



## Article

# Enhancing Diagnostic Accuracy for Skin Cancer and COVID-19 Detection: A Comparative Study Using a Stacked Ensemble Method

Hafza Qayyum <sup>1,\*</sup> , Syed Tahir Hussain Rizvi <sup>2</sup> , Muddasar Naeem <sup>3,\*</sup> , Umamah bint Khalid <sup>1</sup> , Musarat Abbas <sup>1</sup> and Antonio Coronato <sup>3</sup>

<sup>1</sup> Department of Electronics, Quaid-i-Azam University, Islamabad 45320, Pakistan

<sup>2</sup> Department of Electrical Engineering and Computer Science, University of Stavanger, 4036 Stavanger, Norway

<sup>3</sup> Università Department of Computer Engineering, Università Telematica Giustino Fortunato, 82100 Benevento, Italy

\* Correspondence: qayyumhafza@gmail.com (H.Q.); m.naeem@unifortunato.eu (M.N.)

**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 (ML) and ensemble techniques to improve the accuracy and efficiency of medical image diagnosis for critical diseases such as COVID-19 (grayscale images) and skin cancer (RGB images). In this paper, a stacked ensemble learning approach is proposed to enhance the precision and effectiveness of diagnosis of both COVID-19 and skin cancer. The proposed method combines pretrained models of convolutional neural networks (CNNs) including ResNet101, DenseNet121, and VGG16 for feature extraction of 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 architectures. The feature vectors obtained through transfer learning are then fed into base-learner models consisting of five different ML algorithms. In the final step, the predictions from the base-learner models, the ensemble validation dataset, and the feature vectors extracted from neural networks are assembled and applied as input for the meta-learner model to obtain final predictions. The performance metrics of the stacked ensemble model show high accuracy for COVID-19 diagnosis and intermediate accuracy for skin cancer.

**Keywords:** COVID-19; skin cancer; stacking; feature vector; ensemble learning; machine learning algorithms



**Citation:** Qayyum, H.; Rizvi, S.T.H.; Naeem, M.; Khalid, U.B.; Abbas, M.; Coronato, A. Enhancing Diagnostic Accuracy for Skin Cancer and COVID-19 Detection: A Comparative Study Using a Stacked Ensemble Method. *Technologies* **2024**, *12*, 142. <https://doi.org/10.3390/technologies12090142>

Academic Editors: Juvenal Rodriguez-Resendiz, Gerardo I. Pérez-Soto, Karla Anhel Camarillo-Gómez, Saul Tovar-Arriaga and Tamás Haidegger

Received: 19 June 2024

Revised: 16 August 2024

Accepted: 23 August 2024

Published: 27 August 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In recent times, there has been notable and efficient advancement in the domain of the automated analysis of medical images [1]. Modern imaging techniques rely on images with a high resolution to give radiologists multifaceted views to aid in clinical diagnoses, precise predictions, and patient treatment. Ultrasound, endoscopy, X-ray, computed tomography (CT), and magnetic resonance imaging (MRI) are the most common methods for capturing medical images [2]. Currently, numerous studies have emerged concerning the categorization and identification of illnesses through medical imaging. Even though these models have demonstrated promising outcomes, the medical domain still demands enhanced precision [3].

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 [4]. The focus of the proposed study is on utilizing stacked ensemble learning techniques to enhance the accuracy of COVID-19 and skin cancer detection. The main aim of this research is to develop a generalized model using an ensemble learning methodology resulting in precise and accurate predictions for grayscale and RGB medical images.

Artificial intelligence (AI) involves methods and algorithms for performing tasks smartly by learning from previous data or examples, like planning and learning from language [5]. Machine learning (ML) and deep learning (DL) are pivotal branches of AI. ML methods involve training algorithms to learn from data and make predictions and decisions based on patterns identified during training [6]. Deep learning (DL), a subset of ML, employs intricate neural networks with multiple layers to automatically extract complex features from data [7]. This facilitates advancements in tasks like image recognition and natural language processing [8]. The use of ML and DL technologies for the detection and diagnosis of COVID-19 disease has significant effects and has been used in several investigations [9]. These AI-empowered techniques have a considerable tangible capacity for providing an accurate and efficient intelligent system for detecting and estimating the severity of COVID-19. The performance of AI may further be improved by considering safety features of underlying environments [10]. Moreover, AI-based systems can be combined with other technologies such as 5G, cloud storage, and the Internet of Things (IoT) for other variants of COVID-19 epidemics to eliminate geographical issues in the rapid estimation of disease severity, lower treatment costs, and perform epidemic management and immediate epidemic control [11]. Both ML and DL play vital roles in transforming industries by enabling computers to learn and adapt from experience, enhancing their performance over time [12].

The purposeful design and mixing of many models, such as learners, to address specific computational intelligence challenges is known as ensemble learning (EL) [13]. It is a method for combining numerous models to achieve better generalization ability [14]. The ensemble machine learning technique combines different classification algorithms, called base learners (also called base models), to produce a single improved classification model. The main idea is that 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 learner's error in prediction. The predicted output of the base learners is utilized as input for the meta-learner. This generalized ability and the accuracy of prediction results obtained using this technique beat the results of a single ML setup [15]. Ensemble strategies have changed with time to improve model generalization in learning. These strategies may be divided into three categories, bagging, boosting, and stacking, as follows:

- 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 from 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 to merge the predictions from multiple predictors.
- Boosting is an ML algorithm that involves training multiple models sequentially, with each subsequent model focusing on correcting the errors made by the previous model. Boosting stands as a robust approach that effectively mitigates overfitting [14].
- The stacked ensemble ML 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 [15].

#### *Contribution of Research Work*

This research work proposed a novel approach for the classification of COVID-19 and skin cancer. The main contributions and the uniqueness of this work are as follows:

- To ensure a reliable evaluation of the proposed method, a customized distribution strategy is implemented for sampling each dataset. Departing from the standard data split method, the approach involves a balanced division of data at each stage of model

development. This tailored strategy demonstrates its efficacy in enhancing the overall performance of the model.

- Additionally, the most effective and high-performing CNN variants with a default input size of  $224 \times 224$  (DenseNet-121, ResNet101, VGG16, and a combination of these networks) are utilized as feature extractors. These pretrained models are configured to exclude their fully connected layers to make them suitable for feature extraction. The extracted high-level features from these models are combined as a feature vector that can be valuable for subsequent classification tasks and that also reduces model complexity due to the removal of computationally and memory-intensive fully connected layers.
- The generated feature vector is then further used as input of base-learner models to train them. Five different base-learner models (support vector machine (SVM), linear regression (LR), decision tree (DT), random forest (RF), and naive Bayes (NB)) are used in this work.
- The predictions from the base-learner models are integrated with the initial feature vector and the ensemble validation set to generate a very rich and informative fused vector. This fused feature vector is finally fed to the meta-learner model, which provides a very precise final prediction. Five different meta-learner models (RF, DT, LR, NB, and SVM) are compared by training on the same fused feature vector.
- Most of the state-of-the-art methods apply only the base-learner prediction as the input of the meta-learner. The proposed ensemble technique combines different types and levels of features to improve the generalization capability of the final predictors.

This manuscript is organized as follows: Section 1 introduces the field of ensemble learning and the research question. Section 2 discusses work related to COVID-19 and skin cancer; this section also discusses the concept of ensemble learning. Section 3 presents the proposed pipeline, including the datasets, preprocessing methods, deep convolutional neural network architectures, ensemble learning strategies, and machine learning algorithms. Section 4 reports the experimental results, and a discussion of these results can be found in Section 5. Section 6 concludes this paper and give insights on future work.

