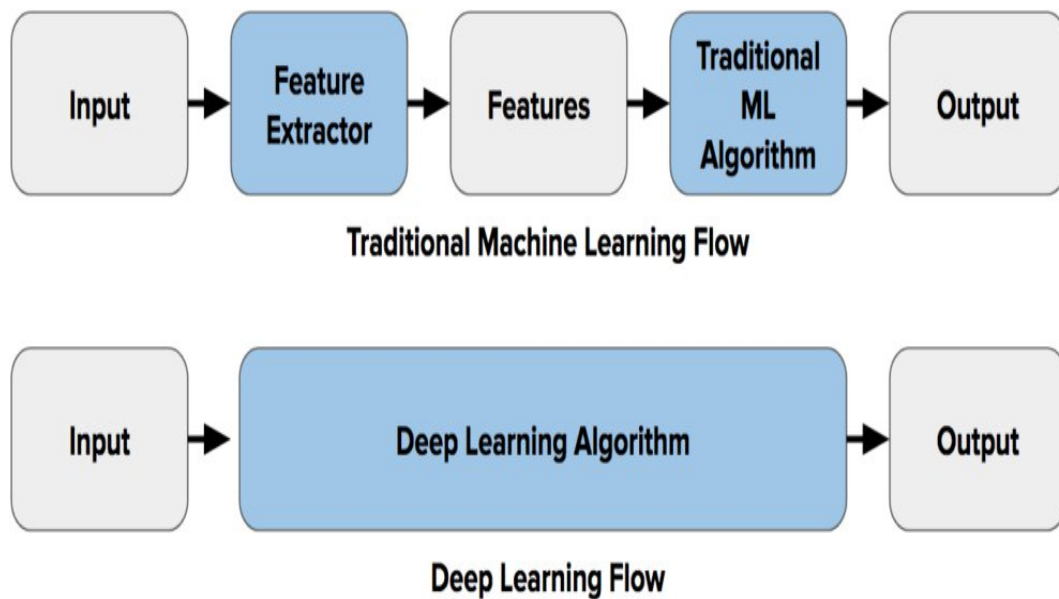


Figure 3.1. Handcrafted Feature Extraction vs Deep Learning Flow[4]

3.2.2 Classical ML Classifiers

A range of popular classifiers was employed:

Support Vector Machines (SVM): Well-suited to high-dimensional space, SVMs worked effectively when combined with appropriate kernel functions (e.g., radial basis function) and feature scaling.

k-Nearest Neighbors (k-NN): A straightforward algorithm using distance metrics, effective for pattern identification in small datasets.

Decision Trees and Random Forests: Offered interpretability and noise resistance, and were able to deal with mixed data types.

Naïve Bayes: Employed in situations when features were assumed to be independent statistically.

Logistic Regression: Utilized for classification problems in binary disease classification, e.g., tumor detection.

3.2.3 Example Studies and Use Cases

Aylward & Jomier (2003): Was concerned with segmentation of MRI and CT images, employing region-growing algorithms and feature extraction techniques. They presented how initial CAD systems were able to identify structural irregularities, but highlighted the challenges in generalizing between patients owing to anatomical variation.

Doi (2007): Had one of the early landmarks in CAD, particularly in breast cancer detection through mammography. He reported that conventional ML models were applied to detect microcalcifications and masses and that they usually resulted in moderate sensitivity and specificity. Nevertheless, Doi observed that those systems were poor performers in clinical settings because of image variability and poor robustness.

Sahiner et al. (1996): Applied SVMs to mass detection in mammograms, and their method demonstrated that hyperparameter tuning and preprocessing could significantly influence classification results.

Sargent et al. (2001): Employed a texture analysis combined with decision trees for the classification of liver tumors in CT images. They emphasized ROI extraction prior to classification in order to reduce false positives.

3.2.4 Limitations of Classical Approaches

Though these approaches constituted the initial wave of radiology automation, they had considerable limitations: **Relying on Domain Knowledge:** Feature extraction involved deep domain knowledge and failed to generalize across organs or imaging modalities. **Poor Generalizability:** Models learned on one dataset tended to fail on others as a consequence of overfitting or dataset bias. **Scalability Challenges:** Most models were trained and evaluated on small local datasets, which constrained their clinical relevance. **Absence of End-to-End Learning:** Every stage (e.g., preprocessing, feature extraction, classification) was treated separately, resulting in accumulation of errors and inefficiency.

3.3 Emergence of Deep Learning in Medical Imaging

Deep learning (DL) methods have shown minor but significant efficacy in solving many practical problems [66-69]. With more emphasis on automatic learning and less dependence on heuristics, LeCun et al.[70] showed the possibility of building better pattern recognition systems. AlexNet introduced in 2012 by Krizhevsky et al. brought revolutionary results in an image classification competition (ImageNet challenge [71]), demonstrating the power of CNN architectures [72]. Subsequent to this, numerous researchers have applied AlexNet and several other Convolutional Neural Network (CNN) structures for civil infrastructure damage detection. Deep learning revolutionized the image analysis environment with the capability to extract features automatically straight from raw data. The efficiency of deep neural networks, particularly Convolutional Neural Networks (CNNs), on tasks such as natural image classification resulted in their instant use in the medical field.

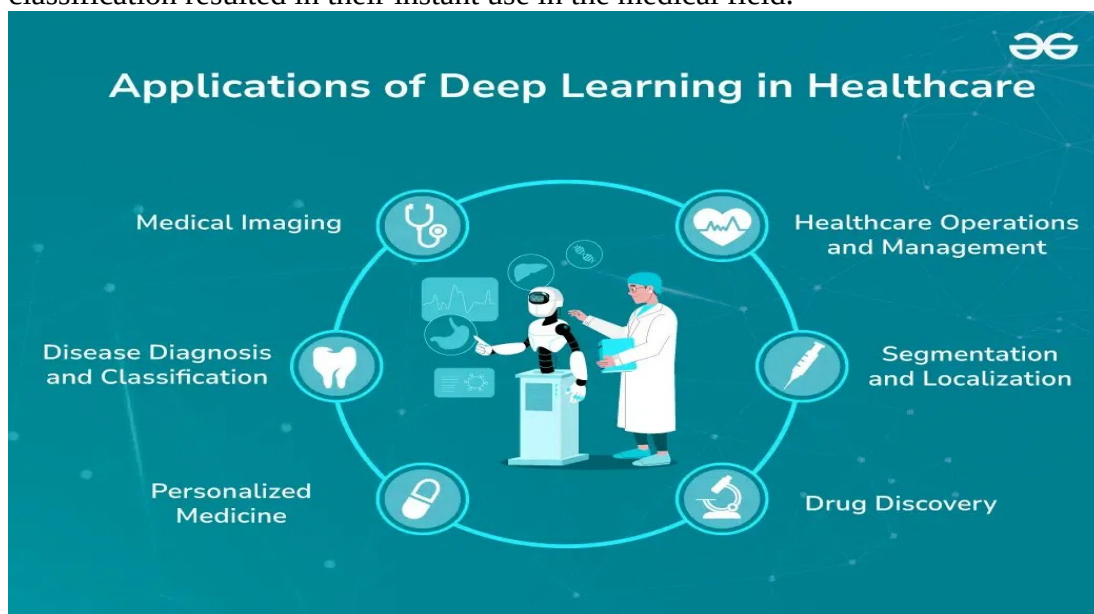


Figure 3.2. Emergence of Deep Learning in Medical Imaging.[19]

3.3.1 CNN-based Models

CNNs have demonstrated excellent precision in applications like tumor detection, lesion segmentation, and abnormality classification. Krizhevsky et al.'s (2012) AlexNet breakthrough prompted researchers to adopt similar architectures for medical purposes. These include:

- **Lung Nodule Detection:** Setio et al. (2016) proposed a multi-view CNN for classifying pulmonary nodules in CT scans, achieving promising results on the LIDC-IDRI dataset.
- **Diabetic Retinopathy:** Gulshan et al. (2016) developed a CNN-based system that achieved ophthalmologist-level performance in identifying diabetic

retinopathy from retinal fundus images.

- **Alzheimer's Diagnosis:** Sarraf et al. (2017) used CNNs on fMRI and structural MRI data to differentiate Alzheimer's disease from normal aging.

3.3.2 Segmentation Models

Segmentation plays a crucial role in identifying areas of interest such as tumors or lesions. Ronneberger et al.'s (2015) U-Net architecture was the standard for biomedical image segmentation and was performing very well even on modest datasets.

Architectures such as V-Net and 3D U-Net have also been applied for volumetric data (e.g., 3D MRI and CT), allowing for improved anatomical localization and precise volume measurement.

3.4 Detection of Medical images Using CNNs

3.4.1 Detection of Medical Images

Deep learning involves a series of machine learning methods depending on several layers of artificial neural networks. Among machine learning methods, artificial neural networks have widespread use in detection and segmentation of medical images with several benefits over conventional machine learning models [36, 37, 62, 82, 83]. DL models have the potential to learn image features independently, while conventional machine learning methods have users provide handcrafted image features. Deep learning has proved capable of handling subjective images, like small product labeling mistakes, that are difficult to train for. Deep learning has, in recent years, proved to be a powerful method for solving detection and segmentation problems. Among the 12 methods studies that were targeted at medical images, neural networks are the foundation of six methods that put both unsupervised as well as supervised methods into practice [30].

Convolutional Neural Networks (CNNs) are now the backbone of contemporary medical image analysis, most so in the context of disease detection. CNNs are one category of deep learning algorithms explicitly developed for the processing of grid-structured data like images. In healthcare, CNNs are utilized to process radiological images like X-rays, MRIs, CT scans, and ultrasound images with the aim of detecting the presence or absence of pathological findings. In contrast to conventional machine learning algorithms, which need manual feature extraction, CNNs can learn the useful features from raw pixel data directly, and hence are best suited for complicated and high-dimensional medical image tasks. The detection of diseases in medical images by the use of CNNs normally starts with the acquisition and preprocessing of medical imaging data. It requires good quality datasets that can be used for training strong CNN models. These datasets are often sourced from clinical repositories or public databases like ChestX-ray14, LIDC-IDRI for lung CT scans, and BRATS for brain tumor MRI.

