Vitiligo image classification using pre-trained Convolutional Neural Network Architectures, and its economic impact on health care

Vitiligo bildklassificering med hjälp av förtränade konvolutionella neurala nätverksarkitekturer och dess ekonomiska inverkan på sjukvården

M Rami Alsaid Suliman
Nour Bashar

Examensarbete inom
Datateknik och ekonomi
Grundnivå, 15 hp
Handledare på KTH: Peter Sillén
Examinator: Ibrahim Orhan
TRITA-CBH-GRU-2022:050
KTH
Skolan för kemi, bioteknologi och hälsa
141 52 Huddinge, Sverige

Abstract

Vitiligo is a skin disease where the pigment cells that produce melanin die or stop functioning, which causes white patches to appear on the body. Although vitiligo is not considered a serious disease, there is a risk that something is wrong with a person's immune system. In recent years, the use of medical image processing techniques has grown, and research continues to develop new techniques for analysing and processing medical images. In many medical image classification tasks, deep convolutional neural network technology has proven its effectiveness, which means that it may also perform well in vitiligo classification. Our study uses four deep convolutional neural networks in order to classify images of vitiligo and normal skin. The architectures selected are VGG-19, ResNeXt101, InceptionResNetV2 and Inception V3. ROC and AUC metrics are used to assess each model's performance. In addition, the authors investigate the economic benefits that this technology may provide to the healthcare system and patients. To train and evaluate the CNN models, the authors used a dataset that contains 1341 images in total. Because the dataset is limited, 5-fold cross validation is also employed to improve the model's prediction. The results demonstrate that InceptionV3 achieves the best performance in the classification of vitiligo, with an AUC value of 0.9111, and InceptionResNetV2 has the lowest AUC value of 0.8560.

Keywords: Vitiligo, deep CNN architectures, Image classification, pre-trained models, dataset, AUC, economic impact.

Sammanfattning

Vitiligo är en hudsjukdom där pigmentcellerna som producerar melanin dör eller slutar fungera, vilket får vita fläckar att dyka upp på kroppen. Även om Vitiligo inte betraktas som en allvarlig sjukdom, det finns fortfarande risk att något är fel på en persons immun. Under de senaste åren har användningen av medicinska bildbehandlingstekniker vuxit och forskning fortsätter att utveckla nya tekniker för att analysera och bearbeta medicinska bilder. I många medicinska bildklassificeringsuppgifter har djupa konvolutionella neurala nätverk bevisat sin effektivitet, vilket innebär att den också kan fungera bra i Vitiligo klassificering. Vår studie använder fyra djupa konvolutionella neurala nätverk för att klassificera bilder av vitiligo och normal hud. De valda arkitekturerna är VGG-19, RESNEXT101, InceptionResNetV2 och Inception V3. ROC- och AUC mätvärden används för att bedöma varje modells prestanda. Dessutom undersöker författarna de ekonomiska fördelarna som denna teknik kan ge till sjukvårdssystemet och patienterna. För att träna och utvärdera CNN modellerna använder vi ett dataset som innehåller totalt 1341 bilder. Eftersom datasetet är begränsat används också 5-faldigt korsvalidering för att förbättra modellens förutsägelse. Resultaten visar att InceptionV3 uppnår bästa prestanda i klassificeringen av Vitiligo, med ett AUC -värde på 0,9111, och InceptionResNetV2 har det lägsta AUC -värdet på 0,8560.

Nyckelord: Vitiligo, djupa CNN-arkitekturer, bildklassificering, förtränade modeller, dataset, AUC, ekonomisk påverkan.

Contents

1. Introduction

Approximately 70 million people worldwide are affected by vitiligo. This disease may be seen as a cosmetic problem that has its psychological problems and consequences. Researchers looked at 1098 vitiligo patients and found that nearly 20% had at least one autoimmune condition [1]. Further, for patients with vitiligo, there seems to be an increased risk of thyroid disease (autoimmune) with age, compared to individuals without the condition. Vitiligo can affect anyone regardless of their gender, race, or ethnicity. Additionally, a study conducted on 150 vitiligo patients reveals that younger patients are more likely to experience depression [2]. It is unfortunate that there is currently no cure for this disease as the cause of the disease is poorly understood [3]. However, early detection may allow some treatments to prevent the disease from spreading to other parts of the body.

Artificial intelligence is efficient at handling large amounts of data in a short period of time, and its capacity to reduce costs [5] makes it important for businesses to reduce energy use and boost efficiency. In medical care, the huge and rapid use of AI helps in analysing complex models, detecting errors or differences, and improving the diagnostics accuracy [4]. “An artificial intelligence never gets tired and can work unsocial hours. Doctors will instead get the time to look more closely at the complicated cases and meet patients” says Max Gordon [6]. Researchers at the University of North Carolina Lineberger Comprehensive Cancer Center employed an AI product to discover alternative treatment options for individuals for more than 1000 patients whose tumours revealed genetic abnormalities [7].

Convolutional neural network (CNN) is a form of machine learning that is commonly used to analyse images by extracting image features for recognition or classification tasks. It is also used by many computer vision algorithms [8].

1.1 Problem statement

Machine learning applications in healthcare are becoming more and more popular. That is because of the power of ML applications in biomedical data analysing. The Deep Convolutional Neural Network is an application of machine learning that has gained prominence in recent years in terms of medical image recognition, due to its high performance in image classification and object detection. Several studies, for instance [16] and [42], have utilised deep CNN models for medical image analysis, mostly in cancer and other diseases. However, few studies have been carried out in the field of vitiligo detection and classification, which needs to be investigated as vitiligo patients are more likely to have at least one autoimmune condition, which may increase the risk of developing cancer [9]. Beside affecting the health impact on a wide range of body tissues and organs [10], early detection of vitiligo using CNN could help save some lives and prevent suffering for others by giving the doctors a chance to handle complicated cases [6]. Economical aspects include finding out how much resources, such as time, energy, and costs CNN can save health care in general and dermatologists in their clinical practice.

1.2 Goals of the project

The purpose of this study is to measure the performance of four deep CNN architectures in classifying vitiligo images. The CNN architectures are used and evaluated for vitiligo classification applications in dermatology. A second goal is to investigate the economic benefits of this technology in health care, as it can reduce time and costs in medical care by providing accurate diagnoses for patients, which may also be seen as a health quality aspect for the patient. The CNN architectures that are used in the experiment have been chosen based on their performance in image classification in previous studies [12, 37, 38, 39]. These architectures are: VGG-19, ResNeXt 101, InceptionResNetV2 and InceptionV3. Even though there are other deep CNN architectures that can perform well in image classification, the selected models have obtained the highest accuracy over the last year in ImageNet competition. All the selected architectures are pre-trained on 1000- class ImageNet dataset [33].

The mentioned CNN architectures will be trained on a dataset that contains two labels which are vitiligo and normal skin images. After training phase these architectures must be evaluated on another set of data that contains the same labels as the first one.

ROC curves and AUC metrics are used to evaluate performance of the different architectures. These ROC curves visualise the performance of the four architectures in the classification task, whereas AUC represents the ability of each architecture to classify the labels. According to ROC and AUC results, economic and technological impact will be discussed.

1.3 Limitation

In this study, CNN architectures were trained on a dataset of images representing vitiligo and normal skin. Due to the use of CPU processors rather than graphics processing units (GPUs) in this experiment, the training of models took an average of 600 minutes per model. With more powerful hardware, such as a GPU, CNN model training may be accomplished in a short period of time. Using CPUs resulted in some computer crashes during training, necessitating the repeating of multiple tests. Not having access to a GPU is a limitation in quantity for important model training that may also affect quality in results.

Another issue is the shortage of medical imaging datasets on the internet. Most open-source datasets on the internet are related to skin cancer. Unfortunately, contacting two clinics did not result in access to additional vitiligo images from their database. To improve the predictions of the CNN models, we used other techniques that could help in cases where the dataset is limited.

2. Background

In computer science and within artificial intelligence (AI) a computer, robot, or other machine may use a method called machine learning (ML) applying advanced mathematics and statistics. The purpose is to analyse, classify, categorise, and memorise, to learn and be able to perform tasks that are normally performed by humans, because they require human awareness, intelligence, and judgement. Even though AI cannot perform all the tasks that a normal human can execute, some AI models can match humans at specific tasks [13]. Typical neural networks were tested in basic lesion detection and classification tasks versus human observers. A class of artificial intelligence called convolutional neural networks (CNN) outperformed humans in general. Nevertheless, CNN is more comparable to a human observer when images have high noise [44].

In digital radiography images, the noise is caused mainly either because of the scattered radiation caused by the x-ray machines (second radiation) or the scattered radiation caused by an object getting to the film [48].

While in digital photography images, noise is caused by photons, read, light conditions, or random electrical disturbances in the camera. There are various forms of noise within an image, and it is, in most cases, unrelated to the scene content. It appears as an unwanted grainy structure on the image, which decreases details and makes the image look coarser. There are generally two types of imaging noise: luminance and chroma [49].

Machine learning models are now faster and more accurate than humans in certain limited fields. Robots can also learn by performing experiments using deep learning (DL) algorithms. This means that robots in limited applications can learn efficiently and then work for long periods of time with high quality performance in comparison with humans [45]. A deep learning technique stimulates the human brain to perform tasks that can be done by humans.

2.1 Applications of artificial intelligence in health care

The use of AI in the medical field is quite widespread. Its use has spread the diagnosis of several vital human body systems, such as the nervous system, the cardiovascular system, the digestive system, and the skeletal system. Medical image analysis algorithms use artificial intelligence (AI) to diagnose diseases such as:

Brain disorders: such as stroke, which is one of the most common causes of death and disability worldwide, and its impact on the healthcare system is vast. Research has proven recently that stroke lesions can be segmented using multimodal MRI images with a high dice rate of 0.84 and 0.59 respectively and using Res-CNN with

0.742 as an average [14]. The dice coefficient is used to quantify the effectiveness of image segmentation methods by comparing the similarity of objects from zero to one.