## 2. Related Work

Several methods have been proposed in the literature to categorize diseases through medical imaging, with some offering solutions for disease identification. The classification of medical images is a growing field that primarily relies on DL and ML techniques. This section reviews the latest research on medical image classification, especially focusing on datasets related to COVID-19 and skin cancer.

### 2.1. COVID-19

A pretrained VGGNet model and SVM algorithm are employed in [16] to classify X-ray images into three distinct categories: COVID-19, healthy images, and pneumonia. This provided an accuracy of 95.81%.

Similarly, the authors in [17] employed an ensemble architecture to classify CT scan images. Their approach combined various deep learning models, including VGG16, ResNet50, VGG19, ResNet50V2, InceptionV3, InceptionResNetV2, MobileNet, and Xception. A VGG model having 19 layers achieved the highest accuracy of 92%, while all other models were confined to below 90% accuracy.

### 2.2. Skin Cancer

The work in [18] provided a comprehensive systematic review of the latest research on the usage of ML algorithms for dermoscopic image categorization of skin cancer. Overall, the review provided valuable insights into the latest research on using ML for skin cancer diagnosis. It can be a useful resource for researchers and practitioners in the field. The results of these studies vary depending on the used algorithm, dataset, and evaluation

metrics. However, the review explored that the recent ML and DL models show high potential in skin lesion classification.

The paper [19] introduced a novel approach to combine ML and DL techniques for skin cancer detection that outperformed both expert dermatologists and other state-of-the-art deep learning models. Input features for the classifiers were synthesized from image processing techniques such as contourlet transform and local binary pattern histogram. Pretrained CNNs were used for transfer learning and automatic feature extraction. This method achieved good detection accuracy for benign and malignant forms of cancer.

Moreover, the authors in [20] proposed a deep neural network-based system for skin lesion classification. They used five state-of-the-art architectures as base models, including ResNeXt, SeResNeXt, ResNet, Xception, and DenseNet. They used average and weighted average ensemble models. This method achieved excellent performance for highly imbalanced seven-class datasets. The ensemble models showed higher distinguishing capability compared to others. The average ensemble model achieved a recall score of 93%, and the weighted average ensemble model achieved a recall score of 94%.

### 2.3. COVID-19 and Skin Cancer

A reproducible pipeline for medical image classification was employed by [21] that they used to analyze the effect of various ensemble learning approaches on the performance. That research work used a dataset related to skin and COVID-19 on various ensemble learning techniques. The paper utilized an ensemble learning approach that combines multiple deep CNNs to improve the accuracy of classification. The outcomes indicated that the stacking exhibited a substantial enhancement in performance, with a notable increase of up to 13% in the F1-score.

A novel approach was proposed in [15] for feature extraction by combining transfer learning from the DenseNet model and six handcrafted techniques to catch more comprehensive and intricate features. In the evaluation of the ISIC dataset and COVID-19 CT Scan data, RF exhibited the highest efficiency for skin detection, and LR outperformed the other algorithms across every evaluation metric during COVID-19 detection.

During the literature review, we observed that there is no nasal association between COVID-19 and skin cancer. However, from a clinical point of view, both are highly associated with the immune system. COVID-19 affects the immune system, and people with low immunity are more prone to the recurrence of melanoma. Melanoma is an aggressive type and can recur even after surgery. During the COVID-19 pandemic, many skin infections were reported which were initially considered as a melanoma [22]. Later, the findings reveal that it was not a melanoma.

### 2.4. Ensemble Learning

Traditional machine learning methods fail when dealing with complex data because they cannot capture multiple characteristics and underlying structures. Building an efficient model for knowledge discovery is crucial in data mining. Ensemble learning aims to integrate data fusion, data modeling, and data mining into a unified framework. In ensemble learning, a set of features is first extracted using various transformations. Finally, ensemble learning combines the valuable knowledge from the extracted results to achieve knowledge discovery and improve predictive performance through adaptive voting schemes [23].

The ensemble approach to ML uses several homogeneous or heterogeneous algorithms for classification, known as base learners. The base learners are called weak learners and are obtained by a base learning algorithm based on training data to create a classification model [24]. The outputs obtained from base learners are then fed as input to the meta-learners. The meta-learner 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, as depicted in Figure 1.

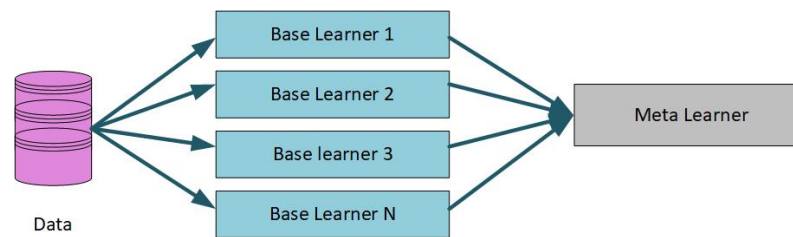


**Figure 1.** Ensemble learning approach.

### 2.5. Types of Ensemble Learning

There is no definitive taxonomy of ensemble learning. Usually, four methods used for ensemble learning are described: bagging, boosting, stacking, and error-correcting output codes [25].

- **Bagging:** It is the first effective and the simplest method of ensemble learning. It was originally designed for classification and is usually applied to decision tree models, but it can be used for regression. The method involves using multiple versions of a training set through the process of bootstrapping. Each of these datasets is used to train a different model. The outputs of the models are combined by averaging (for regression) or voting (for classification) to create a single output. Bagging is only effective for non-linear models.
- **Boosting:** This model-averaging method is a widely used ensemble technique. It can be used for both classification and regression tasks. In this method, weak classifiers are created iteratively, with each one trained on a dataset where misclassified points are given more weight. The final model is created by combining the outputs of the successive models using voting or averaging. Boosting can sometimes fail to generate a classifier as accurate as a single classifier built from the same data, indicating overfitting.
- **Stacking:** It is a distinct way of combining multiple models that introduces the concept of a meta-learner. It is less widely used compared to bagging and boosting. Stacking is normally used to combine models of different types. It splits the training set into two disjoint sets. The method trains several base learners on the first part and tests on the second part. The predictions of the base learners are taken as the inputs to train the meta-learners.
- **Error-correcting output codes:** It is a technique that enhances the performance of classification algorithms in multiclass learning scenarios. In this approach, the multiclass dataset is broken down into multiple independent two-class problems. The algorithm is then applied to each of these problems, and the outputs from the resulting classifiers are combined to make the final prediction.

## 3. Methodology Overview

The main objective of the proposed method is to maximize the performance of the required task of detection by training various models on a particular dataset. In the proposed approach, a combined feature extraction stage and a stacked ensemble technique are employed to address the challenges of identifying COVID-19 and skin cancer disease.

A comprehensive overview of the different sequential stages of the proposed methodology is illustrated in Figure 2. Commencing with image acquisition, the input data are first passed to the feature extraction stage, employing various pretrained learning models. Subsequently, a feature vector is generated through the combination of features derived from these extractions, which is then passed to the base learners. These base learners are trained on this feature vector, encompassing five diverse ML algorithms: RF, DT, LR, NB, and SVM. 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.