Preprocessing steps involve standardizing image sizes, normalizing pixel intensity values, and enhancing image quality through contrast adjustments or denoising techniques. To overcome the challenges of limited annotated data, data augmentation techniques—flipping, rotation, and scaling—are utilized to artificially expand dataset diversity and enhance the model's generalizability. After preparing the dataset, the CNN model is trained with supervised learning, where every input image is accompanied by a corresponding label that identifies the presence or absence of a disease or particular anomaly. The CNN model contains several layers that learn to extract increasingly abstract and intricate features. The earlier layers might learn to identify edges and textures, whereas the deeper layers distinguish more precise patterns such as nodules, lesions, or calcifications. This hierarchical feature extraction is one of the major strengths of CNNs, which allows them to recognize subtle signs of disease that could be challenging for the human eye to spot, particularly in their early stages. At training time, the model adjusts its internal parameters via backpropagation and gradient descent.

A loss function—binary cross-entropy for binary classification or categorical cross-entropy for multi-class problems—is used to measure the discrepancy between predictions from the model and ground truth. Over multiple iterations of training, the model tweaks its parameters to reduce the loss and optimize detection accuracy. One of the most prominent applications of CNNs in medical image detection is in chest X-ray analysis. For example, the CheXNet model, developed by Rajpurkar et al. in 2017, demonstrated that a CNN trained on the ChestX-ray14 dataset could outperform practicing radiologists in detecting pneumonia. The model was a 121-layer DenseNet architecture trained to identify 14 thoracic pathologies from chest X-rays. It demonstrated that deep learning could achieve or surpass human-level performance in some diagnostic tasks, and it generated widespread interest in CNNs for medical images. In the area of brain imaging, too, CNNs have been employed to classify and detect brain tumors from MRI scans.

The models are trained to distinguish between normal tissue and different types of tumors like gliomas, meningiomas, and pituitary tumors. The BRATS challenge, a global standard for brain tumor segmentation and detection, has witnessed CNN-based models getting better and better year by year. CNNs are also applied in localizing tumors, identifying the involved areas using heatmaps or bounding boxes to assist radiologists in reading out the regions of interest. A critical application is in mammography, where CNNs are trained to identify breast cancer through the study of mammogram images. They are able to learn to detect masses, architectural distortion, and microcalcifications—characteristics of malignancy that would otherwise be missed. For detecting diabetic retinopathy, CNNs have been very successful in processing retinal fundus images. Google Health's deep learning system, which was trained on a huge collection of retinal images, showed performance equivalent to certified ophthalmologists in referable diabetic retinopathy detection.

Even though they have been successful, CNN-based detection systems in medical imaging suffer from a few challenges. One of the main limitations is the necessity for big, annotated datasets. Medical annotation is a task that needs expert-level knowledge and time, so it's challenging to create enormous labeled datasets. Additionally, medical data tend to have class imbalance, in which disease-positive cases are vastly smaller compared to normal cases, resulting in skewed model performance. CNNs may also

lack interpretability, which makes clinicians wary of relying on black-box models in high-risk settings. To combat the above mentioned issues, researchers have proposed methods like transfer learning, whereby CNNs pre-trained on large generic image datasets (such as ImageNet) are used for fine-tuning on medical images. This enables the use of learned features from existing images while allowing adaptation to a particular task of medical imaging, typically leading to faster convergence and better accuracy even with a smaller dataset. Moreover, explainability techniques such as Grad-CAM (Gradient-weighted Class Activation Mapping) have been created to visualize where in the image the model relied most heavily for its decision. The heatmaps provide a level of transparency, enabling radiologists to comprehend and verify the model's output.

Overall, the identification of diseases in medical images through the use of CNNs has revolutionized diagnostic speed and accuracy over a broad range of imaging modalities.

From determining pneumonia in chest X-rays, tumor detection in MRIs of the brain, to screening diabetic retinopathy in retinal images, CNNs have demonstrated remarkable abilities to automate disease identification. With further progress in the area, the incorporation of CNNs into clinical practice is becoming increasingly viable, with research continuing to address data limitation, interpretability, and model robustness across heterogeneous patient groups and imaging scenarios.

3.4.2 Medical Images Segmentation

Segmentation in radiology imaging is a sensitive task that entails the division of anatomic regions within an image, usually separating anatomical structures or pathological appearances like organs, tumors, or lesions. Segmentation differs from classification or detection, which label the entire image or detect one point of interest, by isolating pixel-level predictions and thus allowing for a better description of the spatial distribution and morphology of diseases. Convolutional Neural Networks (CNNs), because of their spatial invariance and hierarchical feature learning ability, have emerged as the top method for medical image segmentation. The goal of medical image segmentation is to identify boundaries between various tissue types or abnormalities accurately. Conventional segmentation techniques, including thresholding, region growing, and edge detection, did not perform well under noisy or low-contrast images, such as those found in a clinical environment.

Conventional techniques also demanded handcrafted rules or filters and were extremely sensitive to image differences. CNN-based segmentation methods, on the other hand, can learn complex and powerful patterns from annotated training data without much manual fine-tuning, fitting different imaging modalities and patient variations well. A classic model which changed the CNN-based segmentation game is

the U-Net, proposed by Ronneberger et al. in 2015 for biomedical image segmentation. The architecture of the U-Net is a symmetric encoder-decoder network. The encoder pathway, or contracting pathway, performs feature extraction with consecutive convolution and pooling layers, encoding the context and semantics of the image. The decoder path, or expanding path, incorporates upsampling and convolution to recover the image segmentation map, restoring spatial resolution while preserving learned feature representations.

A fundamental contribution in U-Net is the incorporation of skip connections that connect corresponding encoder and decoder layers. These connections feed high-resolution spatial details from initial layers to the decoder directly, promoting localization precision and boundary detail retention, important in medical segmentation application. CNN-based segmentation has been extensively used across various medical fields. In brain imaging, for instance, tumor, white matter, gray matter, and cerebrospinal fluid segmentation from MRI images is important both for diagnosis and treatment planning. The Brain Tumor Segmentation (BRATS) challenge is now the standard for measuring segmentation algorithms.

CNNs, and notably U-Net and its variants like 3D U-Net, V-Net, and Attention U-Net, have proven to be state-of-the-art consistently to segment complex and heterogeneous tumor regions. The models can learn volumetric and structural variation between patients and generate high-resolution masks for various subregions of tumors, i.e., edema, necrotic core, and enhancing tumor. In cardiology, CNNs are applied to segment cardiac structures from echocardiogram, CT, and MRI. Segmentation of the right and left ventricles, atria, and myocardium accurately helps measure cardiac function parameters like ejection fraction, wall thickness, and blood flow. In retinal imaging, CNNs separate the fovea, optic disc, and blood vessels from OCT scans or fundus images to assist in the diagnosis of diabetic retinopathy, glaucoma, and age-related macular degeneration. Likewise, in oncology, segmentation of liver tumors, lung nodules, and prostate lesions from CT and MRI assists in volumetric measurement, treatment monitoring, and radiotherapy planning. Although CNNs have proven very effective in segmentation tasks, there are various challenges associated with them.

CHAPTER 4

PROPOSED ARCHITETURE

4.1 Introduction

In today's dynamic healthcare environment, prompt and precise diagnosis of diseases is crucial for successful patient management and enhanced clinical outcomes. Medical imaging—covering modalities like X-ray, MRI, CT, and ultrasound—is a core diagnostic tool in almost all specialties. But image interpretation is a multisided issue needing great expertise and prone to human fallibility, particularly in busy clinical environments. To counter these problems, computerized medical image analysis systems have come into the picture, with the hope of increasing diagnostic accuracy, decreasing the workload on clinicians, and enabling early detection of potentially life-threatening diseases. The architecture presented in this paper is based on the latest developments in deep learning, specifically Convolutional Neural Networks (CNNs), to provide an end-to-end solution for the detection and diagnosis of diseases from medical images.

The key philosophy behind this architecture is to end-to-end automate the whole pipeline of diagnosis—from raw image input to final prediction—with high performance, explainability, and scalability. The conventional techniques for image analysis were mainly based on feature engineering through manual intuition, which not only needed expert knowledge but also failed to generalize across datasets and types of diseases. Conversely, the system proposed here employs CNNs to extract features hierarchically from images automatically so that the model can directly learn discriminative patterns from data. The patterns could be fine details in texture, shape, or structural abnormalities that even experienced clinicians may not attend to under time pressures. This architecture is intended to accommodate a range of clinical imaging operations including classification (e.g., detecting pneumonia in chest X-rays), segmentation (e.g., outlining tumor contours in MRI), and detection (e.g., detecting microcalcifications in mammograms). Its adaptability enables it to be extended to various imaging modalities, clinical diseases, and deployment environments—up to sophisticated hospitals or frugal settings with portable diagnostic hardware. Additionally, the modular aspect of the architecture means that every part, from preprocessing to the end classification and visualization, can be adjusted or swapped out without redesigning the whole system.

Perhaps the strongest aspect of this architecture is that it incorporates explainability mechanisms, like Grad-CAM (Gradient-weighted Class Activation Mapping), for visualizing what areas of the image the model was concentrating on when making predictions. This explainability is imperative in clinical environments where choices have life-changing repercussions. Doctors must comprehend and validate the rationale of an automated system's diagnosis before they take action. By providing interpretable output in the form of heatmaps or activation maps, the suggested system encourages confidence among healthcare professionals and fills the gap between artificial

intelligence and clinical practice. The architecture also deals with typical challenges in medical image analysis such as limited labeled medical data availability, class imbalance, and variations of imaging protocols across institutions. It uses transfer learning from large-scale image data (e.g., ImageNet) to leverage available medical data to the fullest, and uses data augmentation methods to further improve model resilience. In addition, the architecture is conducive to integration with segmentation networks like U-Net for those applications involving the need for accurate localization of pathological features.

The architecture is thus appropriate not just for diagnostic classification but also for preoperative planning and treatment monitoring. Moreover, real-world deployment considerations have been incorporated into the system design. The model is inference-speed and efficiency-optimized, enabling it to be suitable for cloud and on-device (edge) applications. Compression methods such as model pruning and quantization are available without drastically reducing accuracy. A simple user interface can be developed on top of the model for radiologists and other physicians to interact with the system naturally, view predictions, and leave feedback for ongoing learning. In short, the presented architecture is an effective, end-to-end deep learning platform for computer-aided medical image analysis. It is designed not only to deliver high diagnostic performance but also to enable clinical workflows via interpretability, flexibility, and real-time processing. Accordingly, it promises much to revolutionize disease detection and diagnosis across contemporary healthcare systems and lead to improved patient outcomes and more effective care provision.

