

Research Article

Assessment of a Deep-Learning System for Colorectal Cancer Diagnosis Using Histopathology Images

Purna Kar¹ , Sareh Rowlands^{2,*} 

¹School of Architecture Design and the Built Environment, Nottingham Trent University, Nottingham, United Kingdom

²College of Engineering, Mathematics, and Physical Sciences, University of Exeter, Exeter, United Kingdom

Abstract

Colorectal Cancer is one of the most common and lethal forms of cancer hence, an early and accurate detection is crucial. Traditional manual diagnosis is a tedious and time-consuming job susceptible to human errors; therefore, it is imperative to use computer-aided detection systems to interpret medical images for a quicker and more accurate diagnosis. In recent years deep-learning approaches have proved to be efficacious in predicting cancer from pathological images. This study assesses several deep-learning techniques for cancer diagnosis on digitized histopathology images, amongst which GoogLeNet and Xception emerged as the most effective, with GoogLeNet exhibiting slightly better precision in identifying cancerous tissues. Building on these findings the study proposes a new model (Xception+) by borrowing the idea from Xception architecture, which outperforms existing architectures with an accuracy of 99.37% for cancer diagnosis and 94.48% for cancer-grade classification. The primary inference of our research is assisting pathologists in detecting colorectal cancer from pathological images faster and more accurately. With notable accuracy and robustness, our proposed model has significant potential to analyze pathological images and detect the patterns associated with other types of cancer. Our study holds promise for driving the advancement of innovative medical diagnostic tools, aiding pathologists and medical practitioners in expediting cancer diagnosis processes.

Keywords

Artificial Intelligence, Cancer Detection, Colorectal Cancer, Convolutional Neural Networks, Deep Neural Networks

1. Introduction

According to the American Institute of Cancer Research (AICR) [1], Colorectal Cancer (CRC) is the third most common form of cancer after lung and breast cancers, contributing to almost 10% of the total cancer cases worldwide. It is the second most common form in women and the third most in men. AICR reported [1] that in 2020 there were more than 1.9 million new cases of cancer and more than 8,50,000 deaths globally with a mortality rate of CRC being 9% of all cancer-related deaths. However, as per the AICR's [1] find-

ings, the five-year survival rate for early detection (regionalized stage) in the USA is as high as 70%, making it vital for early and accurate detection. The diagnosis of CRC demands a thorough visual examination of digitized whole-slide images (WSIs) of Hematoxylin & Eosin (H&E)-stained histology images which is an extremely monotonous task and is prone to errors as the size of the cancer cells are minute and can be easily overlooked, this difficulty has been highlighted in the study [2]. As per the AICR report [1] the cases of CRC

*Corresponding author: S.Rowlands@exeter.ac.uk (Sareh Rowlands)

Received: 6 August 2024; **Accepted:** 2 September 2024; **Published:** 20 September 2024



Copyright: © The Author(s), 2024. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

worldwide are expected to rise by 60% over the next 15 years.

Therefore, the need for the diagnosis will also increase rapidly which would prove disastrous if pathologists only relied on manual examinations. Thus, it is essential to take the help of computer-aided detection (CAD) systems to improve precision and diminish the time and manual effort.

The current study compares several state-of-the-art deep-learning architectures for cancer detection and tissue classification. The study is an effort to exploit the capabilities of Artificial Intelligence to facilitate automated diagnosis to alleviate the burden on pathologists thus enabling faster and improved patient outcomes in the fight against cancer.

In the subsequent sections of the report, the relevant background studies and related work have been discussed, following which the techniques and our proposed model have been outlined. The later sections are dedicated to the presentation and discussion of our results. The final section discusses the key conclusions drawn from this research and the future scope.

2. Background and Motivation

With the advent of computer vision disciplines, there has been a huge improvement in CAD; state-of-the-art deep neural networks have replaced the traditional feature extraction and classification methods. Traditional methods comprise two steps, first, an image descriptor is used to encode the texture and patterns in an image called 'features' into a feature matrix. Then this feature matrix is used in a supervised machine learning-based classifier to classify the images into cancerous and non-cancerous classes. Various studies [3, 4] have incorporated the traditional supervised methods for detection, but the time taken to extract features and classify images is very high and the accuracy is very mediocre. Artificial Neural Networks have revolutionized the field of machine learning and computer vision; Convolutional Neural Networks (CNNs) have been used in most image processing problems [5]. The performance of these CNN-based models has also vastly improved from the traditional methods. Nowadays, CNN-based deep-learning models are used in most computer vision problems like image classification [5]. They are also gaining popularity in bio-medical fields, several studies [6-9] have used deep-learning techniques which have proven to be highly effective. Of late, deep-learning techniques have proved to be very efficient in analyzing pathological images for various oncology and clinical studies for cancer. Many authors performed comparative studies [4, 10, 11] of the CNN architectures for cancer diagnosis. Several studies on post-cancer diagnosis have incorporated deep-learning algorithms for grade classification [12-14], tumor cell detection [15, 16], gland segmentation [13], and even speculation of patient survivorship [17].

One shortcoming of the deep-learning models is the requirement of a massive amount of labeled data to train the model. In the context of pathological datasets, this is a huge

challenge as labeled datasets are in short supply. Proper annotation of clinical images is expensive as it requires visual scrutiny by pathologists which is a very strenuous and time-consuming task. Privacy is of utmost concern and all ethical and privacy policies must be followed; one must ensure that no data or information can be traced to individual patients. Despite the challenges, deep-learning methods are extensively used in several bio-medical problems [18] due to highly accurate diagnosis.

In this study, several CNN-based algorithms were used for Cancer Diagnosis and Cancer Grade Classification. Firstly, the models were used for cancer detection by predicting whether images have cancerous tumors present in them or not. Secondly, tissue classification was performed, where the algorithm determined the tissue class of the pathological images. Furthermore, a new CNN-based model was proposed by slightly modifying the Xception architecture (henceforth called Xception+), and its accuracy was compared with the known models. Finally, cancer grade classification was performed which involves determining the grade of colon cancer from histopathology tissue slides utilizing the top-performing models such as GoogLeNet, Xception, and our proposed model (Xception+).

3. Materials and Methods

This study has two primary goals. Firstly, do a comparative study of several deep learning algorithms to perform cancer diagnosis and tissue classification of digitized H&E-stained images of CRC. Secondly, propose a new model based on the Xception architecture and compare the results with the standard models.

3.1. Data and Resources

The dataset primarily used in this study is the 'NCT-CRC-HE-100K' [19], which contains digitized histopathology images; they are H&E-stained tissue sections. It has 100,000 non-overlapping image patches of CRC and normal tissue parts. The tissue classes are defined as Adipose (ADI), background (BACK), debris (DEB), lymphocytes (LYM), mucus (MUC), smooth muscle (MUS), normal colon mucosa (NORM), cancer-associated stroma (STR), colorectal adenocarcinoma epithelium (TUM). The tissue class 'TUM' represents the cancerous class. For the Cancer detection problem, the Training and Validation dataset contains 30,000 images (split in a 7:3 ratio); i.e., 15,000 images of cancer and normal tissues respectively. The Test dataset contains 7200 images; 3600 images of cancer and normal tissues respectively. For the Tissue Classification Problem, the Training and Validation dataset contains 18,000 images (split in a 7:3 ratio), with 2000 images of each tissue class. The Test dataset contains 4050 images; 450 images of each tissue class. All the aforesaid datasets have no class imbalance. For cancer grade classification, we have used the dataset [14], which has 139

images, consisting of 71 normal, 33 low-grade, and 35 high-grade cancer images.