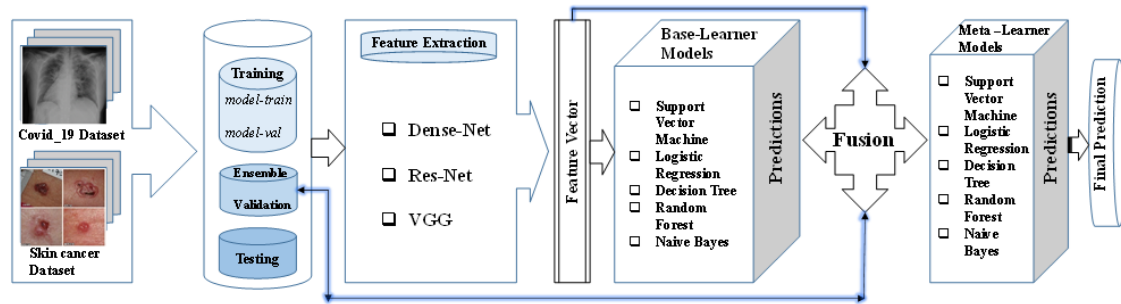


**Figure 2.** Proposed architecture.

### 3.1. Datasets

We utilized publicly accessible image datasets for the performance evaluation of the proposed methodology: SARS-CoV-2 CT scan dataset [26] and ISIC Archive dataset [27]. Visual examples showcasing images from each dataset, encompassing different classes, are shown in Figures 3 and 4.

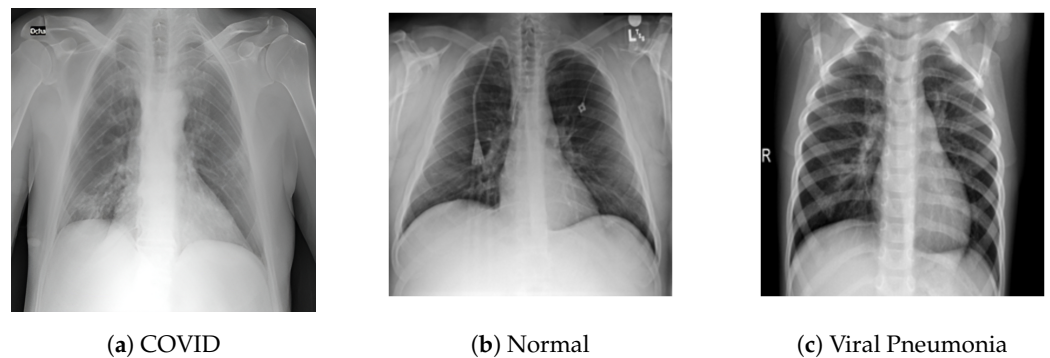


**Figure 3.** Sample images from the SARS-CoV-2 X-ray dataset illustrating instances of COVID-19 cases, healthy control cases and viral pneumonia.

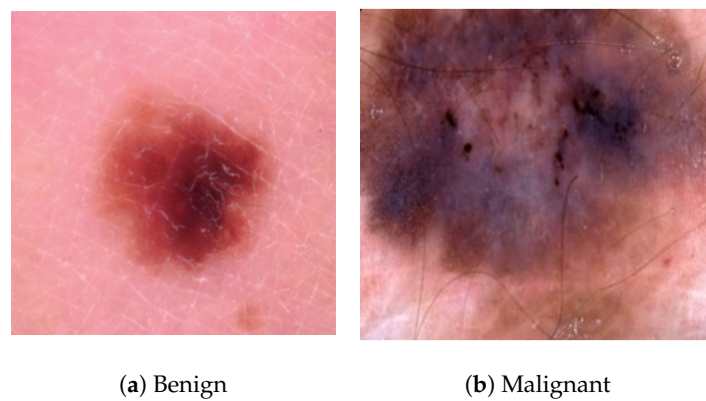


**Figure 4.** Sample images from the ISIC Archive dataset showing both benign and malignant cases.

#### 3.1.1. COVID-19 Radiography Database

X-ray images are important for visualizing medical issues, and they are used a lot in healthcare. They are also used instead of certain tests for COVID-19. Researchers from different places like Qatar, Doha, Dhaka, Bangladesh, Pakistan, and Malaysia collected X-ray images of people's chests who had COVID-19, as well as some who were healthy and others with viral pneumonia. This dataset has 2905 grayscale pictures. There are 220 pictures of people with COVID-19, 1345 with viral pneumonia, and 1340 healthy pictures. The distribution of these data for our experimentation is shown in Figure 5.

| 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 5.** Overview of the datasets used along with their descriptions and how the samples were distributed.

### 3.1.2. 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 performed 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 [21]. The ISIC Archive collection includes 3297 skin cancer images, 1800 of which are classified as benign and 1497 of which are malignant. An overview of the total number of samples and the distribution of the samples are shown in Figure 5.

### 3.2. Data Splitting

To ensure an assessment of the proposed scheme, we employed the following distribution strategy for the sampling of each dataset, as shown in Figure 6. For the training of base-learner models, 75% of the total dataset was divided into two parts (80% for training of base-learner models (called ‘model-train’) and 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.

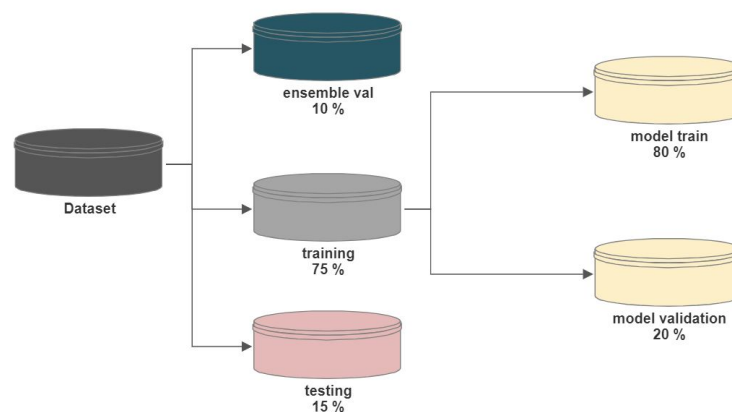


**Figure 6.** Dataset distribution strategy.

### 3.3. Data Preprocessing

Preprocessing plays a crucial role in computer vision applications, serving various purposes such as performing noise reduction, highlighting relevant image features for recognition tasks, and aiding in the training of learning models. In this study, a straightforward approach involving the normalization of pixel intensities within the [0, 1] range was employed. This preprocessing step is essential to ensure the model’s convergence during the training phase [28]. The configuration for image data preparation includes the techniques such as zoom, brightness, and normalization. Furthermore, the image sizes were scaled down to  $224 \times 224$  pixels, the standard input size of selected feature extractors.

Labels were assigned to three classes in the context of COVID-19 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.

### 3.4. Feature Extraction Technique

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 (DenseNet-121, ResNet101, and VGG16) were utilized. These models were already trained on the ImageNet dataset. These models were fine-tuned on our dataset using previous weights; this technique is called transfer learning.

Transfer learning focuses on the idea of preserving the acquired knowledge from one problem and leveraging it for solving distinct yet related problems. TL enables us to efficiently employ pretrained deep learning models that have been trained on extensive and publicly accessible datasets. We used pretrained VGG16, DenseNet-121, and ResNet101 models, loaded using TensorFlow and Keras. We configured these models for our approach by excluding their fully connected layers, rendering them suitable for feature extraction. These extracted high-level features are used as valuable inputs for subsequent classification tasks using base-learner models.