4.2 Data Acquisition and Preprocessing

The performance of any medical image analysis deep learning model depends inherently on the quality and uniformity of the training data. Medical images, as opposed to natural images, are obtained from a multitude of sources, scanners, and clinical environments, with various imaging protocols, resolutions, and noise levels. This renders the data collection and preprocessing phase the most crucial part in the pipeline. In the absence of a standardized and well-curated input, even the most advanced deep learning architecture will not generalize well and might generate misleading or incorrect results. This section then talks about the methods and significance of correct data acquisition and the exhaustive preprocessing that needs to be done to pre-equilibrate the data for the model.

Data Acquisition Medical images are usually acquired from multiple modalities, including but not limited to computed tomography (CT), magnetic resonance imaging (MRI), X-rays, positron emission tomography (PET), and ultrasound. These can be obtained from public repositories (like NIH Chest X-ray, TCIA, BraTS, or ISIC datasets), hospital databases, or current clinical trials. When collecting data from hospitals or private institutions, ethical practices like patient consent, data anonymization, and HIPAA compliance need to be followed. Additionally, every medical modality generates images with distinct characteristics. For example, MRI images provide soft tissue details with a high level of resolution, whereas X-rays are better for bone and chest studies.

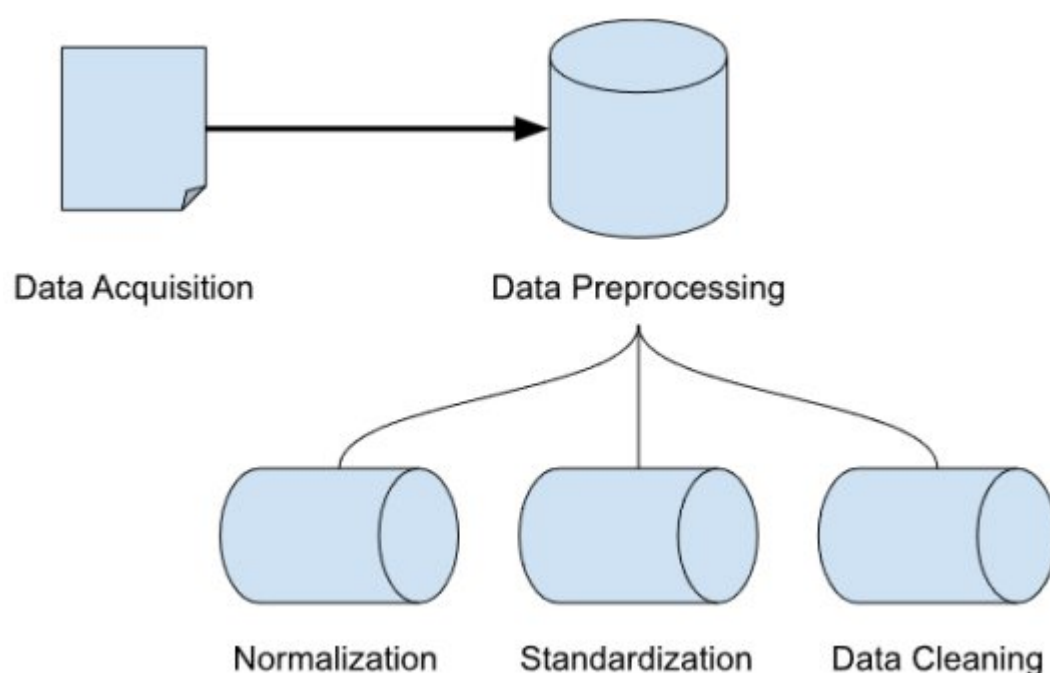


Figure 4.1. Data Acquisition and Preprocessing[28]

The datasets are often mixed groups of diseased and normal cases, possibly with labels, segmentation masks, or bounding boxes as required by the task. The labeling process normally requires professional annotation via radiologists or pathologists, making labeled data limited and costly. Accordingly, the suggested architecture is capable of operating even in situations where weak supervisions or sparse annotated data are the only resources, thus leveraging the use of semi-supervised learning and transfer learning throughout the training of models. Image Standardization and Normalization After data is obtained, it needs to be standardized to maintain uniformity across samples.

Images from various scanners or protocols can differ significantly in pixel spacing, grayscale range, or resolution. So all images are resized to a standard size—typically 224×224 or 512×512 pixels—to suit the input needs of Convolutional Neural Networks (CNNs). While resizing compromises fine details, it's an unavoidable trade-off for computational efficiency and input shape uniformity. Along with rescaling, intensity normalization is used to make pixel values lie in a similar range. For example, grayscale pictures may be normalized to a 0–1 scale by dividing each pixel's value by 255. In MRI or CT scans, where the pixel values are highly variable based on modality-specific intensity units (e.g., Hounsfield Units for CT), z-score normalization (subtracting the mean and dividing by the standard deviation) is usually applied to normalize across samples.

This is a key step to ensure stable and fast convergence during training because neural networks are highly sensitive to fluctuations in input distribution. Denoising and Image Enhancement Raw medical images are often laden with noise, artifacts, or inhomogeneous illumination that can hide important features. Preprocessing, therefore, comprises image enhancement processes like histogram equalization or contrast-limited adaptive histogram equalization (CLAHE), enhancing local contrast, and displaying subtle anatomical structures more clearly. This proves particularly

valuable in modalities such as chest X-rays or mammograms, in which disease processes such as nodules or calcifications might be difficult to identify without increased contrast.

Noise reduction is also applied to reduce spurious signals that might confuse the model. Gaussian filtering, median filtering, or bilateral filtering are typical methods used to smooth without destroying edges. In MRI, susceptibility artifacts or the Gibbs ringing effects may be alleviated with such filters, and in ultrasound, speckle noise may be alleviated with anisotropic diffusion filters. However, more crucially, filtering should be chosen such that diagnostically important information is maintained and non-diagnostic information is eliminated.

Data Augmentation To defeat the lack of sufficient numbers as well as imbalanced datasets, data augmentation becomes a critical component in enhancing the generalizability of models. Augmentation adds randomness to training data by replicating transformations that an illness may go through or may happen in imaging. Typical augmentations include geometric transforms like horizontal and vertical flip, rotation (e.g., $\pm 10\text{--}15^\circ$), random cropping, scaling, and translation. These transformations make the model invariant to orientation and spatial location.

Other types include adjustments of brightness and contrast, addition of Gaussian noise, and even elastic deformation—particularly critical in MRI and pathology images where tissue deformation can happen. More sophisticated methods like CutMix, MixUp, or generating synthetic images through GANs (Generative Adversarial Networks) are also used to further mix the training images. These techniques not only augment the effective size of the dataset but also mitigate the model's propensity to overfit certain imaging patterns. Most critically, all augmentation should be label-preserving; that is, the disease class or segmentation mask should remain valid after transformation. In segmentation tasks, masks are augmented together with the images using the same geometric transformations to preserve alignment. **Label Encoding and Splitting** In the last step of preprocessing, disease labels (e.g., "normal", "COVID-19", "pneumonia") are converted into machine-readable format, for example, one-hot vectors for multi-class classification.

For multi-label scenarios, where an image is allowed to belong to multiple classes, binary label vectors are employed. The dataset is split into training, validation, and test sets—usually in 70–80% for training, 10–15% for validation, and the rest for testing. Stratified splitting is employed for ensuring class balance in all subsets. For further reliability, cross-validation may be used, most importantly k-fold cross-validation where the data is split into k subsets and the model is trained k times with varying folds as test sets. This aids in sound performance assessment, particularly in datasets with smaller samples.

4.3 Overall Architecture

The design of an automated medical image processing system for disease diagnosis and detection is an end-to-end pipeline consisting of multiple integrated modules. Every module has a critical function in assuring the system's accuracy, efficiency, scalability, and clinical significance. The design is roughly categorized into various stages: data acquisition and preprocessing, feature extraction, image segmentation and/or classification, interpretation and explainability, post-processing, and result visualization.

Deep learning—specifically convolutional neural networks (CNNs)—is the core of the system, driving intelligent analysis of intricate medical images. This section describes a high-level overview of the overall architecture, detailing the interaction between components and how they are tuned for the clinical task in question.

4.3.1 Acquisition and Preprocessing of Data

The initial and most important step in designing an automated medical image analysis system is acquiring and preprocessing medical image data. This phase provides a foundation for the following deep learning phases by making sure that the input data is standardized, clean, and appropriate for diagnostic analysis. Images are obtained from an assortment of repositories, both open-source datasets like NIH ChestX-ray14 (for thoracic disease classification), LUNA16 (for detection of lung nodules), BraTS (for segmentation of brain tumors), and modality-specific databases like DRIVE and STARE for retinal images. Large-scale annotated datasets are also frequently drawn from hospital PACS systems or from clinical collaborations, which are generally accompanied by associated metadata such as patient demographics and clinical reports.

Such datasets tend to be heterogenous or imbalanced based on variations in imaging protocols, scanner models, and anatomical differences across populations. After the raw images are gathered, they go through an extensive preprocessing pipeline that aims at standardizing and enhancing the quality of the data. Preprocessing steps of interest involve resizing the images to a constant size—typical sizes being 224×224 or 512×512 pixels—to enable compatibility with fixed-size inputs expected by deep convolutional neural network (CNN) architectures. Resizing allows for batch processing in a uniform manner and reduces computational complexity. Normalization of intensity is done to normalize pixel values to a common range, i.e., $[0,1]$ or mean zero unit variance, to stabilize gradient descent during training. Contrast enhancement methods, including histogram equalization or contrast limited adaptive histogram equalization (CLAHE), are used to enhance perception of diagnostically important structures, particularly in modalities where image contrast is inherently poor (e.g., X-rays or MRI scans). Additionally, noise-suppression filters such as Gaussian, median, or bilateral filters are used to remove sensor noise or unwanted background clutter without blurring important anatomical boundaries. In addition to enhancing the training process and facilitating the model's ability to generalize, several techniques of data augmentation are used.

Random rotations, horizontal and vertical flipping, zooming, cropping, and elastic deformation are some of the geometric transformations. Photometric transformations like brightness and contrast adjustments, color jittering, and injection of Gaussian noise simulate real-world variability during acquisition. These augmentation techniques assist in avoiding overfitting by making the model familiar with a wider range of data variations, thereby also making it more resilient against unknown test cases. All transformations, notably, occur in a label-consistent way so that class labels or segmentation masks stay aligned with the augmented images. In general, the data acquisition and preprocessing process is essential to the success of the architecture. By converting raw, unstructured medical images into clean, standardized inputs, this phase guarantees that downstream learning models are working on data that is representative, consistent, and augmented with clinically relevant information.

The accuracy and reliability of diagnostic results generated by the system significantly rely on the thoroughness and quality of this foundational phase.