Pytorch [20] libraries have been used to build deep-learning architectures. Local Interpretable Model-agnostic Explanations (LIME) [21] has been used for explainable AI.

3.2. Cancer Diagnosis using Deep-Learning Methods

In this experiment several popular deep-learning architectures like AlexNet [5], GoogLeNet [22], Inception V3 [23],

ResNet [24], MobileNet [25], Xception [26], DenseNet [27], and ResNeXt [28] were incorporated. A single experiment has been divided into two phases, the training- validation phase where the deep learning model is trained and validated, and the testing phase, where the model is tested on unseen data. After completion of the test phase, the entire experiment is repeated 5 times, and the mean and standard deviation of the test accuracies are recorded. The below steps have been followed for the deep-learning algorithms:

- 1) In the training-validation phase images are resized (224x224), normalized, and converted to tensors.

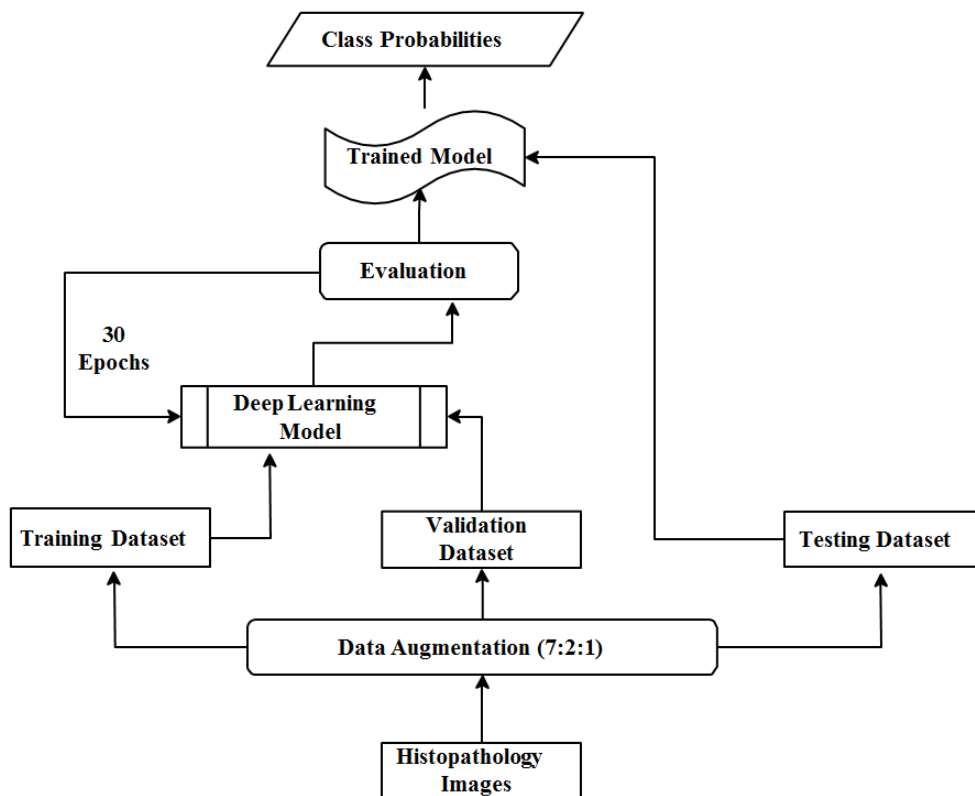


Figure 1. High Level Project Design for deep-learning models.

The dataset is split into training and validation datasets in the ratio 7:3.

- 2) The datasets are loaded to the respective data loaders in batches of 10.
- 3) A deep-learning algorithm is trained on the training images. The cross-entropy loss and the loss gradients are computed for each batch and the weights of all learnable parameters are updated.
- 4) Post-training, the model is used to predict the classes for images in the validation dataset, and the accuracy is noted.
- 5) The previous two steps are repeated for 30 epochs. The weights of the model with the best validation accuracy are saved.
- 6) In the test phase, the test dataset is loaded in batches of

50. The saved model (from step 5) is used to predict the classes of test images and the average accuracy of the prediction is recorded.

- 7) Steps (1) through (6) are repeated 5 times and the mean and standard deviation of the accuracy obtained are recorded.

The High-Level Diagram of the above steps is illustrated in Figure 1.

For cancer detection, the below 4 metrics were used to evaluate the performance of the different models.

- 1) Accuracy represents the number of correctly classified images over the total number of images. The formula is given by:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

- 2) Precision is the positive predictive value and is given by the formula:

$$Precision = \frac{TP}{TP + FP}$$

- 3) Recall also known as sensitivity or true positive rate. This should be high for a good classifier. The formula is given by:

$$Recall = \frac{TP}{TP + FN}$$

- 4) F1 score metric considers both precision and recall, this is a better metric than accuracy and is defined as:

$$F1score = \frac{2 * (Precision * Recall)}{Precision + Recall}$$

Accuracy is the ratio of correctly predicted observations to the total observations (useful when there is no class imbalance) whereas Precision is the ratio of correctly predicted observations to the total positive observations, for this problem it answers how many have cancer? The Recall is the ratio of correctly predicted positive observations to all observations in the actual class. This metric answers the question from all the cancer cases how many were labelled accurately? The F1 score is the weighted average of Precision and Recall, this metric is useful if there is an imbalanced dataset. These metrics are also called “Confusion Metrics” which uses the counts for TN (True Negative), FN (False Negative), FP (False Positive), and TP (True Positive).