### 3.5. Fusion Technique

The process of combining several feature vectors produced from diverse methodologies in disciplines such as computer vision and other ML applications is known as feature fusion. Previously extracted deep features from three different pretrained models (DenseNet121, ResNet101, and VGG16) were used in this context. We used these distinct feature vectors and their combinations to generate a feature vector. Different base-learner models were trained on this vector. Finally, the meta-learner model was trained on the fused feature vector, made up of a concatenated feature vector derived from the predictions of five base-learner models as well as ensemble validation.

### 3.6. Classification Models

In this research work, we employed five foundational ML classification algorithms (RF, DT, NB, LR, and SVM) as base-learner models. Afterwards, we used a stacking technique at level 1 which combined all five of these algorithms. Ensemble validation was used at level 2 for the stacking of meta-learners. This implementation was carried out using the Sklearn library.

### 3.7. Stacked Ensemble Learning

The ensemble approach to ML uses several 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-learner models. The meta-learner is an algorithm that learns to minimize the loss of base-learner models. The prediction output of base learners is used as input to the meta-learner as depicted in Figure 1.

In this research work, we used a first-level stacking strategy for the base-learner phase using five ML algorithms. Furthermore, we used second-level stacking in the meta-learner phase. Previous works solely used base-learner predictions to direct the meta-learner's final decision. In this work, we combined the predictions of the base learners with ensemble validation data to serve as input for the meta-learner. All five base-learner algorithms were used and evaluated for the meta-learner to make a final prediction. In this analysis,

we excluded both boosting and bagging techniques. Boosting is not feasible to apply for image classification because of the dramatic increase in training hours and the complexity of the model, which makes it harder to interpret. According to [21], stacking is a highly effective technique for image classification that can lead to a substantial improvement in performance. Additionally, cross-validation-based bagging has also demonstrated a significant enhancement in performance in some research works, closely competing with stacking. This suggests that stacking is the preferred approach, with bagging being a strong contender.

Another research work [29] verified the superiority of the stacking method as the optimal choice for achieving superior ensemble-based results. Therefore, in this work, we utilized stacking while paying close attention to high-performing strategies, aligning with the findings of both studies and ensuring the originality in this work.

### 3.8. Performance Evaluation Metrics

In this research, we employed various performance evaluation metrics to assess the effectiveness of the models. These metrics included accuracy ( $Ac$ ), precision ( $Pr$ ), recall ( $Re$ ), and the  $F1$ -score. Accuracy describes 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 (1). Precision, as defined in Equation (2), represents the positive predictive value, which represents the fraction of true positives divided by the overall number of actual true instances. Conversely, recall, which is also referred to as sensitivity, is expressed in Equation (3); it measures the fraction of true positives relative to the total number of predicted true instances. Finally, Equation (4) outlines the formula for the  $F1$ -score, a metric that strikes a balance between precision and recall by computing their harmonic mean. We used the Scikit-learn library for evaluation purposes.

$$Ac = \frac{TruePositive + TrueNegative}{TruePositive + FalsePositive + FalseNegative + TrueNegative} \quad (1)$$

$$Pr = \frac{TruePositive}{TruePositive + FalsePositive} \quad (2)$$

$$Re = \frac{TruePositive}{TruePositive + FalseNegative} \quad (3)$$

$$F1\text{-score} = 2 \cdot \frac{Pr \cdot Re}{Pr + Re} \quad (4)$$

## 4. Results and Analysis

In this section, we present a comprehensive analysis of the experimental outcomes achieved through the application of the proposed scheme on two distinct datasets: SARS CoV-2 CT scans and the ISIC dataset. The proposed methodology comprises two pivotal stages: the base-learner and the meta-learner stages. Across both datasets, we conducted training and evaluation processes on five distinct algorithms, namely, RF, DT, LR, SVM, and NB, within both the base-learner and meta-learner phases with different feature extraction models.

The initial step in the proposed method involves the construction of a feature space. This is accomplished by using three distinct pretrained deep learning models: DenseNet-121, ResNet-101, and VGG16. Additionally, we explored the combinations of these models, namely, DenseNet-121 and ResNet-101; DenseNet-121 and VGG16; ResNet-101 and VGG16; and the combination of all three models, DenseNet-121, ResNet-101, and VGG16. We performed feature extraction using these various pretrained models and combinations of these to assess their impact on classification accuracy. The results of these evaluations, specifically the accuracy of the five classification algorithms, are visually represented in Tables 1 and 2 for COVID-19 and in Tables 3 and 4 for skin cancer detection, respectively.

**Table 1.** Accuracy of five classification algorithms for different combinations of pretrained learning feature extraction models and base-learner prediction performance results on COVID-19 dataset. **Bold represents the best-performing configuration.**

| Models | VGG16        | DENSENET121 | RESNET101 | DENSENET121 +<br>RESNET101 | DENSENET121 +<br>VGG16 | RESNET101 +<br>VGG16 | DENSENET121 +<br>RESNET101 +<br>VGG16 |
|--------|--------------|-------------|-----------|----------------------------|------------------------|----------------------|---------------------------------------|
| SVM    | 95.63        | 91.72       | 95.40     | 92.64                      | 94.25                  | 91.49                | 94.48                                 |
| LR     | <b>96.78</b> | 94.71       | 96.09     | 95.40                      | 96.55                  | 94.48                | 95.86                                 |
| DT     | 83.91        | 87.13       | 85.06     | 88.74                      | 82.53                  | 86.44                | 81.15                                 |
| RF     | 92.18        | 91.03       | 92.18     | 91.03                      | 91.03                  | 91.95                | 89.89                                 |
| NB     | 91.26        | 84.83       | 90.34     | 90.57                      | 91.26                  | 90.34                | 91.03                                 |

**Table 2.** Accuracy of five classification algorithms for different combinations of pretrained deep learning feature extraction models and meta-learner prediction performance results on COVID-19 dataset. **Bold represents the best-performing configuration.**

| Models | VGG16 | DENSENET121 | RESNET101    | DENSENET121 +<br>RESNET101 | DENSENET121 +<br>VGG16 | RESNET101 +<br>VGG16 | DENSENET121 +<br>RESNET101 +<br>VGG16 |
|--------|-------|-------------|--------------|----------------------------|------------------------|----------------------|---------------------------------------|
| SVM    | 97.01 | 95.17       | <b>97.24</b> | 94.94                      | 96.78                  | 96.78                | 96.32                                 |
| LR     | 97.01 | 94.71       | 96.09        | 95.40                      | 96.55                  | 96.09                | 96.32                                 |
| DT     | 94.02 | 91.03       | 95.17        | 89.43                      | 94.25                  | 90.57                | 91.49                                 |
| RF     | 96.55 | 93.56       | 96.78        | 94.02                      | 96.55                  | 94.48                | 93.79                                 |
| NB     | 96.05 | 92.64       | 93.33        | 90.57                      | 93.33                  | 90.34                | 95.63                                 |

