COVID-19 AND SKIN CANCER DETECTION USING A STACKED ENSEMBLE APPROACH

Submitted by
HAFZA QAYYUM
04101913026

Supervisor
Dr. Musarat Abbas

Department of Electronics,
Quaid-I-Azam University, Islamabad, Pakistan

This dissertation is submitted for the degree of
Bachelor of Science (B.S.)

2019-2023

Certificate

It is certified that the work presented in this dissertation is accomplished by Hafza Qayyum under the supervision of Dr. Musarat Abbas at Quaid–I–Azam University, Islamabad, Pakistan.

Supervisor:

Dr. Musarat Abbas
Associate Professor
Department of Electronics,
Quaid-I-Azam University, Islamabad

Submitted through:

Prof.Dr.Qaisar Abbas Naqvi
Chairman
Department of Electronics,
Quaid-I-Azam University, Islamabad

Acknowledgments

I sincerely thank Almighty Allah, the Most Gracious and the Most Merciful, Who has rewarded me in ways far beyond anything I could have imagined. I praise Prophet Muhammad (S.A.W.W), whose life inspires all humans.

I would like to extend my profound appreciation to my supervisor, Dr. Mussrat Abbas, for his assistance, inspiration, and direction during the entirety of my B.S. studies. His expertise and experience have enhanced my knowledge and research-related abilities. This research would not have been possible without his gracious guidance and supportive suggestions. His guidance and belief in me have been instrumental in my academic growth and success.

I would like to express my appreciation to Dr. Naveed Qureshi for his contribution to the successful completion of this research work. Sincerely thanks to Ms. Umamah Bint Khalid for her continuous support, advice, encouragement, and valuable suggestions during the entire period of my research work. Moreover, I am very thankful to my friends Arooj Kanwal Yaseen and Amna Shoaib for their constant support and encouragement.

I offer my deepest gratitude to all the teachers at the Department of Electronics who encouraged and taught me in my B.S. coursework.

I want to express my sincere gratitude to my parents, and my family for their unwavering love and assistance during my education.

HAFZA QAYYUM

Dedication

To my parents, who supported me throughout my years of education.

Abstract

In recent years, Covid-19 and skin cancer have become two prevalent illnesses with severe consequences if untreated. This research represents a significant step toward leveraging machine learning and ensemble techniques to improve the accuracy and efficiency of medical image diagnosis for critical diseases like Covid-19 (grayscale images) and skin cancer (RGB images). To enhance the precision and effectiveness of diagnosing both Covid-19 and skin cancer, a stacked ensemble learning approach has been used here. This proposed method combines deep neural network (CNN) pre-trained models for feature extraction, utilizing ResNet101, DenseNet121, and VGG16 for grayscale (COVID-19) and RGB (skin cancer) images. The performance of the model is evaluated using both individual CNNs and a combination of feature vectors generated from ResNet101, DenseNet121, and VGG16 neural network architectures. The feature vectors obtained through transfer learning are then fed into base-learner models consisting of five machine-learning algorithms. In the final step, the predictions from the base-learner models, ensemble validation dataset, and the feature vectors extracted from neural networks are ensembled and applied as input for the meta-learner model to obtain final predictions. The performance metrics of the stacked ensemble model reveal high accuracy for Covid-19 diagnosis and intermediate accuracy for skin cancer.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Problem Area and Motivation

In recent times, there has been notable and swift advancement in the domain of automated analysis for medical images [12]. Healthcare assistance systems are ushering in a novel era, leveraging the capabilities of information technology within the surge of extensive data and the interconnected nature of the Internet of Things [14]. Over time, a multitude of computer-aided systems have been created to aid in the identification and diagnosis of human illnesses through medical imagery. Modern imaging techniques rely on images with a high resolution to give radiologists multifaceted views, aiding in clinical diagnoses, precise predictions, and patient treatment. Ultrasound, endoscopy, X-ray computed tomography(CT), and magnetic resonance imaging (MRI) are among the prevalent modes of medical image capture. Currently, numerous studies have emerged concerning the categorization and identification of illnesses through medical imaging. Even though the mentioned models have demonstrated promising outcomes, the medical domain still demands enhanced precision.

The development of more efficient diagnostic tools with improved patient outcomes over time is the main cause of this proposed research [17].

1.2 Problem Statement

- The ongoing covid-19 pandemic and skin cancer have emphasized the need for accurate and efficient diagnostic tools to identify infected individuals and prevent further transmission.
- The goal is to create a reliable, rapid, and cost-effective solution that can aid healthcare professionals in making informed decisions and controlling the spread of the virus and cancer effectively.
- The main focus of this research is on utilizing stacked ensemble learning techniques to enhance the accuracy of covid-19 and skin cancer detection.

1.3 Aim and Objectives

The main aim of this research is to develop a generalized model that provides precise and accurate predictions for grayscale and RGB images in medical datasets using an ensemble learning method.

1.4 Outline

This work is divided into nine chapters. In chapter 2, the background theory of artificial intelligence is presented, followed by its types: machine and deep learning. In chapter 3, discusses all about ensemble learning. A overview of covid-19 and skin cancer related works is given in Chapter 4. And Chapter 5 will discuss about the method and materials used for this research. Chapter 6 present the methodology employed to perform this research and present the main public databases containing images.The result and discussion of this work are provided in Chapter 7 and 8. Conclusion and future work is in Chapter 9.

Chapter 2

Background

2.1 Artificial Intelligence

Artificial Intelligence (AI) involves machines doing tasks that humans do smartly, like planning and learning from language. There are two dominant concepts of AI: Machine Learning (ML) and Deep Learning (DL).

Artificial Intelligence (AI) is a dynamic field within computer science dedicated to creating intelligent systems capable of performing tasks that traditionally demand human-like intelligence. These tasks encompass a wide spectrum, including decision-making, speech recognition, language translation, and visual perception. A notable contribution to the AI landscape is the Turing Test, conceptualized by Alan Turing. This test gauges a machine's capacity to exhibit intelligence comparable to human behavior. In the Turing Test, a human judge converses in normal language with both a machine and a person while remaining unaware of which one is which. Should the evaluator struggle to differentiate between the machine and human responses reliably, the machine is deemed to have passed the Turing Test, signifying a significant step toward demonstrating artificial intelligence [6].

Machine Learning (ML) and Deep Learning (DL) are pivotal branches of artificial intelligence. ML involves training algorithms to learn from data, making predictions and decisions based on patterns. DL, a subset of ML, employs intricate neural networks with multiple layers to automatically extract complex features from data. This facilitates advancements in tasks like image recognition and natural language processing. Both ML and DL play vital roles in transforming industries by enabling computers to learn and adapt from experience, enhancing their performance over time.

The practical applications of AI are extensive and impactful. Natural Language Processing (NLP) is an arena where AI techniques are harnessed to understand and analyze human language, giving rise to applications like chatbots, virtual assistants, and language translation. In robotics, AI algorithms serve as the driving force behind the precise control and execution of various tasks, spanning manufacturing, assembly, and exploration endeavors. Expert systems, another AI application, leverage advanced techniques to provide expert-level advice and decision-making within specialized domains like medicine, finance, and law. Game playing takes on a new dimension with AI, as algorithms enable the creation of game-playing agents that can rival human players in complex games such as chess, Go, and poker.

Computer vision, yet another realm enriched by AI, empowers systems to decode and interpret visual data, resulting in applications like facial recognition, object detection, and the development of autonomous vehicles. Perhaps among the most transformative facets of AI is Machine learning(ML), where algorithms learn from data to enhance their performance over time. This capability fuels applications like fraud detection, recommendation systems, and predictive maintenance, driving innovation across various industries. Collectively, these applications showcase the profound impact of AI, revolutionizing the way we interact with technology and reshaping industries across the spectrum.