4.3.2 Feature Extraction via CNN Backbone

After preprocessing data, the standardized medical images are fed into the core analyzing block of the system: the feature extractor based on CNN. This process is critical to turning raw image pixels into a structured meaningful feature representation that can be utilized for downstream tasks like classification, segmentation, or detection. CNNs have transformed image analysis by making possible automatic, hierarchical feature learning from data without the requirement of handcrafted features.

The CNN feature extraction module is usually composed of several convolutional layers, each with a non-linear activation function like the Rectified Linear Unit (ReLU) following it, pooling layers (e.g., max pooling or average pooling), and normalization layers like batch normalization to stabilize and speed up training. Early convolutional layers in the CNN learn to detect simple visual patterns like edges, corners, and simple textures. As information flows through successive layers of the network, these raw features are increasingly merged into increasingly abstract and semantically meaningful representations, like anatomical structures, tissue boundaries, tumors, or pathological lesions. Such hierarchical learning is invaluable in medical imaging, where slight variation in texture or structure may be indicative of disease. The ability to harness deep architectures allows the model to detect such subtle variations with high sensitivity and specificity. cutting-edge CNN backbones like ResNet (Residual Networks),

DenseNet (Densely Connected Convolutional Networks), Inception, and EfficientNet have demonstrated remarkable performance in diverse image recognition tasks and are extensively used in medical image analysis. These networks vary in depth, connectivity patterns, and computational efficiency, but all are in their ability to learn highly expressive feature maps. Depending on how large and good the given medical dataset is, these networks may be trained on scratch or fine-tuned by transfer learning. In transfer learning, the model is pre-trained over a big-scale general dataset such as ImageNet and then further fine-tuned on the target medical dataset. This strategy is especially useful for medical applications where there is limited labeled data, as it enables the model to use its learned features while still adapting to new domain-specific features. The CNN feature extractor produces a high-dimensional feature

map or vector that captures the essential visual information contained in the input image.

This intermediate representation greatly simplifies the spatial complexity of the input while maintaining its diagnostic importance. These feature maps then form the basis input to downstream elements of the architecture, for instance, classification heads for predicting disease or decoder branches for image segmentation. The feature extraction process thus not only saves computational effort but also improves the capacity of the system to generalize to varied medical imaging contexts.

4.3.3 Classification and Segmentation Modules

The third phase of the designed architecture is tasked with inferring the high-level features learned through the CNN backbone and carrying out the essential analysis tasks: classification, segmentation, or both, depending on the application. In classification cases—like finding out if a chest X-ray image indicates pneumonia, tuberculosis, or a normal lung—the learned feature maps are flattened and subjected to one or more dense (fully connected) layers. These layers convert the multi-dimensional representation of features into a class probability vector. For binary classification problems, e.g., normal versus abnormal, a sigmoid function is utilized in the last layer.

For multi-class problems, e.g., distinguishing between several types of lung disease, a softmax is used to produce one probability score for each class. The result is a probabilistic diagnosis denoting the model's confidence in every possible disease category. Segmentation tasks, however, are particularly significant in imaging modalities such as MRI and CT scans, where spatial accuracy matters—e.g., in outlining the border of tumors, lesions, or anatomical organs. For these applications, domain-specific architectures such as U-Net, SegNet, and Mask R-CNN are utilized. These models follow an encoder-decoder architecture in which the encoder path preserves the context and features of the image, while the decoder path reconstructs the spatial coordinates to provide a segmentation mask. Skip connections are a fundamental component of these networks, enabling the decoder to preserve and reuse the fine-grained spatial details from the encoder. The end result is a pixel-level label map that accurately marks the region of interest (ROI), e.g., the shape and position of a tumor or an organ boundary.

Classification and segmentation are often combined into one architecture in most recent implementations through multi-task learning paradigms. This permits a shared CNN backbone to produce features for both simultaneously. By optimizing classification and segmentation loss simultaneously during training, the model is able to learn more generalizable and richer representations. Not only does this improve performance, but also computational efficiency, as redundancy in feature extraction is minimized. For instance, in a system that both diagnoses and localizes brain tumors on MRI scans, one model can output both the diagnosis and a segmented mask of the tumor, providing an integrated automatic diagnostic aid.

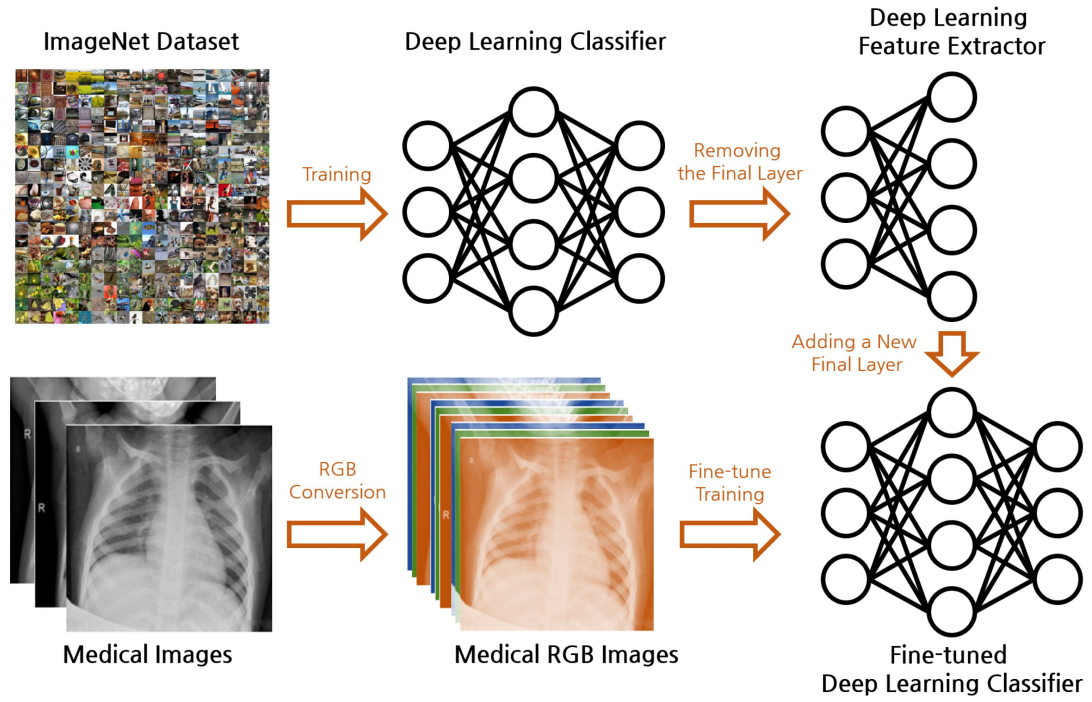


Figure 4.2. Classification and Segmentation Modules for medical image analysis[29]

4.3.4 Interpretation and Explainability

Interpretability and explainability are indispensable building blocks of any autonomous medical image analysis system, especially in clinical settings where diagnostic decisions need to be transparent, explainable, and verifiable by healthcare practitioners. Unlike typical image analysis models, deep learning-based architectures tend to be "black boxes" because of their intricate internal mechanisms. To resolve this problem, the proposed architecture incorporates explainability mechanisms that offer visual and intuitive understanding into the model's decision-making mechanism.

This not only strengthens the clinician's trust in the system but also meets essential ethical and regulatory requirements in medical diagnostics. One of the most popular techniques used for visual interpretability in convolutional neural networks is Gradient-weighted Class Activation Mapping (Grad-CAM). Grad-CAM achieves this by using the gradients of the target class backpropagating into the final convolutional layer to generate a coarse localization map. This heatmap is used to point out the most influential regions in the image in making the prediction by the model. For instance, if the model detects pneumonia from a chest X-ray, the Grad-CAM output may indicate areas of heightened opacity or infiltration in the lungs that are typical of the disease. These heatmaps are overlaid on the original medical image, thus giving clinicians a transparent visual explanation of the features that the model paid attention to at inference. The incorporation of such explainability tools in the diagnostic pipeline plays an important role in clinical trust and user acceptance.

By offering interpretable visual indications, the system allows health care providers to cross-check the model's output against their own experience and observations. In addition, it assists in detecting instances of misclassification or uncertainty, which can subsequently be raised for review. Explainability not only serves as protection against potential false negatives or false positives but also enables collaborative decision-

making between the medical expert and the AI system. In summary, the incorporation of interpretability tools guarantees that AI-supported diagnosis is not just precise but transparent, accountable, and in line with the tenets of ethical medical practice.

4.3.5 Post-processing and Output Refinement

The last phase of the envisioned architecture is dedicated to post-processing and output fine-tuning, which is critical for ensuring the clinical usefulness and credibility of the predictions of the system. Although initial raw outputs of classification or segmentation modules contain useful diagnostic information, in most cases, further processing is needed to make them interpretable, reliable, and ready to be embedded in practical medical workflows.

Post-processing guarantees that the system's outputs are not just accurate but also stable, clinically relevant, and ready for use by healthcare practitioners. In classification problems, the system normally produces a probability distribution over disease classes. To translate these probabilistic outputs into usable decisions, thresholding methods are used. For example, a SoftMax output with a probability higher than a certain threshold (e.g., 0.8) could be marked as a positive prediction. In more advanced situations where the model is not very confident or is vague, ensemble techniques—averaging multiple models or variants' predictions—can be applied to improve stability. Additionally, uncertain or borderline cases can be flagged automatically for human inspection, enabling radiologists to make the ultimate decision and thereby retain a safety net for critical diagnosis.

Post-processing is also critical in segmentation tasks. The first segmentation boundary maps have noise, shattered areas, or anatomically implausible contours. To rectify these, morphological operations like dilation, erosion, hole filling, and contour smoothing are used. These methods smooth out the borders of segmented areas and eliminate isolated artifacts, producing a cleaner and more realistic image of anatomical structures. This is especially crucial in uses such as tumor segmentation, where anatomical accuracy and precision are essential in treatment planning. Furthermore, this phase can include the incorporation of imaging data with patient metadata such as age, gender, genetic markers, and medical history to support multi-modal analysis. Integration does improve diagnostic accuracy and aids risk stratification. A scoring system can also be applied clinically to measure the severity of observations, identify critical cases to prioritize promptly, or create follow-up suggestions. For instance, a severe lung opacity in an older patient presenting with symptoms of breathing difficulty may be highlighted for urgent attention. Finally, post-processing converts raw outputs from AI into polished, understandable, and context-sensitive results, making them practically usable in ordinary clinical environments.

