Journal of Knowledge Learning and Science Technology
ISSN: 2959-6386 (Online)
2024, Vol. 4, No. 3, pp. 200–212
DOI: https://doi.org/10.60087/jklst.v3.n4.p200

Research Article

Convoluted neural network and transfer learning algorithm for improved brain tumor classifications in MRI

Diya Raina[1], Aarna Dawange[2], Thamaru Bandha[3], Anureet Kaur[4], Rakshit Wasekar[5], Kritika Verma[6], Saloni Verma[7], Karan Dhingra[8]

[1]Independent Researcher, Maharashtra, India
[2]Independent Researcher, Maharashtra, India
[3]Independent Researcher, Telangana, India
[4]Independent Researcher, Punjab, India
[5]Independent Researcher, Maharashtra, India
[6]Department of Computer Science, Syracuse University, New York, United States of America
[7]Department of Biomedical Engineering, Cornell University, Ithaca, New York, United States of America
[8]Department of Biomedical Engineering, University of Ottawa, Ontario, Canada

Abstract

Artificial intelligence (AI) has made significant use cases to improve patient care, particularly in medical image analysis. This study aims to develop a deep-learning model for disease classification in medical images and compare its performance in four-class MRI and two-class X-ray classification tasks. We utilize Convolutional Neural Networks (CNNs) for diagnosing pneumonia from chest X-rays and various tumors from brain MRIs, leveraging transfer learning to improve performance. Transfer learning, which reuses pre-trained models like VGG-16, is more efficient than building models from scratch. The VGG-16 model, pre-trained on over a million ImageNet images, achieved 92.7% accuracy. By fine-tuning, we reached 93.6% accuracy. Data augmentation techniques, such as flipping, rotation, and brightness adjustments, further enhance classification accuracy and performance.

Keywords

Artificial Intelligence, Medical Devices, Machine Learning, Image Classification, Convoluted Neural Networks

*Corresponding author: Aarna Dawange

Email addresses:

diyaraina27@gmail.com (Diya Raina), aarnasd08@gmail.com (Aarna Dawange), bandhathamaru@gmail.com (Thamaru Bandha), shehbajsinghmannan343@gmail.com (Anureet Kaur), rakshitwasekar@gmail.com (Rakshit Wasekar), krverma@syr.edu (Kritika Verma) sv458@cornell.edu (Saloni Verma), k2dhingr@gmail.com (Karan Dhingra)

Received: 17-08-2024; **Accepted:** 22-09-2024; **Published:** 03-09-2024

Journal of Knowledge Learning and Science Technology
ISSN: 2959-6386 (Online)
2024, Vol. 4, No. 3, pp. 200–212
DOI: https://doi.org/10.60087/jklst.v3.n4.p200

1. Introduction

Medical image analysis has been an important aspect in many ways, helping the doctors and the diagnosis for further medical processing. It reveals the internal structures hidden under the skin and detects the type of disease that occurred, the crucial treatment can be done [1]. In the year 2017 when we received the approval to introduce artificial intelligence (AI) into the medical field for detecting distinct images it helped in a more concise way with an increase in accuracy, improving the quality and data processing. Day by day the technology has been improved making it possible to control the numerous diagnoses within a short period and helping the doctors in a great way providing detailed information about each problem that occurred [2]. Machine learning (ML) is an AI algorithm that allows the system to learn from the data and provides accurate predictions about the abnormality by identifying the patterns from training sets [3]. Deep learning (DL) is the subset of ML that conforms to whether the results were accurate or predicted without any human intervention. Convolutional neural network (CNN) has been an impactful artificial network helping in image analysis, image registration, image segmentation, feature extraction, and classification [4]. From the year 2014, many image algorithms have been developed in the medical field to enhance perfect diagnostic results [5]. In 2006 the concept of deep learning was proposed with an artificial neural network (ANN) model through several layers. Scientists such as Bzdok and Ioannidis, Litjens et al, Razzak et alKourou et al, and many more have worked on the development of neural networks, provided the importance of DL models and their roles and have studied different backgrounds and their tremendous results of the related images [6]. They also identified the challenges in using the networks and discussed the importance of exploration, inference, and prediction in the fields of neuroscience and biomedical science [7]. The scientists have highlighted the potential of deep learning algorithms and even the challenges that took the frame, they have analyzed several challenges associated with the use of these algorithms such as pixel or data quality, identifying relevant health information and privacy concerns, and even found a route to the solution to each of the problems facing in the field [8]. They found the importance of integrating the data types such as neuroimages, genetics, and many more. The deep learning models have become the most prominent topic around the machine learning area that could help in almost every factor of human life, even displaying future births [9].

Digital image processing and medical image analysis can significantly support medical diagnosis by providing the necessary tools for automatic detection, extracting significant information, and accurate measurement of visible abnormalities [10]. However, amid the high expectations of the accuracy and efficiency that AI can bring to medicine, many challenges have yet to be overcome to integrate the new generation of CAD tools into clinical practice and to minimize the risk of unintended harm to patients [11]. The challenges related to medical images include automation challenges, security challenges, clinical challenges, technical challenges, and challenges that arise during the various stages of image processing including segmentation challenges and pre-processing challenges. However as deep learning methods have achieved state-of-the-art performance over different medical applications, its use for further improvement can be the major step in the medical computing field. Therefore, there is a significant need to fill the gap of missing a comprehensive overview of these challenges [12].

Errors or uncertainties in image processing cannot be avoided. Some of them are common in image processing such as those related to image acquisition (limited resolution, distortion). Others are related to the interpretation of the images by humans, as well as those related to the limited capability of the used techniques and methods [13]. Medical image analysis by human experts is relatively limited due to image complexity, the existence of wide disparities across diverse experts, and fatigue. Automated tools for image acquisition, diagnosis, enhancement, and interpretation based on machine learning algorithms provide accurate and efficient solutions for improving medical image processing. Hence, Despite all the given challenges, digital image processing technologies provide the most effective medical image processing, which helps in disease diagnosis as well as many treatments [14]. Digital image processing extracts meaningful information from images, automates image-based tasks, and enhances the visual quality of images, making them clearer, sharper, and more informative. It can automate many image-based tasks, such as object recognition, pattern detection, and measurement. Digital image processing algorithms can process images much faster than humans, making it possible to analyze large amounts of data in a short amount of time and can also provide more accurate results, especially for tasks that require precise measurements or quantitative analysis [15]. DL techniques using neural networks have been used for an array of applications including supporting clinicians in achieving diagnostic perfection [16]. Following this, a high-tech algorithm or an image analysis tool is developed that can be used to diagnose basic diseases using X-ray or MRI pictures. This model is built using the weights from the VGG-16 pre-trained model i.e. by the mode of transfer learning. Also, a rationale behind using this pre-trained model and general networks that aid in improving accuracy is provided in this research article.