Cardiac disease: huge scale deployments with high-quality outcomes have been achieved. According to Cohen's kappa metric, after training and testing the CNN-BiLSTM network, utilising the k-fold cross-validation scheme of 10, the results attained in terms of accuracy, sensitivity, and specificity were between 97.87–99.58 percent with an AUC (area under the curve) of 0.998 on average [15].

Dermatology: the use of AI in this field is largely based on the analysis and classification of images using various architectural models. A system that may diagnose basal cell carcinoma with 98.1 percent accuracy by analysing histopathology pictures captured by microscopical inspection to study symptoms of the disease. Additionally, melanoma skin cancer seems to be the most successfully diagnosed disease, followed by ulcer and psoriasis, with high accuracy rates as dermatology diseases. Numerous research studies demonstrate the broad effectiveness of applying AI to differentiate noncancerous pigmented nevi called benign nevi from melanoma. Skin disorders such as inflammatory conditions, forecasting skin sensitivity elements, distinguishing various forms of acne, and many other diseases are all common uses of dermatological applications [19].

2.2 Image classification based on deep learning

A computer's analysis of a single image to determine what class it belongs to is what image classification is all about. The deep learning, DL, field is an extension of machine learning, but it is more complex than machine learning. In deep learning, the architecture is composed of many interconnected node layers, similar to neural networks in the brain. Each layer has a weight and a threshold associated with it. Generally, each deep learning architecture has three types of layers: input layer, hidden layers, and output layer. There are many different architectures of DL, such as RNN, CNN. The Convolutional neural network CNN will be used to conduct the experiments for this thesis.

In order to implement the classification process on a specific DL model, the model must first be trained using a dataset of images from the target class. The classification process begins by identifying low-level features such as dots, lines, colours, and edges first. Then, these low-level features are processed layer by layer to form high-level features such as objects and shapes. The result is a distributed representation of data representing which class the image belongs to. Typically, a model can only learn low-level properties from the input image, which includes textures, colours, and edges. The model becomes increasingly abstract and capable of analysing high features as the number of layers increases [16].

2.3 Convolutional neural networks

Deep learning is a method of data processing that is composed of many complex layers that stimulate the human neural network. Recently, deep learning has made significant strides in the areas of computer vision, speech recognition, natural language processing, audio recognition, and bioinformatics. The convolutional neural network (CNN) is one of the most widely used image analysis models among the models of deep learning [17].

The CNN architecture recognizes the input image as an array of pixels whose size is dictated by the image resolution, i.e., it senses the image's height, width, and dimension. The dimension could be 3 for an RGB matrix or 1 for a grayscale matrix. As a result, the pictures may be an array of RGB matrices or an array of grayscale matrices [18].

There are multiple convolutional layers, pooling layers, and fully connected layers (FC) that are interconnected in the CNN model, as seen in Figure 2.1 below. When trained or tested on a dataset, the model will apply this series of layers to the input image in order to categorise the image at the end of the process. Fundamentally, CNN models operate in the manner described in Figure 2.1.

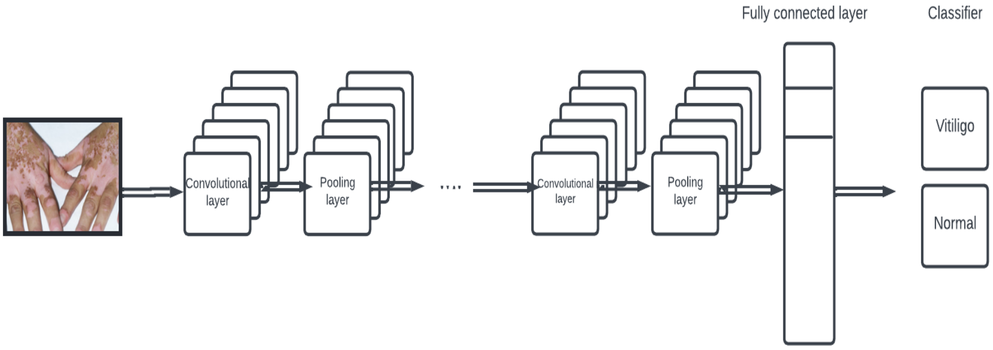

Figure 2.1: Convolutional neural network topology.

2.4 Training the CNN model

Supervised learning is a type of machine learning that makes use of labelled data, and image classification is a type of supervised learning. This means that the CNN model is trained on a labelled dataset in order to consistently classify and predict outcomes [20]. The purpose of the training process is to optimize the CNN model's efficiency by minimizing loss. The loss function is a function that estimates the difference between the model's output and the labels, and it is used to provide feedback on the model's performance during the classification process. During the training process, the layer's weight stores the operations applied to the input data, containing the knowledge that this layer has learned about the data through the backpropagation algorithm. A flowchart representing the training process for a CNN model is shown in Figure 2.2. As the CNN model is trained on more datasets, the layer's weights are gradually fine-tuned and the loss is gradually reduced, resulting in improved model performance [21].

To train the CNN model, the dataset is split into training and testing sets, and the model is trained on the training set in order to learn how to classify different objects.

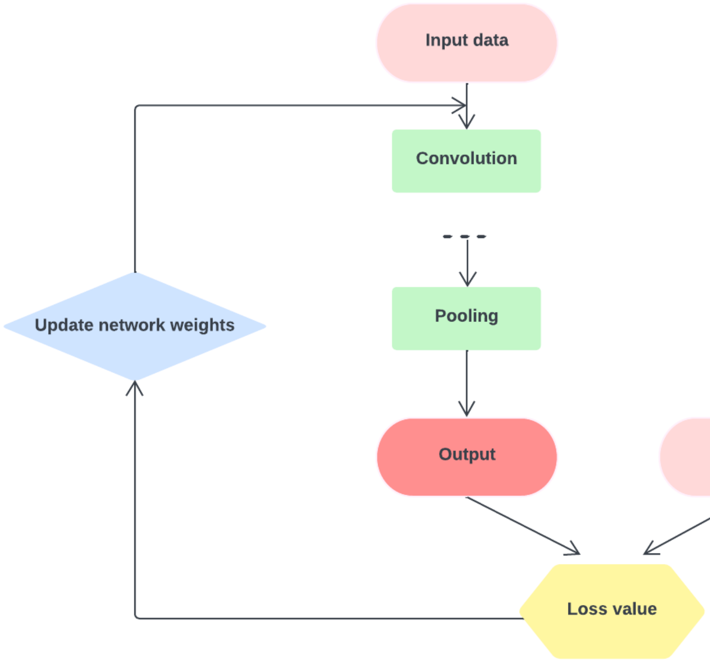

Figure 2.2: a description of the CNN model training process.

As seen in Figure 2.2, the training process involves two steps: forward

propagation and error back propagation. Forward propagation is the process of extracting features from input data using the convolutional and pooling layers, then storing the operations that the layers perform on it in the layer's weight, and then obtaining predictions from the output layer. The loss value is calculated after obtaining the output by finding the difference between the actual label data and the model predictions. Back propagation refers to the process of transferring prediction errors backwards to update the weights of the network.

2.5 Improvements to different deep CNN architectures

The architecture of CNNs is intended to resemble or technically replicate functions in the human brain. LeNet was the first CNN architecture developed by LeCun and other researchers in 1998. That architecture was designed to recognize handwritten digits [27]. Following that, several improvements to deep CNNs have been made since the development of the ALexNet, the first CNN architecture capable of performing well on picture classification and recognition tasks, with an extended depth of five layers. The improvements to CNN design focused on factors such as hyperparameter optimization, architectural patterns, layer connections, and other factors to make the network more efficient at performing image classification [24]. Optimizing the hyperparameters of the CNN model involves finding a combination of hyperparameters that improves its performance; an architecture pattern design enhances the design of the CNN network for better performance; layer connections define the way the node layers are connected.

The efficacy of CNNs in image classification has prompted researchers to set their sights on enhancing the architecture design of CNNs. At the 2014-ILSVRC competition, a new CNN architecture, VGG, took second place due to high accuracy, following GoogLeNet (Inception 1), which won the competition [24]. The VGG was constructed based on the results obtained after ZfNet won the 2013-ILSVRC competition. VGG has 19 layers, whereas AlexNet and ZfNet, which were released before VGG, had 5 and 8 layers, respectively. The architecture's depth was increased to stimulate the relationship that exists between the architecture's depth and its performance capacity [29].

VGG19, InceptionV3, ResNeXt101 64x4d, and Inception-ResNet-V2 are the CNN models examined for the experiments in this paper.

2.6 Economic benefits of using image classification in healthcare

The average life expectancy is increasing worldwide, and this requires an increasing amount of health care due to health issues that come with age. Therefore, dermatology departments will likely have a lack of available physicians. The costs of the national healthcare systems have increased over many decades, which may be seen as an obstacle and a challenge for financial sustainability worldwide [23].

Using AI techniques such as CNN in dermatology means that the skin is examined by pre-trained software with a high accuracy level and similar precision as an expert dermatologist.

In qualitative research that was conducted on 48 participants, with 33% having a history of melanoma, 33% having experience of nonmelanoma skin cancer exclusively, and 33% having no experience of skin cancer, 75% of the respondents would advise to use AI to their families, while 94% of the respondents emphasised the relevance of human-artificial intelligence collaboration [22]. Dermatologists may prefer to use AI as this technology will allow them to spend more time with patients and work easier.

Artificial intelligence can help organizations increase revenue and save resources and thereby contribute to sustainable solutions. The use of AI will help enhance the workflow in many ways since it saves time and resources as well as adding high accuracy [11]. However, the use of artificial intelligence in business will have different impacts, depending on the nature of a company. According to an investigation related to medical care [25], the usage of artificial intelligence was shown to be connected with a slight improvement in results. Companies and organisations could achieve sustainable results by using different AI models for different purposes. Google, for instance, saves 40% of energy on cooling costs for its data centres with the use of AI. IBM also improved weather forecasting accuracy by 30% with the help of AI. [26]. This does not mean that the use of AI in health care will replace competent skilled human resources such as dermatologists. More likely AI may assist dermatologists in the field of skin and health diagnostics with quick and accurate results. This gives the profession the ability to deliver better skincare [28].