Artificial Intelligence (AI) is progressively finding applications in the biomedical field, particularly in areas such as diagnostics, drug exploration, and fundamental life science inquiries. In recent times, AI-driven imaging technologies have shifted from academic endeavors to commercially-driven initiatives. Solutions now exist that can identify various eye and skin conditions, detect cancers, and assist in the precise measurements required for clinical assessments. Certain of these systems rival the diagnostic capabilities of expert pathologists and radiologists, alleviating mundane tasks like quantifying cell divisions in cancerous tissue. However, while the implementation of automated systems in some spheres raises ethical considerations, the deployment of AI for risk scoring in healthcare is becoming increasingly prevalent [7].

2.2 Machine Learning

Machine Learning (ML) falls under the umbrella of artificial intelligence (AI), enabling software applications to enhance their prediction accuracy without requiring direct, explicit programming. Instead, ML algorithms leverage historical data as an input to forecast future output values. Machine learning involves using algorithms to automatically get useful information from data. When training, the ML model takes input data and finds patterns or important features. This helps it learn how to link attributes with each piece of data or group data into clusters. Later, the model can use this learning to predict things about new, unseen data [5].

2.2.1 Types of Machine Learning

Machine learning is commonly classified based on how algorithms improve their prediction precision. Four fundamental types of machine learning exist supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Data scientists decide on an algorithm based on the nature of the data they intend to predict.

Supervised Learning

In Supervised learning, training data includes desired outputs. It involves input variables (x) and an output variable (Y), where an algorithm learns the mapping function from input to output: Y = f(X) When the model encounters new input data (x), it can readily predict the corresponding output variables (Y) for that data. Most of the supervised learning algorithms are used as classifiers.

Unsupervised Learning

In Unsupervised Learning, training data does not include desired outputs. It involves algorithms that train on unlabeled data. Unsupervised learning is when you only have input data (X) and no expected answers. The main aim is to understand the patterns or arrangement within the data, helping us learn more about the data itself.

Semi-Supervised Learning

Semi-supervised learning is an intermediate ground between Supervised and Unsupervised learning. This method in machine learning combines aspects of the two previous approaches. Scientists give the algorithm lots of labeled training data, but the model also looks at the data by itself to learn about it.

Reinforcement Learning

Reinforcement learning is rewards from a sequence of actions. Reinforcement learning is like a trial-and-error way of learning. The machine gets points as a reward or penalty for its actions. This algorithm focuses on how the learning agent interacts with the environment. The learning agent balances exploring and making the most of what it already knows.

2.2.2 Machine learning Algorithms

Here are the primary ML algorithms used for classification.

Decision Trees

A decision tree is structured by rules and takes on the form of a flowchart, resembling a tree and mirroring human logic, which aids in its comprehensibility. The standout feature of a decision tree is its user-friendly nature, owed to its mechanism that allows easy interpretation. This method effectively merges both categorical and numerical data, although it does face certain problems and difficulties. For instance, it can encounter challenges like classifying data with many dimensions or addressing imbalanced data scenarios [1]. Decision Trees (DT) prove their versatility in machine learning by accommodating both categorical and numerical variables. Unlike other methods, they don't need assumptions about data distribution or classifier structure. This makes them suitable for classification, regression, and multi-output tasks. Their accuracy and efficiency shine in managing large and intricate datasets.

Random Forest (RF)

The foundation of the RF technique is the algorithm of a decision tree. The forest is created by combining many DT algorithms. The RF classifier is composed of a set of M trees, a quantity that users can define. When given an input vector ($\times$), each tree in the ensemble casts a vote to determine the most common class for allocation. This approach involves assessing attributes using techniques like information gain, gain ratio, entropy, and the Gini Index. This study, employs the Gini Index, a metric that gauges how attributes are mixed about classes. One of RF's advantages compared to decision trees is its ability to expand without pruning, allowing it to manage generalization errors even with an increase in the number of trees [17].

Support Vector Machine (SVM)

SVM stands as a statistical learning model employed within the framework of supervised learning. It falls under the category of linear classification models, suitable for scenarios where the data is lower in dimension, and exhibits simplicity that allows for easy separation in the data space. Nevertheless, SVM also exhibits capability in handling intricate, high-dimensional, nonlinear classifications. It achieves this by utilizing kernel functions that transform the original data into a higher-dimensional space, enabling effective separation [1].

When used for sorting things into categories, SVMs create special lines called hyperplanes. These lines act like clear boundaries, dividing different groups of things. These hyperplanes are like decision guides that assign things to their correct groups. SVMs work by finding the best possible line that creates the biggest separation between groups, making sure that the line is at its best by keeping enough space from the closest points [5] [17].

SVMs have become popular over time, especially in dealing with medical pictures [17]. They're good at figuring out patterns, like in predicting if someone has cancer or not [1].

Logistic Regression (LR) Logistic regression stands out as a highly favored statistical machine-learning technique used for making predictions. It is widely applied to solve both binary and multiclass problems. LR employs a logistic function, also known as a sigmoid function, to estimate the probability of class labels based on the provided attributes [17].

Naïve Bayes

Naïve Bayes is a straightforward classification method in supervised learning, rooted in probabilistic theory. It proves highly effective, particularly when coupled with feature selection functions. Its superiority over other robust methods stems from its simplicity, speed, and capacity to achieve high accuracy. This technique finds utility across various domains, including disease diagnosis, RNA sequence classification, and spam filtering [1].

2.3 Deep Learning

In recent times, the domain of automated medical image analysis has experienced significant expansion. The adoption of deep neural networks has emerged as a prevailing and extensively employed approach for tasks involving computer vision. When it comes to tasks in computer vision, such as image classification, deep CNN stands as the state-of-the-art, offering matchless accuracy and resilience [12]. However, determining the appropriate hyperparameter settings and selecting the right architecture significantly relies on the specific computer vision task [11]. Deep learning stands as the most efficient machine learning solution, demonstrating a high level of intelligence to independently learn from data through intricate architectures that mimic the structure of the human brain. CNNs first appeared in the beginning of the 2000s, taking inspiration from the Visual cortex neural networks, and have risen to prominence as highly effective and extensively utilized tools in the realm of computer vision. Within CNNs, diverse layers serve distinct functions, encompassing some with trainable parameters and others that execute predefined functions. Among the frequently employed layer types in CNN architectures are the following [5].

Convolutional Layers Convolutional layers play a significant role in the architecture. They can grasp patterns within specific local regions. This has notable implications: the patterns learned exhibit invariance to translation, and the learning process extends to hierarchical spatial patterns. Consequently, CNN can adeptly acquire intricate visual concepts as its network depth deepens. Within convolutional layers, arrays of filters traverse the input image, executing convolution operations and producing feature maps for onward transmission to subsequent layers.

Normalization Layers Normalization layers fulfill the role of normalizing input data through a dedicated function that operates solely during forward propagation and doesn't involve any trainable parameters. However, the prevalence of these layers has decreased in recent years.

Regularization Layers Regularization layers serve the purpose of mitigating overfitting by introducing an element of randomness, causing a fraction of neurons to be disregarded during each training iteration. Among the most recognized regularization methods is dropout.

Pooling Layers These layers perform feature map subsampling, preserving essential information within them. This aids in curbing the model's parameters and computational demands, contributing to efficiency. Pooling filters, notably max pooling and average pooling, scan the received feature maps, much like the convolutional layers, but devoid of any trainable parameters. The objective remains to minimize the computational load without compromising crucial content.

Fully Connected Layers (FC) Fully connected layers establish connections between every neuron in the layer and all of the activation functions of the preceding layers. The first FC gets feature maps created from the final convolutional layer or pooling layer as input. The ultimate layer among FCs functions as a classifier within the CNN structure.

2.4 Pre-Trained Models and Transfer Learning

There are many pre-trained models for Convolutional Neural Networks (CNNs) available, and some of the most often used ones include AlexNet, ResNet, DenseNet, Xception, and Inception-ResNet VGG, and EfficientNet. Table 2.1 provides a summary of the important details and settings for each of these models.