2. Methods

By leveraging a large amount of annotated data, DL models can learn intricate patterns and relationships within medical images, facilitating accurate detection, localization, and diagnosis of disease and abnormalities [17]. Creating an image diagnosis of X-rays and MRI images can support clinicians in achieving diagnostic perfection. However deep learning methods are highly effective when the number of available samples is large [18]. Existing research says that through deep learning the highest performance reached a top curve that X-rays and MRI images can identify and detect multiple images only after training on extensive data sets. In 2012, Krizhevsky et al proposed a CNN with five convolutional layers and 3 fully connected layers (named “AlexNet”) containing over 60 million weights for training 1.2 million images with annotations and achieved in ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [11] that classified over 1000 classes of images. DL enables the quantitative analysis of medical images, extracting meaningful measurements and biomarkers. For example, estimating tumor volume, measuring decreased progression time, or assessing treatment response by qualifying changes in medical images; this quantitative information assists clinicians in making objective assessments, monitoring disease progression, and tailoring treatment plans [19]. For data-driven learning, large-scale and well-annotated datasets with representative data distribution characteristics are crucial to learning more accurate or general models. The CNN models trained upon this database serve as the backbone for significantly improving many object detection and image segmentation problems [20]. One of the primary advantages of deep learning over traditional ML methods is its capacity to learn features from raw data automatically, allowing it to capture underlying patterns, making it powerful with large, complex datasets. Deep learning has also facilitated significant advancements in various tasks, including but not limited to image and speech recognition, comprehension of natural language, and the development of capabilities [19, 21].

2.1. Dataset and Image Parameters

For X-ray & MRI, there are many datasets available to conduct training and testing through including but not limited to VinBigData for chest X-rays, Re3Data for general research, CT Data Set for cancer research and the National Institute of Health (NIH) Database for chest X-rays. In this study, we used Brain MRIs for Tumor Classification (1,311 images) and Chest X-ray Images (Pneumonia) (5,863 images) as described in Table 1. A total of 40932676 image parameters were used for the model with, 26217988 trainable parameters, and 14714688 non-trainable parameters.

<table>
<tr><th>Dataset</th><th>Modality</th><th>Size</th><th colspan="2">Split Ratio (train/test)</th><th>Resolution</th></tr>
<tr><td colspan="2">Brain MRIs for Tumor Detection</td><td>MRIs</td><td>1,311</td><td>80/20</td><td>224x224</td></tr>
<tr><td>Chest X-Ray Images</td><td>X-rays</td><td>5,863</td><td colspan="2">80/20</td><td>224x224</td></tr>
</table>

Table 1. Dataset parameters used in the study.

2.2. Evaluation Index and Metrics

A typical medical image analysis system (here, VGG-16) is evaluated by using different key performance measures such as accuracy, F1-score, precision, recall, sensitivity, specificity, and dice coefficient [22]. Mathematically, these classification metrics are calculated as,

$$F1 - score = 2 \times \frac{(precision \times recall)}{(precision + recall)} \tag{1}$$

$$Precision = \frac{TP}{(TP+FP)} \tag{2}$$

$$Recall = \frac{TP}{(TP+TN)} \tag{3}$$

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (4)$$

$$Sensitivity = \frac{TP}{(TP+FN)} \quad (5)$$

$$Specificity = \frac{TP}{(TP+FP)} \quad (6)$$

$$Dice\ Score = \frac{2 \times |P \cap GT|}{|P|+|GT|} \quad (7)$$

Where,

TP = true positive, represents the number of cases correctly recognized as defected.

FP = false positive, represents the number of cases incorrectly recognized as defected.

TN = true negative, represents the number of cases correctly recognized as non-defected.

FN = false negative, represents the number of cases incorrectly recognized as non-defected.

P = prediction as given by the system being evaluated for a given testing sample.

GT = ground truth of the corresponding testing sample.

Table 2 describes the applications and descriptions of evaluation metrics.

Metric	Description	Application
Accuracy	Proportion of correct predictions	Overall Performance
F1 Score	Combines Precision and Recall	Indicates Reliability of the model
Recall	Quantifies amount of positive class predictions	Identifies the model's ability to make correct predictions
Precision	Indicates how often a machine makes true positive predictions	Identifies the model's ability to make correct predictions
Sensitivity	True Positive Rate	Identifies how many positive instances the model was able to identify accurately
Specificity	True Negative Rate	Identifies how many negative instances the model was able to identify accurately
Dice Score	Measure of similarity between two data sets	Evaluates the similarity between a predicted segmentation mask and the ground truth segmentation mask [23]

Table 2. *Applications of evaluation metrics.*

2.3. Architecture

The application of machine learning to MRI and X-ray data facilitates the transformation of these datasets into mathematical models, enhancing data analysis efficiency. These models categorize the data into three distributions: (a) Positive skew (most values cluster around values less than the mean value). (b) Symmetrical Distribution (equal frequencies of data above and below the central value). (c) Negative skew (most values cluster around values greater than the mean value), all of which operate under the principles of probability and statistics, enabling a deeper understanding of the data through mathematical problem-solving. By leveraging statistical measures such as mean, mode, and median, these models achieve higher accuracy in data analysis [24].

Transfer learning (TL) stems from cognitive research that uses the idea that knowledge is transferred across related tasks to improve performance on a new task [25]. Benefitting from the development of deep learning, the analysis of medical images, which used to be a challenging, yet exhausting task carried out manually by physicians, has also experienced fast development. Small-scale data can't guarantee the performance of the developed systems, while large-scale data is usually unavailable due to expensive costs in the collection and storage process [26]. To allow a fast transition from one domain to another for reuse, experts and researchers have extensively delved into transfer learning, which is an efficient and low-cost learning technique [27]. Applying these concepts to machine learning and neural networks, a network can be trained by transferring the weights and specifications of another pre-trained model. The exception is the last fully connected layer whose number of nodes depends on the number of classes in the dataset. A common practice is to replace the last fully connected layer of the pre-trained CNN with a new fully connected layer that has as many neurons as the number of classes in the new target application [28]. The model aims to use Transfer Learning using VGG16 for Image Classification of Brain MRIs and Chest X-rays. The top/output layer of the VGG16 is removed to add various other layers about the requirements of classifying the medical images.