AI was predicted to be worth $6.6 billion in 2021 before the spread of COVID-19, but in 2021 the spending on AI reached $57 billion globally and is predicted to be worth $190 billion in 2025. Furthermore, AI is expected to add to the gross domestic product for many nations [31]. AI costs may differ depending on the complexity of the model and solution used. A minimal viable product might cost anywhere between $8,000 and $15,000, while a comprehensive bespoke unique AI system might cost anything from $20,000 to $1,000,000 or even more [30].

2.7 Previous work

The main cause of vitiligo is an autoimmune disorder that stops the melanin-producing cells from functioning. This disease is widespread and can affect people of any age or gender [3]. There is no specific area of the body that could be infected, but the disease most commonly affects the hands and the face. Unfortunately, there is no cure for vitiligo, however, some treatments may slow down or stop it from spreading to other parts of the body when it is detected early.

Recent studies have examined the application of machine learning in detecting diseases. These studies were mostly based on analysing medical images. Deep learning's applications have proven successful in detecting, segmenting, and classifying medical images since 2006. Currently, CNN models are the most commonly used deep learning applications in medical image analysis, due to their high performance in image classification and object detection [42].

In [46], an experiment to classify vitiligo using a CNN pre-trained model was conducted, in which four CNN pre-trained models (Resnet50, Vgg16, Xception, and InceptionV3) were used. The authors trained their CNN models on a dataset that consisted of 30000 images and used 8677 images to evaluate the models' performance. They applied the typical data augmentations for classification tasks, which are Rotating, Flipping, Shifting and batch size 32, with a learning rate 0.001. The batch size represents the number of data samples processed before the network updates its weights, and the learning rate is the hyperparameter that configures how much the network weights after detecting a training loss. One of the metrics utilised to assess the performance of the classification was AUC. Results showed that Xception scored an AUC of 0.911, Inception a score 0.912, VGG-16 a score 0.917% and ResNet 50 a score 0.922%, which resulted in that ResNet 50 achieved the best performances in vitiligo classification based on that approach.

In [36], authors used ResNet 50 to classify images of vitiligo across different datasets. They applied four trials, the first of which was to apply classification to the original dataset without pre-processing it. While the last trial employed ResNet 50 in order to classify a dataset that was pre-processed using two methods. When compared with classifying the original dataset in the first attempt, the results of the last try after data processing were better. After the final attempt, the accuracy, sensitivity, and specificity were 85.69 percent, 88.39 percent, and 79.40 percent, respectively. While the first attempt yielded for similar measures 76.37, 80.73, and 66.20 percent, respectively.

There are few experiments on classification of vitiligo, compared with other skin diseases. This is because CNNs are still a relatively new technology in classification tasks, in addition to the lack of data for medical images. However, deep CNN architectures are used in medical image processing. In an article about Implication of Convolutional Neural Networks in the Classification of Vitiligo [47] four pre-trained deep CNN models Inception-V3, VGG-16, VGG-19, and SqueezeNet were implemented for feature extraction. These four pre-trained models were used to convert images into a limited set of variables containing just relevant and critical information, which was then used to carry out the classification task by other ML methods.

3. Method

This chapter describes the various methods used to conduct the experiment of training CNN models, as well as the techniques used to optimise the models' performance in terms of image classification. Additionally, the evaluation and measuring methods used to compare the performance of the models are presented in this chapter.

The experiments were conducted using the programming language Python and the PyTorch deep learning library. Several steps were taken to produce reasonable results for CNN models in classifying vitiligo, from collecting the data to evaluating the performance of the different models. A theoretical study based on gathering information from previous studies and attempting to connect information about AI and CNN applications with the economy has been used to answer the research's economic questions.

3.1 Data collection

The dataset used to train and evaluate the models is collected from multiple public dermatological atlas websites, namely DermNet, DermNet NZ, AtlasDerm, DermIS, SD-260, Kaggle, and DanDerm, and it was already collected and provided by a third party [32]. The dataset was divided into three categories: train, test, and validation; however, the test category was not required for this experiment and was therefore merged with the train. There are 1341 images in total, 717 of which were classed as vitiligo images, and 543 of which were created to represent depigmented or hypopigmented lesions in which patches of skin are lighter than the body's skin tone. This dataset is sufficient as long as the chosen models have already been pre-trained on the ImageNet 1000-class dataset [33]. The models were trained and validated on the entire dataset using the k-cross-validation method, which will be described further below.

3.2 5-folds cross-validation

The k-fold cross-validation method is a method to improve model predictions when only a limited dataset is available [34]. In this case, k-fold cross-validation is an effective method for estimating the models' performance on new data due to the small dataset that was available. To use this method, the dataset was divided into K subsets, and the method was repeated K times for each model. For each iteration, the model utilises k-1 folds for training and the K:th fold for validation. Figure 3.1 shows the scenario in which the dataset is divided into 5 folds and then repeated 5 times to apply the 5-folds cross-validation. Stratification guarantees that each fold may represent the complete dataset, allowing for parameter optimization and improving the model's ability to classify new images [35].

3.3 Deep CNNs architecture selection and implementation

A theoretical pre-study was conducted in order to select the CNN architectures that can be used for image classification. In the end, the following architectures were chosen: InceptionV3, ResNeXt 101, InceptionResNetV2, and VGG-19. The reasons for selecting these architectures vary. ResNeXt 101 ranked second in the 2016 ILSVRC classification competition with a top-five error rate of 4.1% [12, 37]. InceptionResnetV2 was also selected because it performed best in a prior vitiligo segmentation study in 2019 [38]. While InceptionV3 was chosen because it is part of the Inception family, an efficient architecture that achieves great performance at a cheap computational cost [39]. Additionally, the VGG-19 was selected because of its good performance in several image classification projects, particularly in the medical field.

PyTorch was employed to carry out the experiment. It is a widely used open-source framework for machine learning, particularly in deep learning application research. The architectures that have been used are imported from the Touchvision library in PyTorch, where the models are pre-trained on 1000 class ImageNet dataset [33].

3.4 Data augmentation and transformation

The purpose of this study is to investigate the performance of different CNN architectures in vitiligo image classification, and the dataset is an important factor for this study. Our dataset was enough, but a bigger dataset would help the CNN models make better and more accurate predictions. Therefore, data augmentation tools were applied to make the dataset richer. Colour modification, rotation, flipping, resizing, rescaling, cropping, zooming, and so on are the most often used transformation techniques in image classification [40]. We applied different data augmentation and transformations to the images during the training and validation of the models. The transformations used are RandomResizedCrop, which crops a random piece of the image, RandomHorizontalFlip, which flips the image horizontally, and normalisation, which is rescaling the data to ensure similar distribution across each input. We used the mean = [0.485, 0.456, 0.406] and standard deviation = [0.229, 0.224, 0.225], which are suggested by Pytorch and determined by ImageNet, for normalisation. To implement these transformations, the transforms module of the Torchvision library was used [33]. The input image size has been set to 224x224 for VGG-19 and ResNeXt 101, and 299x299 for InceptionV3 and inceptionResnetV2.

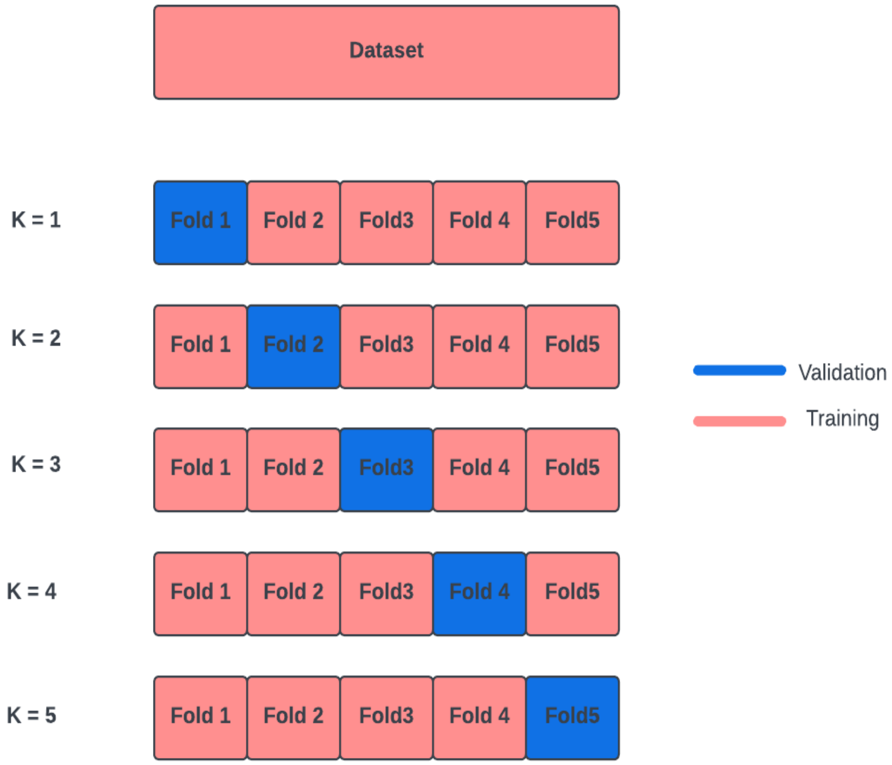

Figure 3.1: 5-fold cross-validation algorithm description

3.5 Hyperparameter settings

Hyperparameters are considered to be settings of the model's learning process since they affect predictions of CNN architecture. Among these parameters are batch size, epochs, learning rate, and optimizer. The batch size determines how many images will be passed through the neural network per iteration, while the epoch number determines how many times the entire dataset will be passed through the neural network. The optimizer is a method for updating the network's parameters and weights in order to reduce prediction errors.

The optimal hyperparameter settings were estimated using a widely used method called guesstimating. Guesstimating refers to estimating hyperparameters through many tests and trials. In many CNN applications, the hyperparameters were set using the guesses of the initial values and choosing the optimal values based on previous predictions' values [41].