Architetture	**Year**	**Parameters**	**Input Size**
AlexNet	2012	62.3 M	256×256
InceptionV3	2015	23.8 M	299 × 299
ResNet101	2015	44.7 M	224 × 224
DenseNet121	2017	20.2 M	224 × 224
Xception	2017	22.9 M	299 × 299
Inception-ResNet	2017	55.8 M	299 × 299
VGG16	2014	134.2 M	224 × 224

Table 2.1: Summary of frequently used CNN architectures:

As mentioned earlier, training deep learning models with randomly initialized parameters can be challenging since a large volume of annotated data is required, which is frequently unavailable. Transfer learning offers a valuable solution by allowing us to leverage knowledge stored in pretrained CNN models trained on extensive datasets like ImageNet. This approach often yields good results in the original domain where the model was trained.

In this specific case, transfer learning is primarily used for feature extraction. Notably, select pretrained models such as DesNet, ResNet, VGG, or combinations of these models. These models are chosen partly because they require an input image size of 224x224 pixels. This particular size was adopted because already converted images from 229x229 pixels to 224x224 pixels, making it convenient to utilize these specific pretrained models for this task.

Chapter 3

Ensemble Learning

3.1 Outline

Ensemble learning has become increasingly popular in the field of machine learning. Despite the rise of learning, ensembles continue to be favored by researchers, corporate data scientists, and even those participating in data science competitions. The key reason, for their success lies in their ability to enhance the predictions made by learners while also offering stability through variance. Moreover, ensembles have the potential to scale up effectively and typically require parameter tuning. [13].

3.2 Background

For numerous decades, the machine learning research area has used multiple classifiers/regressors in parallel or sequential approaches for training. However, enhancing current ensemble approaches in terms of speed as well as accuracy remains a frequent subject at recent renowned conferences. [13]. Ensemble modeling techniques are widely employed in data mining solutions. The task of determining the disease based on a person's symptoms is known as disease diagnosis, which can be quite challenging, due to the presence of ambiguous symptoms and signs. Accurate identification of the disease is vital, for treatment. Numerous researchers have developed machine learning algorithms to effectively diagnose diseases. These algorithms enable the creation of models that can predict diseases and their corresponding treatments. Disease prediction is an area of research given the amount of available data [8].

3.3 Introduction

The purposeful design and mixing of many models, such as learners or experts, to address specific computational intelligence challenges is known as ensemble learning (EL) [19]. It is a method for combining numerous models to achieve better generalization ability [4]. Ensemble learning is a method for merging numerous predictions to form a single inference. Traditionally, it involves building an ensemble of multiple predictions originating from different models. Recent novel approaches, define ensemble learning in the context of learning as combining knowledge or predictions for a single result. This kind of information or prediction may come from numerous different models or from a single model. [12]. The ensemble machine learning method is a technique that combines many classification algorithms, called Base-Learners (Base Models) to create a single improved classifier. The essential idea is the final prediction is made by the meta-learner (Meta Model) based on the base learners. The meta-learner is an approach that works to reduce the base learners' error in prediction. The predicted output of the base learners is utilized as input for the meta-learner. The ability to be generalized and the accuracy of prediction results obtained using this technique beat the results of a single machine learning setup. [17]. Ensemble learning is a classifier enhancement strategy that combines each of the contributions of sub-models to deal with an identical classification issue. The base learners are homogeneous or heterogeneous classification algorithms that are combined to generate a single, better-performing classification model. The meta-learner is a model that learns to correct the prediction of the base learners and makes the final prediction based on their vote. The ensemble approach results in better prediction accuracy than single learners, and its generalizability capability usually reduces the variance in the prediction, ensuring the most stable and best possible prediction is made. The meta-model takes the output of the sub-models as input and learns to merge their predictions to make the final prediction, which is better than each of the base classifiers [4].

3.4 Ensemble Strategies

Ensemble strategies have changed with time to improve model generalization in learning. These strategies may be categorized into three categories: Bagging, Boosting, and Stacking.

3.4.1 Bagging

Bagging, often referred to as bootstrap aggregating, serves as a commonly employed method for creating ensemble-based algorithms. The core concept behind bagging involves creating a set of independent datasets, each having the same size and distribution as the original data. By working with these sets of observations, an ensemble predictor is constructed, outperforming a single predictor built solely on the original data.

Bagging introduces two key steps to the original models: firstly, the creation of bagging samples and their subsequent presentation to the base learner models, and secondly, the approach employed to merge the predictions from multiple predictors. Bagging samples can be generated either with or without replacement. The method of combining the outputs from the base predictors can differ; typically, majority voting is employed for classification tasks, while averaging is used in regression scenarios to generate the ensemble output [4].
Homogeneous model ensembles refer to a collection of models that share identical algorithms, hyper-parameters, or architectures [12]. The bagging technique encompasses the process of training numerous models on distinct subsets of the training data. Subsequently, these models' predictions are merged to formulate a conclusive prediction.
The Bagging method relies on enhancing the training dataset sampling process and is a widely used homogeneous ensemble learning approach. Unlike a conventional standard training/validation split that generates only one model, Bagging involves training multiple models on randomly selected subsets of the dataset. It utilizes k-fold cross-validation on the dataset to create k models, making it a bagging technique for ensemble learning. [12]. It generated a diverse set of models, each trained on distinct subsets of the training data. This method not only improved the effective utilization of the provided training data but also enhanced the prediction's reliability [11]. The bagging approach aims to diminish variance and enhance the model's accuracy.

3.4.2 Boosting

The boosting technique is employed within ensemble models to transform an inadequate learning model into a more proficient one with enhanced generalization capabilities [4]. Boosting is a machine learning algorithm that involves training multiple models sequentially, with each model trying to correct the errors of the previous model.
The boosting method involves training a weak learner, making predictions, identifying the training samples that were incorrectly classified, and subsequently training another weak learner using an updated training set that includes the previously misclassified instances from the prior training iteration.
Boosting stands as a robust approach that effectively mitigates overfitting. Nevertheless, certain drawbacks are associated with boosting. One such limitation is its vulnerability to outliers, as each classifier is compelled to rectify errors from previous iterations, leading to a notable reliance on outliers. Additionally, the method's scalability is challenging, posing difficulties in its expansion to larger scales. The objective is to minimize bias and enhance the model's accuracy.
The integration of boosted deep belief networks (DBNs) into facial expression recognition brings together the boosting technique and multiple DBNs through an objective function. This combination leads to the creation of a robust classifier with enhanced capabilities [4]. Due to the huge increase in training time, boosting is not feasible for image classification. [12].

3.4.3 Stacking

The ensemble machine learning method is a technique that combines multiple classification methods that are homogeneous or heterogeneous, known as base learners, to produce a single superior-performing classification model. The key idea of this method is that the meta-learner generates its final predictions according to the base-learner predictions. The meta-learner is a model that aims to reduce the prediction mistakes of the base learners. It takes as input the prediction outputs generated by the base learners. As a result, ensemble technique outperforms a single machine learning algorithm in terms of generalizability and accuracy in prediction [17].
Ensembling typically enhances prediction stability and minimizes variance by combining multiple models. The meta-model receives predictions from

individual base-learners, learning to effectively blend these inputs into a final prediction that surpasses the performance of each base-classifier [10].In contrast to simple algorithm approaches, the ensemble of various models demonstrated considerable improvements in overall performance. This type of ensemble learning is more complicated. The stacking approach works by stacking another machine-learning algorithm on top of these predictions. [12]. Stacking is a bias-reducing technique [4].
This approach, incorporated a stacking technique when employing five distinct machine learning algorithms as base learners. Unlike previous methods that relied solely on base learners' predictions for the meta-learner's final decision, this research combined these predictions with the original features and ensemble validation data as input for the meta-learner. Furthermore, the same five classification algorithms used as base learners were also leveraged and assessed for the meta-learner. This innovative approach aims to enhance model performance by leveraging both base learners' predictions and original data features in the stacking process.