[vgg = VGG16(input_shape = [224, 224, 3], weights = "imagenet", include_top=False)]

Keras was used as a Python interface for neural networks. Scikit-Learn (sklearn) for data library for predictive data analysis. NumPy (nmp) for support towards multidimensional arrays, mathematical functions, and numerical computing.

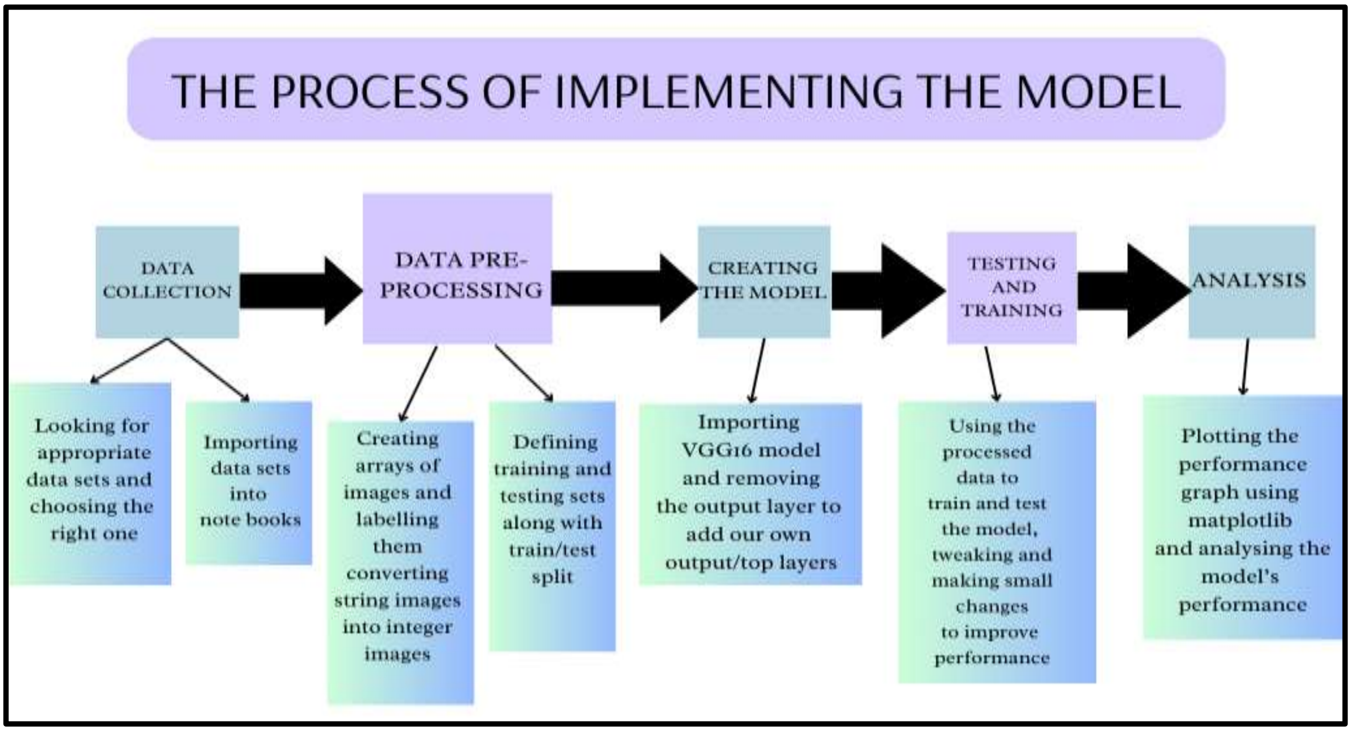

Figure 1. *An approach used to create a model for image classification of Brain MRIs and Chest X-rays to aid Health Care Providers in disease diagnosis.*

TensorFlow (tsf) for creating and deploying ML models and interacting with the OS. OpenCV (cv2) for computer Vision and ML. Matplotlib for creating Visualizations in Python and tqdm for displays iteration based as progress bars. VGG16 has a uniform architecture and it has been pre-trained on the ImageNet dataset with millions of images and thus can effectively be used for Transfer Learning. The VGG16 model is a deep-layer model with 16 weight layers. It contains the following layers [29]: 13 convolution layers, 2 fully connected layers, 1 softmax classifier, which is further broken down into, Layer 1,2: Convolutional layers, Filter Size: 3x3, Image Size is changed into 224x224x64. Output passed through, mac pooling layer (stride:2), layer 3-4 convolution layers, 124 kernel filters (3x3), followed by a max pooling layer (stride:1), layers 8-13 2 sets of convoluted layers (kernel size 3x3), 512 kernel filters each, followed by a max pooling layer (stride:1), layer 14-15 fully hidden layers (4096 units) and layer 16 softmax output layer (1000 units). This architecture was introduced by Karen Simonyan and Andrew Zisserman [30]. To enhance the performance of our machine learning model via Transfer Learning, we undertook a systematic approach involving the removal of existing output layers and the integration of supplementary layers. This adjustment was primarily targeted toward reducing the risk of overfitting, thereby enhancing the model's generalization capabilities. Concurrently, this modification was designed to maximize both training and validation accuracies, ensuring that the model maintains high performance even when confronted with novel data inputs. The layers along with the rationale behind their use include:

Layer type	Output shape	Param #
input_1 (InputLayer)	[(None, 224, 224, 3)]	0

block1_conv1 (Conv2D)	[(None, 224, 224, 64)]	1792
block1_conv2 (Conv2D)	[(None, 224, 224, 64)]	36928
block1_pool (MaxPooling2D)	[(None, 112, 112, 64)]	0
block2_conv1 (Conv2D)	[(None, 112, 112, 128)]	73856
block2_conv2 (Conv2D)	[(None, 112, 112, 128)]	147584
block2_pool (MaxPooling2D)	[(None, 56, 56, 128)]	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 1024)	25691136
dropout (Dropout)	(None, 1024)	0
dense_1 (Dense)	(None, 512)	524800
dense_2 (Dense)	(None, 2)	1026

***Table 3*.** A total of 40931650 (156.14 MB) parameters were tested including 26216962 (100.01 MB) trainable parameters and 14714688 (56.13 MB) non-trainable parameters.