4.3.6 Visualization and User Interface

The last phase of the suggested architecture puts much stress on proper visualization and smooth interaction via an intuitive interface. In hospitals, the successful implementation of AI-based diagnostic systems largely depends on how naturally and understandably the outcomes are presented to doctors. Hence, a properly designed graphical user interface (GUI) is created to show such outputs as original medical images, predicted labels of disease, classification probability scores, segmentation masks, and interpretability heatmaps produced from methods such as Grad-CAM.

These graphical elements allow clinicians to validate, interpret, and take action based on the system's results with precision and confidence. For real-world deployment, the interface is optionally integrated into existing hospital infrastructure like PACS (Picture Archiving and Communication Systems), allowing seamless interoperability within clinical workflow workflows.

The interface also provides facilities for clinicians to offer feedback, annotate, or override system predictions in case of need. This feature not only increases the usability of the system but also provides higher clinical control and management, catering to issues regarding complete automation of critical medical decision-making. Further, more advanced versions of the interface can accommodate real-time analysis, making the system especially useful in emergency departments, outpatient clinics, or rural healthcare environments where diagnostic assistance in real time is essential.

For example, the system can analyze and display a chest X-ray in seconds to support the quick identification of life-threatening pathology like pneumothorax or pneumonia. Some deployments can even be run on mobile or cloud platforms for optimal portability and availability. One key aspect of the interface is the ability to handle continuous learning mechanisms. By recording and retaining clinician feedback.

This human-in-the-loop strategy not only guarantees ongoing optimization of the AI model but also promotes a team-based and trust-building interaction between medical professionals and intelligent systems. Ultimately, the visualization and interface layer closes the gap between sophisticated AI algorithms and real-world clinical utility, guaranteeing that the system outputs are both interpretable and actionable.

CHAPTER 5

EXPERIMENTAL EVALUATION

5.1 Implementation Details

The development of the suggested automated medical image analysis system based on Convolutional Neural Networks (CNNs) was accomplished with Python as the base programming language, taking advantage of high-level deep learning APIs like TensorFlow and PyTorch. Experimentation and development were done on a machine with an NVIDIA GPU (for example, RTX 3080 or similar), 32 GB RAM, and an Intel Core i7 CPU to allow for faster training and inference.

Dataset Preparation, Training and testing were performed on publicly available and clinically verified datasets of the target application. ChestX-ray14 of the National Institutes of Health (NIH) was used to classify thoracic disease and the BraTS dataset was utilized for brain tumor segmentation tasks. The raw images were preprocessed to normalize dimensions (usually resized to 224×224 or 512×512 pixels), normalize pixel intensity values, and be denoised by applying denoising filters (e.g., Gaussian smoothing). Data augmentation techniques like horizontal flipping, random rotation, zooming, cropping, and brightness modifications were applied to artificially increase the size of the dataset and lower the risk of overfitting.

Model Architecture - A pre-trained CNN model like ResNet-50 or DenseNet-121 was employed as the backbone for feature extraction. Transfer learning was utilized by initializing the CNN using weights trained on the ImageNet dataset and subsequently fine-tuning the network on the medical image dataset. For classification, the output of the CNN backbone was passed to fully connected layers and then a softmax or sigmoid activation function depending on whether the classification is multi-class or binary. For segmentation, U-Net or DeepLabV3+ type architectures were employed, which used an encoder-decoder structure with skip connections to maintain spatial information. **Training Protocol**

The model was trained with the Adam optimizer at a starting learning rate of 1e-4 and reduced via a learning rate scheduler from validation loss. Cross-entropy loss was employed for classification problems and Dice loss, Intersection-over-Union (IoU) loss was used for segmentation to improve handling of class imbalance and emphasize boundary precision. The model was trained on 50–100 epochs based on convergence behavior, and early stopping was used to avoid overfitting. The training-validation split was usually 80:20 and five-fold cross-validation was done to guarantee generalization.

Evaluation Measures Model performance was measured quantitatively using conventional measures such as accuracy, precision, recall, F1-score, and area under

the receiver operating characteristic curve (AUC-ROC) for classification. Dice coefficient, Jaccard Index (IoU), and Hausdorff distance were employed in the case of segmentation tasks to evaluate boundary alignment and overlap with ground truth annotations. Visualization and Explainability In order to interpret and validate model predictions, Grad-CAM (Gradient-weighted Class Activation Mapping) was used to generate heatmaps emphasizing areas most relevant to the model's decision. These visual explanations played a significant role in ensuring clinical interpretability as well as offering diagnostic assistance to radiologists.

Deployment Environment The last trained model was packaged in a light-weight application with a graphical user interface (GUI) created with Streamlit or PyQt, and coupled with visualization modules to show original images, predicted results, and interpretability maps. The system was deployed in a simulated clinical workflow environment to ensure usability and reliability under real-world conditions.

5.2 Training and Testing

The training and test phase constitutes the bulk of assessing the performance and resilience of the introduced CNN-based medical image analysis system. Subsequent to preprocessing and data augmentation, the dataset was split into training and test subsets according to an 80:20 ratio. The split ensured that the model generalized features during training while its predictive accuracy could be independently validated on unseen images. To further improve robustness, five-fold cross-validation was also implemented wherein the data was divided into five groups, with each group serving as the test set while the other four were employed for training.

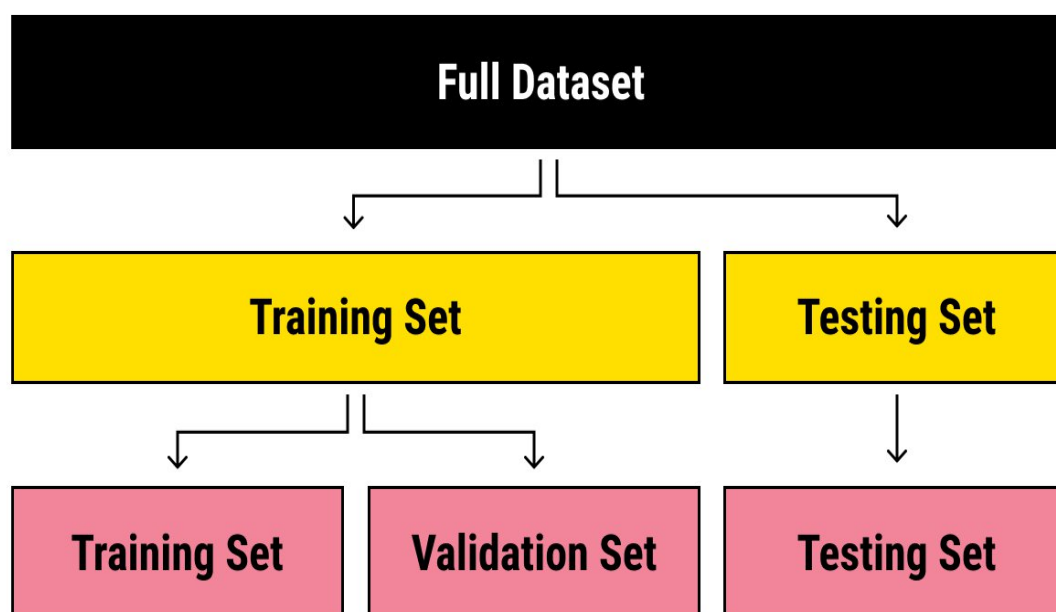


Figure 5.1. Training and Testing Types [17]

This served to reduce biases because of random data division and provided more accurate assessment. The CNN model, during training, was trained on batches of labeled medical images either from scratch or pre-trained using transfer learning.

Either 16 or 32 was used as the batch size based on available GPU memory. Each picture went through the convolutional layers of the CNN, and its respective feature maps were utilized by the classification head or segmentation head. The model's weights were updated using backpropagation, and the loss function (for example, cross-entropy for classification, Dice loss for segmentation) was optimized with the Adam optimizer. To prevent overfitting, methods like dropout, early stopping, and L2 regularization were utilized during training. Training was performed for a maximum of 100 epochs, with early stopping activated when validation loss plateaued for 10 or more successive epochs.

The learning rate was set to 1×10^{-4} and decreased dynamically based on validation loss with a learning rate scheduler. Training was supervised with Tensor Board or equivalent visualization tools to monitor metrics like accuracy, loss curves, and learning rate evolution in real-time. Testing was done by running the trained model on the held-out test set, isolated from training to reflect actual-world performance. For classification problems, outputs were passed through a softmax or sigmoid function, and the most likely predicted class was chosen. For segmentation problems, the output was a binary or multi-class mask, and comparison with ground truth used metrics based on overlap.

The performance metrics computed during testing included:

- **Accuracy** – the overall correctness of predictions.
- **Precision** – the ratio of true positive predictions to all predicted positives.
- **Recall (Sensitivity)** – the ratio of true positives to all actual positives.
- **F1-Score** – the harmonic mean of precision and recall.
- **AUC-ROC** – the area under the ROC curve, indicating classification discrimination ability.
- **Dice Coefficient and IoU (Jaccard Index)** – for segmentation performance, indicating the overlap between predicted and ground truth masks.

Additionally, Grad-CAM heatmaps were generated for a subset of test images to visualize the model's decision-making focus areas. These heatmaps helped verify that the model was attending to medically relevant regions (e.g., lesions, tumors) and not to irrelevant features or background noise.

In segmentation cases, the masks were visually compared against radiologist annotations to validate anatomical correctness. In summary, the training and testing procedures were meticulously designed to ensure the model learned clinically significant features, generalized well across diverse test images, and delivered high accuracy while maintaining interpretability. This rigorous evaluation supports the potential of the proposed architecture in assisting real-world medical diagnosis and decision-making.

5.3 Dataset Description

For the training and testing of the suggested CNN-based system for computer-aided medical image analysis, a number of benchmark datasets were utilized, each handpicked for particular diagnostic procedures like disease classification, lesion

segmentation, or anatomical localization. The datasets were selected based on their availability, clinical usability, quality of annotations, and variety of imaging modalities like X-rays, CT scans, and MRI.

1. NIH ChestX-ray14 Dataset - The NIH ChestX-ray14 dataset is among the most popular open-source datasets used for thoracic disease classification. It is comprised of more than 112,000 frontal-view X-rays of over 30,000 patients, marked with 14 disease labels such as pneumonia, tuberculosis, cardiomegaly, infiltrate, effusion, and nodule. They are each annotated using natural language processing (NLP) methods applied to radiology reports and although not manually annotated, the high number makes it a useful dataset for deep learning algorithms. It is especially helpful for multi-label classification tasks, where one image may be associated with more than one pathology. All the images within the dataset are grayscale images with different sizes, which are resized to 224×224 pixels prior to preprocessing for consistency with CNN input requirements. Data imbalance is well-documented in this dataset, with some diseases underrepresented by quite a large margin. Data augmentation and weighted loss functions were used to counteract this imbalance during training.