3.5 Applications

In recent times, the remarkable progress in the power of computers and machine learning (ML) has spurred numerous innovations and advancements across various research domains and aspects of human existence. Over the past few years, the evolution of machine learning has brought about fundamental transformations in nearly every facet of our daily lives. Notably, two prominent approaches, ensemble learning and deep learning, have emerged as dominant forces within the field of machine learning. Ensemble algorithms, in particular, have demonstrated exceptional performance across a wide range of ML applications and have established themselves as well-established tools in real-world scenarios. Their superior performance has made ensemble methods the preferred choice in many machine learning competitions. [9].

Chapter 4

Related Work

Numerous efforts have been made to categorize diseases through medical imaging, with some offering solutions for disease identification. The classification of medical images growing field primarily relies on deep learning and machine learning techniques. This section reviews the latest research pertaining to medical image classification, especially focusing on datasets related to covid-19 and skin cancer.

4.1 Covid Related Works

With a pre-trained VGGNet model and support vector machine algorithm, Article [15]proposes X-ray images are classified into three distinct categories: covid, healthy images, and pneumonia. and provides an accuracy of 95.81%.

Article [12] discusses a reproducible pipeline for medical image classification that analyzes the effect on the performance of various ensemble learning approaches. The paper uses a dataset related to skin and covid-19 on various ensemble learning techniques. The paper proposes an ensemble learning approach that combines multiple deep CNNs to improve the accuracy of covid-19 classification. The results show that Stacking emerged as the ensemble learning approach that demonstrated the most significant performance improvement, and In terms of accuracy and efficiency, the proposed method exceeds other methods.

In the study, [16] employs an ensemble architecture to classify CT scan images. Their approach combines various deep learning models, including VGG16, ResNet50, VGG19, ResNet50V2, InceptionV3, InceptionResNetV2, MobileNet, and Xception.

In [17] Article, a novel approach was proposed for feature extraction by combining transfer learning from the DenseNet model and six handcrafted techniques to catch more comprehensive and intricate features. Additionally, a distinct stacked ensemble technique in the first level was introduced, which incorporates RF, GBM, LR, SVM, and KNN algorithms in both base-learner and meta-learner phases to improve the classifier model's performance. In the evaluation of the Covid-19 CT Scan data, LR outperforms the other algorithms across every evaluation metric, achieving an accuracy of 99.73%.

4.2 Skin Cancer Related Works

Article [5] provides a comprehensive systematic review of the latest research on using machine learning algorithms for dermoscopic image categorization of skin cancer. The article also discusses the limitations and challenges in implementing machine learning for skin cancer diagnosis in clinical settings. Overall, this review provides valuable insights into the latest research on using machine learning for skin cancer diagnosis and can be a useful resource for researchers and practitioners in the field. The results of these studies vary depending on the algorithm used, the dataset used, and the evaluation metrics used. However, the review found that recent machine learning and deep learning models show high potential in skin lesion classification.

The paper [12] discusses a reproducible pipeline for medical image classification that analyzes the effect on the performance of various ensemble learning approaches. The paper uses a dataset related to skin cancer: the ISIC dataset, which hosts the largest publicly available collection of quality-controlled images of skin lesions. The paper proposes an ensemble learning approach that combines multiple deep convolutional neural networks to improve the accuracy of skin lesion classification. The proposed approach is evaluated on the two datasets and compared to other state-of-the-art methods. The results show that the proposed approach outperforms other methods in terms of accuracy and efficiency.The outcomes indicated that Stacking exhibited the most substantial enhancement in performance, with a notable increase of up to 13% in the F1-score.

In [17] Article, a novel approach was proposed for feature extraction by combining transfer learning from the DenseNet model and six handcrafted techniques to catch more comprehensive and intricate features. Additionally, a distinct stacked ensemble technique in the first level was introduced, which incorporates RF, GBM, LR, SVM, and KNN algorithms in both base-learner and meta-learner levels to improve the classifier model.In the evaluation of the ISIC dataset, RF exhibits the highest efficiency with an accuracy of 89.09%.

Chapter 5

Method and Materials

5.1 Literature Review

A Literature Review is regarded essential at the start of any research because it allows one to gain knowledge about the subject of interest. Furthermore, knowledge and expertise are gathered regarding past contributions to the topic, how the research was conducted, and the significant problems and issues encountered. It is necessary to demonstrate that have an in-depth knowledge of previous research on this topic. The primary goal of the literature study undertaken in this thesis is to identify and evaluate the benefits of the various methodologies used. Other goals of performing a literature review include learning about the procedures employed and identifying datasets and assessment measures that were used.

5.1.1 Search Terminology and Terms

The search terms/keywords that were utilized are given here. "Ensemble Learning", "Stacked ensemble model", "Machine Learning", "Ensemble Methods", "Skin Cancer Classification ", "Transfer Learning", and "Covid-19".
For the article selection phase,IEEE Xplore, arXiv, and ScienceDirect and other digital libraries were used to collect articles.

5.1.2 Inclusion Criteria

Inclusion criteria used in the selection included (i) openly available works, (ii) articles in English, (iii) Ensemble learning research papers, (iv) works based on dermoscopic pictures, and (v) papers published from 2018 to 2023.

5.1.3 Exclusion Criteria

Exclusion criteria included (i) review papers, (ii) papers published in any language other than English, (iii) studies that were not complete with results, (iv) publications that dealt only with segmentation, and (v) articles that did not use publicly available datasets.

5.2 Data Collection

While collecting data, there are several factors that need careful consideration. These include the specific requirements of work, the size of the data set, the number of classes within it, and the relevant features. Therefore, there is a need to be mindfully select data that is well-cited, aligns with the best practices from influential papers, demonstrates strong performance, freely accessible, and is tailored to the model's needs.
The data sets selected for this study were sourced from the "Kaggle" repository, which offers a collection of publicly available data sets suitable for research and studies.
The Kaggle repository plays a pivotal role as a prominent platform hosting an extensive array of datasets. These datasets are generously contributed by the worldwide data science community. Covering a multitude of domains and subject matters, they present researchers and professionals with an expansive selection for investigation and scrutiny. Additionally, the vibrant participation and collaboration within Kaggle's community substantiate the authenticity and excellence of these datasets. This collaborative environment ensures that the datasets conform to stringent benchmarks of quality and utility, meeting the exacting requisites for utilization in research and academic pursuits.

5.2.1 Data Types

In the field of healthcare data analysis, there are three main types of data that are really important: Data that is statistical, image-based, or sequential.

- Statistical data is all about numbers that come from looking at things, testing things, or measuring things. This kind of data helps doctors and researchers make smart choices about treatments and health policies.
- Image-based data is like pictures from medical machines such as X-rays or MRIs. These pictures help doctors spot problems like tumors or broken bones, so they can treat patients accurately.
- Lastly, sequential data is like a timeline of information. It's often found in medical records and monitoring data. When doctors study patterns in this data, they can understand how diseases develop over time and plan the best treatments.

All these data types work together, like in studying genes to find out about diseases. When they're used together, doctors can make better decisions and improve patient care [14].

5.3 Software Environment

5.3.1 Development

Integrated development environment Software IDEs are software tools designed to provide a specialized environment for working with machine learning tasks. These environments offer features tailored to developing, testing, and deploying machine learning models efficiently.They typically include capabilities for coding, data visualization, model training, and analysis, all aimed at streamlining the machine learning development process.ML IDEs contribute significantly to enhancing productivity and collaboration among data scientists and researchers in the field of machine learning.