(a) Dense Layers: To add fully connected layers with either 512/1024 neurons. (b) Dropout Layer: To drop random neurons by a decided fraction into each training cycle to help prevent overfitting. (c) Flatten Layer: To convert images to 1D Feature Vectors.

```
[ def lw(bottom_model, num_classes):

top_model=bottom_model.output

top_model=Flatten()(top_model)

top_model=Dense(1024,
activation='relu')(top_model)top_model=Dropout(0.5)(top
_model)

    top_model=Dense(512,
    activation='relu')(top_model)
    top_model=Dense(num_classes,
```

activation='softmax')(top_model)

return top_model]

2.4. Image classification

Image classification involves assigning different labels to our datasets [31]. In our case, the classes and labels in the MRI dataset include: Glioma Tumor (G), Meningioma Tumor (M), Pituitary Tumor (P), No Tumor (NT). The classes and labels in the Chest X-ray dataset include Pneumonia (P) and Normal (N).

2.4.1. Preprocessing

Preparing raw images from our dataset for subsequent analysis by enhancing their quality using various techniques is known as Image Preprocessing [32, 33]. The method of preprocessing we used is Label Encoding. It converts categorical columns to numerical ones so that they can be fitted in ML models [34]

[lbrain = preprocessing.LabelEncoder()

Y1_train = lbrain.fit_transform(Y1_train)

Y1_test = lbrain.fit_transform(Y1_test)

Y1_train = tsf.keras.utils.to_categorical(Y1_train, num_classes=4)

Y1_test = tsf.keras.utils.to_categorical(Y1_test, num_classes=4)

Y1_train = nmp.array(Y1_train)

X1_train = nmp.array(X1_train)

Y1_test = nmp.array(Y1_test)

X1_test = nmp.array(X1_test)]

2.4.2. Testing and training data

The training dataset includes images that are used to train the model and help the model learn to predict the labels. The test dataset involves images whose labels are predicted by the model as a way to test the model and answer the research question [35].

The test train split involves dividing our data into two datasets for training and testing purposes. We used a randomized or cross-validated test train split which is the most widely used method [36]. This involves dividing our data into two arrays, in our case 'X' and 'Y', which will store the image data and the corresponding labels respectively. After creating the X and Y arrays we split our dataset into X_train (containing images from the training dataset), Y_train (containing labels from the training dataset), X_test (containing images from the test dataset), and Y_test (containing labels from the test dataset). We also specify test_size (ratio of test dataset to train dataset) and random_state [37].

This involves fitting the training and testing data into the model we prepared. In this, we specify our X and Y train sets, and our X and Y test sets as our validation data. The validation data is the data that can give an estimate of the model's skills by testing it and comparing the predicted labels with the actual labels [38]. We also specify the number of epochs which means the number of times the learning algorithm will work through the entire training dataset [39].

We used MatplotLib which is a Python plotting library to create a graph of our model [40]. This gives us a visual depiction of our Validation & Training Accuracy and Validation & Training Loss and helps us to better understand the performance of the model.

2.5. Transfer learning and data augmentation

We observed slight overfitting (suggested by higher training accuracy but lower validation accuracy and higher validation loss) and potential bias in the model due to reduced ability to learn newer features from unseen data [41]. Potential solutions to these limitations include 1) batch normalization layers to help stabilize and increase the speed of the training process (L2 regularization). This helps in making the weight distribution based on the coefficients and prevents any one feature from being dominant, 2) data augmentation to enhance the quantity and quality of the training data for use in deep learning training [42]. In the task of image classification, popular data augmentation techniques include flipping, cropping, rotating, distortion, color distortions, blurring, and many more.

Augmented-generated images retain their original label and are used as additional training data. Data augmentation targets issues that come along with a training dataset that is too small, which leads to overfitting [43].

Improving a deep CNN from scratch is an extremely hectic and impractical task because of its rigorous requirements of large amounts of labeled training data and high expertise to ensure proper convergence. A good alternative is to fine-tune a CNN that has already been pre-trained using a large set of labeled natural images [44]. Consider VGG-16 with L layers where the last 3 layers are fully connected layers. Also, let αl denote the learning rate of the l^th layer in the network. We can fine-tune only the last (new) layer of the network by setting αl = 0 for l 6= L. This level of fine-tuning corresponds to training a linear classifier with the features generated in layer L−1. Likewise, the last 2 layers of the network can be fine-tuned by setting αl = 0 for l 6= L, L − 1. This level of fine-tuning corresponds to training an artificial neural network with 1 hidden layer, which can be viewed as training a nonlinear classifier using the features generated in layer L−2. Similarly, fine-tuning layers L, L−1, and L − 2 is essentially equivalent to training an artificial neural network with 2 hidden layers. Including the previous convolution layers in the update process further adapts the pre-trained CNN to the application at hand but may require more labeled training data to avoid overfitting [45].

3. Results

The brain MRI dataset had four classes in total, namely: Glioma (G), Meningioma (M), Pituitary (P) and No Tumor (NT). We trained and tested the model on a total of 1,311 images. Admittedly due to the smaller size of the dataset and higher number of classes, we got mixed results. The final epoch results suggest slight overfitting which can be resolved by tweaking the dropout layer values, adding an L2 regularization, and testing with larger datasets as shown in Figure 1 and Table 4.

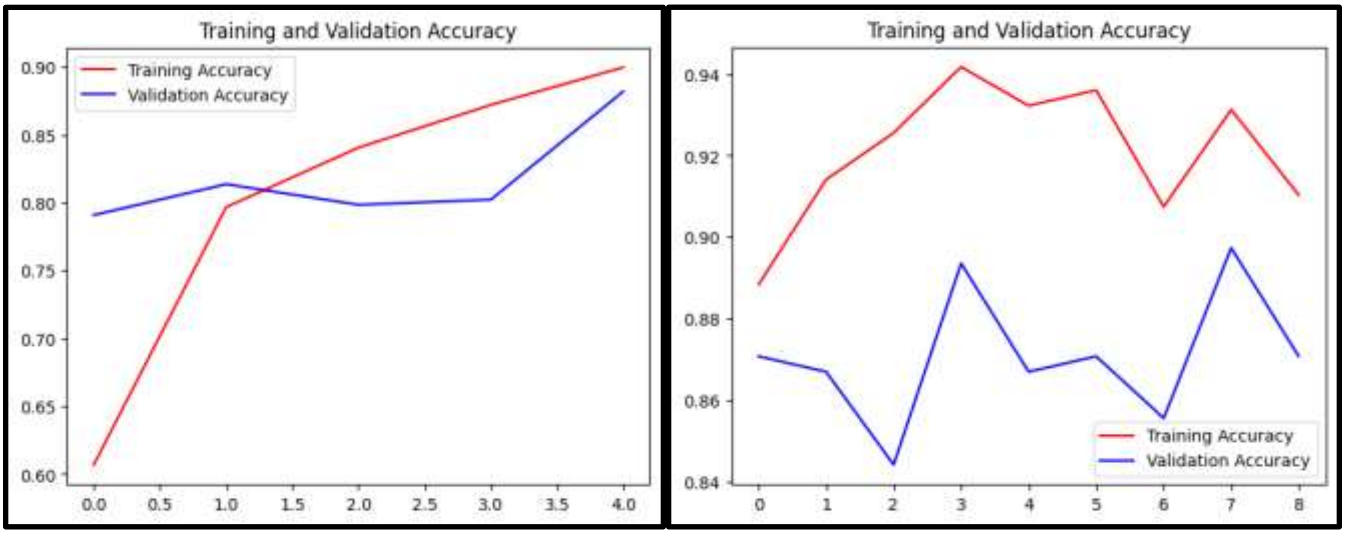

Figure 1. Training and Validation Accuracy Graphs for Brain MRI Image Classification.