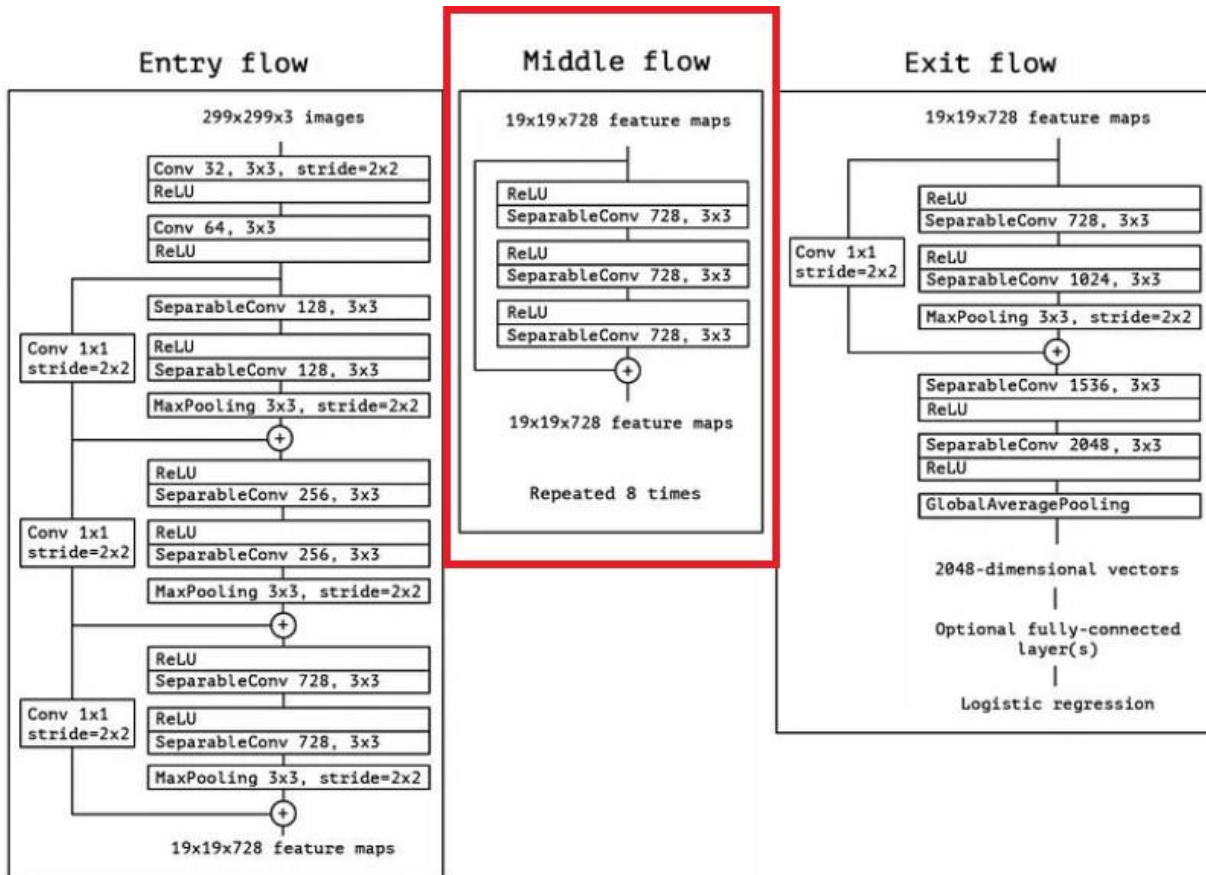


Figure 2. Architecture of the Xception model. [26]. The layers outlined in red have been omitted, the resultant layers form our proposed model (Xception+).

3.3. Proposed Model (Xception+)

In this experiment, a modified Xception [26] architecture (Xception+) was proposed and the accuracy in cancer diagnosis and tissue classification was noted. Figure 2, portrays the original architecture of the Xception. The Xception [26]

uses modified Depthwise Separable Convolutions, where the pointwise convolutions (1x1 convolution) are followed by the depthwise convolutions (nxn spatial convolutions). The architecture is divided into 3 flows, The Entry Flow, The Middle Flow, and the Exit Flow. The Entry Flow consists of a 3x3 convolutional layer with 32 filters and a stride of 2x2, followed by a 3x3 convolutional layer with 64 filters and finally

modified depthwise separable convolution layers with 128, 256, and 728 filters, with 1x1 convolutional layers followed by 3x3 Max Pooling. The entry layer is responsible for extracting low-level features from the image. The Middle Layer comprises eight repeated blocks, where each block is made up of depthwise separable convolutions with 728 filters and a 3x3 kernel, it is responsible for extracting complex and higher-level features. Finally, the Exit block consists of Separable convolution with 728, 1024, 1536, and 2048 filters, all with 3x3 kernel, followed by Average Pooling and a Dense Layer, thus refining the extracted features and performing final predictions. The proposed model, Xception+ has been designed to omit the entire Middle Flow comprising 8 blocks. In Figure 2 the blocks outlined with red rectangle have been excluded, the remainder network with only the Entry and Exit Flow is the proposed model (Xception+).

3.4. Cancer Grade Classification

In this task, a grade classification of cancer from histo-

pathology tissue slides of CRC has been performed. Since the dataset [14] has only 139 images, a patch generation technique [29] was used to increase the size and diversity of our dataset. In this technique a single image is cropped into multiple smaller non-overlapping images or patches, each patch inherits the label of the original image and is considered a unique image. The patches generated are of 3 grades (classes) namely High, Low, and Normal. The dataset is split into training and validation datasets in the ratio 7:3. A deep-learning algorithm (GoogLeNet, Xception, and Xception+) is trained for 50 epochs on these images. Once all batches of images are trained, the model is used to predict the classes of images using the validation dataset, and the best validation accuracy is noted.

4. Results

This section presents the results of the experiments and the performance of the proposed model.

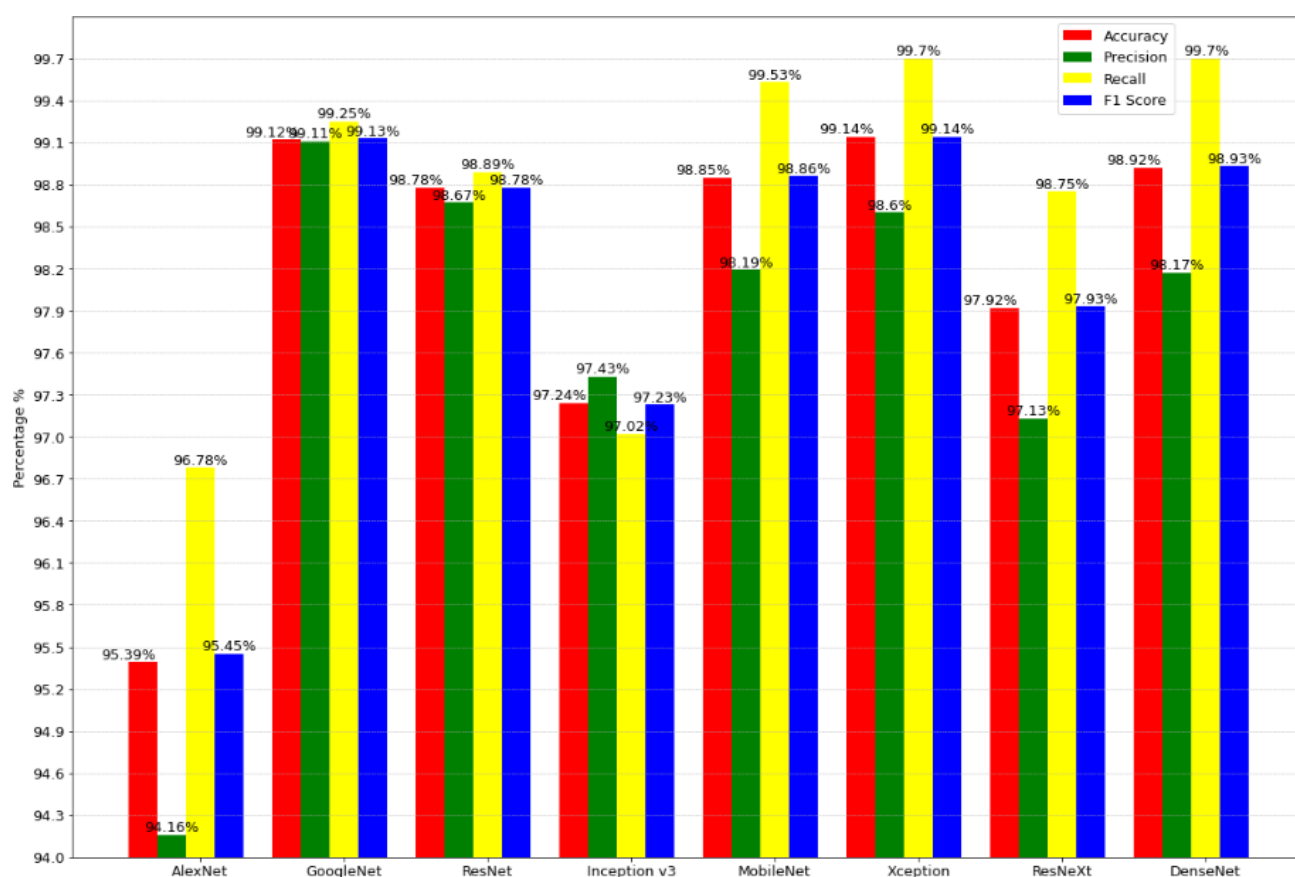


Figure 3. Cancer Detection Problem: Performance comparison of the deep-learning algorithms.