Software	**Description**
Google Colab	Google Colab is a cloud-based notebook environment that allows users to run Jupyter notebooks directly on Google's infrastructure. It provides free access to GPUs and TPUs, enabling users to perform resource-intensive computations without the need for powerful local hardware. It's particularly convenient for collaborative work, as notebooks can be easily shared and worked on in real-time.
VS Code	Visual Studio Code (VS Code) is a versatile code editor that can be extended to support various programming languages, including machine learning. With a range of extensions and plugins, developers can create customized environments for ML tasks. Its flexibility, combined with features like integrated Git support and debugging, makes it a suitable choice for machine learning projects.
Jupyter Notebook	Jupyter Notebook is a popular interactive environment that combines code, visualizations, and explanatory text in a single document. Its modular structure allows users to execute code in small, manageable chunks and immediately visualize results, making it great for exploration and sharing insights. An interactive console serves as a suitable platform for comprehensively recording the entire computational procedure, encompassing code execution, documentation,and result visualization.
Python	It's an easy but strong programming language to pick up. It has fancy ways of handling data and helps you step into more advanced programming. Think of it as a handy tool for quick tasks, used in lots of different kinds of projects.

Table 5.1: Description of Integrated Development Environment Software (IDEs)

5.3.2 Frameworks

Libraries	Description
Numpy	It is a quick, simple, and versatile open source library that serves as the foundation for machine learning libraries such as sckit-learn. Tools for mathematical computations such as algebra functions, and so on are provided. Provides a diverse set of computing platforms and devices.
Pandas	It's a free and openly accessible source, way to manipulate and analyze data.It is a Python-based set up that is rapid, simple to use, and adaptable.
Scikit	It is freely available open-source ML package based on scipy, and numpy. It offers efficient as well as straightforward methods for predictive data analysis.
TensorFlow	TensorFlow is a freely available library designed for rapid numerical computation. Initially created by Google Brain's Machine Intelligence team for ML and neural network research, it now serves a broad range of applications.
Keras	Keras is a freely available Python library that provides a platform for creating artificial neural networks(ANNs), and it functions as a frontend for the TensorFlow library.

Table 5.2: Description of python Libraries

5.4 Data Splitting

Machine learning algorithms extract patterns from data and apply them to create predictions on previously unseen data. It is essential to assess models' performance on data that the models have not encountered during training. Data-splitting methods come into play here. Data splitting occurs when the given data is divided into two or more subsets in order for a model to be trained, tested, and validated.

Typically, two or three components of the main dataset are created.

- Train Data: The data used to train the machine learning model is known as train data. It is made up of input features and their associated output labels. The algorithm learns patterns from existing data and applies them to predict new, previously unknown data.

- Validation Data: Validation data is used to assess the performance of the model during training. This information is used to fine-tune the model's hyperparameters and avoid overfitting. The validation data should be indicative of the unknown data encountered by the model in the real world.

- Test Data: Test data evaluate the final performance of the model after training and adjusting. To avoid data leaking, this data should be kept fully separate from the training and validation data.

Chapter 6

Methodology

Numerous research studies have been conducted to identify diseases utilizing various medical imaging techniques, such as the detection of COVID-19 skin cancer. The main goal was to maximize performance by using a particular dataset to train various models. In this context, a combined feature extraction approach and a stacked ensemble technique were employed to address the challenges of identifying COVID-19 and skin cancer ailments.

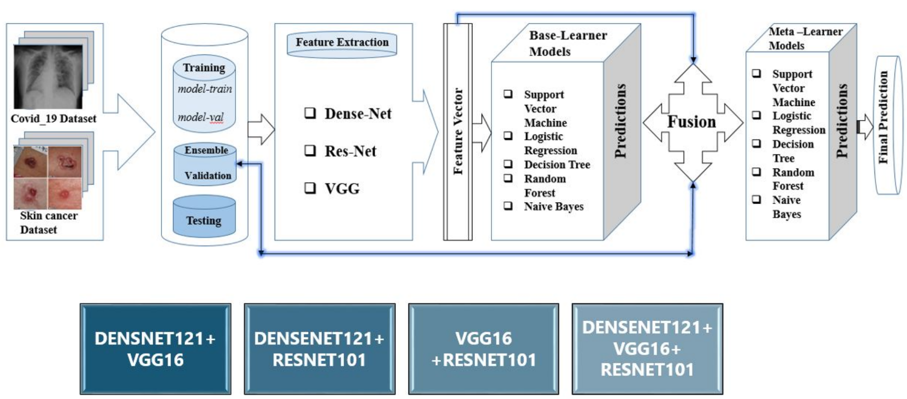


Figure 6.1: Proposed architecture.

Illustrated in Figure 6.1 is a comprehensive overview of the proposed methodology's sequential stages. Commencing with image acquisition, the process proceeds to feature extraction employing various pre-trained deep learning methodologies. Subsequently, a feature vector is generated through the amalgamation of features derived from these extractions, which is then inputted into the base learners. These base learners are trained on the feature vector, encompassing five diverse machine learning algorithms: random forest, decision trees, logistic regression, support vector machine, and naive Bayes. Conclusively, the predictions originating from the base learners are integrated with the initial feature vector and the ensemble validation set. The final prediction is executed through the application of the stacking ensemble technique.

6.1 Data Description

Employed a pair of publicly accessible image datasets for this proposed methodology: SARS-CoV-2 CT scan dataset [2] and ISIC Archive dataset [3]. The tabulated data in Figure 6.2 highlights the class distribution across both datasets. Visual examples showcasing images from each dataset, encompassing all classes, can be observed in Figures 6.3 and 6.4.

6.1.1 Covid-19 Radiography Database

X-ray pictures are really important for looking at medical issues, and they're used a lot in healthcare. They're even used instead of certain tests for covid-19. Researchers from different places like Qatar, Doha, Dhaka, Bangladesh, Pakistan, and Malaysia collected X-ray pictures of people's chests who had covid-19, as well as some who were healthy and others with viral pneumonia. They got these pictures from different places like a radiology database [12]. The collection has about 2,905 gray-scale pictures. There are 220 pictures of people with covid-19, 1,345 with viral pneumonia, and 1,340 healthy pictures.

6.1.2 Skin Lesion Images for Melanoma Classification

Melanoma is a serious health issue where colored spots appear on the skin. It causes a lot of people to get sick, with over 300,000 new cases every year, and sadly, many people die from it. Dermoscopy is a way to find melanoma early. This can be done by experts looking closely at the skin, or by using special cameras that take really detailed pictures. The International Skin Imaging Collaboration (ISIC) has a big collection of pictures of skin spots that people can use to learn and study [12]. The ISIC Archive collection includes 3297 skin cancer images, 1800 of which are classified as benign and 1497 of which are malignant. Overview of the datasets used along with their descriptions and how the samples were distributed.

Dataset	Modality	Classes	Number of Samples	Model Train	Model validation	Ensemble val	Testing
COVID	X-ray	3	2905	1743	435	290	437
ISIC	Dermoscopy	2	3297	1978	494	329	490

Figure 6.2: Overview of the datasets used along with their descriptions and how the samples were distributed.

