COMP 6321 - FINAL REPORT

Abstract

This project comprises two main tasks. Task 1 involves comparing the performance of two models on the NCT-CRC-HE-100K dataset: VGG11 trained from scratch and VGG11 pretrained on ImageNet with fine-tuning. The latter achieves the best results with 96% accuracy, underscoring the efficacy of transfer learning for improving performance. Task 1 also incorporates hyperparameter grid search and investigates the importance of data normalization. In Task 2, the exploration of transfer learning without fine-tuning is extended to two distinct datasets. Features are extracted using VGG11 models from Task 1 and bare VGG11 pretrained with ImageNet. Dimensionality reduction via t-SNE is followed by classification using KNN and SVM, with KNN eventually outperforming SVM. Results are validated through accuracy, precision, recall metrics, and confusion matrices. We reveal the sensitivity of transfer learning when domains differ, and how the final layers of a CNN model play a more significant role in high-level encoding.

1. Introduction

1.1. Problem statement

With the ability to extract complex visual features and patterns, Convolution Neural Networks (CNNs) emerged as the backbone of modern computer vision. Thanks to the democratization of graphic processing units, CNNs have got a foundational role in several domains, opening the door to advanced applications in fields like virtual assistants, robots, and healthcare [1].

Image classification, a fundamental task in computer vision, involves categorizing images based on their content. Despite its importance, the process comes with certain challenges. Selecting an architecture that aligns with the complexity of the problem can be difficult, given the diverse performance of architectures in various applications. Moreover, hyperparameter tuning is both manual and time-consuming, with potential errors leading to overfitting or underfitting. Deep learning, in general, relies on large datasets, but labelled datasets are often scarce in certain domains. To address this issue, transfer learning (TL) can be adopted to leverage knowledge from extensive datasets. Nevertheless, transferring learning across domains is hindered by feature disparities.

This project is guided by two primary tasks. In Task 1, we train from scratch a VGG11 model on a colorectal cancer image dataset, striving for high accuracy, precision and recall. Then, we contrast the model's performance with that of a VGG11 pretrained on ImageNet. The pretrained network is anticipated to bring better performance as it has already acquired valuable features from a large dataset. We use a grid search for hyperparameter optimization and we test the effect of data normalization. t-SNE visualization is also applied to the output features. In Task 2, we use those two versions of VGG11 as encoders to extract features from two additional distinct datasets. On top of these features, we apply t-SNE. Eventually, we classify the t-SNE extracted features using two supervised methods: K-Nearest Neighbours (KNN) and Support Vector Machines (SVM). The two tasks aim to provide insights into the effectiveness of TL when transitioning from one domain to another.

Challenges encountered include the prolonged training process requiring GPU usage and the time-consuming grid search, hindering the project workflow. As expected, fine-tuning the pretrained VGG11 on the colorectal cancer dataset produced superior results compared to training it from scratch, achieving an accuracy of 96%. Furthermore, when employing both versions of VGG11 for feature extraction on additional datasets, the application of KNN demonstrated better performance in classifying the t-SNE extracted features on the third dataset, achieving an accuracy of 99.25%.

1.2. Literature review

CNN-based image classification holds a significant importance for several applications, owing to the capability to automatically extract complex patterns. In this regard, LeNet-5 [2] is recognized as the initial milestone, which used for classifying hand-written digits. Over time, deeper networks, such as GoogLeNet [3] and VGGNet [4] have emerged, demonstrating a considerable improved performance. However, training deep learning models needs extensive data to achieve optimal outcomes. In particular, acquiring medical image datasets proves challenging, as the collection and annotation processes are laborious and require domain expertise [5]. To overcome this challenge, TL is very often utilized, which is the case in Kermany et al. study that achieved a 92% accuracy on a limited dataset of pneumonia X-ray images [6].