Epoch round →	***Initial***	***Final***
Validation accuracy	0.8821	0.8707
Training accuracy	0.8998	0.9103
Validation loss	2.2090	2.9372
Training loss	1.4086	1.6084

Table 4. Training accuracy of the model.

The chest X-ray dataset had a total of 5,863 images and two labels, listed as: Pneumonia (P) and Normal (N). After fitting the datasets into the model we ran three epochs, that is, we went through the algorithm to train and test the model thrice and the results for each epoch are described in Table 5. In the results described in Figure 2, loss refers to a measure of the difference between actual labels and predicted labels. Due to a large dataset on few labels, the model performed very well on the training accuracy and even better on the validation accuracy with a marginal difference.

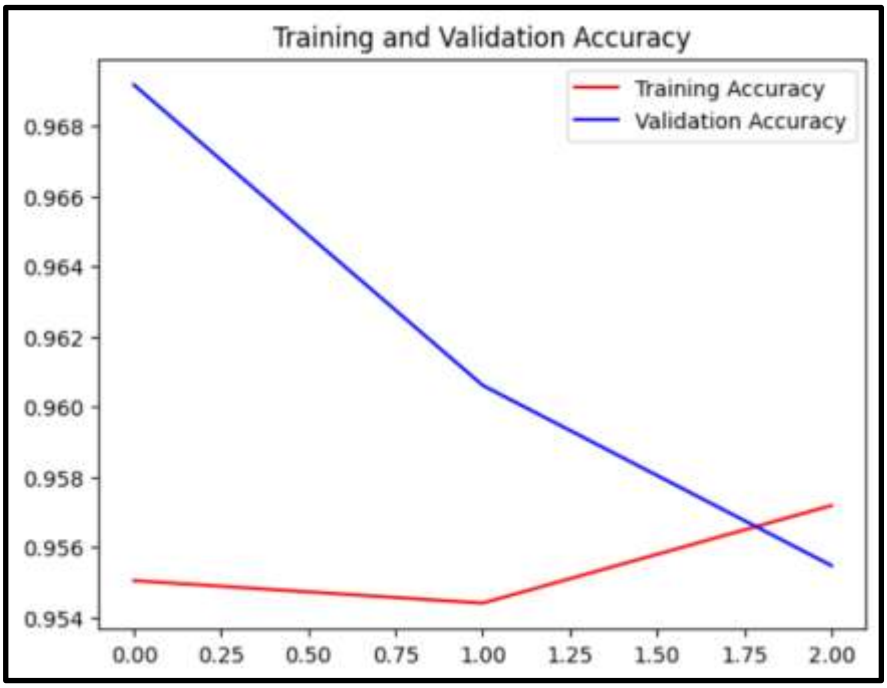

Figure 2**.** Performance chest X-ray image classification as an Accuracy-Epoch graph.

Epoch round →	*1st*	*2nd*	*3rd*
Validation accuracy	96.92%	96.06%	95.55%
Training accuracy	95.51%	95.44%	95.72%
Validation loss	0.1377	0.2517	0.1527
Training loss	0.2772	0.1937	0.1792

Table 5**.** Training accuracy of the model.

4. Discussion

Conventionally, data mining and machine learning algorithms are engineered to address problems in isolation. These algorithms are employed to train the model in separation on a specific feature space and the same distribution. Depending on the business case, a model is trained by applying a machine learning algorithm for a specific task. A widespread assumption in the field of machine learning is that training data and test data must have identical feature spaces with the underlying distribution. On the contrary, In the real world, this assumption may not hold and thus models need to be rebuilt from scratch if features and distribution change. It is an arduous process to collect related training data and rebuild the models. In such cases, Transferring Knowledge or transfer learning from disparate domains would be desirable. Transfer learning is a method of reusing a pre-trained model's knowledge for another task. Transfer learning can be used for classification, regression, and clustering problems. We have used one of the pre-trained models – VGG - 16 with Deep Convolutional Neural Network to classify images [46].

VGG-16 is an object detection and classification algorithm that can classify more than 1000 images of Peculiar categories with more than 90% accuracy. It is a type of CNN that is one of the best computer vision models to date due to the advantages and satisfaction it gives to its users as it is renowned for its simplicity and effectiveness, as well as its ability to achieve strong performance on various computer vision tasks, including image classification and object recognition. The model's architecture features a stack of convolutional layers followed by max-pooling layers, with progressively increasing depth. This design enables the model to learn intricate hierarchical representations of visual features, leading to robust and accurate predictions. It also comes with several challenges that need to be tackled. For instance, 1) It is very slow to train (the original VGG model was trained on the Nvidia Titan GPU for 2-3 weeks). 2) The size of VGG-16 trained imageNet weights is 528 MB. So, it takes quite a lot of disk space and bandwidth which makes it inefficient. 3) 138 million parameters lead to exploding gradients problem [47]. Gratefully, there are a variety of ways to enhance the performance of the VGG-16 model during training and validation. Apply data augmentation techniques to increase dataset diversity and reduce overfitting. It's also crucial to ensure the dataset's cleanliness, correct labeling, and appropriate division into training, validation, and test subsets. Experiment with different learning rates and optimizers and consider the use of learning rate schedulers if necessary. Employ regularization methods like dropout and L2 regularization to tackle overfitting issues. Check the training process carefully, implement early stopping, and adjust the batch size as needed. Lastly, make sure the hardware resources are utilized effectively, and explore ensemble methods to potentially enhance model performance [48]. These strategies will surely help to overcome the low accuracy challenge that comes with the VGG-16 model. Further advancements include resnets that can be introduced to prevent exploding gradients problem that occurs in this model. The integration of AI with augmented reality and advancements in real-time image classification are opening new avenues for user engagement and accessibility. AutoML platforms are democratizing image classification, making it accessible to non-experts, and fostering innovation across various sectors. Looking forward, the field of image classification is poised for further breakthroughs, with technologies like generative adversarial networks (GANs) and advances in unsupervised learning opening new possibilities for even more sophisticated image analysis. However, as these technologies evolve, so do the ethical challenges they present. The journey ahead involves not only technological innovation but also the cultivation of a robust ethical framework that governs the use of AI using a system based approach [49]. The journey of image classification is an ongoing adventure marked by giant technological strides that have redefined our interaction with the digital world. As we look towards the future, the importance of ethical AI development cannot be overstressed. Balancing innovation with responsibility will be key to unlocking the full potential of image classification, ensuring it contributes positively to society and industry.