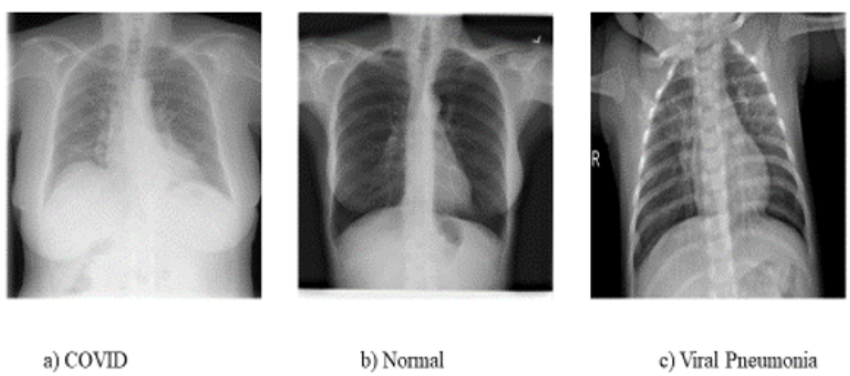
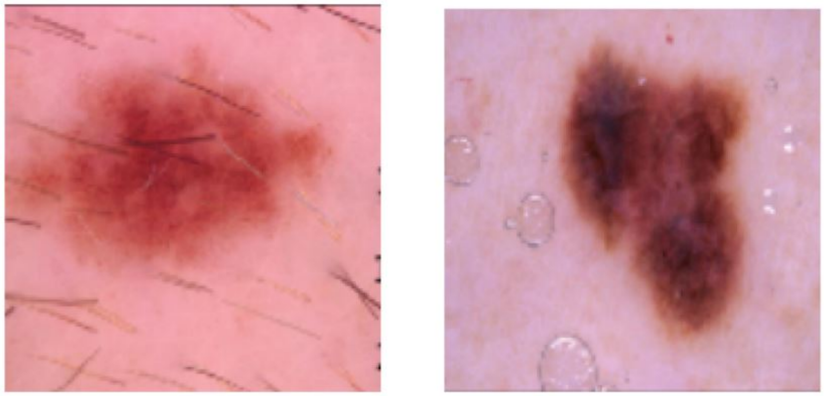


Figure 6.3: Sample images from the SARS-CoV-2 X-ray dataset illustrating instances of covid-19 cases, viral pneumonia, and healthy control cases.

6.2 Data Splitting

To ensure a trustworthy assessment of this models, employing the subsequent distribution strategy for sampling each dataset: For base learners training, 75% of the total dataset split into 80% for training(called ‘model-train’) and



(a) Benign (b) Malignant

Figure 6.4: Sample pictures from the ISIC archive dataset showing both benign and malignant cases.

20% for validation (called 'model-val'). For possible training of the meta-model, an additional 10% of the total dataset was set aside (called 'ensemble-validation'). The remaining 15% of the overall dataset was sampled as a testing set (called 'testing') for the final predictions.

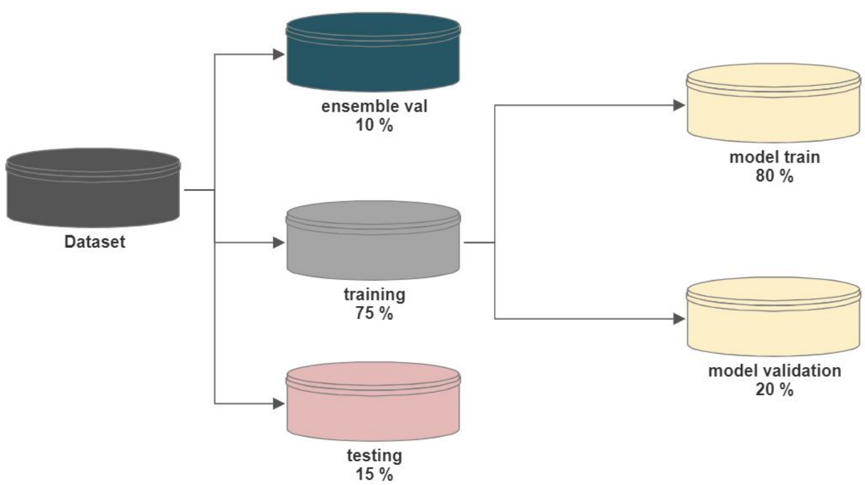


Figure 6.5: Dataset distribution strategy:

6.3 Data Pre-processing

Pre-processing plays a crucial role in computer vision applications, serving various purposes such as noise reduction, highlighting relevant image features for recognition tasks, and aiding in the training of deep learning models. In this study, a straightforward approach involving the normalization of pixel intensities within the [0, 1] range has been employed. This pre-processing step is essential to ensure the model's convergence during the training phase [18].
Keras provides the ImageDataGenerator, which clarifies the configuration for image data preparation which includes the following techniques: zoom, brightness, and normalization. Furthermore, the image sizes were scaled down to the model architecture standard input sizes of 224x224 pixels,during the resizing. Assign labels to three classes in the context of COVID classification: 'COVID,' 'Normal,' and 'Viral Pneumonia,' which are represented numerically as [0, 1, 2]. These labels are used to categorize various circumstances within the dataset.
In the context of skin cancer classification, there are two categories: 'benign' and ' malignant,' which are labeled numerically [0, 1]. These numeric identifiers are used to distinguish between benign (non-cancerous) and malignant (cancerous) skin lesions.

6.4 Feature Extraction Technique

Convolutional neural networks (CNNs) have become the prevailing method for performing tasks such as feature extraction, segmentation, and classification in the field of image processing. In this research work, some of the most effective and high-performing CNN variants, namely DenseNet-121, ResNet101, and VGG16, were utilized. These models have been trained using the ImageNet dataset, a technique known as transfer learning(TF).
Transfer Learning focuses on the idea of preserving the acquired knowledge from addressing one problem and leveraging it for solving distinct yet related problems. TL technology enables us to efficiently employ pre-trained deep learning models that have been trained on extensive and easily accessible datasets.
Pre-trained VGG16, DenseNet-121, and ResNet101 models have been loaded using TensorFlow and Keras. These models, originally trained on the ImageNet dataset, have been configured to exclude their fully connected layers,

rendering them suitable for feature extraction. These extracted high-level features from the images can now serve as valuable inputs for subsequent classification tasks using base learner models.

6.5 Fusion Technique

The process of combining several feature vectors produced from diverse methodologies in disciplines such as computer vision and other machine learning applications is known as feature fusion.Previously extracted deep features from three different pre-trained models—DenseNet121, ResNet101, and VGG16—in this context. These distinct feature vectors and their combinations have been combined, and predictions from the base learner model are based on them. The feature set of the meta-model is made up of a concatenated feature vector derived from the predictions of five base learner models, as well as ensemble validation.

6.6 Classification Models

The research employed five foundational machine learning classification algorithms, namely random forest, decision tree, logistic regression, support vector machine, and naive bayes, as its base learners. Afterwards, decided to use a stacking technique at level 1 which included all five of these algorithms. Section 2.2.2 of the thesis provides a concise overview of the implemented classification methods, with the implementation carried out using the Sklearn library.

6.7 Stacking

The ensemble approach to machine learning uses a number of homogeneous or heterogeneous algorithms for classification, known as Base Learners, which work together to create a classification model that achieves superior performance. The fundamental concept is the idea that the meta-learner produces its final prediction depending on the base-learners. The meta-learners is an algorithm that learns to minimize the base learners' prediction mistakes. The prediction output of base learners is used as input to the meta-learner.

In the base-learner phase, this study used a first-level stacking strategy using five machine-learning algorithms. According to the researchers' knowledge, earlier works solely used base learner predictions to direct the meta-learner's final decision. In this work, combine the predictions of the base learners, with the features and ensemble validation data to serve as input, for the meta-learner. All five base-learner algorithms are used and evaluated for the meta-learner to make a final prediction.

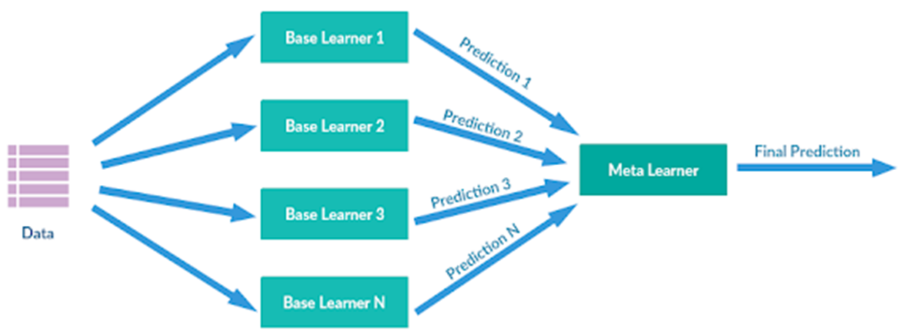


Figure 6.6: Stacked ensemble approach:

In this analysis, exclude both Boosting and Bagging techniques. However, Boosting is not feasible to apply for image classification because of the dramatic increase in training hours, and complexity of the model which makes it harder to interpret. According to the [12], Stacking has proven to be a highly effective technique for image classification, leading to a substantial improvement in performance. Additionally, Cross-validation-based Bagging has also demonstrated a significant enhancement in performance, closely competing with Stacking. This suggests that Stacking is the preferred approach, with Bagging being a strong contender.
Building upon this insight, the latest publication [8] reaffirms the superiority of the Stacking method as the optimal choice for achieving superior ensemble-based results. Therefore, in this article, I have utilized Stacking while paying close attention to high-performing strategies, aligning with the findings of both studies and ensuring originality in my work.