Among several CNN architectures, the VGG family has demonstrated remarkable success in various image classification tasks. According to [7], VGG architectures dominate medical image studies, accounting for 50% of the breast cancer research. In addition, they account for 26% of the number of TL studies using ultrasound images. When it comes to skin lesions, fundoscopy, and Optical Coherence Tomography images, VGG is also the most utilized, comprising 44%, 42%, and 49% of the literature, respectively. For instance, Khan et al. used TL with four models: DenseNet121, ResNet50, VGG16, and VGG19 in order to diagnose X-ray images for COVID-19. Both VGG models demonstrated superior performance compared to the other two models [8]. Another study that explored the effectiveness of TL for classifying a breast cancer dataset shows that VGG16 with logistic regression classifier yielded the best performance among all tested models with 92.6% accuracy [9].

As part of medical imaging, the NCT-CRC-HE-100K dataset is crucial for studying colorectal cancer in medical imaging and serves as a reference for computer vision studies. Numerous papers have applied VGG architectures using this dataset. Anju et al. achieved 97.9% accuracy by fine-tuning VGG16 on two datasets, including NCT-CRC-HE-100K [10]. Moreover, Kumar et al. introduced CRCCN-Net, and compared its performance with VGG16, DenseNet121, Inception-ResNetV2, and Xception on NCT-CRC-HE-100K. Notably, VGG16 outperformed the other models, except for CRCCN-Net itself [11].

While TL proves enhanced performance, it still faces the challenge of disparity between the source dataset and the target one. In our context, CNNs undergone pretraining on ImageNet are intended for medical image classification. However, a notable concern arises due to the substantial dissimilarity between images within ImageNet and those of medical imagery. TL from non-medical domain to medical domain seems to be controversial [12–14]. For example, a study has showed that lightweight models trained from scratch perform nearly as well as ImageNet-transferred models [15].

In light of this literature review, we opt for a VGG architecture to be trained/fine tuned on the NCT-CRC-HE-100K dataset. Specifically, we use VGG11 since, to the best of our knowledge, it has not been explored on this dataset before. We investigate the effectiveness of TL with fine-tuning (Task 1) and without fine-tuning (Task 2).

2. Methodology

2.1. Datasets

Three datasets will serve to validate and assess the performance of our models. **NCT-CRC-HE-100K** is a 100K-image dataset featuring 224×224 pixel patches from H&E stained histological images of human colorectal cancer and normal tissue, including 9 classes. The non-overlapping images with a 0.5-micron per pixel resolution were manually extracted from 86 H&E stained human cancer tissue slides. The dataset can be accessed via [16]. In this project, we used a reduced version containing 6K images with only 3 classes MUS for smooth muscle, STR for cancer-associated stroma, and NORM for normal colon mucosa.

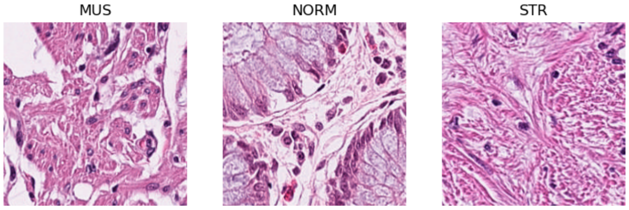

Figure 1. NCT-CRC-HE-100K — Representative samples from each class

Digital Pathology and Artifacts is another dataset introduced in [17], containing 120K of 300×300 image patches, with 50K tumor tissue patches, 50K non-neoplastic glandular prostate tissue patches, and 20K non-glandular tissue patches. The dataset was collected using six datasets from four different institutions digitized by different scanner systems. The dataset can be accessed via [18]. Here, we will be using only 6K images for the same 3 classes: Prostate Cancer Tumor Tissue, Benign Glandular Prostate Tissue, and Benign Non-Glandular Prostate Tissue.

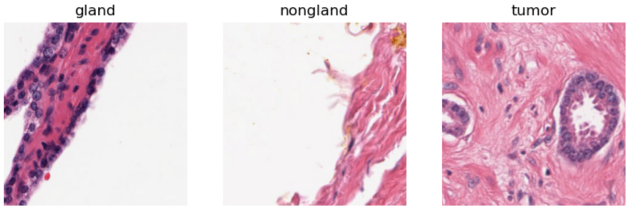

Figure 2. Digital Pathology and Artifacts — Representative samples from each class

The third dataset, **Animal Faces-HQ**, was introduced in [19] with 16K JPG images categorized into 3 classes. These images have a resolution of 512×512 pixels, with each class containing about 5K images. Data were collected from *Flickr* and *Pixabay* websites. It is accessible via [20]. In our project, we will use only 6K images for the same 3 classes: Cats, Dogs, and wildlife animals.