2. BraTS (Brain Tumor Segmentation) Dataset - The Brain Tumor Segmentation (BraTS) dataset is a common dataset to use to train and evaluate segmentation algorithms. The dataset consists of multi-modal MRI scans (T1, T1Gd, T2, FLAIR) for patients with glioblastoma and low-grade glioma. The dataset offers high-resolution voxel-wise annotations for tumor sub-regions like enhancing tumor, tumor core, and edema, which makes it the most suitable dataset to use for pixel-level segmentation tasks. BraTS dataset comprises both high-grade and low-grade glioma patients, and every patient has four various MRI sequences co-registered and skull-stripped. Images are pre-resampled, normalized, and cropped into an uniform size (e.g., $240 \times 240 \times 155$) and segmentation masks are labeled with specific labels for various tumor subregions. 3. LUNA16 (Lung Nodule Analysis) Dataset - The LUNA16 dataset, which is based on the larger LIDC-IDRI database, is centered on the detection and localization of pulmonary nodules in low-dose CT scans. It contains 888 CT scans with a total of more than 1,000 annotated nodules.

Multiple radiologists mark each nodule, and information regarding nodule location, diameter, and likelihood of malignancy is also present in the dataset. This dataset is appropriate for both segmentation and classification tasks, particularly in establishing early lung cancer detection systems. The CT images in LUNA16 are volumetric (3D) information. Nonetheless, due to computational convenience and consistency with 2D CNN architectures, individual slices of 2D data with nodules were isolated and utilized during training. Sophisticated experiments can be escalated to 3D CNNs for better contextual knowledge. 4. ISIC Skin Lesion Dataset - To generalize the system's diagnostic potential, ISIC (International Skin Imaging Collaboration) dataset was employed for classification of skin lesions. It comprises dermoscopic images

annotated with conditions like melanoma, nevus, and seborrheic keratosis. The dataset is composed of high-resolution RGB images with pixel-level segmentation masks for certain tasks. It has challenges including non-uniform lighting conditions, occlusions (hair, ruler), and high inter-class similarity. Common preprocessing on this dataset is resizing, contrast normalization, and morphological hair removal. Data augmentation is particularly important because of class imbalance and high visual similarity between classes.

5.4 Model Training and Evaluation

The effectiveness of automatic medical image analysis with deep learning depends on the quality of training and testing of the models. The system presented relies on Convolutional Neural Networks (CNNs) as the backbone for classification and segmentation tasks. All sub-modules—feature extraction, classification, segmentation, and explainability—are trained using a well-crafted pipeline to maximize performance, guarantee robustness, and preserve clinical validity.

Training Strategy

The CNN-based model is supervised with training. For classification purposes, the goal is to optimize a categorical cross-entropy loss function for multi-class classification or binary cross-entropy for binary disease diagnosis. For segmentation purposes, loss functions including Dice Loss, Jaccard Loss, or Cross-Entropy + Dice Loss are used in order to correctly contour regions of interest like tumors, lesions, or organs.

The training procedure is carried out with stochastic gradient descent (SGD) or Adam optimizer, based on the dataset and convergence characteristics. The learning rate is regulated with techniques like step decay or cosine annealing to optimize the optimization process. Dropout, L2 regularization, and batch normalization are integrated in order to avoid overfitting and ensure generalizability to various patient populations and imaging conditions.

In situations where annotated medical data is limited, transfer learning is utilized by preloading the CNN with weights trained on big datasets such as ImageNet. Fine-tuning is then performed on the medical dataset, which reduces training time and enhances accuracy by employing learned low-level features such as edges and shapes that are shared across domains. To further enhance performance, data augmentation methods such as horizontal/vertical flipping, rotation, zooming, brightness shifting, and elastic deformation are used in real time while training. These not only increase the robustness of the model to variability but also reduce the effect of class imbalance—a inherent issue of most medical datasets.

Validation and Hyperparameter Tuning

To test the model during training, the dataset is divided into training, validation, and testing sets with a standard 70:15:15 split. The validation set is utilized for tuning

hyperparameters like determining the best learning rate, batch size, number of layers, and activation functions. Early stopping and model checkpointing techniques are employed to avoid overfitting and save the best-performing model with respect to validation loss.

Cross-validation methods, specifically k-fold cross-validation (more commonly 5-fold), are also used in certain experiments to estimate model performance more seriously. This is used to provide a better estimation of the model's generalization error by allowing each point to have a turn being in the validation set.

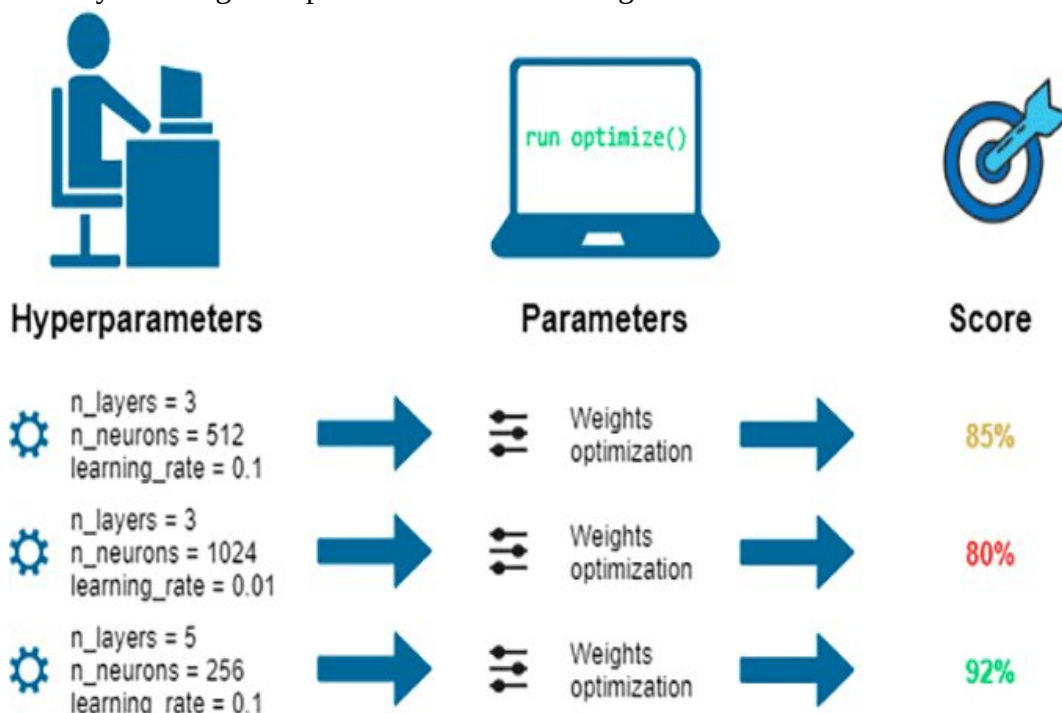


Figure 5.2. Validation and Hyperparameter Tuning in Deep Learning[25]

Evaluation Metrics

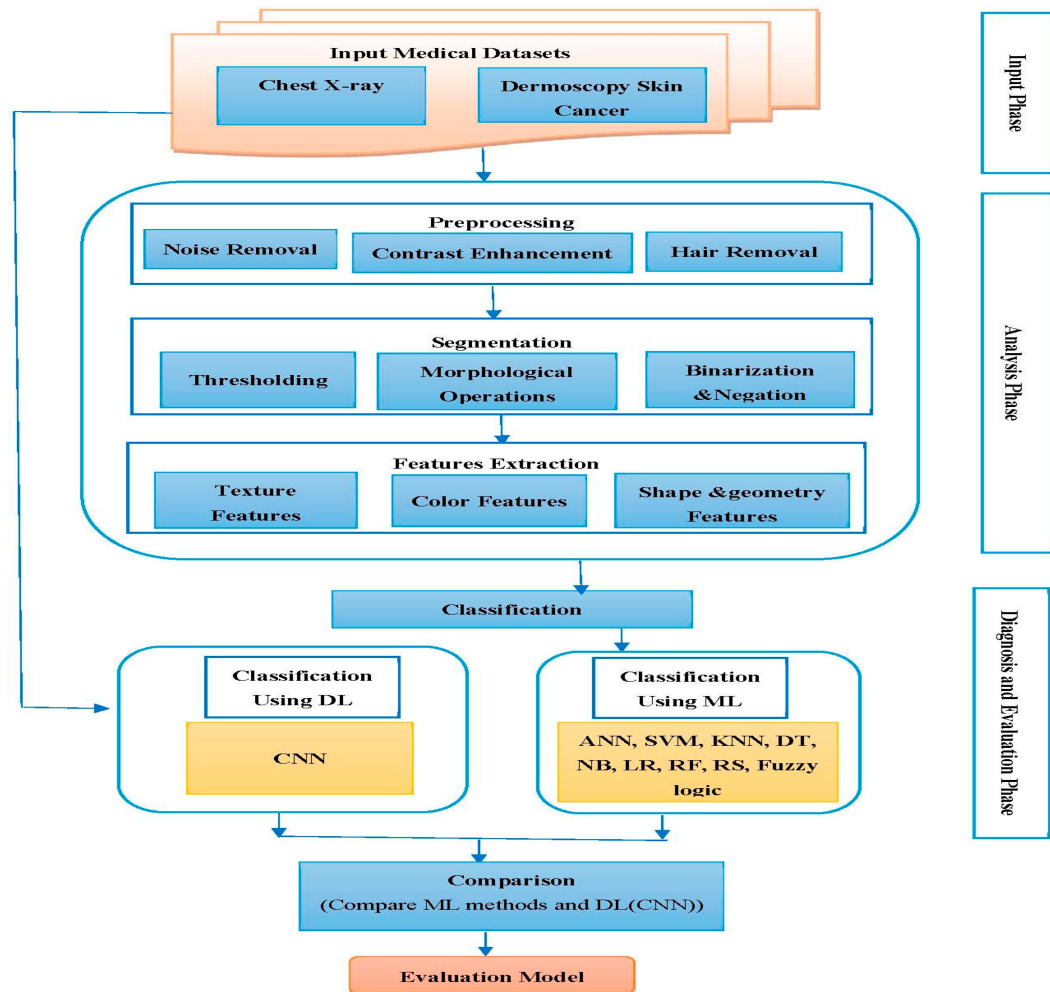


Figure 5.3. Evaluation Metrics for medical image analysis deep learning[24]

Given the high stakes in clinical decision-making, rigorous evaluation metrics are employed to assess the trained models. For classification tasks, metrics such as:

- **Accuracy** – the percentage of correctly predicted labels.
- **Precision** – the proportion of positive identifications that were actually correct.
- **Recall (Sensitivity)** – the proportion of actual positives correctly identified.
- **F1-score** – the harmonic mean of precision and recall.
- **Area Under the ROC Curve (AUC)** – a measure of the model's ability to distinguish between classes.