4.1. Performance of Deep-Learning Models

Two optimizers were used in the training phase of the deep-learning models: Adam (uses features of AdaGrad and

RMSProp algorithms) and SGD (extension of Gradient Descent). Optimizers are algorithms for stochastic gradient descent for training deep-learning models. Table 1, shows the performance of each model using the two optimizers. It has been observed the performance of the models using SGD

optimizers is slightly better in most cases. The values of the hyperparameters have been experimentally determined and have been kept constant for all models. The following values were used:

- 1) Learning rate=0.0005
- 2) Weight decay=0.001 (only for Adam Optimizer)
- 3) Momentum=0.9 (only for SGD)
- 4) Epochs=30

Since there is no class imbalance, the accuracy, and F1 score are approximately similar. Figure 3, illustrates the performance of the models for Cancer Detection. It has been observed that the GoogLeNet architecture has the highest precision of 99.11%, whereas the mean accuracy, recall and F1 score of Xception are the highest (99.25%, 99.70%, and 99.14% respectively). Overall, Xception has the best performance among all the deep-learning models. Receiver Operating Characteristic Curve (ROC) has been plotted for only 3 models Xception (the best performing), AlexNet (the worst performing), and Inception v3 (medium performing) for easy comparison and visualization, refer Figure 4. The ROC curve for a binary classifier shows the performance of the model at various threshold settings. The ROC curve is plotted with TPR (True Positive Rate) vs FPR (False Positive Rate). TPR

is nothing but Recall, FPR is defined by the formula:

$$\frac{FP}{FP + TN}$$

AUC stands for the Area under the Curve, it is a performance metric widely used for classification problems, the higher the AUC the better is the model at predicting the classes. As seen in Figure 4, the AUC for Xception is 1.0 which is the ideal value, this implies that this model is highly accurate in classifying the images.

For the tissue classification problem, the accuracy of the models in classifying the 9 tissue classes has been compared, refer to Tables 1 and 2 for the detailed results. Figure 5 illustrates the performances of the models for classifying each tissue class; it has been observed that GoogLeNet has the highest accuracy for classification across all tissue classes followed by Xception. Figure 6, compares the overall performance of each deep-learning model for both the tasks, i.e., Cancer detection and tissue classification. The average, mean, and standard deviation of the test accuracies from 5 experiments have been tabulated in Table 2.

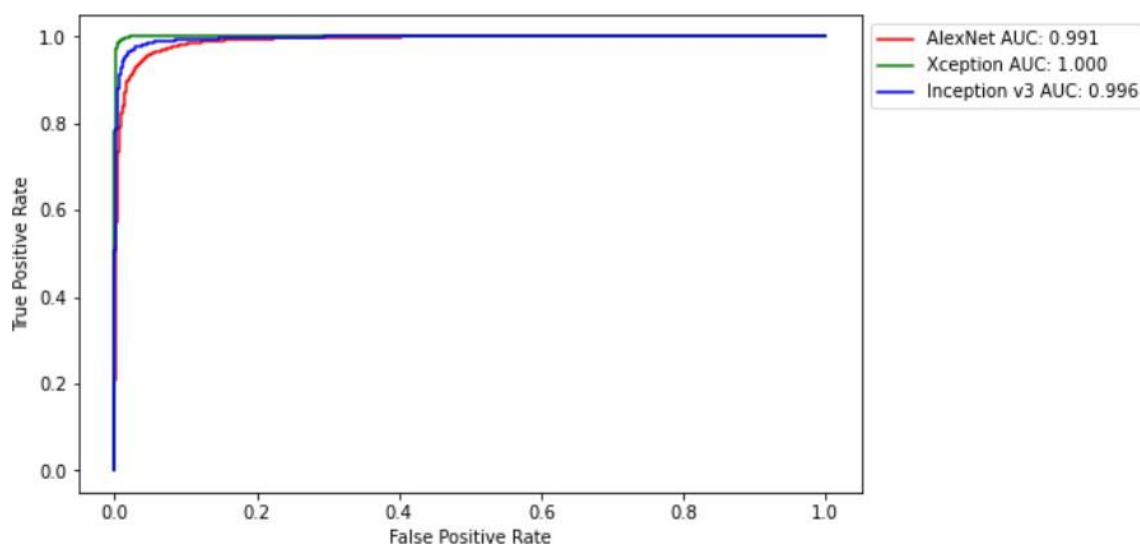


Figure 4. Cancer Detection Problem: ROC curve for few deep-learning models.

Table 1. Accuracy of Different models.

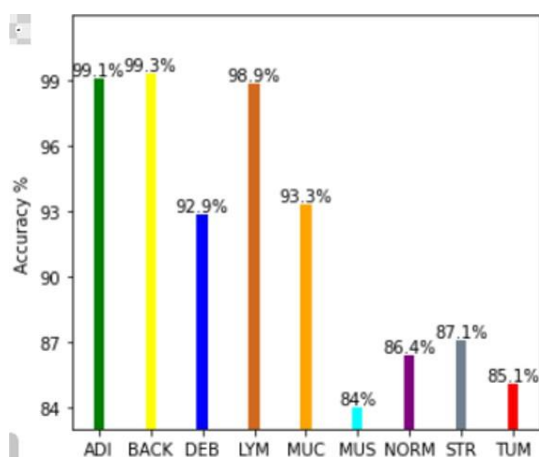
Model	Cancer detection		Tissue classification	
	Adam Optimizer	SGD Optimizer	Adam Optimizer	SGD Optimizer
AlexNet	95.39%	94.04%	91.80%	87.96%
GoogLeNet	99.00%	99.12%	97.88%	98.86%
ResNet	98.78%	98.72%	97.04%	96.94%
Inception v3	97.24%	95.81%	93.92%	96.44%

Model	Cancer detection		Tissue classification	
	Adam Optimizer	SGD Optimizer	Adam Optimizer	SGD Optimizer
MobileNet	98.85%	98.19%	96.22%	94.37%
Xception	98.47%	99.14%	96.25%	97.16%
ResNeXt	96.25%	97.92%	93.63%	93.33%
DenseNet	97.00%	98.92%	94.30%	96.37%

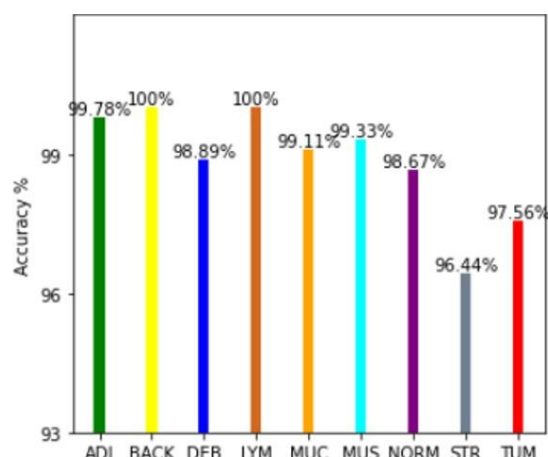
4.2. Performance of Proposed Model: Xception+