Figure 3. Animal Faces-HQ — Representative samples from each class

Table 1 gives a statistical overview about the datasets with a comparison between the original versions (OV) and the reduced ones (RV).

Dataset		**Data points**	**Image size**	**№ of classes**
NCT-CRC-HE-100K	OV	100 K	224×224	9
	RV	6 K	224×224	3
Digital Pathology and Artifacts	OV	120 K	300×300	3
	RV	6 K	300×300	3
Animal Faces-HQ	OV	16 K	512×512	3
	RV	6 K	512×512	3

Table 1. Statistical nuances between original and reduced data

2.2. CNN-based classification

Task 1 focuses on the NCT-CRC-HE-100K dataset, during which we train a VGG11 model from scratch, and concurrently, we fine-tune a pretrained VGG11 model with weights trained on ImageNet. The goal is to compare the performance of a non-pretrained model with a pretrained one based on accuracy, precision, recall scores, and t-SNE graph. VGG11 is an 11-layer CNN with seven 3x3 kernel convolution layers, each followed by ReLU activation. It includes 5 max pooling operations, reducing feature maps by a factor of 2. The initial layer has 64 channels, doubling after each max pooling until reaching 512, while subsequent layers maintain a constant channel count. For training VGG11 from scratch, a grid search hyperparameter determines the optimal learning rate and batch size. When fine-tuning the pretrained version, the parameters yielding the best results in the initial model are employed. For the fine-tuning process, it's important to highlight that we keep all weights frozen, except for the last 5 layers that undergo training. Additionally, a Dense layer is introduced to tailor the model for handling 3 classes. During the training of both models, we employ early stopping as a technique to interrupt the training process if there is no enhancement in validation loss over 5 consecutive epochs. This helps prevent overfitting and accelerates the training procedure.

2.3. Feature Extraction

Task 2 involves the Digital Pathology and Artifact as well as Animal Faces-HQ datasets. We use the VGG11 tuned previously on NCT-CRC-HE-100K, and a VGG11 pretrained only on ImageNet as encoders without classification heads in order to extract feature representations from the aforementioned datasets. Then, the extracted features are assessed with t-SNE visualization. In Task 2, we utilize two versions of VGG11:

- VGG11-I: VGG11 that gave the best performance in Task 1, which is the one pretrained on ImageNet and fine-tuned on NCT-CRC-HE-100K.
- VGG11-II: VGG11 solely pretrained on ImageNet

Finally, those extracted features are fed to a t-SNE for dimensionality reduction. We employ KNN and SVM algorithms for the classification of the features extracted by t-SNE.

3. Results

To start, the dataset is resized to the model input size of 224x224 and then randomly partitioned into training (80%), validation (10%), and testing (10%) subsets. During the training of VGG11 from scratch, we perform a hyperparameter grid search. We launch the model with varying learning rates (lr) of 0.1, 0.01 and 0.001 as well as different batch sizes (bs) of 128, 64 and 32. It is important to mention that, at this stage, the input data lack normalization. The findings of this experiment, as illustrated in Table 2, indicate that the best outcome yielded an 84% test accuracy, 0.29 test loss, 0.87 test precision, and 0.84 test recall, achieved with a learning rate of 0.001 and a batch size of 32. Additionally, it is noted that increasing the batch size and learning rate lead to a decline in results.

lr	**0.001**			**0.01**			**0.1**		
bs	**32**	**64**	**128**	**32**	**64**	**128**	**32**	**64**	**128**
ta	87%	84%	61%	48%	33%	29%	36%	33%	29%
tl	0.29	0.36	1.07	1.08	1.10	1.11	1.09	1.11	1.10
tp	0.87	0.84	0.6	0.31	0.11	0.089	0.13	0.11	0.089
tr	0.87	0.84	0.61	0..48	0.33	0.29	0.36	0.33	0.29