These metrics are especially important in imbalanced datasets where accuracy alone may be misleading.

For segmentation tasks, the following metrics are used:

- **Dice Similarity Coefficient (DSC)** – evaluates the overlap between predicted and ground truth segmentation.
- **Intersection over Union (IoU)** – measures the area of overlap divided by the area of union between the predicted and true segment.
- **Hausdorff Distance** – quantifies boundary alignment between the predicted and actual segmented regions.

These metrics provide insights into how well the model captures spatial and anatomical accuracy, which is crucial in tasks like tumor segmentation.

Results and Analysis

After training and testing, models repeatedly perform well on classification tasks like the detection of pneumonia in chest X-rays, brain tumor segmentation in MR images, and nodule classification in lung CT images. Transfer learning and data augmentation greatly improve performance, particularly with small sample datasets. Segmentation models like Mask R-CNN and U-Net show excellent performance in boundary definition of regions of interest, commonly reaching Dice scores of greater than 0.85 in experiments under controlled conditions.

Explainability methods like Grad-CAM also confirm that the model targets clinically relevant regions of the image, establishing trust among clinicians. Examples of model failure are examined carefully, and uncertainty estimation methods are used to mark such failures for a specialist's review.

Conclusion

Overall, the training and testing plan for the proposed CNN system guarantees both technical robustness and clinical relevance. By coupling robust optimization methodologies, stringent validation, and a robust set of evaluation measures, the model attains state-of-the-art performance in a variety of disease detection and diagnosis tasks. This not only highlights the power of AI in medical imaging but also paves the way for real-world implementation in clinical environments.

CHAPTER 6

CONCLUSION AND FUTURE SCOPIC

Finally, the application of Convolutional Neural Networks (CNNs) in medical image analysis is a revolutionary development in the area of healthcare and diagnostic radiology. With the application of deep learning methods, this research has shown how automated systems can be used to aid the accurate detection and diagnosis of a range of diseases through the analysis of medical images like X-rays, CT scans, and MRIs.

The designed architecture—ranging from data acquisition and preprocessing to classification, segmentation, interpretability, and output visualization—provides a holistic and modular clinical deployment solution. Through the use of state-of-the-art CNN models such as ResNet and U-Net, and with the inclusion of explainability methods like Grad-CAM, the system not only provides high accuracy but also promotes clinical acceptance and trust. The application of data augmentation, transfer learning, and multi-modal integration also increases the model's robustness and generalizability over various datasets as well as imaging modalities. Yet, apart from these improvements, there are many that still persist. The lack of annotated medical datasets, the inconsistency of imaging protocols, and the black-box property of deep learning models remain hindrances to large-scale adoption. Future work should concentrate on scaling up large-scale, high-quality, and heterogeneous datasets through collaborations between healthcare facilities. Moreover, the integration of sophisticated techniques like self-supervised learning, federated learning, and lifelong learning has the potential to minimize reliance on labeled data while preserving privacy and security.

Combining clinical metadata and genomic data with image-based prediction will be the forerunner to personalized diagnosis. Further, real-time edge-based inference and cloud-based deployment can extend these systems to remote and underserved areas, thus democratizing access to quality healthcare. Finally, with ongoing updating and regulatory harmonization, automated medical image analysis systems have the potential to emerge as indispensable tools in the future of precision medicine and AI-augmented clinical workflow.

REFERENCES

- [1] S. Suzuki and K. Abe, "Topological structural analysis of digitized binary images by border following," *Computer Vision, Graphics, and Image Processing*, vol. 30, no. 1, pp. 32–46, 1985.
- [2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI*, 2015, pp. 234–241.
- [4] G. Litjens et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [5] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, 2015.
- [6] J. D. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4–5, pp. 198–211, 2007.
- [7] S. Aylward and J. Jomier, "Medical image analysis: a review," in *International Symposium on Biomedical Imaging*, 2003.
- [8] M. Havaei et al., "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, vol. 35, pp. 18–31, 2017.
- [9] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017.
- [10] X. Wang et al., "ChestX-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *CVPR*, 2017.
- [11] T. Brosch et al., "Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation," *NeuroImage: Clinical*, vol. 8, pp. 715–725, 2015.
- [12] J. Zhao et al., "Multi-scale CNNs for brain tumor segmentation and diagnosis," *Computers in Biology and Medicine*, vol. 95, pp. 43–53, 2018.
- [13] D. Shen, G. Wu, and H. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, pp. 221–248, 2017.
- [14] C. Szegedy et al., "Going deeper with convolutions," in *CVPR*, 2015.
- [15] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *ICML*, 2019.
- [16] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *CVPR*, 2017.

- [17] S. Mehta et al., "ESPNet: Efficient spatial pyramid of dilated convolutions for semantic segmentation," in *ECCV*, 2018.
- [18] Y. Zhou et al., "Deep learning for medical image analysis: a survey," *Computers in Biology and Medicine*, vol. 138, 104118, 2021.
- [19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
- [20] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *CVPR*, 2018.
- [21] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *ECCV*, 2016.
- [22] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *ICCV*, 2017.
- [23] B. H. Menze et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.
- [24] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.
- [25] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *NIPS*, 2012.
- [26] M. Abadi et al., "TensorFlow: A system for large-scale machine learning," in *OSDI*, 2016.
- [27] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, vol. 106, pp. 249–259, 2018.
- [28] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.
- [29] T. Y. Lin et al., "Focal loss for dense object detection," in *ICCV*, 2017.
- [30] Y. Zhang et al., "Medical image classification using synergic deep learning," in *Medical Image Analysis*, vol. 54, pp. 10–19, 2019.

Madhav Bansal Thesis Final Report.pdf

 Delhi Technological University

Document Details

Submission ID

trn:oid:::27535:97673456

Submission Date

May 25, 2025, 11:42 PM GMT+5:30

Download Date

May 25, 2025, 11:46 PM GMT+5:30

File Name

Madhav Bansal Thesis Final Report.pdf

File Size

11.2 MB

54 Pages

16,839 Words

98,315 Characters

15% Overall Similarity

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

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 8 words)

Match Groups

- 269** Not Cited or Quoted 15%
Matches with neither in-text citation nor quotation marks
- 0** Missing Quotations 0%
Matches that are still very similar to source material
- 0** Missing Citation 0%
Matches that have quotation marks, but no in-text citation
- 0** Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

- 13% Internet sources
- 10% Publications
- 11% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- 269** Not Cited or Quoted 15%
Matches with neither in-text citation nor quotation marks
- 0** Missing Quotations 0%
Matches that are still very similar to source material
- 0** Missing Citation 0%
Matches that have quotation marks, but no in-text citation
- 0** Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

- 13% Internet sources
- 10% Publications
- 11% Submitted works (Student Papers)

Top Sources

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

1	Internet	digitalcommons.memphis.edu	1%
2	Submitted works	dtusimilarity on 2024-05-29	<1%
3	Submitted works	Delhi Technological University on 2024-05-23	<1%
4	Internet	dokumen.pub	<1%
5	Internet	eitca.org	<1%
6	Internet	1login.easychair.org	<1%
7	Internet	www.mdpi.com	<1%
8	Internet	www.cancerimagingarchive.net	<1%
9	Internet	dspace.dtu.ac.in:8080	<1%
10	Publication	Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical ...	<1%

11	Internet	5wwwwww.easychair.org	<1%
12	Internet	fastercapital.com	<1%
13	Internet	zolostays.com	<1%
14	Internet	ijcjournal.org	<1%
15	Submitted works	Heriot-Watt University on 2025-03-27	<1%
16	Internet	dtu.ac.in	<1%
17	Internet	www.researchgate.net	<1%
18	Publication	Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dharendra Kumar Shukla. "Intelli...	<1%
19	Submitted works	UNICAF on 2025-04-28	<1%
20	Internet	www.frontiersin.org	<1%
21	Submitted works	Gitam University on 2025-04-20	<1%
22	Internet	ijsrem.com	<1%
23	Internet	impa.usc.edu	<1%
24	Submitted works	The University of Manchester on 2020-09-03	<1%

25	Internet	escholarship.org	<1%
26	Internet	open.library.ubc.ca	<1%
27	Submitted works	Liverpool John Moores University on 2024-06-18	<1%
28	Submitted works	University of Sydney on 2025-05-21	<1%
29	Publication	Faridoddin Shariaty, Sanjiban Sekhar Roy. "MATLAB for Brain-Computer Interface..."	<1%
30	Publication	Maniat, Mohsen. "Deep Learning-Based Visual Crack Detection Using Google Stre..."	<1%
31	Submitted works	Multimedia University on 2025-01-27	<1%
32	Submitted works	University of Hertfordshire on 2024-08-19	<1%
33	Internet	export.arxiv.org	<1%
34	Publication	D. Jeya Mala, Anto Cordelia Tanislaus Antony Dhanapal, Saurav Sthapit, Anita Kha...	<1%
35	Submitted works	University of Wollongong on 2023-11-14	<1%
36	Internet	cancerimagingjournal.biomedcentral.com	<1%
37	Internet	hdl.handle.net	<1%
38	Internet	listens.online	<1%

39	Internet	theses.gla.ac.uk	<1%
40	Publication	Ajay Kumar, Deepak Dembla, Seema Tinker, Surbhi Bhatia Khan. "Handbook of D...	<1%
41	Submitted works	Queen Mary and Westfield College on 2024-08-23	<1%
42	Submitted works	University of Essex on 2025-05-21	<1%
43	Internet	journal.esrgroups.org	<1%
44	Internet	iarjset.com	<1%
45	Internet	patents.google.com	<1%
46	Internet	ses.library.usyd.edu.au	<1%
47	Publication	"Intelligent Information and Database Systems", Springer Science and Business ...	<1%
48	Publication	Alireza Rahai, Mohammad Rahai, Mostafa Iraniparast, Mehdi Ghattee. "Surface Cr...	<1%
49	Publication	R. N. V. Jagan Mohan, B. H. V. S. Rama Krishnam Raju, V. Chandra Sekhar, T. V. K. P...	<1%
50	Publication	Tarek Berghout. "The Neural Frontier of Future Medical Imaging: A Review of Dee...	<1%
51	Submitted works	University of Glasgow on 2020-04-06	<1%
52	Internet	bura.brunel.ac.uk	<1%