6.8 Performance Evaluation Metrics

The research employed various performance evaluation indicators to evaluate the models' performance. These metrics included accuracy (Ac), precision (Pr), recall (Re), and the F1-score. Accuracy gauges the overall reliability of the model's performance by computing the ratio of correctly classified class labels to the overall number of data points in the dataset, shown in Equation (6.1). Precision as defined in Equation (6.2), represents the positive predictive value, which represents a fraction of true positives divided by the overall number of actual true instances. Conversely, recall, also referred to as sensitivity and expressed in Equation (6.3), measures the fraction of true positives relative to the total number of predicted true instances. Finally, Equation (6.4) outlines the formula for the F1-score, a metric that strikes a balance between precision and recall by computing their harmonic mean. And used the Scikit-learn library for evaluation purposes.

$$Ac = \frac{TruePositive + TrueNegative}{TruePositive + FalsePositive + FalseNegative + TrueNegative} \tag{6.1}$$

$$Pr = \frac{TruePositive}{TruePositive + FalsePositive} \tag{6.2}$$

$$Re = \frac{TruePositive}{TruePositive + FalseNegative} \tag{6.3}$$

$$F1 - score = 2.\frac{Pr.Re}{Pr + Re} \tag{6.4}$$

Chapter 7

Results and Analysis

In this subsection, delve into a comprehensive analysis of the experimental outcomes achieved through the application of the proposed architecture on two distinct datasets: SARS CoV-2 CT scans and the ISIC. The proposed methodology encompasses two pivotal stages: the base-learner and the meta-learner stage. Across both datasets, conduct training and evaluation processes on five distinct algorithms, namely RF (Random Forest), DT (Decision Trees), LR (Logistic Regression), SVM (Support Vector Machine), and NB (Naive Bayes), within both the base-learner and meta-learner phases with different feature extraction models.
The initial step in this proposed method involves the construction of a feature space. This is accomplished by leveraging three distinct pre-trained deep learning models: DenseNet-121, ResNet-101, and VGG16. Additionally, explore combinations of these models, namely DenseNet-121 and ResNet-101, DenseNet-121 and VGG16, ResNet-101 and VGG16, as well as the combination of all three models: DenseNet-121, ResNet-101, and VGG16. This work carries out feature extraction using these various pre-trained models and combinations to assess their impact on classification accuracy. Evaluation is conducted using the two datasets employed in this study. The results of these evaluations, specifically the accuracy of the five classification algorithms, are visually represented in Figure 7.1 and Figure 7.2.
Additionally, we examine how both the individual pre-trained models and their combined usage for feature extraction influence the performance of all five classification algorithms. Thus, it can be observed that the combination of VGG16 and DenseNet121 yields superior performance on Logistic Regression compared to the other pre-trained model.

MODELS	VGG16	DENSNET121	RESNET101	DESNET121+ RESNET101	DENSENET121 +VGG16	RESNET101+ VGG16	DENSENET121 +VGG16+ RESNET101
SVM	96	91	95	93	95	91	91
LR	96.78	94	96	95	97.47	94	94
NB	91	84	90	91	91	90	91
DT	84	87	85	91	84	87	86
RF	92	91	92	89	91	91	92

Figure 7.1: Accuracy of five classification algorithms for different combinations of pre-trained deep learning feature extraction models on covid-19 dataset:

Figure 7.1 demonstrates the effectiveness of the base-learner models in detecting COVID-19 using the SARS-CoV-2 CT scan database. An accuracy of 96%, 97%, 91%, 84%, and 92% is observed in the base-learner stage from SVM, LR, NB, DT, and RF using VGG16. During the base-learner phase with the DesNet121 model, the recorded accuracies for different classifiers were 91%, 94%, 84%, 87% and 91% for SVM, LR, NB, DT, and RF, respectively. In the initial training phase with the ResNet101 model 95%, 96%, 90%, 85% and 92% for SVM, LR, NB, DT, and RF. Moreover, the performance of all pre-trained models and their combination in the base-learner stage is also an excellent outcome; like, LR scores the highest in all the implemented pre-trained model feature extraction techniques; for example, LR scores 94%, 96%, 97%, 95%, 97%, 94% and 94% accuracy is recorded in the DenseNet-121, ResNet-101, and VGG16. Additionally, DenseNet-121 and ResNet-101, DenseNet-121 and VGG16, ResNet-101 and VGG16, as well as the combination of all three models: DenseNet-121, ResNet-101, and VGG16.Certainly, in summary, Logistic Regression consistently yielded high accuracy on applied pre-trained models. However, when further analyzed, it was observed that Logistic Regression achieved its highest 97.47% accuracy on a specific combination of pre-trained models that is DesNet-121 and VGG16 as compared to other feature extraction methods; for example 96.78% for VGG16. A following experiment was carried out utilizing the ISIC Archive dataset in Figure 7.2, with the goal of identifying cancer cases. This experiment followed the same technique as the previous one on the SARS-CoV-2 CT scan dataset. However, in contrast, to outcomes observed in the SARS-CoV-2 CT scan dataset, it's important to note that the LR algorithm

MODELS	VGG16	DENSNET121	RESNET101	DESNET121+ RESNET101	DENSENET121 +VGG16	RESNET101+ VGG16	DENSENET121 +VGG16+ RESNET101
SVM	83	82	87	81	84	82	80
LR	82	83	88	85	81	84	82
NB	80	78	79	79	78	78	77
DT	70	76	71	70	71	72	76
RF	80	82	84	83	81	84	79

Figure 7.2: Accuracy of five classification algorithms for different combinations of pre-trained deep learning feature extraction models on ISIC dataset :

doesn't consistently outperform the other five algorithms during the base-learner phase. This highlights the fact that there isn't a single classifier that consistently excels in all scenarios or across various datasets. The varying performance of different algorithms across datasets can be attributed to the specific traits of the algorithms employed and the inherent characteristics of the datasets themselves.
In research, LR demonstrated superior performance in the context of the SARS-COV-2 CT scan dataset, while RF, LR, and SVM excelled in the ISIC Archive datasets. LR tends to yield results when there are variables that introduce noise as compared to the variables that provide explanatory information. It also performs well when the number of variables is equal, to or less, than the number of factors. SVM's strength lies in its ability to handle high-dimensional data effectively. Conversely, RF exhibits higher true and false positive rates as the number of explanatory data points in a dataset improves. The main objective in utilizing two distinct datasets to assess this methodology was to evaluate how these algorithms perform in different scenarios. Upon closer examination, it became evident that Logistic Regression and Support Vector Machine achieved their highest accuracy rates, reaching 88% and 87%, respectively, when applied to a particular pre-trained model, namely ResNet-101. This performance outshone other feature extraction techniques and classification algorithms.

Chapter 8

Discussion

The primary objective of this current study is to develop a comprehensive model capable of delivering precise and accurate predictions for both grayscale and RGB images within medical datasets. Achieve this using an ensemble learning approach and assess the model's performance in detecting both COVID-19 and skin cancer diseases. In pursuit of this goal, we have enhanced both the feature extraction and classification components, which are pivotal aspects of the field of medical image processing.