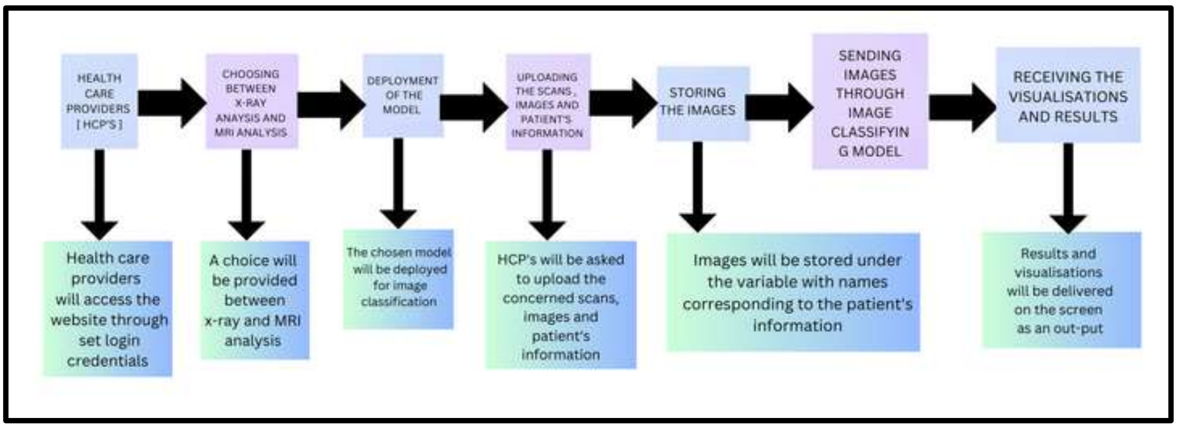

Figure 3. Mapping out the various steps involved in the model used in professional settings by HCPs

5. Conclusion

Here, we employ a transfer-learning system architecture to classify medical images of Brain MRI scans and Chest X-rays to detect various diseases such as Glioma, Meningioma and Pituitary Tumor in MRIs and Pneumonia in X-rays along with identification of normal scans as well. The VGG16 is a model with 16 convolutional layers with an accuracy of almost 92.7%, as it has been trained on the ImageNet dataset. The dataset used in this project has been relatively smaller in size with restrictions to only certain kinds of diseases. Using a larger amount of images with even more classes corresponding to numerous diseases would help by increasing variations in the scans and assist the model in recognizing more and more diseases. Though improvements to machine learning models are continuously made, using pre-trained models along with added specifications to tailor them to the highest achievable efficiency concerning the thrust of a project is a novel method to improve accuracy and increase convenience when it comes to analyzing and classifying images. With the help of certain technical changes, the model can be further trained to enhance performance and can be deployed for HCPs to assist them in identifying abnormal scans and diagnosing patients. ML models in the healthcare industry, though still in the early stages, have proved to be assets. From disease identification to diagnosis, it could resolve the problem of inaccurate diagnoses due to human errors and improve the quality of healthcare by making it possible to have timely and correct treatments available to patients. These models will largely reduce the burden on medical practitioners and assist them with quick identification of any abnormalities present. ML models coupled with human intelligence could pave the way for drastic improvements in healthcare systems all over the world. In conclusion, the application of transfer learning in medical imaging, as demonstrated in various studies, underscores its transformative potential in enhancing diagnostic accuracy and efficiency, similar to advancements seen in pathogen detection and glucose sensing technologies, thereby paving the way for significant improvements in healthcare delivery [50] [51] [52].

Author Contributions

All authors contributed to providing critical feedback and helped shape the research. DR and AD developed the theory, constructed the machine learning models, and conducted training and testing. TB, RW, and AK collected theoretical data. DR and AD interpreted and analyzed results from training and testing. TB delved into existing research, and AK delved into fine-tuning, evaluation, challenges, and future improvement aspects. TB and DR created the flow charts RW, DR and AD collected datasets for image analysis. All authors made equal contributions to the writing and formatting of the paper. KV, SV, KD conceptualization, writing- reviewing and editing.

Conflicts of Interest

The authors declare no competing financial interests or conflicts of interest.

References

[1] M. Tsuneki, "Deep learning models in medical image analysis," Journal of Oral Biosciences, Mar. 2022, doi: https://doi.org/10.1016/j.job.2022.03.003.

[2] H. I. Greenwood, K. Dodelzon, and J. T. Katzen, "Impact of Advancing Technology on Diagnosis and Treatment of Breast Cancer," The Surgical Clinics of North America, vol. 98, no. 4, pp. 703–724, Aug. 2018, doi: https://doi.org/10.1016/j.suc.2018.03.006.

[3] L. Balyen and T. Peto, "Promising Artificial Intelligence-Machine Learning-Deep Learning Algorithms in Ophthalmology," Asia-Pacific Journal of Ophthalmology, vol. 8, no. 3, 2019, doi: https://doi.org/10.22608/apo.2018479.

[4] A. Shrestha and A. Mahmood, "Review of Deep Learning Algorithms and Architectures," IEEE Access, vol. 7, pp. 53040–53065, 2019, doi: https://doi.org/10.1109/access.2019.2912200.

[5] Chilmakuri, L.; Mishra, A. K.; Shokeen, D. .; Gupta, P. .; Wadhwa, H. H.; Dhingra, K. .; Verma, S. A Wearable EMG Sensor for Continuous Wrist Neuromuscular Activity for Monitoring. J. Knowl. Learn. Sci. Technol. 2024, 3 (4), 148-159. https://doi.org/10.60087/jklst.v3.n4.p148.

[6] Y. Wu and J. Feng, "Development and Application of Artificial Neural Network," Wireless Personal Communications, vol. 102, no. 2, pp. 1645–1656, Dec. 2017, doi: https://doi.org/10.1007/s11277-017-5224-x.