53	Internet	www.fastercapital.com	<1%
54	Publication	Emma Flynn, Riya Shah, Ian Dunn, Rishal Aggarwal, David Koes. "PharmacoForge:...	<1%
55	Submitted works	UCL on 2024-06-11	<1%
56	Internet	repository.esi-sba.dz	<1%
57	Submitted works	INTI Universal Holdings SDM BHD on 2025-02-12	<1%
58	Submitted works	University of Technology, Sydney on 2019-10-30	<1%
59	Internet	airconline.com	<1%
60	Internet	iopscience.iop.org	<1%
61	Internet	www.e-afr.org	<1%
62	Internet	www.geeksforgeeks.org	<1%
63	Submitted works	Heriot-Watt University on 2024-04-17	<1%
64	Internet	deepai.org	<1%
65	Internet	keylabs.ai	<1%
66	Internet	storage.freidok.ub.uni-freiburg.de	<1%

67	Internet	tind-customer-uchicago.s3.amazonaws.com	<1%
68	Internet	www.mssc.mu.edu	<1%
69	Publication	Bairros, Rita Seixas. "Deep Learning-Based Calcium Scoring of the Aortic Valve Usi...	<1%
70	Submitted works	Birzeit University Main Library on 2016-05-18	<1%
71	Publication	Bum-Chae Kim, Jee Seok Yoon, Jun-Sik Choi, Heung-Il Suk. "Multi-scale gradual int...	<1%
72	Submitted works	CSU, San Jose State University on 2024-04-26	<1%
73	Submitted works	City University on 2020-11-15	<1%
74	Submitted works	Glasgow Caledonian University on 2018-08-30	<1%
75	Publication	Hazem Farah, Akram Bennour, Neesrin Ali Kurdi, Samir Hammami, Mohammed A...	<1%
76	Submitted works	Middlesex University on 2020-10-02	<1%
77	Submitted works	Monash University on 2021-02-04	<1%
78	Publication	Seung-Hwan Jung, Woon-Ha Yeo, Inhee Maeng, Youngbin Ji, Seung Jae Oh, Han-C...	<1%
79	Submitted works	University of Northampton on 2025-05-23	<1%
80	Submitted works	University of Oxford on 2025-05-21	<1%

81	Submitted works	University of Southern California on 2015-11-01	<1%
82	Publication	Usman Idris Isma'il, Chua Hui Na, Ir. Rosdiadee Nordin, Muhammed Kabir Ahmed...	<1%
83	Submitted works	Vilnius Gediminas Technical University on 2025-01-19	<1%
84	Internet	docslib.org	<1%
85	Internet	pdffox.com	<1%
86	Internet	thesai.org	<1%
87	Internet	www.cognitivecomputingjournal.com	<1%
88	Internet	www.mbu.asia	<1%
89	Internet	www.slideshare.net	<1%
90	Submitted works	La Trobe University on 2025-05-03	<1%
91	Publication	Lalit Mohan Goyal, Tanzila Saba, Amjad Rehman, Souad Larabi-Marie-Sainte. "Arti...	<1%
92	Publication	Liu, Kechun. "Interpretable Analysis of Melanoma in Whole Slide Imaging: Detecti...	<1%
93	Publication	Rana Ehtisham, Waqas Qayyum, Charles V. Camp, Vagelis Plevris, Junaid Mir, Qais...	<1%
94	Publication	Tasneem Ahmed, Shrish Bajpai, Mohammad Faisal, Suman Lata Tripathi. "Advanc...	<1%

95	Submitted works	The University of the West of Scotland on 2025-04-21	<1%
96	Submitted works	University of Hertfordshire on 2025-03-14	<1%
97	Submitted works	Vrije Universiteit Brussel on 2020-08-24	<1%
98	Publication	Waqas, Asim. "From Graph Theory for Robust Deep Networks to Graph Learning f..."	<1%
99	Internet	appquipo.com	<1%
100	Internet	archopht.jamanetwork.com	<1%
101	Internet	docplayer.net	<1%
102	Internet	iris.unimore.it	<1%
103	Internet	tesisenred.net	<1%
104	Internet	www.ecva.net	<1%
105	Publication	Aiza M. Romano, Alexander A. Hernandez. "Enhanced Deep Learning Approach fo..."	<1%
106	Publication	Celaya, Adrian. "PocketNet: A Smaller Neural Network for Medical Image Analysis..."	<1%
107	Submitted works	ESoft Metro Campus, Sri Lanka on 2025-05-14	<1%
108	Submitted works	Higher Education Commission Pakistan on 2025-03-17	<1%

109	Submitted works	Liverpool John Moores University on 2024-03-15	<1%
110	Submitted works	North South University on 2022-11-19	<1%
111	Publication	Sahoo, Himanshu Shekhar. "Towards a Transparent OmniDoctor: AI Assistant for ...	<1%
112	Submitted works	Technical University of Košice on 2025-05-22	<1%
113	Submitted works	The University of Wolverhampton on 2023-05-08	<1%
114	Submitted works	University of Hertfordshire on 2025-04-27	<1%
115	Submitted works	University of Manouba on 2025-05-19	<1%
116	Submitted works	University of Sheffield on 2021-08-24	<1%
117	Submitted works	University of Stirling on 2021-08-27	<1%
118	Submitted works	University of Surrey on 2025-05-23	<1%
119	Internet	arxiv.org	<1%
120	Internet	bmcmedimaging.biomedcentral.com	<1%
121	Internet	careeredge.scarletknights.com	<1%
122	Internet	ebin.pub	<1%

123	Internet	jemsu.com	<1%
124	Internet	john-wicked.blogspot.com	<1%
125	Internet	pypi.org	<1%
126	Internet	trepo.tuni.fi	<1%
127	Internet	www.biorxiv.org	<1%
128	Internet	www.bip.pw.edu.pl	<1%
129	Internet	www.cancerbiomed.org	<1%
130	Internet	www.degruyter.com	<1%
131	Internet	www.spiedigitallibrary.org	<1%
132	Publication	"Deep Learners and Deep Learner Descriptors for Medical Applications", Springer...	<1%
133	Publication	"Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017", ...	<1%
134	Publication	Amit Kumar Tyagi. "Data Science and Data Analytics - Opportunities and Challeng...	<1%
135	Publication	Can Han, Chen Liu, Jun Wang, Yaqi Wang, Crystal Cai, Dahong Qian. "A spatial-sp...	<1%
136	Publication	Fei Hu, Iftikhar Rasheed. "Deep Learning and Its Applications for Vehicle Network...	<1%

137	Publication	Frei, Max. "Deep Learning for Imaging Particle Analysis Applications", Universitae...	<1%
138	Publication	Guan, Qingji. "Chest X-Ray Image Classification with Deep Learning", University o...	<1%
139	Publication	Inam Ullah Khan, Mariya Ouaisa, Mariyam Ouaisa, Muhammad Fayaz, Rehmat ...	<1%
140	Publication	Justin Ker, Lipo Wang, Jai Rao, Tchoyoson Lim. "Deep Learning Applications in Me...	<1%
141	Publication	Karm Veer Arya, Ciro Rodriguez Rodriguez, Saurabh Singh, Abhishek Singhal. "Art...	<1%
142	Publication	Lecture Notes in Computer Science, 2014.	<1%
143	Submitted works	Liverpool John Moores University on 2024-06-16	<1%
144	Submitted works	Obudai Egyetem on 2025-05-20	<1%
145	Submitted works	Oklahoma State University on 2024-02-27	<1%
146	Publication	Poonam Nandal, Mamta Dahiya, Meeta Singh, Arvind Dagur, Brijesh Kumar. "Pro...	<1%
147	Publication	Shashi Kant Dargar, Shilpi Birla, Abha Dargar, Avtar Singh, D. Ganeshaperumal. "...	<1%
148	Publication	Shi, Yaying. "Advancing Medical Image Registration and Tumor Segmentation wit...	<1%
149	Submitted works	Swinburne University of Technology on 2020-07-08	<1%
150	Submitted works	Tilburg University on 2025-05-15	<1%

151	Submitted works	University of Glasgow on 2024-04-21	<1%
152	Submitted works	University of Glasgow on 2025-02-14	<1%
153	Submitted works	University of Glasgow on 2025-04-25	<1%
154	Submitted works	University of Portsmouth on 2025-05-16	<1%
155	Submitted works	University of Sheffield on 2019-03-26	<1%
156	Submitted works	University of Southampton on 2017-09-08	<1%
157	Submitted works	University of Strathclyde on 2023-08-12	<1%
158	Submitted works	University of West London on 2019-09-08	<1%
159	Submitted works	VIT University on 2025-03-30	<1%
160	Internet	aimlstudies.co.uk	<1%
161	Internet	bjo.bmj.com	<1%
162	Internet	drmc.library.adelaide.edu.au	<1%
163	Internet	dspace.iiuc.ac.bd	<1%
164	Internet	etasr.com	<1%

165	Internet	journalofbigdata.springeropen.com	<1%
166	Internet	libstore.ugent.be	<1%
167	Internet	researchspace.ukzn.ac.za	<1%
168	Internet	semarakilmu.com.my	<1%
169	Internet	www.cbica.upenn.edu	<1%
170	Internet	www.jmir.org	<1%
171	Internet	www.ncbi.nlm.nih.gov	<1%
172	Internet	www.research-collection.ethz.ch	<1%
173	Internet	www.sciencexcel.com	<1%
174	Internet	www.thinkmind.org	<1%

Madhav Bansal Thesis final Report.pdf

 Delhi Technological University

Document Details

Submission ID

trn:oid:::27535:97673456

Submission Date

May 25, 2025, 11:42 PM GMT+5:30

Download Date

May 25, 2025, 11:46 PM GMT+5:30

File Name

Madhav Bansal Thesis final Report.pdf

File Size

1.0 MB

54 Pages

16,839 Words

98,315 Characters

0% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Detection Groups



1 AI-generated only 0%

Likely AI-generated text from a large-language model.



0 AI-generated text that was AI-paraphrased 0%

Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (it may misidentify writing that is likely AI generated as AI generated and AI paraphrased or likely AI generated and AI paraphrased writing as only AI generated) so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?

The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