**Table 3.** Accuracy of five classification algorithms for different combinations of pretrained deep learning feature extraction models and base-learner prediction performance results on ISIC dataset. **Bold represents the best-performing configuration.**

| Models | VGG16 | DENSENET121 | RESNET101    | DENSENET121 +<br>RESNET101 | DENSENET121 +<br>VGG16 | RESNET101 +<br>VGG16 | DENSENET121 +<br>RESNET101 +<br>VGG16 |
|--------|-------|-------------|--------------|----------------------------|------------------------|----------------------|---------------------------------------|
| SVM    | 83.40 | 82.39       | 86.64        | 81.17                      | 83.60                  | 82.19                | 80.16                                 |
| LR     | 82.19 | 83.00       | <b>87.85</b> | 84.82                      | 80.77                  | 84.41                | 81.98                                 |
| DT     | 70.04 | 75.91       | 71.05        | 69.84                      | 71.26                  | 72.27                | 76.32                                 |
| RF     | 80.97 | 81.98       | 84.21        | 82.79                      | 81.17                  | 83.81                | 78.95                                 |
| NB     | 79.35 | 77.94       | 79.51        | 78.74                      | 78.14                  | 78.14                | 76.52                                 |

**Table 4.** Accuracy of five classification algorithms for different combinations of pretrained deep learning feature extraction models and meta-learner prediction performance results on ISIC dataset. **Bold represents the best-performing configuration.**

| Models | VGG16 | DENSENET121 | RESNET101    | DENSENET121 +<br>RESNET101 | DENSENET121 +<br>VGG16 | RESNET101 +<br>VGG16 | DENSENET121 +<br>RESNET101 +<br>VGG16 |
|--------|-------|-------------|--------------|----------------------------|------------------------|----------------------|---------------------------------------|
| SVM    | 84.50 | 83.28       | 86.79        | 82.37                      | 83.60                  | 84.19                | 80.85                                 |
| LR     | 82.37 | 83.59       | <b>87.89</b> | 85.72                      | 82.37                  | 83.59                | 81.99                                 |
| DT     | 77.81 | 76.90       | 82.67        | 79.94                      | 76.90                  | 79.94                | 77.81                                 |
| RF     | 80.99 | 81.98       | 84.27        | 83.37                      | 81.72                  | 83.91                | 78.98                                 |
| NB     | 54.71 | 54.71       | 54.71        | 54.71                      | 54.71                  | 54.71                | 54.71                                 |

Table 1 demonstrates the effectiveness of the base-learner models in detecting COVID-19 using the SARS-CoV-2 CT scan database. First, we examined the influence of both the individual pretrained models and their combined usage for feature extraction on the performance of all five classification algorithms. Thus, it can be observed that the individual VGG16 and combination of VGG16 with DenseNet121 yield superior performance on LR compared to the other pretrained models. LR exhibits superior performance relative to other ML algorithms during the base-learner stage. Accuracies of 95.63%, **96.78%**, 83.91%, 92.18%,

and 91.26% are observed in the base-learner stage from SVM, **LR (Best)**, DT, RF, and NB using VGG16. During the base-learner phase with the DenseNet121 model, the recorded accuracies for different classifiers are 91.72%, 94.71%, 87.13%, 91.03%, and 84.83% for SVM, LR, DT, RF, and NB, respectively. In the initial training phase with the ResNet101 model, results are 95.40%, 96.09%, 85.06%, 92.18%, and 90.34% for SVM, LR, DT, RF, and NB. Moreover, the performance of all pretrained models and their combination in the base-learner stage is also an excellent outcome; the LR scores were the highest in all the implemented pretrained model feature extraction techniques.

Certainly, in summary, LR consistently yielded high accuracy on applied pretrained models in base-learner stage. However, when further analyzed, it was observed that LR achieved its **highest accuracy, 96.78%**, for VGG16 as compared to other feature extraction methods; the second highest was a 96.55% accuracy on a specific combination of pretrained models, that is, DenseNet-121 and VGG-16.

Table 2 illustrates the effectiveness of the meta-learner models for detecting COVID-19 using base-learner prediction and ensemble validation as input. As Table 2 shows, the algorithm's performance for all cases in the meta-learner phase is better than in the base-learner phase. This implies that using the stacking ensemble method is a promising technique to improve classification accuracy. Moreover, the performance of all pretrained models and their combination in the meta-learner stage is also an excellent outcome. SVM and LR scores the highest in all the implemented pretrained model feature extraction techniques.

In summary, SVM consistently yielded high accuracy on applied pretrained models in the meta-learner phase. However, it can also be observed that SVM achieved its **highest accuracy, 97.24%**, for ResNet-101 as compared to other feature extraction methods.

Figure 7 illustrates the evolution of the accuracy of ML algorithms in each feature extraction method from the base-learner to the meta-learner stage.

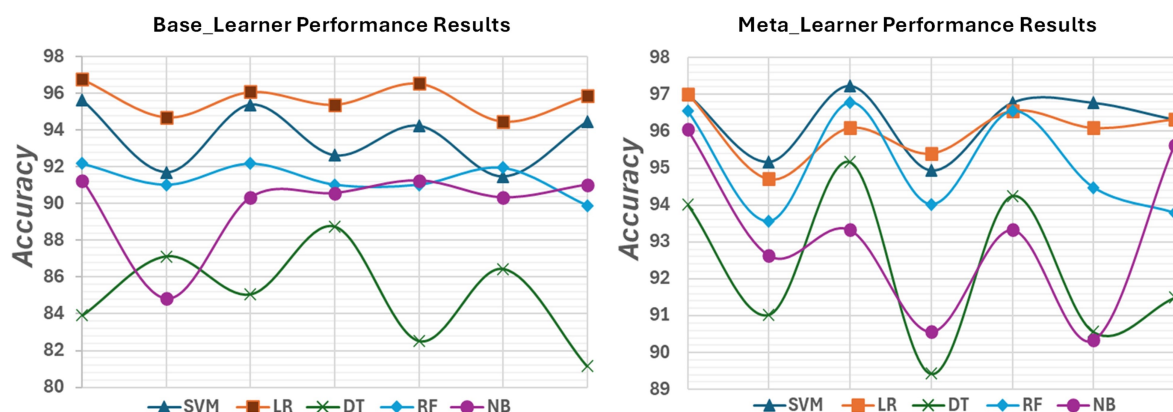


**Figure 7.** Graphical description of the performance results for all implemented ML algorithms using the SARS-CoV-2 CT scan.

The same experiment was carried out utilizing the ISIC Archive dataset to identify skin cancer cases at the base-learner level. Results of this experiment are listed in Table 3. In the case of the base-learner model, LR again demonstrated superior performance compared to other combinations of models, similar to the SARS-CoV-2 CT scan dataset; furthermore, SVM also excelled in the ISIC Archive dataset. LR tends to perform better when there are fewer noisy variables compared to explanatory factors or when the number of noisy variables is equal to or lower than the number of explanatory factors. SVM's strength lies in its ability to handle high-dimensional data effectively. The main objective in utilizing two distinct datasets to assess this methodology was to evaluate how the algorithms perform in the same scenarios. Upon closer examination of base-learner models, it became evident that **LR** and **SVM** achieved their highest accuracy rates, reaching **87.85%** and **86.64%**, respectively, when applied to a particular pretrained model, namely, ResNet-101. This performance surpasses other feature extraction techniques and classification algorithms.

Table 4 illustrates the effectiveness of the meta-learner models in detecting skin cancer and shows that the algorithm's performance in the meta-learner phase is better than in the base-learner phase for every utilized algorithm except NB. This implies that using the two-level stacking method is a promising technique to improve classification accuracy, but the meta-learning stage is unsuitable for the NB algorithm. Moreover, SVM and LR score the highest in all the implemented pretrained model feature extraction techniques. It can be observed that LR achieved its highest accuracy, 87.89%, for ResNet-101 as compared to other feature extraction methods.