Table 2. Grid search test results of VGG11 model using 3 different learning rates (lr) and batch sizes (bs)

Using a learning rate of 0.01 or 0.1 resulted in the gradient being unable to converge to a satisfactory minimum or becoming trapped in a local minimum. This could be attributed to a learning step that is likely too rapid, causing it to skip over the desirable minima. Now that we have identified the optimal hyperparameters, we proceed to reinitiate the training of the model. However, this time, we incorporate input data normalization using mean values of 0.485, 0.456, and 0.406 for red, blue, and green channels, respectively. The standard deviation is set to 0.229, 0.224, and 0.225. These values are obtained through training on the 1.2M ImageNet images and have proven to work well as a general- purpose normalization [21]. Table 3 illustrates the contrast between the model without and with input normalization and the importance of the latter.

	Without Normalization	**With Normalization**
Accuracy	87%	93%
Loss	0.29	0.17
Precision	0.87	0.933
Recall	0.87	0.931

Table 3. Comparison of VGG11 with and without input data normalization

Recognizing normalization's significance and determining optimal hyperparameters, we proceed with transfer learning using VGG11, with weights sourced from training on ImageNet. We freeze all those weights, except those of the last 5 layers that we are going to train and we add a Dense layer to adapt the model to 3 classes. The results give superiority to this model with 96.33% of test accuracy and 0.105 of test loss.

	Training from scratch	**Transfer learning**
Accuracy	93%	96%
Loss	0.17	0.10
Precision	0.933	0.963
Recall	0.931	0.963

Table 4. From-scratch versus transfer learning performance

t-SNE is afterwards employed on both models. Figure 4 is relative to the model trained from scratch, while Figure 5 is related to the pretrained model.

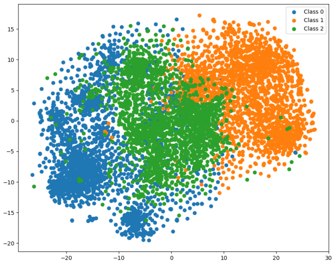

Figure 4. t-SNE visualization on the non-pretrained VGG11

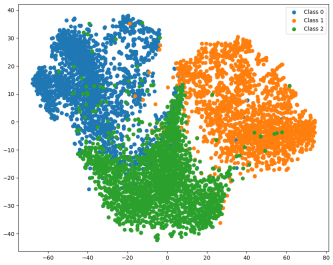

Figure 5. t-SNE visualization on the pretrained VGG11

The from-scratch model still shows class grouping but without clear boundaries, posing a risk of misclassifying tumours as non-tumorous or different tissues. Additionally, there's notable intra-class variation and evident false classifications. In contrast, the pretrained model reveals a refined structure and the overall grouping is more cohesive, aligning with the performance metrics. Those visualizations highlight the importance of TL in reducing misclassifications.

For Task 2, we begin with using VGG11-I (the one giving the best performance in Task 1) on Digital Pathology and Artifacts dataset for feature extraction. As shown by the t-SNE visualization in Figure 6, the t-SNE visualization indicates that features associated with the tumour class are distinguishable from other classes. However, this distinction is less pronounced for the two other classes, namely gland and non-gland. This is indeed corroborated by the results obtained after employing KNN and SVM algorithms. Notably, KNN and SVM demonstrate comparable performance, with KNN slightly surpassing SVM. Specifically, KNN achieves 84.83%, 84.98%, and 84.83% for accuracy, precision, and recall, respectively, while SVM yields 83.25%, 83.5%, and 83.25% for accuracy, precision, and recall, respectively.

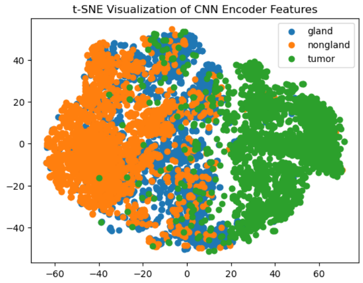

Figure 6. Digital Pathology and Artifacts— t-SNE visualization on VGG11-I

The confusion matrices depicted in Figure 7 and Figure 8 for respectively KNN and SVM, reveal that KNN outperforms SVM not only in overall performance but also on a class-specific level. They also confirm the t-SNE visualization as the tumour class is the most accurately predicted by both algorithms.