The proposed model Xception+ performed exceptionally well for both classification tasks. It achieved an overall average accuracy of 99.37% and 98.22% for cancer detection and tissue classification respectively, which is better than the known architectures, refer to Table 2. Through the experiments, it has been observed that Xception+ performs better than the original architecture. To interpret the proposed ar-

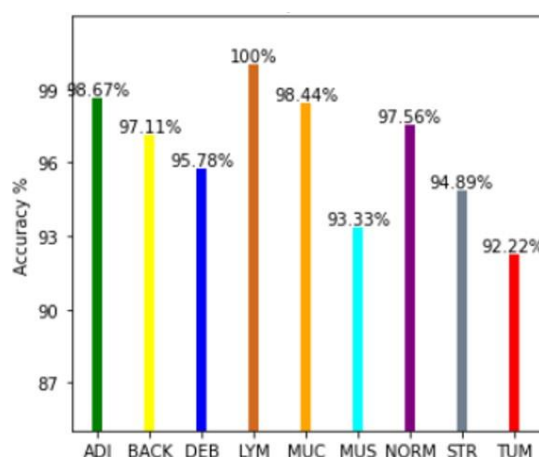
chitecture, LIME [21] was used for visual explanations. LIME aims to explain or approximate a machine-learning model with a local interpretable model to visually explain each prediction. Figure 7 shows the regions of the image that contributed to the classification by the model. For the image 7a, the regions highlighted in green, show the segments in the image the model is looking at for labeling as 'Cancerous'. Whereas, in Figure 7b the regions highlighted in red influenced the labeling of the 'Non-Cancerous' class.



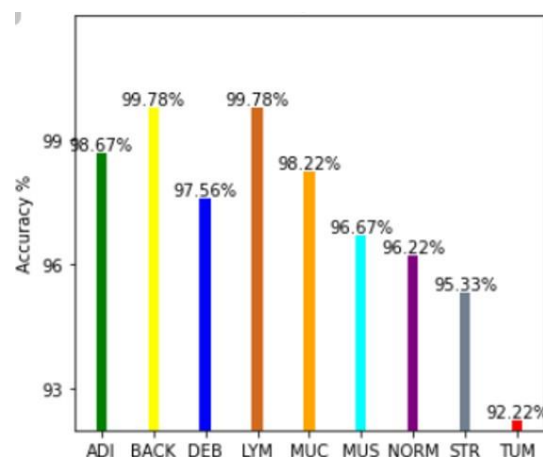
(a) AlexNet



(b) GoogLeNet



(c) Inceptionv3



(d) Xception

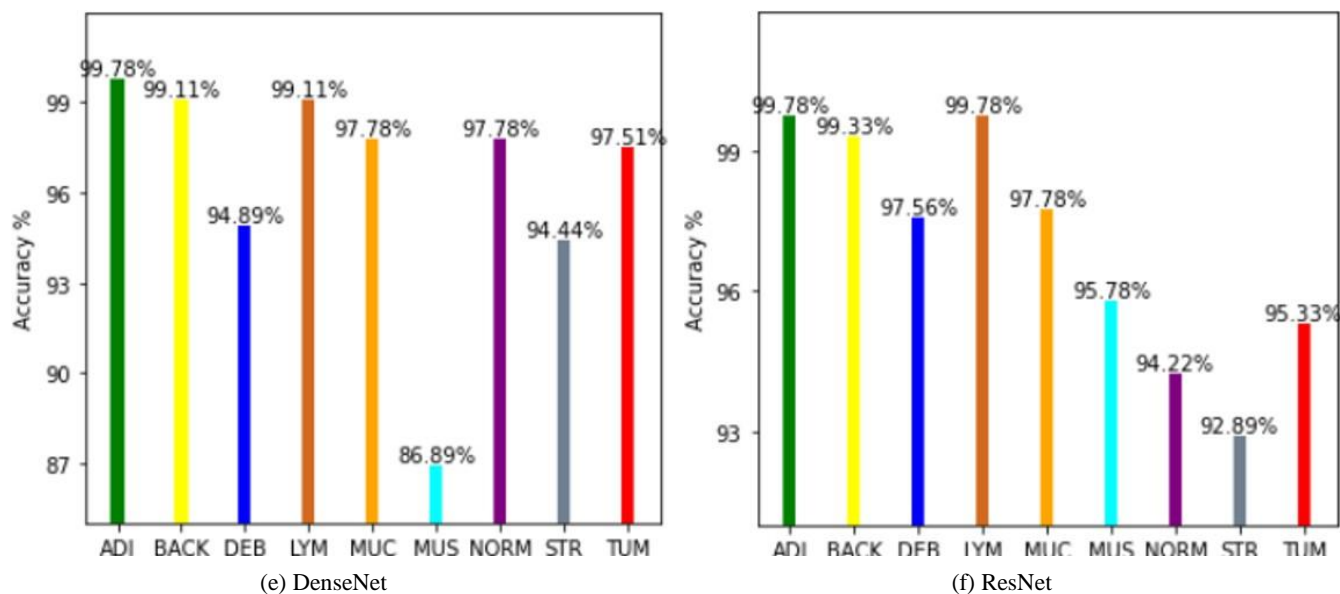


Figure 5. Class-wise Accuracy of Models for Tissue Classification.

4.3. Grade Classification

Due to the huge class imbalance and inadequacy in the dataset, a patch generation technique was employed to increase the dataset size and diversity. By using a step size of 12 on images of dimensions 512x512x3, a total of 15,520 images were generated. The 3 top performing CNN models from cancer diagnosis experiments namely, GoogLeNet, Xception, and Xception+ were used for training and validating these

patches for classification. The highest validation accuracy of GoogLeNet and Xception models was 92.86%. While Xception+ (the proposed model) achieved a higher accuracy of 94.48%. Figure 8 illustrates the per-class accuracies for each of the models. Xception+ demonstrated a superior accuracy in predicting all three grades (High: 85.1%, Low: 94.9%, and Normal: 97.1%) compared to the two other models. In contrast to this study, the accuracy reported by the study by Awan et al. [14] was 91%, indicating the superior performance of the proposed model compared to Awan et al. [14].