Figure 8 illustrates the evolution of accuracy among ML algorithms at each feature extraction, from the base learner to the meta-learner, for the ISIC dataset. It demonstrates that not all ML algorithms exhibit improvement in accuracy. For instance, the accuracy of the NB algorithm decreases as it progresses from the base-learner to the meta-learner stage because of its simple probabilistic approach that assumes independence between features. As the algorithm progresses to more advanced stages like meta-learning, where it may encounter more complex feature interactions, its accuracy further decreases.

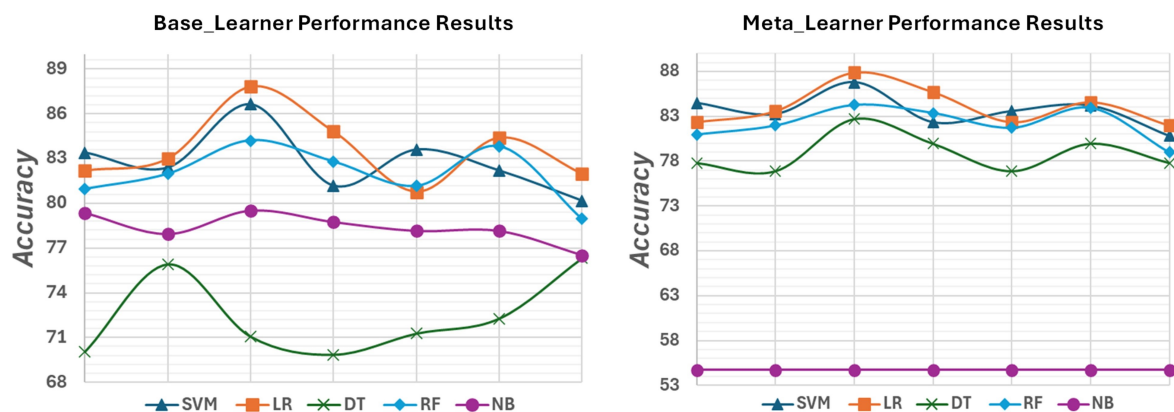


**Figure 8.** Graphical description of the performance results for all implemented ML algorithms using the ISIC archive dataset.

This experiment followed the same technique as the previous one on the SARS-CoV-2 CT scan dataset. However, in contrast to the outcomes observed on the ISIC dataset, it is important to note that the LR algorithm did not consistently outperform the other five algorithms during the base-learner phase. This highlights the fact that there is not 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.

As per Figure 9, when rigorously examined, it shows that in COVID-19 and skin cancer databases, ResNet101 outperformed all models for feature extraction. Likewise, in the field of COVID-19 classification, ResNet101 in conjunction with the SVM algorithm obtained the best accuracy over the other ML algorithms. However, in skin cancer classification, ResNet101 embedding LR showed the best accuracy.

The total training time for the complete analysis can be visualized from the following distribution chart: Experiments relevant to SARS CoV-2 CT scans took a total of 16 hours. Experiments relevant to ISIC took less than 12 h. It has to be noted that the stacking techniques with the ML algorithm and pretrained model do not require extensive additional training time. DenseNet-121 and ResNet-101 revealed a high training time across all pretrained models of 6 hours and 53 min for the SARS CoV-2 CT scans dataset, whereas the VGG16 model had the lowest training time across the SARS CoV-2 CT scans dataset at 16 min. For the skin cancer dataset, the same as in COVID-19 detection, DenseNet-121 and ResNet-101 revealed a high training time across all pretrained models of 9 h and 23 min, whereas the VGG16 model had the lowest training time of 30 min. Further details on

training times for all feature extraction models with distributive datasets are found in Figure 10.

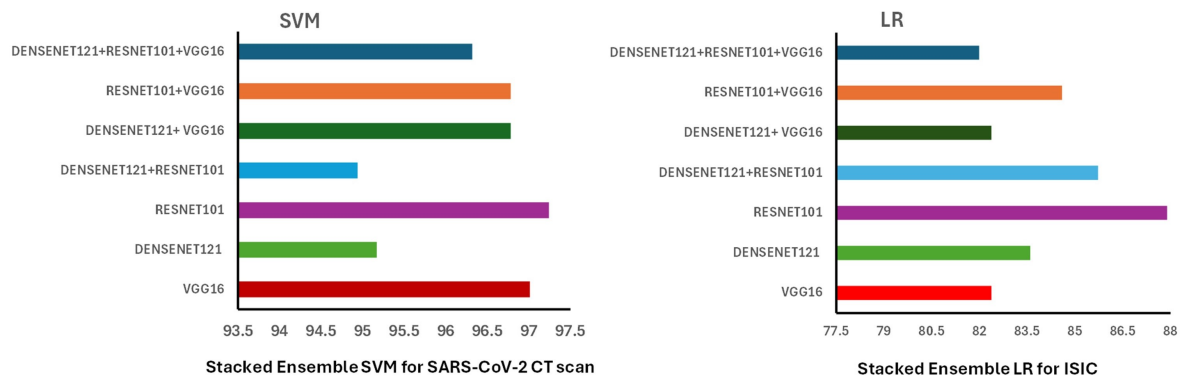


**Figure 9.** Comparative analysis of feature extraction models' performance in COVID-19 and skin cancer datasets.

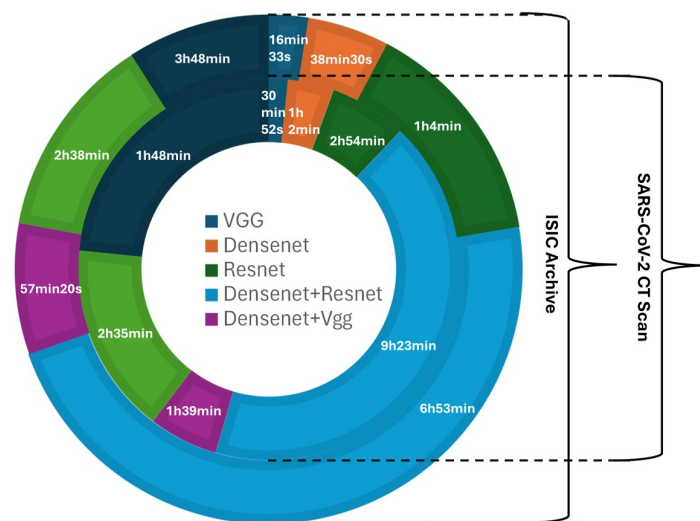


**Figure 10.** Training times for all feature extraction models with datasets of COVID-19 and skin cancer.

Table 5 shows a detailed comparison of all evaluation matrices (accuracy, precision, recall and F1-score) of the final combination of the proposed scheme on the COVID-19 (SARS-CoV-2 CT) and cancer (ISIC) datasets. As mentioned earlier, ResNet101 outperformed all other pretrained models in the task of feature extraction. Furthermore, ResNet101 in conjunction with the SVM algorithm obtained the best accuracy in the field of COVID-19. However, in the case of skin cancer classification, ResNet101 showed the best accuracy when combined with the linear regression model.