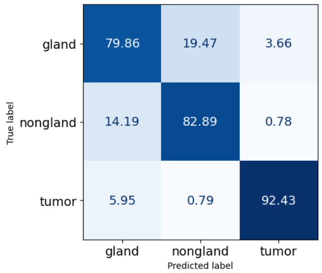

Figure 7. Digital Pathology and Artifacts— KNN Confusion matrix when features extracted with VGG11-I

Now, we replicate the process with VGG11-II, while continuing to focus on the Digital Pathology and Artifacts dataset. The performance of VGG11-II is not as satisfactory as that achieved with VGG11-I. This can be anticipated from the analysis of the t-SNE graph in Figure 9, which reveals no clear boundaries between all classes. Furthermore, KNN once again demonstrated superior performance compared to SVM, achieving an accuracy of 74.75%, precision of 75%, and recall of 74.75%, while SVM yielded an accuracy of 69.58%, precision of 72.39%, and recall of 69.58%.

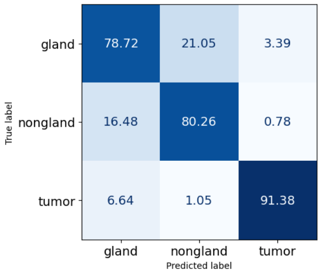

Figure 8. Digital Pathology and Artifacts— SVM Confusion matrix when features extracted with VGG11-I

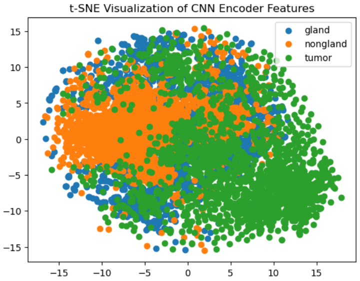

Figure 9. Digital Pathology and Artifacts— t-SNE visualization on VGG11-II

The confusion matrices in Figure 10 and Figure 11, for respectively KNN and SVM, are in consistency with the findings and show that the tumour class stands out as the most accurately predicted.

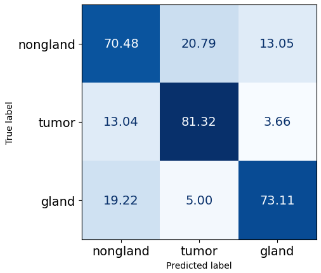

Figure 10. Digital Pathology and Artifacts— KNN Confusion matrix when features extracted with VGG11-II

Actually, it is comprehensible why the VGG11-I performs better than VGG11-II, since the former has gained expertise not only from ImageNet but also from training on NCT-CRC-HE-100K of colorectal cancer data, which exhibits some degree of similarity to Digital Pathology and Artifacts dataset: both involve histological images of tissues from the human body.

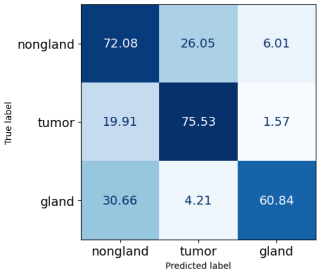

Figure 11. Digital Pathology and Artifacts— SVM Confusion matrix when features extracted with VGG11-II

The application of VGG11-I for feature extraction from the Animal Faces-HQ dataset, shows the model failed to effectively capture discernible patterns within the dataset. This inadequacy is corroborated by the t-SNE visualization in Figure 12, which reveals a distribution of points uniformly scattered across the space without clear demarcations between classes.