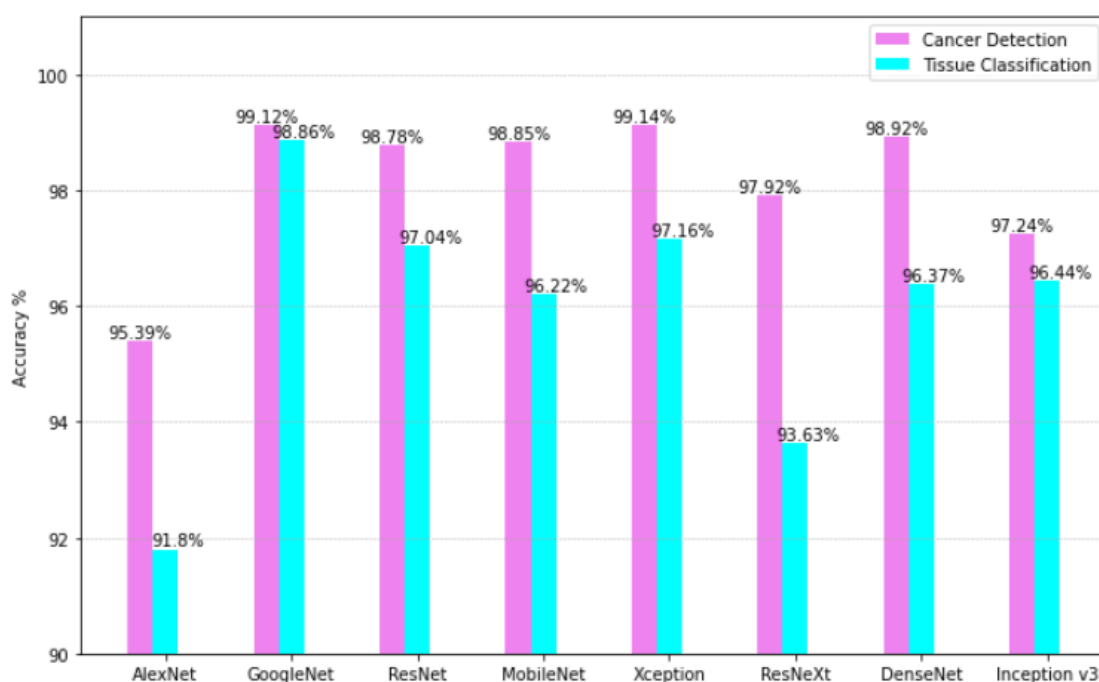


Figure 6. Comparison of accuracy of each model for the two problems.

Table 2. Mean and standard deviation of average test accuracy.

Model	Cancer detection		Tissue classification	
	Mean accuracy	Standard Deviation	Mean accuracy	Standard Deviation
AlexNet	97.96%	0.0877	94.55%	0.1143
GoogLeNet	99.14%	0.0665	98.86%	0.0671
ResNet	98.61%	0.2081	96.61%	0.2201
MobileNet	98.39%	0.2819	96.27%	0.2713
Xception	99.25%	0.1033	97.52%	0.3409
Xception+	99.37%	0.0524	98.22%	0.1896

5. Discussion

As demonstrated by the results, Xception has the best mean accuracy (99.25%) for the Cancer detection problem and GoogLeNet has displayed the best mean accuracy (98.86%) in the tissue classification problem. There might be a few reasons why Xception and GoogLeNet outperformed the rest of the state-of-the-art neural networks. As the network goes deeper it becomes more difficult to train and beyond a certain point the train loss increases which overfits the model and then it does a poor generalization which in turn has a detrimental effect on its performance. Moreover, deeper networks are more prone to the vanishing gradient problem. By increasing more layers with activation functions like sigmoid, the gradients of the loss function tend to zero making the network difficult to train. The gradients of the loss functions for each layer are computed using the backpropagation method which uses the chain rule, if the gradients are very low, with each layer the gradient reduces exponentially and by the

time it propagates to the initial layer, it approaches zero. The main purpose of backpropagation is to find the optimum amount for changing the weights and biases of the learnable parameters. If the gradient of the initial layers is very low then the learnable parameters of these layers will not be updated properly thus the performance will get saturated. The architecture of GoogLeNet is less deep than all the state-of-the-art models hence GoogLeNet is less prone to vanishing gradient problem and the accuracy increases faster than most models. Xception on the other hand uses skip connections, skip connections allow the gradients to propagate to the initial layers with greater magnitude by skipping a few in-between layers thus tackling the vanishing gradient problem. Additionally, in the Xception model, convolutions are not performed across all the channels, making fewer connections thus making the architecture less deep, therefore, the model is more trainable and less prone to overfitting. When compared to the various studies conducted by other researchers, such as [4, 10, 11], the models in this study demonstrated superior results as shown in Table 3.

Table 3. Comparison of similar works.

Model		Cancer detection	Tissue classification
GoogLeNet		99.14%	98.86%
Xception		99.25%	97.52%
Xception+	(Our Proposed Model)	99.37%	98.22%
Inception	v3	97.24%	97.15%
Inception	v3 [4]	90.50%	87.00%
Adaptive	CNN [4]	94.50%	92.00%
Proposed	AI model [10] (same dataset as ours)	98.11%	
Proposed	AI model [10] (different dataset)	99.02%	
Multiple	CNNs [11]	94.11%	94.4%

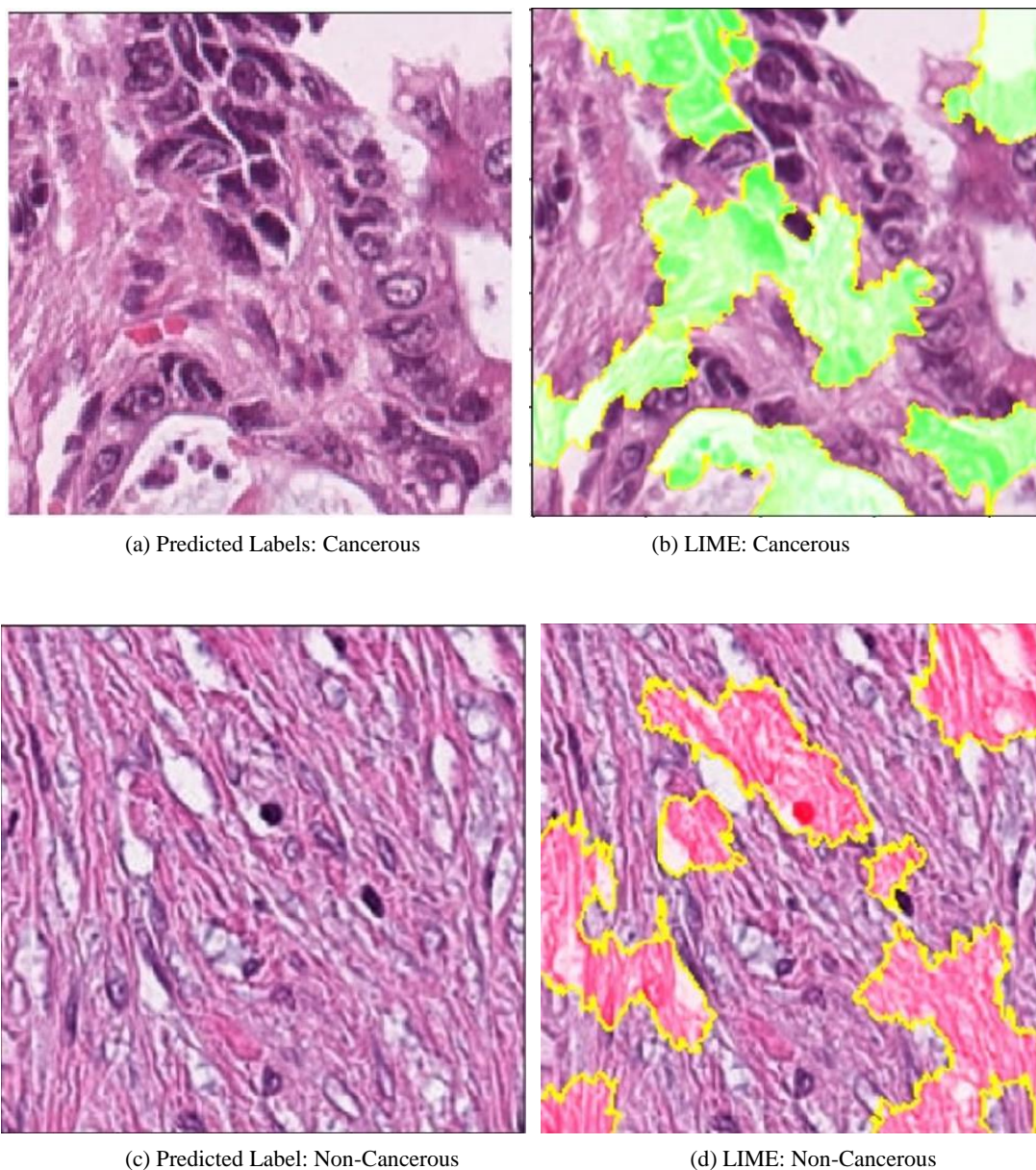


Figure 7. Examples of explainable model predictions for Xception+ using the LIME algorithm.