The number of tests was limited due to the use of a CPU processor. Several different numbers of epochs have been tested – the authors started with 5, then went to 10, then 30 and finally chose 50. By testing different epochs, the authors determined whether the models' performance was improving. By increasing the number of epochs, each architecture layer analyses the same dataset multiple times. This updates its weights each time after obtaining the loss gradient, which results in improved performance from underfitting to normal, but sometimes it also leads to overfitting. Overfitting occurs when the CNN model learns well in the training dataset, which makes it hard for the model to predict new samples not included in the training dataset. In contrast, underfitting occurs when the data that a model was trained on is difficult to classify.

In the training phase, the learning rate determines how much the neural network weights are updated. Two distinct learning rates were examined; 0.0001 and 0.001, but the latter was selected since the earlier resulted in an extremely high loss rate and low accuracy throughout training and evaluation. During each iteration of data input analysis, the learning rate controls how much weights in the model will be updated after the loss gradient has been determined.

The four architectures have been tested with different batch sizes. InceptionResNetV2 has been trained on batch size 1 since it crashes during training with a larger batch size. A later decision was made to change the batch size to 32, as the first batch size had such a high loss rate. For the other architectures, three different batch sizes have been tested, which are 8, 16, and 32; 32 gave the best results with a learning rate of 0.001. Table 3.1 outlines the hyperparameters that were chosen.

Table 3.1: Hyperparameters used to configure the deep CNN models.

Hyperparameter	**Value**
LR	0,001
Epochs	50
Batch size	32
Optimizer	SGD

Table 3.1 provides the CNN hyperparameters used during training on the dataset. The learning rate is what determines how much the model's weights need to be updated after obtaining the loss gradient. The epoch number is the number of training iterations the model will do on the same dataset. Increasing the epoch number will help the model learn more on the dataset, and the model will perform better. Batch size refers to how many data samples are processed over the network at a time. Stochastic gradient descent (SGD) is a widely used optimization algorithm for estimating a model's loss gradient.

3.6 Transfer learning with deep convolutional neural network

Transfer learning is very popular and widely used with machine learning models, especially for medical image analysis. Since training a CNN model from scratch requires a huge amount of data, it can be challenging to gather all the medial data images needed to train the model. In this scenario, transfer learning can be applied to reduce both time and costs. Transfer learning means using of a pre-trained models that already have been trained on a huge dataset to do a specific task, on another task [42]. In this case, because the data is limited, the authors are using a pre-trained models that have already been trained on the ImageNet dataset.

The authors used feature extraction transfer learning which means starting with a pretrained model and only update the final layer weights from which the authors derive predictions. In this approach, the authors replaced the number of output classes in the fully connected layer (the final layer) from 1000 to 2, where 1000 represent the different classes in the ImageNet dataset, while 2 represent the vitiligo class and the normal skin class [33].

3.7 Evaluation metrics

By visualising the performance of the models on the receiver operating characteristic (ROC), the authors were able to assess the performance of the different architectures. The x-axis of the ROC represents the false positive rate FPR while the y-axis represents the true positive rate TPR. The True Positive Rate (TPR) represents the probability that a vitiligo image is correctly classified, while the False Positive Rate (FPR) represents the probability that a non-vitiligo image is wrongly classified. Then the AUC values for each architecture have been calculated over the five folds in order to produce a plot that shows the AUC values for the five folds for all models, where AUC represents the ability of the classifier to classify the data. The AUC -ROC is a popular metric for comparing CNN model performance, particularly in classification tasks. Figure 3.2 illustrates the ROC curve for different cases where the model's performance is good, bad, or perfect. The diagonal represents the random classifier, which is used as a sign to determine if the model is good or bad. If the curve is below the diagonal, then the model performed poorly, while if the curve is above the diagonal, then the model performed well. Classifiers with a zero FPR and a higher TPR value are considered perfect classifiers. The model's capability to classify the labels is reflected in the AUC value, making it an essential metric [43].

In addition, the authors plotted the train and validation loss as a function of the number of epochs in order to investigate how well the training and validation loss performed for the model and determine the optimal number of epochs for the experiment. The training loss is a way to measure how well the CNN model fits the training data by considering its error on the training data. Similarly, the validation loss measures how well the model performs on the validation dataset.

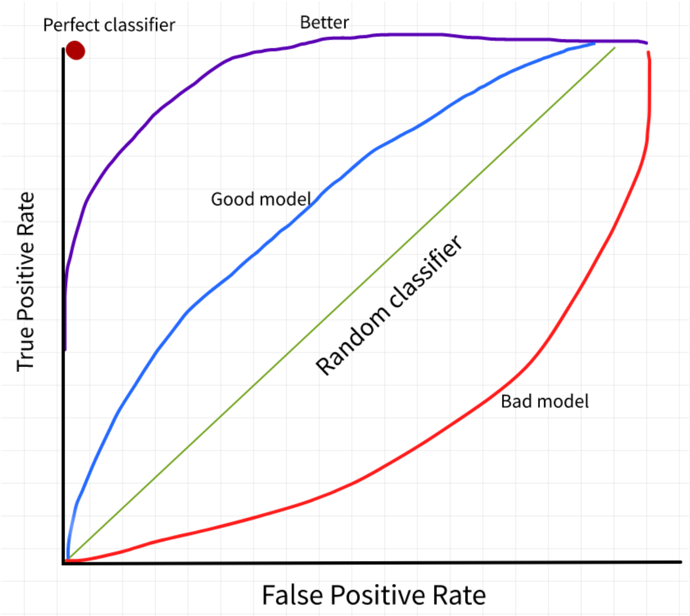

Figure 3.2: Hand drawn ROC curves describing what a perfect, good, and bad model are.

3.8 Determining the economic, social, and ethical impact

This thesis is primarily concerned with comprehending the convolutional neural network (CNN) and evaluating four various CNN architectural models on a single dataset in order to determine the most effective performance in the area of vitiligo image classification. The social and ethical impact of the study is related to whether image classification through CNNs can add positive health aspects for an individual and for clinics as important functions in society. This will be studied and discussed to the extent whether the answer is a yes or no. A further analysis of health outcomes will not be carried out. Related to that potential individual and clinical health impact is a potential economic impact. The economic impact focuses on whether image classification through a CNN model with high accuracy may benefit patients, medical staff, and clinics by saving resources primarily time for both professional clinicians and individuals or patients. It could be argued that commercialisation of image classification using models as CNNs may also be described as a business or investment case, with potential increase in income. We will as part of the study provide an argument if there may be a ’business case’ that could be further investigated. Though, the study will not go further than a statement on that.

3.9 Determining the model's efficiency

Quantitative and qualitative studies were needed to understand how reliable the use of CNN is in the field of image classification in medical care. The quantitative studies helped in discovering the efficiency level of using CNN as a class of artificial intelligence to classify the vitiligo. The efficiency of the CNN model was assessed based on two factors:

The performance: How well the chosen CNN architecture models did regarding the classification of vitiligo images. A quantitative study was done by collecting, comparing, and analysing the performance results such as accuracies, ROC AUC, mean of ROC AUC, and drawing ROC curves of the multi test in order to compare the performance of the selected CNN architecture models. This procedure is very important and necessary since it may help approve or reject the use of CNN in image classification of vitiligo upon the results, because it reveals the possibilities of errors occurring. Such procedures may help decision makers adopt such technology, especially if it is convenient, adds quality to diagnostics and, thereby, is convincing for the majority of patients.

The training time and energy are two important factors to consider when evaluating the CNN model. Using a CPU instead of a GPU processor can significantly increase the training time for deep CNN architectures. In this study, we examine the amount of time spent by each CNN model during the training phase. The training time information is collected after each training for the four models, presented in a table, and visualized in a graph of the training time by each model while training on the five folds.

4. Results

This chapter presents an overview of the results of a study conducted on the classification of vitiligo images using four pre-trained convolutional neural network models. Results include a statistical plot of the ROC and AUC for the four models, a plot of the loss rate as a function of epochs, and a few other relevant ROC statical and numerical plots. The Receiver Operating Characteristic (ROC) curves demonstrate how well each binary classifier performs. Following the plotting of the ROC curves, the Area Under Curve (AUC) can be calculated, and it measures the model's ability to classify the target data.

4.1 ROC curve of the 5-folds cross-validation

The study implemented the training process for the CNN models on the dataset, and then evaluated the models by using the validation datasets in the 5-folds to generate ROC curves. Below are the ROC curves represented over the five-folds, where Figure 4.1 represents the first fold, Figure 4.2 represents the second fold, etc.

In Table 4.1 the AUC values for each model varied depending on the fold. For example, VGG-19 achieved 0.94 in fold 4 and 0.85 in fold 0 which is a big difference, and this observation will be discussed later on. The curve for the VGG-19 model is higher in Figure 4.3 than in the others, and the AUC results in Table 4.1 confirm this finding. It may be that fold 2 samples are easier to classify than the samples in the other folds.

Performance of the InceptionV3 model is best on folds 1 and 2, but worst on fold 0 see Figure 4.1, where all CNN models perform badly compared to other folds.

Interestingly, InceptionResNetV2 model always has the lowest curve when compared to the other models, see Figure 4.4 except for fold 3 where InceptionResNetV2 has an AUC value of 0.7649 which is better than ResNeXT101.

The ResNeXt 101 achieved the highest AUC value in fold 2 when compared to the other models. However, its performance on fold 3 is the lowest out of all folds as well as the performance of all the other models on this fold, see Figure 4.4 and Table 4.1.

Table 4.1: Values of ROC AUC over the five-fold cross-validation for the four CNN architectures.