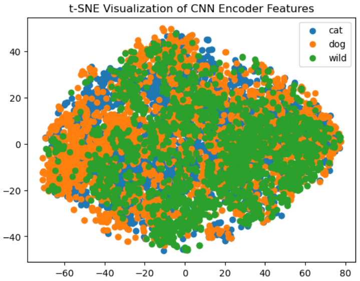

Figure 12. Animal Faces-HQ— t-SNE visualization on VGG11-I extracted features

In terms of classification performance, KNN outperformed SVM, achieving an accuracy of 59% compared to SVM's 48.75%. KNN demonstrates higher accuracy even within individual classes. Notably, SVM exhibited relatively better proficiency in predicting the "wild" than predicting the other classes but KNN demonstrated a superior ability to predict the "cat" class in comparison to classifying the other classes. This is highlighted through the confusion matrices in Figure 13 and Figure 14.

These findings underscore the subtle distinctions in performance between SVM and KNN when confronted with the features extracted from the Animal Faces-HQ dataset. Despite the expectation that the VGG11 model from Task 1— having been trained on the ImageNet dataset containing natural images featuring dogs, cats, and other wildlife classes— would possess the knowledge to

Table 5. Performance comparison of KNN and SVM on both datasets

	Dataset 2				**Dataset 3**			
	VGG11 of task1		VGG11 pretrained		VGG11 of task1		VGG11 pretrained	
	SVM	KNN	SVM	KNN	SVM	KNN	SVM	KNN
Accuracy (%)	83.25	84.83	69.58	74.75	48.75	59.08	98.75	99.25
Precision (%)	83.5	84.98	72.39	75	48.84	59.05	98.75	99.25
Recall (%)	83.25	84.83	69.58	74.75	48.75	59.08	98.75	99.25

effectively classify animals, its performance fell below expectations. Apparently, the last layers of the model are more relevant for high-level encoding.

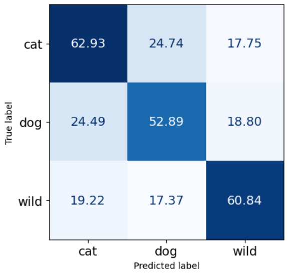

Figure 13. Animal Faces-HQ— KNN Confusion matrix when features extracted with VGG11-I

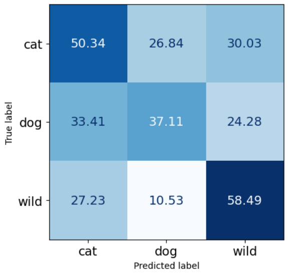

Figure 14. Animal Faces-HQ— SVM Confusion matrix when features extracted with VGG11-I

Utilizing VGG11-II on the animal faces dataset yields highly satisfactory outcomes. We remind that VGG11-II was originally pretrained exclusively on ImageNet, a dataset encompassing a diverse range of dogs, cats, and wild animals that the model learned to classify. Consequently, the obtained results align with expectations. The t-SNE visualization in Figure 15 illustrates less variability among classes, with well-defined boundaries. There are occasional instances of confusion, such as some dogs being misclassified as cats or wild animals. For example, husky dogs resemble quite much to wild wolves. Such confusions occur even for human observers. Furthermore, the t-SNE analysis indicates that the wild class itself can be subdivided into four distinct clusters, a logical outcome considering the broad nature of the wild category.

Once again, KNN demonstrated superior performance compared to SVM, achieving an accuracy, precision, and recall of 99.25%. In contrast, SVM yielded slightly lower results with an accuracy, precision, and recall of 98.75%. Examination of the confusion matrices in Figure 16 and Figure 17 reveals that KNN accurately predicted all wild data, and the dog class exhibited the highest confusion with other classes.

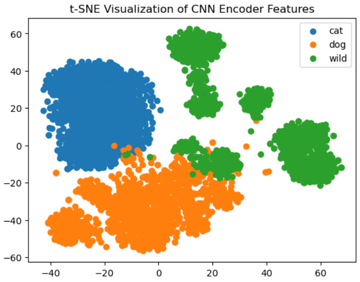

Figure 15. Animal Faces-HQ— t-SNE visualization on VGG11-II extracted features

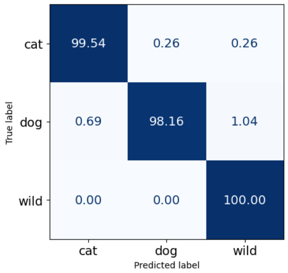

Figure 16. Animal Faces-HQ— KNN Confusion matrix when features extracted with VGG11-II

Throughout the previous experiments, KNN consistently demonstrated superior performance compared to SVM. This can be attributed primarily to the absence of a linear separation between the classes. As observed in the t-SNE visualizations, even in cases where classes are distinguishable, the separation is non-linear. Table 5 provides an overview of the outcomes obtained from SVM and KNN classifications of t-SNE-extracted features across the entirety of Task 2 experiments.