The proposed model, Xception+ has outperformed the other standard models (GoogLeNet and Xception) in cancer diagnosis as well as cancer grade classification. Figure 9 illustrates the confusion matrix showing the performance of the proposed model (Xception+) for each of the problems, as can be seen, the model has very few misclassifications. Large networks are prone to overfitting and incur a high computational cost during training. Omitting layers is effective in reducing computational cost, overfitting, and generalization error thus making the network more efficient. Not every

neuron in the neural network contributes to the output some of them are redundant, and removing these neurons contributes to a smaller and faster network. A resultant smaller network is better at handling overfitting and vanishing gradient problems thus improving the accuracy. The proposed model (Xception+) is more compact and smaller than the actual Xception model and achieves better prediction accuracy. For the cancer grade classification task as well Xception+ performed better than the known models.

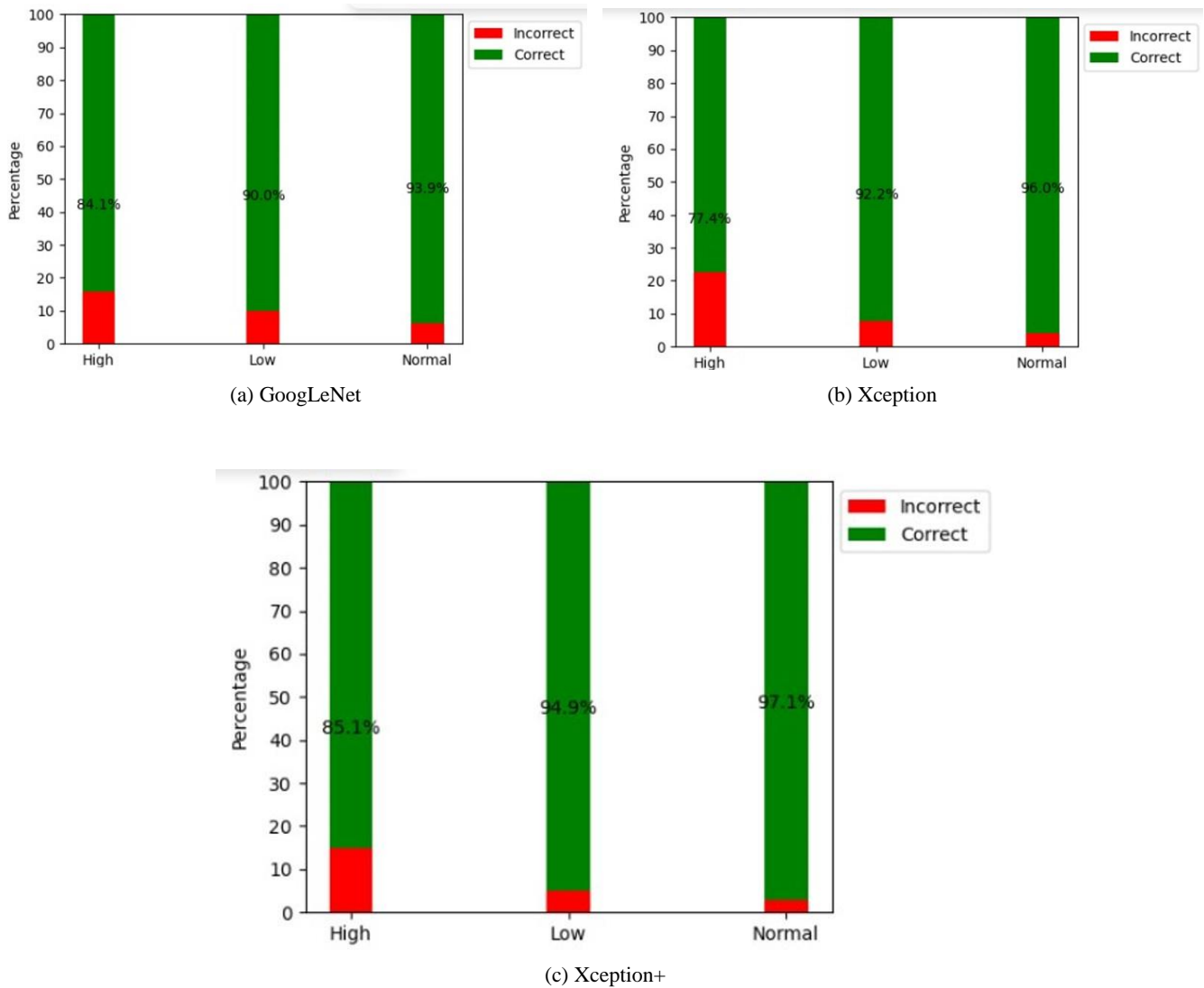
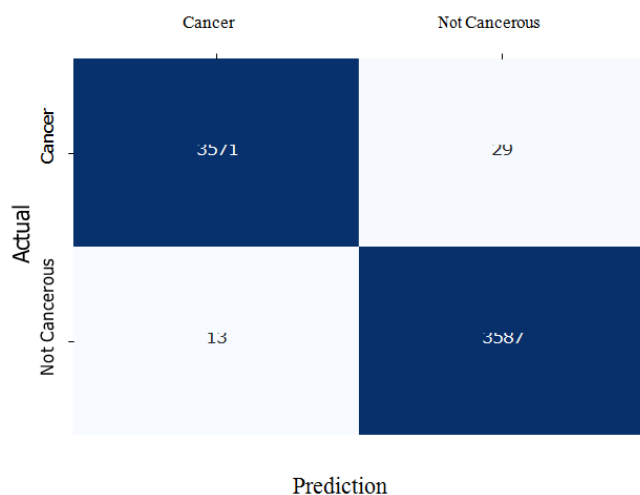


Figure 8. Accuracy for different methods for Cancer Grade Classification.