Fold	VGG-19	ResNeXt 101	InceptionV3	InceptionResNetV2
0	0.8568	0.8723	0.8648	0.8176
1	0.9606	0.9561	0.9819	0.9400
2	0.9854	0.9779	0.9849	0.9730
3	0.8124	0.7604	0.8216	0.7649
4	0.9402	0.9770	0.9504	0.7845

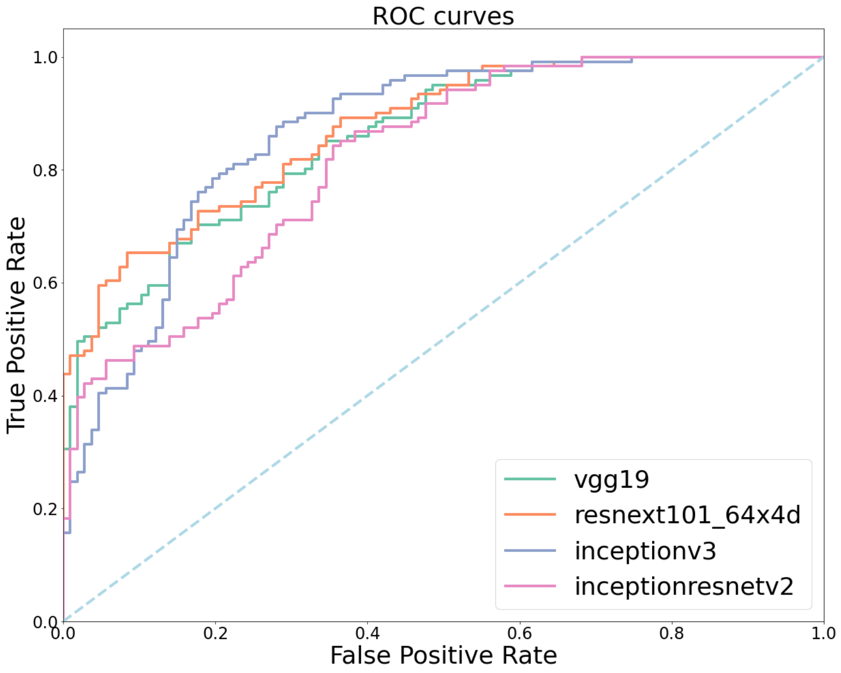

Figure 4.1: ROC curves for the four CNN architectures over fold 0.

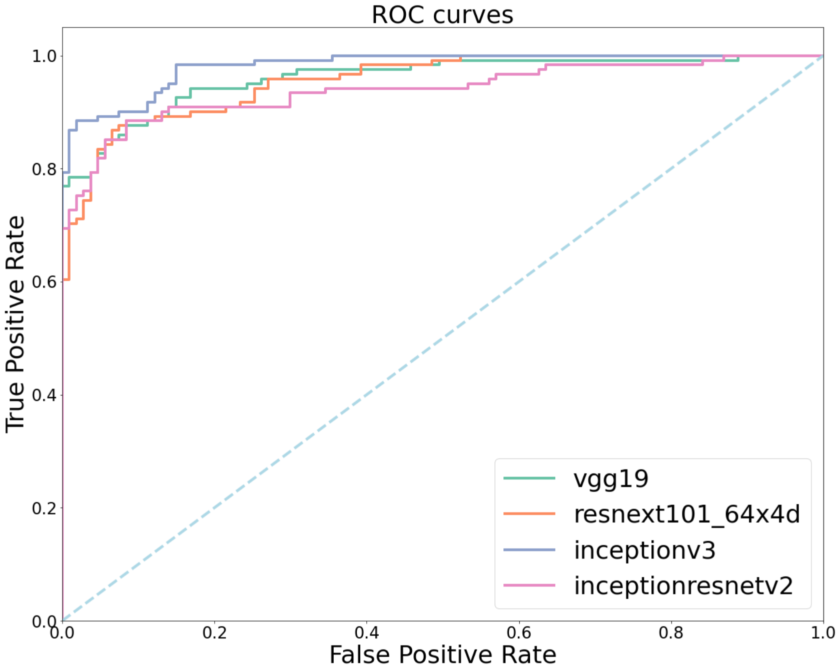

Figure 4.2: ROC curves for the four CNN architectures over fold 1.

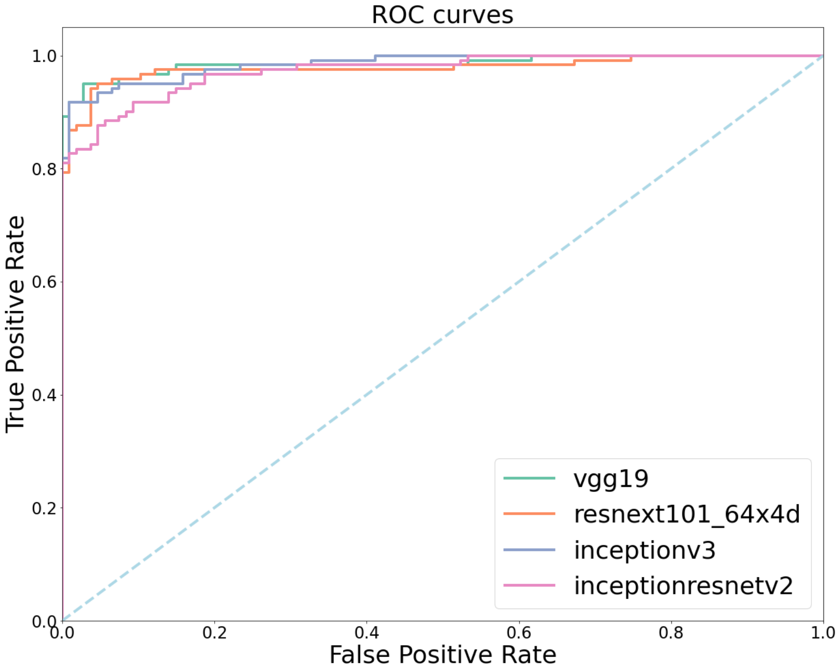

Figure 4.3: ROC curves for the four CNN architectures over fold 2.

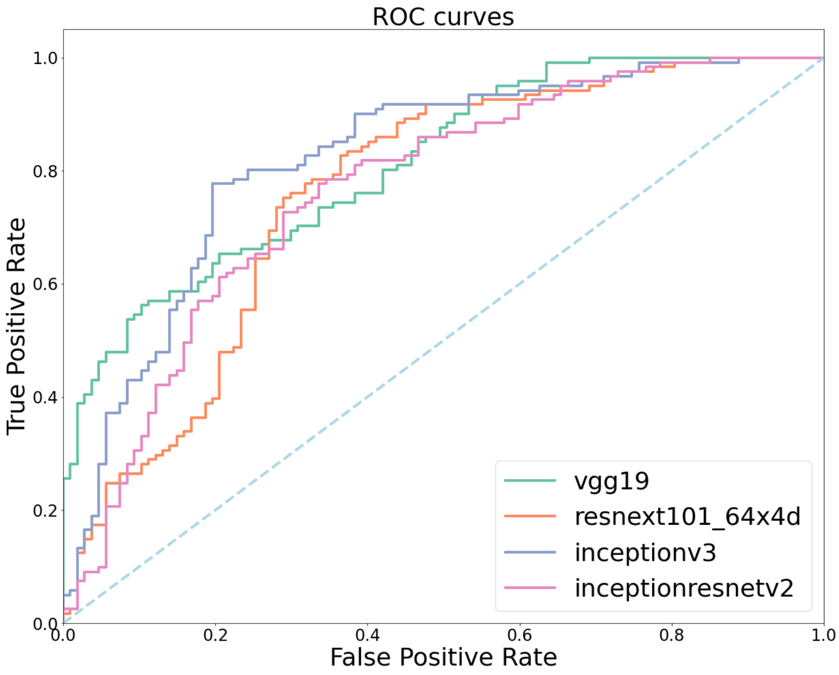

Figure 4.4: ROC curves for the four CNN architectures over fold 3.

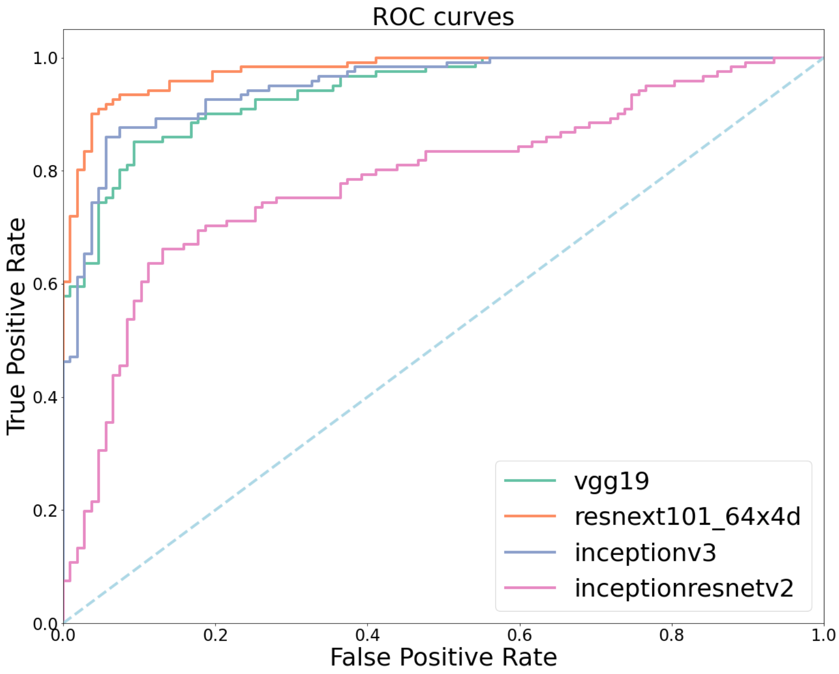

Figure 4.5: ROC curves for the four CNN architectures over fold 4.

Table 4.2: The mean AUC values for the CNN models.

The CNN Model	**Mean ROC AUC**
vgg19	0.9111
resnext101_64x4d	0.9087
InceptionV3	0.9207
inceptionresnetv2	0.8560

The mean AUC values in Table 4.2 were calculated based on the AUC values in Table 4.1. The mean AUC value for InceptionV3 supports our finding that it ranked highest among the CNN models. It has a mean AUC of 0.92, which is considered an excellent AUC value. While InceptionResNetV2 has the lowest mean ROC AUC value with 0.86, this is considered a good value as long as it is between 0.8 and 0.9.