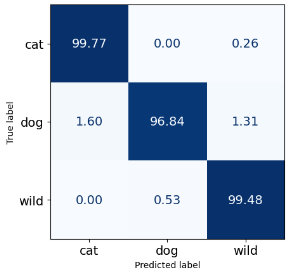

Figure 17. Animal Faces-HQ— SVM Confusion matrix when features extracted with VGG11-II

In summary, VGG11 demonstrated proficiency in capturing patterns within the colorectal dataset, particularly when incorporating transfer learning. The utilization of transfer learning proved beneficial in enhancing models, particularly through fine-tuning on a specific target dataset, as illustrated in Task 1. However, Task 2 highlighted that transfer learning without fine-tuning may be less effective when the target domains exhibit significant differences.

References

[1] M. M. Taye, "Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions," *Computation*, vol. 11, no. 3, 2023. 1

[2] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. 1

[3] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015. 1

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014. 1

[5] S. S. Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *Journal of Big data*, vol. 6, no. 1, pp. 1–18, 2019. 1

[6] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *cell*, vol. 172, no. 5, pp. 1122–1131, 2018. 1

[7] P. Kora, C. P. Ooi, O. Faust, U. Raghavendra, A. Gudigar, W. Y. Chan, K. Meenakshi, K. Swaraja, P. Plawiak, and U. R. Acharya, "Transfer learning techniques for medical image analysis: A review," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 1, pp. 79–107, 2022. 1

[8] I. U. Khan and N. Aslam, "A deep-learning-based framework for automated diagnosis of covid-19 using x-ray images," *Information*, vol. 11, no. 9, p. 419, 2020. 2

[9] R. Mehra *et al.*, "Breast cancer histology images classification: Training from scratch or transfer learning?," *ICT Express*, vol. 4, no. 4, pp. 247–254, 2018. 2

[10] T. Anju and S. Vimala, "Finetuned-vgg16 cnn model for tissue classification of colorectal cancer," in *International Conference on Intelligent Sustainable Systems*, pp. 73–84, Springer, 2023. 2

[11] A. Kumar, A. Vishwakarma, and V. Bajaj, "Crccn-net: Automated framework for classification of colorectal tissue using histopathological images," *Biomedical Signal Processing and Control*, vol. 79, p. 104172, 2023. 2

[12] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016. 2

[13] Y. Bar, I. Diamant, L. Wolf, and H. Greenspan, "Deep learning with non-medical training used for chest pathology identification," in *Medical Imaging 2015: Computer-Aided Diagnosis*, vol. 9414, pp. 215–221, SPIE, 2015. 2

[14] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016. 2

[15] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009. 2

[16] J. N. Kather, N. Halama, and A. Marx, "100,000 histological images of human colorectal cancer and healthy tissue," *Zenodo10*, vol. 5281, 2018. 2

[17] B. Schömig-Markiefka, A. Pryalukhin, W. Hulla, A. Bychkov, J. Fukuoka, A. Madabhushi, V. Achter, L. Nieroda, R. Büttner, A. Quaas, *et al.*, "Quality control stress test for deep learning-based diagnostic model in digital pathology," *Modern Pathology*, vol. 34, no. 12, pp. 2098–2108, 2021. 2

[18] Y. Tolkach, "Datasets digital pathology and artifacts, part 1," June 2021. 2

[19] Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "Stargan v2: Diverse image synthesis for multiple domains," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 2

[20] Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "Animal faces," 5 2020. Available at `www.kaggle.com/datasets/andrewmvd/animal-faces`. 2

[21] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, pp. 6105–6114, PMLR, 2019. 3