[7] A. Prieto et al., "Neural networks: An overview of early research, current frameworks and new challenges," Neurocomputing, vol. 214, pp. 242–268, Nov. 2016, doi: https://doi.org/10.1016/j.neucom.2016.06.014.

[8] Ruthwik Guntupalli, Saloni Verma and Karan Dhingra 2024. Impact of Healthcare Digitization: Systems Approach for Integrating Biosensor Devices and Electronic Health with Artificial Intelligence. American Scientific Research Journal for Engineering, Technology, and Sciences. 98, 1 (Aug. 2024), 246–257, https://asrjetsjournal.org/index.php/American_Scientific_Journal/article/view/10786/2789.

[9] S. S. Han, M. S. Kim, W. Lim, G. H. Park, I. Park, and S. E. Chang, "Classification of the Clinical Images for Benign and Malignant Cutaneous Tumors Using a Deep Learning Algorithm," Journal of Investigative Dermatology, vol. 138, no. 7, pp. 1529–1538, Jul. 2018, doi: https://doi.org/10.1016/j.jid.2018.01.028.

[10] Periketi, P.; Kaur, K. .; Naseer Vaid, F. .; Sree M, Y. .; Madhu, M. .; Verma, S. .; Dhingra, K. . Blood Brain Barrier-on-a-Chip Permeation to Model Neurological Diseases Using Microfluidic Biosensors. J. Knowl. Learn. Sci. Technol. 2024, 3 (4), 78-93. https://doi.org/10.60087/jklst.v3.n4.p78.

[11] H.-P. Chan, R. K. Samala, L. M. Hadjiiski, and C. Zhou, "Deep Learning in Medical Image Analysis," Advances in Experimental Medicine and Biology, vol. 1213, pp. 3–21, 2020, doi: https://doi.org/10.1007/978-3-030-33128-3_1.

[12] H. J. Pandya et al., "A microfluidic platform for drug screening in a 3D cancer microenvironment," Biosensors and Bioelectronics, vol. 94, pp. 632–642, Aug. 2017, doi: https://doi.org/10.1016/j.bios.2017.03.054.

[13] G. Varoquaux and V. Cheplygina, "Machine learning for medical imaging: methodological failures and recommendations for the future," npj Digital Medicine, vol. 5, no. 1, pp. 1–8, Apr. 2022, doi: https://doi.org/10.1038/s41746-022-00592-y.

[14] Kulkarni S, Dhingra K, Verma S., "Applications of CMUT Technology in Medical Diagnostics: From Photoacoustic to

Ultrasonic Imaging", International Journal of Science and Research (IJSR), Volume 13 Issue 6, June 2024, pp. 1264-1269, https://www.ijsr.net/ar-chive/v13i6/SR24619062609.pdf.

[15] Simplilearn, "What Is Image Processing: Overview, Applications, Benefits, and Who Should Learn It," Simplilearn.com, Apr. 29, 2021. https://www.simplilearn.com/image-processing-article.

[16] Gal Levy-Fix, G. J. Kuperman, and Noémie Elhadad, "Machine Learning and Visualization in Clinical Decision Support: Current State and Future Directions," arXiv (Cornell University), Jan. 2019, doi: https://doi.org/10.48550/arxiv.1906.02664.

[17] M. Puttagunta and S. Ravi, "Medical image analysis based on deep learning approach," Multimedia Tools and Applications, Apr. 2021, doi: https://doi.org/10.1007/s11042-021-10707-4.

[18] D. Shen, G. Wu, and H.-I. Suk, "Deep Learning in Medical Image Analysis," Annual Review of Biomedical Engineering, vol. 19, no. 1, pp. 221–248, Jun. 2017, doi: https://doi.org/10.1146/annurev-bioeng-071516-044442.

[19] M. Rana and M. Bhushan, "Machine learning and deep learning approach for medical image analysis: diagnosis to detection," Multimedia Tools and Applications, vol. 82, Dec. 2022, doi: https://doi.org/10.1007/s11042-022-14305-w.

[20] Shin, Hoo-chang & Roth, Holger & Gao, Mingchen & Lu, Le & Xu, Ziyue & Nogues, Isabella & Yao, Jianhua & Mollura, Daniel & Summers, Ronald. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. IEEE Transactions on Medical Imaging. 35. 10.1109/TMI.2016.2528162.

[21] M. Li, Y. Jiang, Y. Zhang, and H. Zhu, "Medical image analysis using deep learning algorithms," Frontiers in Public Health, vol. 11, Nov. 2023, doi: https://doi.org/10.3389/fpubh.2023.1273253.

[22] D. Müller, I. Soto-Rey, and F. Kramer, "Towards a Guideline for Evaluation Metrics in Medical Image Segmentation," arXiv.org, Feb. 10, 2022. https://arxiv.org/abs/2202.05273

[23] "Dice score - OECD.AI," oecd.ai. https://oecd.ai/en/catalogue/metrics/dice-score.

[24] F. A. Zadeh, M. V. Ardalani, A. R. Salehi, R. Jalali Farahani, M. Hashemi, and A. H. Mohammed, "An Analysis of New Feature Extraction Methods Based on Machine Learning Methods for Classification Radiological Images," Computational Intelligence and Neuroscience, vol. 2022, p. e3035426, May 2022, doi: https://doi.org/10.1155/2022/3035426.

[25] Hosseinzadeh Taher MR, Haghighi F, Feng R, Gotway MB, Liang J. "A Systematic Benchmarking Analysis of Transfer Learning for Medical Image Analysis." (Domain Adapt Represent Transf Afford Healthc AI Resour Divers Glob Health (2021)), 2021 Sep-Oct, doi: 10.1007/978-3-030-87722-4_1.

[26] L. Alzubaidi et al., "Towards a Better Understanding of Transfer Learning for Medical Imaging: A Case Study," Applied Sciences, vol. 10, no. 13, p. 4523, Jun. 2020, doi: https://doi.org/10.3390/app10134523.

[27] X. Yu, J. Wang, Q.-Q. Hong, R. Teku, S.-H. Wang, and Y.-D. Zhang, "Transfer learning for medical images analyses: A survey," Neurocomputing, vol. 489, pp. 230–254, Jun. 2022, doi: https://doi.org/10.1016/j.neucom.2021.08.159.

[28] Tajbakhsh, Nima & Shin, Jae & Gurudu, Suryakanth & Hurst, R Todd & Kendall, Christopher & Gotway, Michael & Liang, Jianming. (2016). Convolutional Neural Networks for Medical Image Analysis: Fine Tuning or Full Training?. IEEE Transactions on Medical Imaging. 35. 1-1. 10.1109/TMI.2016.2535302.