4.2 AUC values of the 5-folds cross-validation

Figure 4.6 shows how the CNN models perform differently for different folds in the dataset. For instance, ResNeXt 101 achieves a high AUC value of over 0.95 for three folds, but its performance is very poor for fold 3, where its value is only 0.76. One of the most significant points in Figure 4.6 is that InceptionV3 still has the best AUC value of all the models over the most folds.

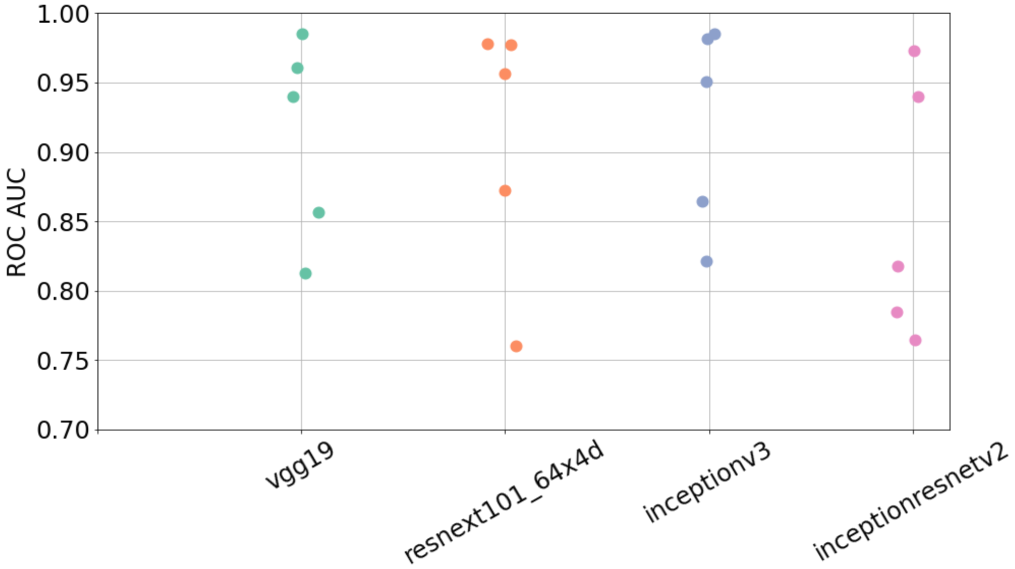

Figure 4.6: AUC values for the four CNN architectures over the five folds.

4.3 Plot of the CNN models' loss during training and validation

The authors examined the CNN models' loss rates during training and validation by plotting them for random folds. Figure 4.7 and Figure 4.8 show the loss rates for the best performing models, namely VGG-19 and InceptionV3. Therefore, the authors were interested in seeing how much their performance improved as they trained on more epochs.

According to Figure 4.7, the loss rate in the validation phase model continues to decrease as the number of epochs increases until it reaches 20 epochs. The train loss rate goes down more which suggests that the performance of the model is getting better.

Figure 4.8, the blue line indicates the performance of the training while the red line represents the performance of the validation. Neither training nor validation performance improve significantly with the increasing epochs number. A different setting of hyperparameters may yield different results.

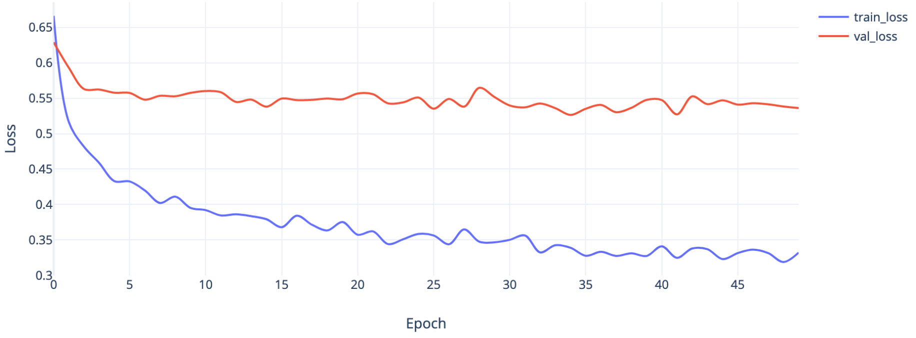

Figure 4.7: VGG-19 loss rate during training and validation on fold 3.

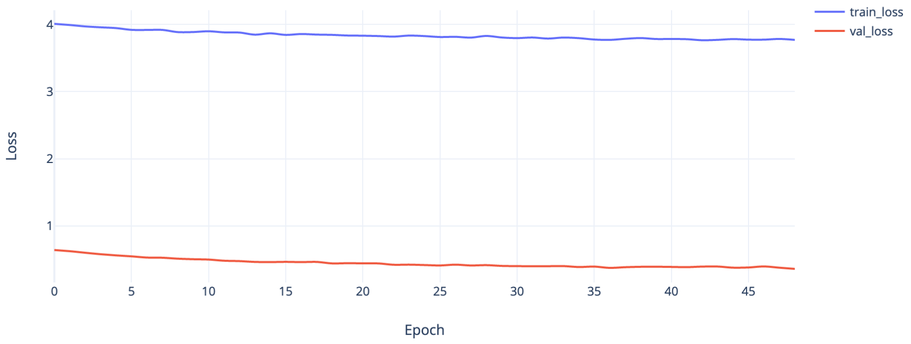

Figure 4.8: InceptionV3 loss rate during training and validation on fold 0.

4.4 The training time for the different models

According to the tests on the four selected CNN models, the InceptionV3 performed best according to the mean ROC AUC of 0.92 in Table 4.2. and lower ROC AUC of 0.82 which is higher than the lowest observed cross-validation for the other tested models. Table 4.9 shows that InceptionV3 is faster at completing the training phase on the five folds than other models. Figure 4.9 confirms this, where the time InceptionV3 took is represented in blue, and it is evident that InceptionV3 is faster in all folds except in fold 1. In fold 1, InceptionResNetV2 completed the training process in 55 minutes less time than InceptionV3.

Table 4.3 Training time for each folder and architectural model in minutes.

Folder	InceptionV3	VGG19	Resnext101	Inceptionresnetv2
0	259	541	578	512
1	305	788	843	250
2	261	525	581	551
3	254	510	568	622
4	250	714	504	457

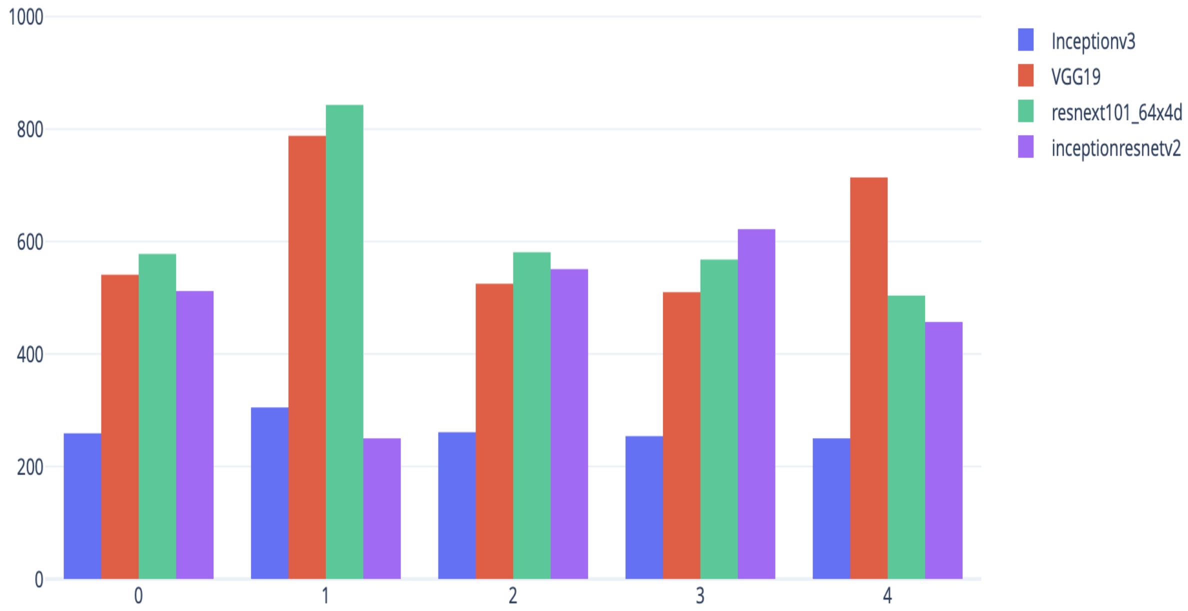

Figure 4.9: Illustrating the training time of each CNN architecture over the five folds.

Furthermore, the mean time of the InceptionV3 is quite low, at 265.8 minutes, while the mean time consumed by the resnext101_64x4d and VGG19 were the highest, at 614 minutes and 615 minutes, respectively. This large difference between the architectural models indicates that the use of InceptionV3 may also substantially lower resource costs.

Using InceptionV3 in dermatology may support doctors in providing a quick and accurate diagnosis. Such procedures may help the clinical staff provide quality customer service. If images are analysed by a system that is able to learn through analysis of datasets of patients' images, that can add to the quality and accuracy of diagnostics. Also, the system will never be tired. So, this may free up time for dermatologists and give them more time to analyse and take care of patients and provide treatments. This will also increase the productivity of the clinical organisation. From an economical aspect, the implementation of an accurate and valid model like InceptionV3 may assist dermatologists in the field of diagnosis of vitiligo and non-vitiligo images. The use of InceptionV3 may contribute to a reliable initial diagnosis.

5. Analysis and discussion

The chapter presents an overview of the study as well as observations on the performance of CNN architectures in vitiligo classification. Moreover, the authors discuss how the chosen methods affect the adjustment of the models' predictions. Further, the economic benefits of this technology in terms of resource costs such as time, and manpower are discussed.

The classification of images using CNN may have many positive health impacts for patients since it speeds up the diagnostic process without affecting the accuracy of the diagnosis because of its feature for self-learning from past analysis. Apart from the quick diagnosis, it may even give dermatologists more time to focus on more complicated or urgent cases. Furthermore, it may help dermatologists to deepen disease analysis in order to understand conditions more and to ensure early detection of the possibility of vitiligo. Besides possibilities of implementing studies to reach reliable treatments. However, the use may also have social negatives effects regarding employment of administrative workers at the clinics. This will depend on the solutions and automation level. But this may as well add opportunities for administrative workers to learn and administer the automation process of image classification.