**Table 5.** Comparison of best-performing feature extractor and final output of meta-learner models on COVID-19 and ISIC datasets. **Bold** represents the best-performing configuration.

| Base Model | SARS-CoV-2 CT Scan Database |              |              |              | ISIC-Archive Dataset |              |              |              |
|------------|-----------------------------|--------------|--------------|--------------|----------------------|--------------|--------------|--------------|
|            | ResNet101                   |              |              |              | ResNet101            |              |              |              |
|            | Accuracy                    | Precision    | Recall       | F1-Score     | Accuracy             | Precision    | Recall       | F1-Score     |
| SVM        | <b>97.24</b>                | <b>97.78</b> | <b>96.89</b> | <b>97.55</b> | 86.79                | 87.45        | 86.99        | 86.23        |
| LR         | 96.09                       | 96.67        | 95.84        | 96.27        | <b>87.89</b>         | <b>88.56</b> | <b>87.29</b> | <b>87.92</b> |
| DT         | 95.17                       | 96.45        | 95.23        | 95.05        | 82.67                | 82.88        | 82.68        | 82.79        |
| RF         | 96.78                       | 97.01        | 96.55        | 96.55        | 84.27                | 84.98        | 84.43        | 84.99        |
| NB         | 93.33                       | 93.99        | 92.56        | 93.23        | 54.71                | 53.21        | 54.27        | 54.23        |

Figure 11 shows the receiver operating characteristic (ROC)–area under curve (AUC) classification metric. It is an excellent metric to observe the performance of a model. If the AUC is near to 1 or 100 percent, then it means that the model has a good measure of separability. A poor model shows an AUC near 0. We can also verify from this graph that for the ISIC dataset, the AUC of the naive Bayes (NB) algorithm is around 0.5 (54.71 percent), which is very poor compared to other models. As mentioned earlier, in the case of progression to more advanced stages like meta-learning, where more complex features are required to be distinguished or identified, the accuracy of NB shows a significant decrease.

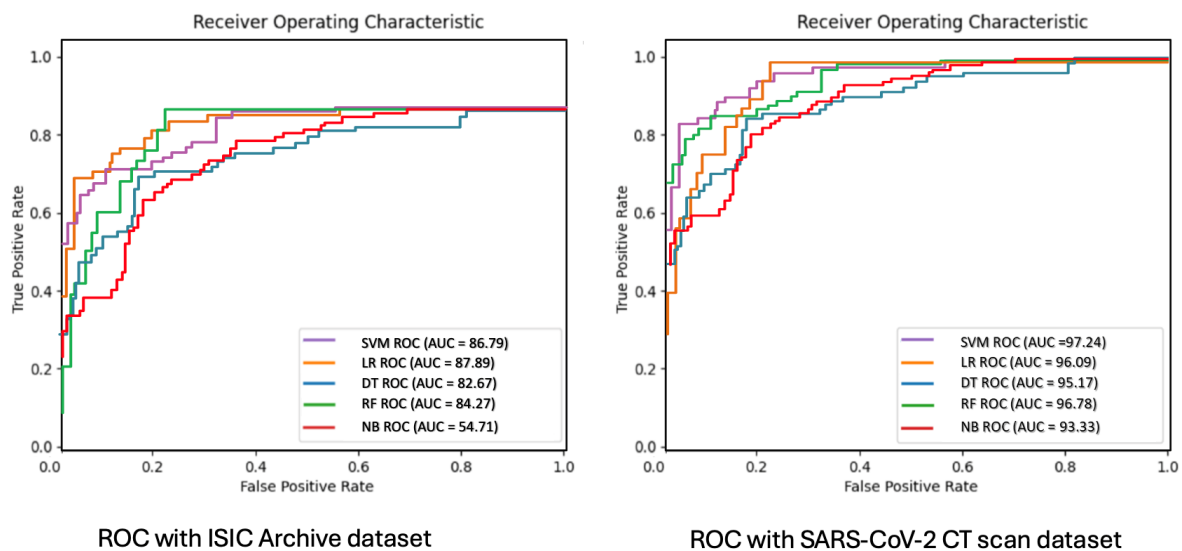


**Figure 11.** ROC-AUC classification evaluation metric of final meta-learners for COVID-19 and ISIC datasets.

## 5. Discussion

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

This approach involves a fused feature extraction technique, combining features extracted from three different pretrained feature extractor methods with various combinations. Additionally, this study 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, training and evaluation procedures used five different algorithms: RF, DT, LR, SVM, and NB. 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 pretrained 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 pretrained 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 pretrained models, ResNet-101 exhibits superior performance in the context of SVM. In this study, SVM showcased exceptional performance when applied to the SARS-CoV-2 CT scan dataset, whereas LR and SVM demonstrated outstanding results when dealing with the ISIC Archive

dataset. Certainly, it is evident that LR achieved the highest accuracy rates with ResNet-101, respectively.

## 6. Conclusions

This research introduces a feature extraction and classification model designed to optimize the accurate detection of COVID-19 and skin cancer within the context of image datasets. This approach for feature extraction techniques harnesses deep learning features obtained from pretrained models. Subsequently, the resulting fused feature vector is integrated into the stacked ensemble approach, particularly when utilizing combinations of two or three pretrained models. Initially, base-learner predictions are made, followed by concatenation of the original feature map and ensemble validation data, ultimately feeding into the meta-learner stage for the final prediction.

The learner was trained and tested with a pair of datasets: the SARS-CoV-2 CT Scan and the ISIC Archive. Employing a stacked method, the SVM technique yielded the highest classification accuracy for the SARS-CoV-2 CT Scan dataset, whereas LR approaches outperformed others for the ISIC Archive dataset. These results were achieved when both datasets yielded feature extraction using ResNet-101.

### Future Work

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

- This work exclusively underwent testing on COVID-19 and skin cancer datasets. Therefore, in order to extend its applicability to further datasets, additional research and investigation are required.
- This study approach relied on five established ML models to construct this stacked ensemble model. Additionally, we utilized pretrained models with a fixed input size of  $224 \times 224$  for feature extraction. Nevertheless, in future research, there is potential to broaden the scope of this method by adapting it to various pretrained CNN architectures that may have different input sizes.

**Author Contributions:** Conceptualization, H.Q., S.T.H.R. and M.N.; methodology, S.T.H.R. and M.N.; software, H.Q. and U.b.K.; validation, S.T.H.R. and M.N.; formal analysis, H.Q. and U.b.K.; investigation, U.b.K. and M.N.; resources, M.A. and A.C.; data curation, H.Q. and S.T.H.R.; writing—original draft preparation, H.Q. and U.b.K.; writing—review and editing, S.T.H.R. and M.N.; visualization, U.b.K., S.T.H.R. and M.N.; supervision, M.A. and A.C.; project administration, M.A. and M.N.; funding acquisition, M.A. and A.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research has been partially supported by the Project of National Relevance “Innovative mathematical modelling for cell mechanics: global approach from micro-scale models to experimental validation integrated by reinforcement learning”, financed by European Union—Next-GenerationEU—National Recovery and Resilience Plan—NRRP-M4C2-I 1.1, CALL PRIN 2022 PNRR D.D. 1409 14-09-2022—(Project code P2022MXCJ2, CUP D53D23018940001) granted by the Italian MUR.

**Data Availability Statement:** The raw data supporting the conclusions of this article will be made available by the authors on request.