This approach involves a fused feature extraction technique, combining features extracted from three different pre-trained feature extractor methods with various combinations. Additionally, introduces a unique stacked ensemble classification method that incorporates the original feature maps, ensemble validation data, and base-learner predictions as inputs for the meta-learner. The experimental results demonstrate that this method achieves the highest performance levels.

In both datasets, performed training and evaluation procedures using five different algorithms: RF (Random Forest), DT (Decision Trees), LR (Logistic Regression), SVM (Support Vector Machine), and NB (Naive Bayes). These algorithms were applied in both the base-learner and meta-learner phases, employing various feature extraction models. This work conducts feature extraction using a variety of pre-trained models and combinations to evaluate how they influence classification accuracy. In the context of the SARS-COV-2 CT scan dataset, this study investigates the impact of using individual pre-trained models and their combined application in feature extraction on the performance of all five classification algorithms. Consequently, it becomes evident that when compared to other pre-trained models, the combination of VGG16 and DenseNet121 exhibits superior performance in the context of LR.In this study, LR showcased exceptional performance when applied to the SARS-COV-2 CT scan dataset, whereas RF, LR, and SVM demonstrated outstanding results when dealing with the ISIC Archive datasets. Certainly, it is evident that Logistic Regression and Support Vector Machine achieved their highest accuracy rates, recording 88% and 87%, respectively. This noteworthy performance was observed when both of these algorithms were applied to a specific pre-trained model, namely ResNet-101.

Chapter 9

Conclusion

This research introduces a feature extraction and classification model designed to optimize the accurate diagnosis of COVID-19 and skin cancer within the context of image data sets. This approach for feature extraction techniques harnesses deep learning features obtained from pre-trained models. Subsequently, the resulting fused feature vector is integrated into the stacked ensemble approach, particularly when utilizing combinations of two or three pre-trained models. Initially, base-learner predictions are made, followed by concatenation the original feature map and ensemble validation data, ultimately feeding into the meta-learner stage for the final prediction. The base learner was trained and tested with a pair of datasets: the SARS-CoV-2 CT Scan and the ISIC Archive. The stacked LR technique produced the greatest classification accuracy for the SARS-CoV-2 CT Scan dataset, whereas the stacked RF, LR, and SVM approaches surpassed others for the ISIC Archive dataset.

9.1 Future Work

This research presented a stacked ensemble methodology for the classification of COVID-19 and skin cancer infections. Certain limitations in the study will be subject to future improvements. Here are the constraints in this work:

- Developing a hybrid model for the purpose of addressing both COVID and skin cancer, and subsequently conducting a comparative analysis with our proposed model.

- This work method has exclusively undergone testing on covid-19 and skin cancer dataset. Therefore, in order to extend its applicability to further datasets, additional research and investigation are warranted.

- This study approach, relied on five established machine learning models to construct this stacked ensemble model. Additionally, we utilized pre-trained models with a fixed input size of 224X224 for deep-learning feature extraction. Nevertheless, in future research, there is potential to broaden the scope of this method by adapting it to various pre-trained CNN architectures that may have different input sizes.

Bibliography

[1] Maha M Alshammari, Afnan Almuhanna, and Jamal Alhiyafi. Mammography image-based diagnosis of breast cancer using machine learning: a pilot study. *Sensors*, 22(1):203, 2021.

[2] Muhammad EH Chowdhury, Tawsifur Rahman, Amith Khandakar, Rashid Mazhar, Muhammad Abdul Kadir, Zaid Bin Mahbub, Khandakar Reajul Islam, Muhammad Salman Khan, Atif Iqbal, Nasser Al Emadi, et al. Can ai help in screening viral and covid-19 pneumonia? *Ieee Access*, 8:132665–132676, 2020.

[3] C Fanconi. Skin cancer: malignant vs. benign-processed skin cancer pictures of the isic archive, 2019. [Online].

[4] Mudasir A Ganaie, Minghui Hu, AK Malik, M Tanveer, and PN Suganthan. Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 115:105151, 2022.

[5] Flavia Grignaffini, Francesco Barbuto, Lorenzo Piazzo, Maurizio Troiano, Patrizio Simeoni, Fabio Mangini, Giovanni Pellacani, Carmen Cantisani, and Fabrizio Frezza. Machine learning approaches for skin cancer classification from dermoscopic images: A systematic review. *Algorithms*, 15(11):438, 2022.

[6] Joost N Kok, Egbert J Boers, Walter A Kosters, Peter Van der Putten, and Mannes Poel. Artificial intelligence: definition, trends, techniques, and cases. *Artificial intelligence*, 1:270–299, 2009.

[7] Michael L Littman, Ifeoma Ajunwa, Guy Berger, Craig Boutilier, Morgan Currie, Finale Doshi-Velez, Gillian Hadfield, Michael C Horowitz, Charles Isbell, Hiroaki Kitano, et al. Gathering strength, gathering

storms: The one hundred year study on artificial intelligence (ai100) 2021 study panel report. *arXiv preprint arXiv:2210.15767*, 2022.

[8] Palak Mahajan, Shahadat Uddin, Farshid Hajati, and Mohammad Ali Moni. Ensemble learning for disease prediction: A review. In *Healthcare*, volume 11, page 1808. MDPI, 2023.

[9] Ibomoiye Domor Mienye and Yanxia Sun. A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access*, 10:99129–99149, 2022.

[10] Mohanad Mohammed, Henry Mwambi, Innocent B Mboya, Murtada K Elbashir, and Bernard Omolo. A stacking ensemble deep learning approach to cancer type classification based on tcga data. *Scientific reports*, 11(1):15626, 2021.

[11] Dominik Müller, Iñaki Soto-Rey, and Frank Kramer. Multi-disease detection in retinal imaging based on ensembling heterogeneous deep learning models. In *German Medical Data Sciences 2021: Digital Medicine: Recognize–Understand–Heal*, pages 23–31. IOS Press, 2021.

[12] Dominik Müller, Iñaki Soto-Rey, and Frank Kramer. An analysis on ensemble learning optimized medical image classification with deep convolutional neural networks. *Ieee Access*, 10:66467–66480, 2022.

[13] Anil Narassiguin. *Ensemble learning, comparative analysis and further improvements with dynamic ensemble selection.* PhD thesis, Université de Lyon, 2018.

[14] Duc-Khanh Nguyen, Chung-Hsien Lan, and Chien-Lung Chan. Deep ensemble learning approaches in healthcare to enhance the prediction and diagnosing performance: the workflows, deployments, and surveys on the statistical, image-based, and sequential datasets. *International Journal of Environmental Research and Public Health*, 18(20):10811, 2021.

[15] Rekha Rajagopal. Comparative analysis of covid-19 x-ray images classification using convolutional neural network, transfer learning, and machine learning classifiers using deep features. *Pattern Recognition and Image Analysis*, 31(2):313–322, 2021.

[16] Nagur Shareef Shaik and Teja Krishna Cherukuri. Transfer learning based novel ensemble classifier for covid-19 detection from chest ct-scans. *Computers in Biology and Medicine*, 141:105127, 2022.

[17] BH Shekar and Habtu Hailu. An efficient stacked ensemble model for the detection of covid-19 and skin cancer using fused feature of transfer learning and handcrafted methods. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 11(3):878–894, 2023.

[18] Pedro Silva, Eduardo Luz, Guilherme Silva, Gladston Moreira, Rodrigo Silva, Diego Lucio, and David Menotti. Covid-19 detection in ct images with deep learning: A voting-based scheme and cross-datasets analysis. *Informatics in medicine unlocked*, 20:100427, 2020.

[19] Dan Xue, Xiaomin Zhou, Chen Li, Yudong Yao, Md Mamunur Rahaman, Jinghua Zhang, Hao Chen, Jinpeng Zhang, Shouliang Qi, and Hongzan Sun. An application of transfer learning and ensemble learning techniques for cervical histopathology image classification. *IEEE Access*, 8:104603–104618, 2020.