5.1 Result analyses

The experiment of comparing the performance of four deep CNN models in the classification of images of vitiligo and normal skin was successful. In general, all models performed well in the classification task, with an AUC value of 0.98 in the best performance and an AUC value of 0.76 in the worst. As long as the AUC value is over 0.5, where 0.5 is considered the AUC value of a random classifier, even the worst model could perform well in the study.

Looking at the ROC plots over the five folds in Figure 4.6, it is interesting how different the performance of the same model is in different folds. That suggests that the dataset is quite heterogeneous. Generally, if the models perform equally well on all folds, then it would not matter what training or validation data was chosen, and therefore the performance of the models would be robust (within the dataset). However, for small datasets, it's more likely that certain images will be easier to predict than others, and if a random validation set contains more of those easy images, the result will be a better performance. We found, for instance, in Table 4.2 that all the CNN architectures achieved a higher AUC value on fold 2 than it did on the other folds, which means that the validation dataset for fold 2 contains easier images to classify.

In [46], Inception was employed in a vitiligo classification study with three other architectures. In that study, the authors applied almost the same data augmentations as we did on our dataset, which are typical for image classification tasks, and the AUC value for inception architecture was 0.912. In our approach, Inception achieved the highest AUC value of 0.920, and these two AUC values are relatively similar. Based on the study, the authors found ResNet 50 to achieve the best performance in terms of vitiligo classification with an AUC of 0.922, which is very close to the best value we achieved with InceptionV3 which is 0.920.

5.2 Training and validation loss rate

In the study, the authors tested the models on various epoch numbers in order to see if increasing the epoch number would decrease the loss rate, which would mean improving the model's performance.

Considering that VGG-19 achieved the second-best AUC value in vitiligo classification, we thought it was interesting to observe its loss rate during training and validation. VGG-19 loss rate during training and validation on fold 3 is represented in Figure 4.7. After around 20 epochs, validation loss appears to not improve much, while training loss rate goes down more, which means that increasing the number of epochs to 50 did not significantly affect the model's performance.

The loss rate curves of InceptionV3 in Figure 4.8, the architecture that achieved the best performance in vitiligo classification, have an impressive loss rate since the training loss seems high and does not improve when epoch numbers increase. However, InceptionV3 performed the best in vitiligo classification, which is what [39] confirms that generally, Inception architectures are efficient. Inception architectures, as predicted, are more accurate and efficient than other architectures because their networks are more complex and have more layers than any other deep CNN architecture in the study.

5.3 Evaluation reliability

The experiment was carried out using methods applied to similar medical image classification tasks [42,39], and the CNN architectures used are already trained on the ImageNet dataset [33]. A number of training tests using different hyperparameters were performed in order to improve the performance of the CNN architectures and ensure that the results were reliable. As an example, the authors ran several tests to determine the right batch size for InceptionResNetV2, because it was stopping training with a high batch size, so we chose batch size 8. Moreover, learning rate was also one of the parameters that improved the model's predictions. To estimate the learning rate 0.001, several different learning rates were attempted. The use of hyperparameter adjustment led to better data predictions in our case, as

[41] confirms by showing that adjusting network parameters can improve accuracy to 99.79%.

The results we found were generally reliable, depending on the methods used and the evaluation metrics used. However, we still see some differences in the models' performance across folds, which may be due to the size and nature of the dataset. Considering the heterogeneous nature of medical images in terms of shape, size, and appearance, they pose a challenge for CNN classification.

5.4 Economic, social, ethical, and environmental aspects

According to the results, InceptionV3 achieved the best record, while Inceptionresnetv2 was ranked runner-up with a mean of 478 training minutes. This indicates that Inception's architecture models give good results in terms of low training costs in [39] and accuracy in comparison with the models VGG19 and Resnext101_64x4d.

The results in general give a good impression of what the technology may contribute to the medical care field. Whereas medical care also needs time to conduct rapid and effective research and experiments, to understand and discover treatments for new and old illnesses, including those with dangerous implications for the immune system, such as vitiligo.

CNN is one of the most important tools to be implemented in the dermatology department, since it gives good accuracy results. This may be one of the reasons that made many recommend it to their loved ones in [22]. Furthermore, population expansion and rising life expectancy necessitate extensive effort in the medical care field. Consequently, the medical care section needs technological developments and support to be able to deliver efficient diagnosis and treatment. CNN is able to provide quick and accurate diagnosis.

From an ethical perspective, the data that is used to analyse images needs to be regulated. Data protection regulations must include the consent of the patient to have his personal information used for public studies and research, and to share this information on a specific platform that collects patient information for the benefit of scientific research and to serve patients in the future.

From a social and economic perspective, CNN could also support business growth in different ways, e.g., by saving time for patients, which is a form of customer value. Besides saving physicians' time and decreasing physician stress, it could also reduce administrative costs. Looking at these factors together, the technology may add positive effects for sustainability. Even though there may be arguments regarding the rise in unemployment and job redundancy or layoffs, idleness may also provide opportunities for learning and improving competence to find other jobs. CNN models cannot replace dermatologists. It will, however, enable dermatologists to work more efficiently and may give them time to focus on patients and complex cases.

From an environmental perspective, CNN's may help humans save various types of resources through its sustainable solutions and its ability to work long hours at the same high level. If a good processor is used, it is able to provide precise and sharp analysis with minimal energy consumption as well. GPU processors, for instance, have the ability to process enormous amounts of data in a short amount of time, which makes them more energy efficient. The CPU processor, on the other hand, consumes more power than the GPU processor due to the long time it needs to perform image analysis tasks. Furthermore, implementing CNN in medical care may serve society in different ways. Because of its ability to analyse, it may assist scientists in understanding new diseases that may come in the future. Besides, assisting in situations of lack of experts and working cadres in the medical section to avoid affecting the physicians because of the increasing number of patients, which may cause stress. Moreover, its ability to work longer than humans may help decrease working hours for workers, which may have positive psychological impacts, especially for countries that have hard climatic conditions such as long, dark winter days in Sweden.

6. Conclusions

This study had two main goals. The first to train VGG-19, ResNeXt 101, InceptionResNetV2 and InceptionV3 on vitiligo and non-vitiligo datasets, and then examine their performance in vitiligo image classification. Based on a limited dataset, the study examined the performance of four leading deep CNN architectures in vitiligo classification. Comparing the performance of these architectures, the study found that InceptionV3 performed the highest in vitiligo classification, but ResNet 50 performed better in another study [46]. There is not much difference between ResNet50's AUC value of 0.922 and InceptionV3's value of 0.920. Since Inception architectures are very complicated and have a lot of parameters, the performance of InceptionV3 could be made better by fine-tuning the model or adjusting the hyperparameters.

The second goal is to determine whether using vitiligo classifiers in health care reduces time and costs and improves the quality of disease diagnosis. Cooperation between physicians and knowledgeable CNN users may benefit physicians, patients, sustainability, and the health care organization. The CNN model may be expensive to set up and train, but it is less costly than human workers because it has auto-learning capabilities and performs at a high level regardless of the amount of time it works. In addition, CNN's costs vary depending on the complexity of the solution. Investments in medical care clinics are worthwhile and worth further analysis, as they may enhance an organisation efficiency and facilitate clinical work and patient quality. It also gives workers opportunities to increase their competence in their jobs.

6.1 Future work

During training, a high loss rate was observed for the InceptionV3 model; yet this model achieved the highest AUC compared to other architectures. We believe that InceptionV3 could perform better than ResNet50 in vitiligo classification if a larger dataset was used to train the models. Thus, we propose investigating the performance of both ResNet50 and InceptionV3 in vitiligo classification and fine-tuning both models in order to improve predictions and obtain higher accuracy.

References

1. Gill L, Zarbo A, Isedeh P, Jacobsen G, Lim HW, Hamzavi I. Comorbid autoimmune diseases in patients with vitiligo: A cross-sectional study. J Am Acad Dermatol [Internet]. 2016 [cited 2022 April 10];74(2):295–302. Available from: https://pubmed.ncbi.nlm.nih.gov/26518171/
2. Kota RS, Vora RV, Varma JR, Kota SK, Patel TM, Ganjiwale J. Study on assessment of quality of life and depression in patients of vitiligo. Indian Dermatol Online J [Internet]. 2019 [cited 2022 April 10];10(2):153–7. Available from: https://pubmed.ncbi.nlm.nih.gov/30984590/
3. Vitiligo facts [Internet]. Global Vitiligo Foundation. 2019 [cited 2022 April 22]. Available from: https://globalvitiligofoundation.org/vitiligo-facts/
4. Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. Gastrointest Endosc [Internet]. 2020 [cited 2022 April 22];92(4):807–12. Available from: https://pubmed.ncbi.nlm.nih.gov/32565184/
5. Can AI help achieve environmental sustainability? [Internet]. Earth.Org - Past | Present | Future. 2021 [cited 2022 April 22]. Available from: https://earth.org/data_visualization/ai-can-it-help-achieve-environmental-sustainable/
6. How AI may solve our health problems [Internet]. .Ki.se. [cited 2022 April 22]. Available from: https://ki.se/en/research/how-ai-may-solve-our-health-problems
7. Insider Intelligence. How the medical field is benefiting from AI in 2022 and beyond [Internet]. Insider Intelligence. 2022 [cited 2022 April 22. Available from: https://www.insiderintelligence.com/insights/artificial-intelligence-healthcare/
8. Taylor M. Computer Vision with Convolutional Neural Networks [Internet]. The Startup. 2020 [cited 2022 April 22]. Available from: https://medium.com/swlh/computer-vision-with-convolutional-neural-networks-22f06360cac9
9. Autoimmune diseases [Internet]. Cancer Treatment Centers of America. 2021 [cited 2022 Apr 24]. Available from: https://www.cancercenter.com/risk-factors/autoimmune-diseases
10. Autoimmune Diseases [Internet]. Cleveland Clinic. [cited 2022 Apr 24]. Available from: https://my.clevelandclinic.org/health/diseases/21624-autoimmune-diseases
11. Appen. How Artificial Intelligence Data Reduces Overhead Costs for organizations [Internet]. Appen. 2022 [cited 2022 May 6]. Available from: https://appen.com/blog/how-artificial-intelligence-data-reduces-overhead-costs-for-organizations/
12. Papers with code - ImageNet benchmark (image classification) [Internet]. Paperswithcode.com. [cited 2022 April 13]. Available from: https://paperswithcode.com/sota/image-classification-on-imagenet
13. Copeland BJ. artificial intelligence. In: Encyclopedia Britannica. 2022. Available from: https://www.britannica.com/technology/artificial-intelligence
14. Liu, X., Gao, K., Liu, B., Pan, C., Liang, K., Yan, L., Ma, J., He, F., Zhang, S., Pan, S., & Yu, Y. (2021). Advances in deep learning-based medical image analysis. Health Data Science, 2021, 1–14. https://doi.org/10.34133/2021/8786793
15. Alkhodari, M., & Fraiwan, L. (2021). Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings. Computer Methods and Programs in Biomedicine, 200(105940), 105940. https://doi.org/10.1016/j.cmpb.2021.105940
16. Hou Y. Breast cancer pathological image classification based on deep learning. J Xray Sci Technol [Internet]. 2020 [cited 2022 Apr 30];28(4):727–38. Available from:https://content.iospress.com/articles/journal-of-x-ray-science-and-technology/xst200658
17. Cai L, Gao J, Zhao D. A review of the application of deep learning in medical image classification and segmentation. Ann Transl Med [Internet]. 2020 [cited 2022 Apr 30];8(11):713. Available from:http://dx.doi.org/10.21037/atm.2020.02.44
18. Prabhu. Understanding of Convolutional neural network (CNN) — deep learning [Internet]. Medium. 2018 [cited 2022 May 1]. Available from:

https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148

19. Gomolin A, Netchiporouk E, Gniadecki R, Litvinov IV. Artificial intelligence applications in dermatology: Where do we stand? Front Med (Lausanne) [Internet]. 2020; 7:100. Available from: http://dx.doi.org/10.3389/fmed.2020.00100
20. Supervised vs. Unsupervised Learning: What's the difference? [Internet]. Ibm.com. [cited 2022 May 3]. Available from: https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning
21. Cao, Huaigang & Wang, Wenbo & Lin, su & Ni, Haiyan & Gerstoft, Peter & Ren, Qunyan & Ma, Li. (2021). Deep transfer learning for underwater direction of arrival using one vector sensor. The Journal of the Acoustical Society of America. 149. 1699-1711. 10.1121/10.0003645. Researchgate.net. [cited 2022 May 3]. Available from:https://www.researchgate.net/publication/349987639_Deep_transfer_learning_for_underwater_direction_of_arrival_using_one_vector_sensor
22. Nelson CA, Pérez-Chada LM, Creadore A, Li SJ, Lo K, Manjaly P, et al. Patient perspectives on the use of artificial intelligence for skin cancer screening: A qualitative study: A qualitative study. JAMA Dermatol [Internet]. 2020;156(5):501–12. Available from: http://dx.doi.org/10.1001/jamadermatol.2019.5014
23. Agah A. Medical applications of artificial intelligence [Internet]. Agah A, editor. Boca Raton, FL: CRC Press; 2013. Available from: https://books.google.at/books?id=tRDSBQAAQBAJ
24. Khan A, Sohail A, Zahoora U, Qureshi AS. A survey of the recent architectures of deep convolutional neural networks. Artif Intell Rev [Internet]. 2020;53(8):5455–516. Available from: http://dx.doi.org/10.1007/s10462-020-09825-6
25. Gomez Rossi J, Rojas-Perilla N, Krois J, Schwendicke F. Cost-effectiveness of artificial intelligence as a decision-support system applied to the detection and grading of melanoma, dental caries, and diabetic retinopathy. JAMA Netw Open [Internet]. 2022;5(3):e220269. Available from: http://dx.doi.org/10.1001/jamanetworkopen.2022.0269
26. Mulhern O. Can AI help achieve environmental sustainability? [Internet]. Earth.Org - Past | Present | Future. 2021 [cited 2022 May 5]. Available from: https://earth.org/data_visualization/ai-can-it-help-achieve-environmental-sustainable/
27. Gupta A. Evolution of convolutional neural network architectures [Internet]. The PEN Point. 2020 [cited 2022 May 4]. Available from: https://medium.com/the-pen-point/evolution-of-convolutional-neural-network-architectures-6b90d067e403
28. De A, Sarda A, Gupta S, Das S. Use of artificial intelligence in dermatology. Indian J Dermatol [Internet]. 2020;65(5):352–7. Available from: http://dx.doi.org/10.4103/ijd.IJD_418_20
29. Shrivastav A. Different types of CNN models [Internet]. OpenGenus IQ: Computing Expertise & Legacy. 2021 [cited 2022 May 3]. Available from: https://iq.opengenus.org/different-types-of-cnn-models/
30. Luzniak K. *Cost of AI in healthcare industry [Internet]. Neoteric. 2021 [cited 2022 May 5]. Available from: https://neoteric.eu/blog/whats-the-cost-of-artificial-intelligence-in-healthcare/*
31. Kuflinski Y. The pros and cons of artificial intelligence: A global outlook [Internet]. Iflexion. 2020 [cited 2022 May 5]. Available from: https://www.iflexion.com/blog/pros-and-cons-artificial-intelligence
32. Zhang L, Mishra S, Zhang T, Zhang Y, Zhang D, Lv Y, et al. Design and assessment of convolutional neural network-based methods for vitiligo diagnosis. Front Med (Lausanne) [Internet]. 2021;8:754202. Available from: http://dx.doi.org/10.3389/fmed.2021.754202
33. Finetuning Torchvision Models — PyTorch Tutorials 1.2.0 documentation [Internet]. Pytorch.org. [cited 2022 April 15]. Available from: https://pytorch.org/tutorials/beginner/finetuning_torchvision_models_tutorial.html
34. Dantas J. The importance of k-fold cross-validation for model prediction in machine learning [Internet]. Towards Data Science. 2020 [cited 2022 May 7]. Available from:

https://towardsdatascience.com/the-importance-of-k-fold-cross-validation-for-model-prediction-in-machine-learning-4709d3fed2ef

35. Sinha J, Manollas M. Efficient deep CNN-BiLSTM model for network intrusion detection. In: Proceedings of the 2020 3rd International Conference on Artificial Intelligence and Pattern Recognition. New York, NY, USA: ACM; 2020.
36. Luo W, Liu J, Huang Y, Zhao N. An effective vitiligo intelligent classification system. J Ambient Intell Humaniz Comput [Internet]. 2020 [cited 2022 May 1];1–10. Available from: https://link.springer.com/article/10.1007/s12652-020-02357-5
37. The evolution of ImageNet for deep learning in computer vision [Internet]. Analytics India Magazine. 2020 [cited 2022 April 12]. Available from: https://analyticsindiamag.com/imagenet-and-variants/
38. Low M, Huang V, Raina P. Automating vitiligo skin lesion segmentation using convolutional neural networks. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE; 2020. p. 1–4.
39. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, Inception-ResNet and the impact of residual connections on learning. arXiv [csCV] [Internet]. 2016 [cited 2022 May 7]; Available from: http://arxiv.org/abs/1602.07261
40. Soni P. Data augmentation: Techniques, benefits and applications [Internet]. Analyticssteps.com. [cited 2022 May 8]. Available from: https://www.analyticssteps.com/blogs/data-augmentation-techniques-benefits-and-applications
41. Tuba E, Bačanin N, Strumberger I, Tuba M. Convolutional Neural Networks Hyperparameters Tuning. In: Artificial Intelligence: Theory and Applications. Cham: Springer International Publishing; 2021. p. 65–84.
42. Ashraf A, Naz S, Shirazi SH, Razzak I, Parsad M. Deep transfer learning for Alzheimer neurological disorder detection. Multimed Tools Appl [Internet]. 2021;80(20):30117–42. Available from: http://dx.doi.org/10.1007/s11042-020-10331-8
43. Narkhede S. Understanding AUC - ROC Curve [Internet]. Towards Data Science. 2018 [cited 2022 April 28]. Available from: https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5
44. De Man R, Gang GJ, Li X, Wang G. Comparison of deep learning and human observer performance for detection and characterization of simulated lesions. J Med Imaging (Bellingham) [Internet]. 2019 [cited 2022 May 9];6(2):025503. Available from: http://dx.doi.org/10.1117/1.JMI.6.2.025503
45. Robot scientist 'works 1,000 times faster' than human researchers [Internet]. Times Higher Education (THE). 2020 [cited 2022 May 9]. Available from: https://www.timeshighereducation.com/news/robot-scientist-works-1000-times-faster-human-researchers
46. Liu J, Yan J, Chen J, Sun G, Luo W. Classification of Vitiligo Based on Convolutional Neural Network. In: Lecture Notes in Computer Science. Cham: Springer International Publishing; 2019. p. 214–23.
47. IRJET Journal. IRJET- implication of convolutional neural network in the classification of vitiligo. IRJET [Internet]. 2020 [cited 2022 May 11]; Available from: https://www.academia.edu/44187410/IRJET_Implication_of_Convolutional_Neural_Network_in_the_Classification_of_Vitiligo?auto=citations&from=cover_page
48. Manson EN, Ampoh A, Fiagbedzi E, Amuasi JH, Flether JJ, Schandorf C, et al. Curr Trends Clin Med Imaging. Curr trends clin med imaging [Internet]. 2019; Available from: https://juniperpublishers.com/ctcmi/pdf/CTCMI.MS.ID.555620.pdf
49. Noise [Internet]. Image-engineering.de. [cited 2022 May 20]. Available from: https://www.image-engineering.de/library/image-quality/factors/1080-noise