[29] Tammina, Srikanth. (2019). Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images. International Journal of Scientific and Research Publications (IJSRP). 9. p9420. 10.29322/IJSRP.9.10.2019.p9420.

[30] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv.org, Apr. 10, 2015. https://arxiv.org/abs/1409.1556.

[31] W. Wang et al., "Medical Image Classification Using Deep Learning," Intelligent Systems Reference Library, pp. 33–51, Nov. 2019, doi: https://doi.org/10.1007/978-3-030-32606-7_3.

[32] J. Deva, "Impact of the Preprocessing Steps in Deep Learning-Based Image Classifications," National Academy Science letters, Jan. 2024, doi: https://doi.org/10.1007/s40009-023-01372-2.

[33] Chandna, R. .; Bansal, A.; Kumar, A.; Hardia, S.; Daramola, O.; Sahu, A.; Verma, K.; Dhingra, K.; Verma, S. Skin Disease Classification Using Two Path Deep Transfer Learning Models. J. Knowl. Learn. Sci. Technol. 2024, 3 (4), 169-187. https://doi.org/10.60087/jklst.v3.n4.p169.

[34] H. J. Pandya et al., "Label-free electrical sensing of bacteria in eye wash samples: A step towards point-of-care detection of pathogens in patients with infectious keratitis," Biosensors and Bioelectronics, vol. 91, pp. 32–39, May 2017, doi: https://doi.org/10.1016/j.bios.2016.12.035.

[35] Rob, "We Need to Talk About train-dev-test Splits," Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Jan. 2021, doi: https://doi.org/10.18653/v1/2021.emnlp-main.368.

[36] Gupte, P.; Dhingra, K.; Saloni , V. Precision Gene Editing Strategies With CRISPR-Cas9 for Advancing Cancer Immunotherapy and Alzheimer's Disease. J. Knowl. Learn. Sci. Technol. 2024, 3 (4), 11-21. https://doi.org/10.60087/jklst.v3.n4.p11.

[37] K. Team, "Keras documentation: Keypoint Detection with Transfer Learning," keras.io. https://keras.io/examples/vision/keypoint_detection/

[38] T. Bylander and L. Tate, "Using Validation Sets to Avoid Overfitting in AdaBoost *." Accessed: Jun. 28, 2024. [Online]. Available: https://cdn.aaai.org/FLAIRS/2006/Flairs06-108.pdf.

[39] J. Brownlee, "What is the Difference Between a Batch and an Epoch in a Neural Network?" Available: https://deeplearning.lipingyang.org/wp-content/uploads/2018/07/What-is-the-Difference-Between-a-Batch-and-an-Epoch-in-a-Neural-Network_.pdf.

[40] Matplotlib, "Matplotlib: Python plotting — Matplotlib 3.1.1 documentation," Matplotlib.org, 2012. https://matplotlib.org/.

[41] M. Iman, H. R. Arabnia, and K. Rasheed, "A Review of Deep Transfer Learning and Recent Advancements," Technologies, vol. 11, no. 2, p. 40, Mar. 2023, doi: https://doi.org/10.3390/technologies11020040.

[42] A. Fawzi, H. Samulowitz, D. Turaga, and P. Frossard, "Adaptive data augmentation for image classification," IEEE Xplore, Sep. 01, 2016. https://ieeexplore.ieee.org/abstract/document/7533048.

[43] L. Perez and J. Wang, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning," arXiv:1712.04621 [cs], Dec. 2017, Available: [1712.04621] The Effectiveness of Data Augmentation in Image Classification using Deep Learning (arxiv.org)

[44] S. Kumaresan, S. Jai, S. S. Kumar, and M. Dev Anand, "Deep learning-based weld defect classification using VGG16 transfer learning adaptive fine-tuning," International Journal on Interactive Design and Manufacturing (IJIDeM), May 2023, doi: https://doi.org/10.1007/s12008-023-01327-3.

[45] N. Tajbakhsh et al., "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1299–1312, May 2016

[46] Pundlik, A.; Verma, S.; Dhingra, K. Neural Pathways Involved in Emotional Regulation and Emotional Intelligence. J. Knowl. Learn. Sci. Technol. 2024, 3 (3), 165-192. https://doi.org/10.60087/jklst.vol3.n3.p.165-192.

[47] M. Chhabra and R. Kumar, "An Advanced VGG16 Architecture-Based Deep Learning Model to Detect Pneumonia from Medical Images," Lecture Notes in Electrical Engineering, pp. 457–471, 2022, doi: https://doi.org/10.1007/978-981-16-8774-7_37.

[48] Kazim, I.; Gande, T.; Reyher, E. .; Gyatsho Bhutia, K. .; Dhingra, K.; Verma, S. Advancements in Sequencing technologies:: From Genomic Revolution to Single-Cell Insights in Precision Medicine. J. Knowl. Learn. Sci. Technol. 2024, 3 (4), 108-124. https://doi.org/10.60087/jklst.v3.n4.p108.

[49] Mehta, A.; Alaiashy, O.; Kumar, P.; Tamilinian, V.; Besong, S.; Balpande, S.; Verma, S.; Dhingra, K. Advancing Model-Based Systems Engineering in Biomedical and Aerospace research:: A Comprehensive Review and Future Directions. J. Knowl. Learn. Sci. Technol. 2024, 3 (4), 133-147. https://doi.org/10.60087/jklst.v3.n4.p133.

[50] M. Safavieh et al., "Paper microchip with a graphene-modified silver nano-composite electrode for electrical sensing of microbial pathogens," Nanoscale, vol. 9, no. 5, pp. 1852–1861, 2017, doi: https://doi.org/10.1039/c6nr06417e.

[51] Peyman GhavamiNejad, Amin GhavamiNejad, H. Zheng, K. Dhingra, M. Samarikhalaj, and Mahla Poudineh, "A Conductive Hydrogel Microneedle-Based Assay Integrating PEDOT:PSS and Ag-Pt Nanoparticles for Real-Time, Enzyme-Less, and Electrochemical Sensing of Glucose," Advanced Healthcare Materials, vol. 12, no. 1, Oct. 2022, doi: https://doi.org/10.1002/adhm.202202362.

[52] S. Odinotski et al., "A Conductive Hydrogel-Based Microneedle Platform for Real-Time pH Measurement in Live Animals," Small, vol. 18, no. 45, Sep. 2022, doi: https://doi.org/10.1002/smll.202200201.