**Conflicts of Interest:** Authors declare no conflict of interest.

## References

1. Kaliraman, B.; Duhan, M. A new hybrid approach for feature extraction and selection of electroencephalogram signals in case of person recognition. *J. Reliab. Intell. Environ.* **2021**, *7*, 241–251. [\[CrossRef\]](#)
2. Umair, A.; Masciari, E.; Ullah, M.H. Vaccine sentiment analysis using BERT + NBSVM and geo-spatial approaches. *J. Supercomput.* **2023**, *79*, 17355–17385. [\[CrossRef\]](#)

3. Hovorushchenko, T.; Moskalenko, A.; Osiadlyi, V. Methods of medical data management based on blockchain technologies. *J. Reliab. Intell. Environ.* **2023**, *9*, 5–16. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Umair, A.; Masciari, E. Sentimental and spatial analysis of covid-19 vaccines tweets. *J. Intell. Inf. Syst.* **2023**, *60*, 1–21. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Naeem, M.; Coronato, A. An AI-empowered home-infrastructure to minimize medication errors. *J. Sens. Actuator Netw.* **2022**, *11*, 13. [\[CrossRef\]](#)
6. Naeem, M.; Coronato, A.; Paragliola, G. Adaptive treatment assisting system for patients using machine learning. In Proceedings of the 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), Granada, Spain, 22–25 October 2019; IEEE: New York, NY, USA, 2019; pp. 460–465.
7. Haseeb, A.; Khan, M.A.; Shehzad, F.; Alhaisoni, M.; Khan, J.A.; Kim, T.; Cha, J. Knee Osteoarthritis Classification Using X-Ray Images Based on Optimal Deep Neural Network. *Comput. Syst. Sci. Eng.* **2023**, *47*, 2397–2415. [\[CrossRef\]](#)
8. Shah, S.I.H.; Coronato, A.; Naeem, M.; De Pietro, G. Learning and assessing optimal dynamic treatment regimes through cooperative imitation learning. *IEEE Access* **2022**, *10*, 78148–78158. [\[CrossRef\]](#)
9. Gheisari, M.; Taami, T.; Ghaderzadeh, M.; Li, H.; Sadeghsalehi, H.; Sadeghsalehi, H.; Abbasi, A.A. Mobile applications in COVID-19 detection and diagnosis: An efficient tool to control the future pandemic; a multidimensional systematic review of the state of the art. *JMIR MHealth UHealth* **2023**, *12*, e44406.
10. Fiorino, M.; Naeem, M.; Ciampi, M.; Coronato, A. Defining a Metric-Driven Approach for Learning Hazardous Situations. *Technologies* **2024**, *12*, 103. [\[CrossRef\]](#)
11. Ghaderzadeh, M.; Asadi, F.; Ramezan Ghorbani, N.; Almasi, S.; Taami, T. Toward artificial intelligence (AI) applications in the determination of COVID-19 infection severity: Considering AI as a disease control strategy in future pandemics. *Iran. J. Blood Cancer* **2023**, *15*, 93–111. [\[CrossRef\]](#)
12. Coronato, A.; Naeem, M. A reinforcement learning based intelligent system for the healthcare treatment assistance of patients with disabilities. In Proceedings of the International Symposium on Pervasive Systems, Algorithms and Networks, Naples, Italy, 16–20 September 2019; Springer: Cham, Switzerland, 2019; pp. 15–28.
13. Xue, D.; Zhou, X.; Li, C.; Yao, Y.; Rahaman, M.M.; Zhang, J.; Chen, H.; Zhang, J.; Qi, S.; Sun, H. An Application of Transfer Learning and Ensemble Learning Techniques for Cervical Histopathology Image Classification. *IEEE Access* **2020**, *8*, 104603–104618. [\[CrossRef\]](#)
14. Ganaie, M.A.; Hu, M.; Malik, A.; Tanveer, M.; Suganthan, P. Ensemble deep learning: A review. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105151. [\[CrossRef\]](#)
15. Shekar, B.; Hailu, H. An efficient stacked ensemble model for the detection of COVID-19 and skin cancer using fused feature of transfer learning and handcrafted methods. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **2023**, *11*, 878–894. [\[CrossRef\]](#)
16. Rajagopal, R. Comparative analysis of COVID-19 X-ray images classification using convolutional neural network, transfer learning, and machine learning classifiers using deep features. *Pattern Recognit. Image Anal.* **2021**, *31*, 313–322. [\[CrossRef\]](#)
17. Shaik, N.S.; Cherukuri, T.K. Transfer learning based novel ensemble classifier for COVID-19 detection from chest CT-scans. *Comput. Biol. Med.* **2022**, *141*, 105127. [\[CrossRef\]](#)
18. Grignaffini, F.; Barbuto, F.; Piazzi, L.; Troiano, M.; Simeoni, P.; Mangini, F.; Pellacani, G.; Cantisani, C.; Frezza, F. Machine Learning Approaches for Skin Cancer Classification from Dermoscopic Images: A Systematic Review. *Algorithms* **2022**, *15*, 438. [\[CrossRef\]](#)
19. Tembhurne, J.V.; Hebbar, N.; Patil, H.Y.; Diwan, T. Skin cancer detection using ensemble of machine learning and deep learning techniques. *Multimed. Tools Appl.* **2023**, *82*, 1–24. [\[CrossRef\]](#)
20. Rahman, Z.; Hossain, M.S.; Islam, M.R.; Hasan, M.M.; Hridhee, R.A. An approach for multiclass skin lesion classification based on ensemble learning. *Inform. Med. Unlocked* **2021**, *25*, 100659. [\[CrossRef\]](#)
21. Müller, D.; Soto-Rey, I.; Kramer, F. An analysis on ensemble learning optimized medical image classification with deep convolutional neural networks. *IEEE Access* **2022**, *10*, 66467–66480. [\[CrossRef\]](#)
22. Shinkai, K.; Bruckner, A.L. Dermatology and COVID-19. *JAMA* **2020**, *324*, 1133–1134. [\[CrossRef\]](#)
23. Goyal, M.; Pandey, M. Ensemble-based data modeling for the prediction of energy consumption in hvac plants. *J. Reliab. Intell. Environ.* **2021**, *7*, 49–64. [\[CrossRef\]](#)
24. Rincy, T.N.; Gupta, R. Ensemble learning techniques and its efficiency in machine learning: A survey. In Proceedings of the 2nd International Conference on Data, Engineering and Applications (IDEA), Bhopal, India, 28–29 February 2020; IEEE: New York, NY, USA, 2020; pp. 1–6.
25. Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J.; Data, M. Practical machine learning tools and techniques. In *Proceedings of the Data Mining*; Elsevier: Amsterdam, The Netherlands, 2005; Volume 2, pp. 403–413.
26. Chowdhury, M.E.; Rahman, T.; Khandakar, A.; Mazhar, R.; Kadir, M.A.; Mahbub, Z.B.; Islam, K.R.; Khan, M.S.; Iqbal, A.; Al Emadi, N.; et al. Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access* **2020**, *8*, 132665–132676. [\[CrossRef\]](#)
27. Fanconi, C. Skin Cancer: Malignant vs. Benign-Processed Skin Cancer Pictures of the ISIC Archive. 2019. Available online: <https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign> (accessed on 15 June 2024).

28. Silva, P.; Luz, E.; Silva, G.; Moreira, G.; Silva, R.; Lucio, D.; Menotti, D. COVID-19 detection in CT images with deep learning: A voting-based scheme and cross-datasets analysis. *Inform. Med. Unlocked* **2020**, *20*, 100427. [[CrossRef](#)]
29. Mahajan, P.; Uddin, S.; Hajati, F.; Moni, M.A. Ensemble Learning for Disease Prediction: A Review. *Healthcare* **2023**, *11*, 1808. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.