6. Conclusion

Among the deep-learning-based models, the best results were exhibited by Xception and GoogLeNet with the precision of GoogLeNet being slightly better than that of Xception, implying that GoogLeNet predicts the cancerous class with better perfection. It is notable that the proposed model Xception+, which is obtained by modifying the Xception network to a smaller size architecture, provides better accuracy in both problems (99.37% and 98.22%). Xception+ also outperformed the accuracy achieved in similar studies conducted by other researchers like [4, 10, 11]. Furthermore, for the task for the grade classification of CRC, the average accuracy achieved by Xception+ was 94.48%, the highest among the

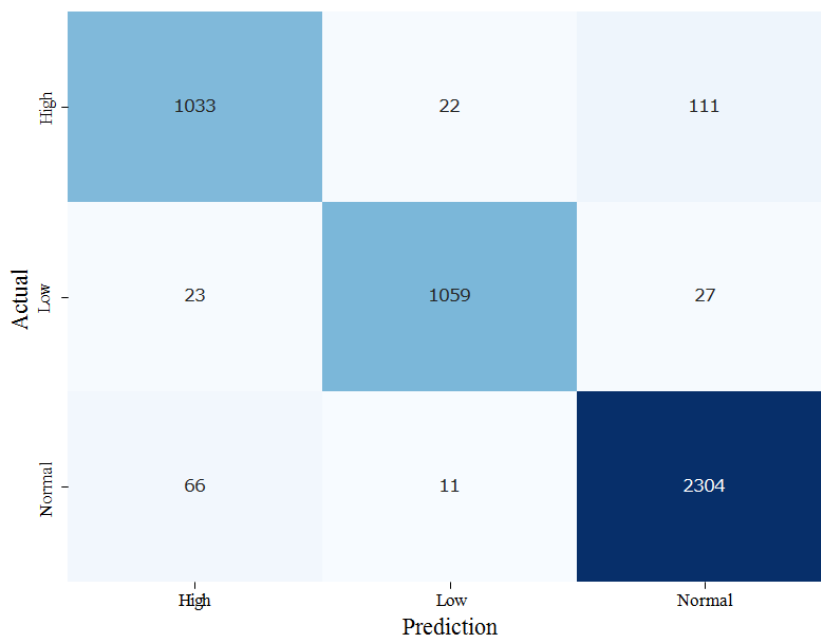
known models and other studies [14]. In the future, we aim to expand this research by incorporating other comparable datasets and observe the performance of Xception+ on datasets comprising histopathology images sourced from various patients, covering conditions such as breast cancer, skin cancer, brain tumors, and others. Additionally, to boost the robustness of the model we intend to incorporate diverse types of pathological images like biopsies, CT scans, and X-rays. There is significant scope for further exploring the effects of Transfer Learning by using pre-trained models trained on similar histology images like breast cancer, skin cancer, etc. Finally, fine-tuning various architectures and analyzing their efficiency could contribute to a deeper understanding of model optimization.



(a) Cancer Diagnosis

Actual \ Prediction	ADI	BACK	DEB	LYM	MUC	MUS	NORM	STR	TUM
ADI	449	0	0	0	0	1	0	0	0
BACK	0	449	0	0	0	0	1	0	0
DEB	0	0	445	0	0	0	0	4	1
LYM	0	0	0	450	0	0	0	0	0
MUC	0	0	1	0	444	0	2	2	1
MUS	0	0	1	0	0	446	0	3	0
NORM	0	0	0	1	4	0	442	1	2
STR	0	0	8	3	0	11	1	423	4
TUM	0	0	3	1	4	0	12	0	430

(b) Tissue Classification



(c) Cancer Grading

Figure 9. Confusion Matrices for showing the performance of Xception+ model.

Our research incorporating deep-learning models for cancer diagnosis has the potential to have a high impact in the field of clinical studies for cancer due to its high efficiency and predictive accuracy. It has the proficiency to analyze complex pathological imaging data and unearth the abstruse patterns or biomarkers of CRC in providing highly accurate outcomes leading to early detection and personalized treatment procedures resulting in improved survival rates. Research such as ours has the potential to propel the development of innovative medical diagnostic tools to assist pathologists and medical professionals in faster cancer diagnosis.

Abbreviations

ADI	Adipose
AICR	American Institute for Cancer Research
AUC	Area Under the Curve
BACK	Background
CAD	Computer-Aided Detection
CNN	Convolutional Neural Network
CRC	Colorectal Cancer
DEB	Debris
FN	False Negative
FP	False Positive

H&E	Hematoxylin and Eosin
LIME	Local Interpretable Model-Agnostic Explanations
LYM	Lymphocytes
MUC	Mucus
MUS	Muscle
NORM	Normal
ROC	Receiver Operating Characteristic
SGD	Stochastic Gradient Descent
STR	Stroma
TN	True Negative
TP	True Positive
TUM	Tumour
WSI	Whole-Slide Images

Ethics

No human trials were involved as part of this study. The authors have used publicly available datasets *NCT-CRC-HE-100K* [19]. Please refer to their Ethics statement for the dataset from the link: <https://zenodo.org/records/1214456>.

Funding

The authors declare no funding was received for this study.

Data Availability Statement

All data for performing the experiments have been uploaded in the location: Dataset. The codes are available on the Github repository: Source Code. The ReadMe file in the GitHub project has detailed instructions on executing the codes, and which datasets to use to reproduce the results. If the above link is inaccessible, all datasets and codes can be produced upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Clinton, S. K., Giovannucci, E. L., Hursting, S. D. The World Cancer Research Fund/American Institute for Cancer Research Third Expert Report on Diet, Nutrition, Physical Activity, and Cancer: Impact and Future Directions, *Journal of Nutrition* 2020; 4th edn.
- [2] Ahmet H. Y., Hassan E., Gokalp C., et al. Classification of Diabetic Rat Histopathology Images Using Convolutional Neural Networks, *International Journal of Computational Intelligence Systems* 2021; 14: 715-722, Available from: <https://doi.org/10.2991/ijcis.d.201110.001>
- [3] Kralr P., Lenc L. LBP features for breast cancer detection, *IEEE International Conference on Image Processing (ICIP)* 2016; 2643-2647, Available from: <https://doi.org/10.1109/ICIP.2016.7532838>
- [4] Malik J., Kiranayaz S., Kunhoth S., et al. Colorectal cancer diagnosis from histology images: A comparative study, *ArXiv* 2019; abs/1903.11210, Available from: <https://doi.org/10.48550/arXiv.1903.11210>
- [5] Krizhevsky A., Sutskever I., Hinton G. E. ImageNet classification with deep convolutional neural networks, *Communications of the ACM* 2017; 60: 84-900, Available from: <https://doi.org/10.1145/3065386>
- [6] Tang M. C. S., Teoh S. S., Ibrahim H., et al. Neovascularization Detection and Localization in Fundus Images Using Deep Learning, *Sensors (Basel)* 2021; 16: 5327, Available from: <https://doi.org/10.3390/s21165327>
- [7] Tang M. C. S., Teoh S. S., Ibrahim H., et al. A Deep Learning Approach for the Detection of Neovascularization in Fundus Images Using Transfer Learning, *IEEE Access* 2022; 10: 20247-20258, Available from: <https://doi.org/10.1109/access.2022.3151644>
- [8] Tang M. C. S., Teoh S. S. Blood vessel segmentation in fundus images using Hessian matrix for diabetic retinopathy detection, *IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* 2020; 0728-0733, Available from: <https://doi.org/10.1109/IEMCON51383.2020.9284931>
- [9] Tang M. C. S., Teoh S. S., Ibrahim H. Retinal Vessel Segmentation from Fundus Images Using DeepLabv3+, *IEEE 18th International Colloquium on Signal Processing and Applications (CSPA)* 2022; 377-381, Available from: <https://doi.org/10.1109/CSPA55076.2022.9781891>
- [10] Wang K. S., Yu G., Xu C., et al. Accurate diagnosis of colorectal cancer based on histopathology images using artificial intelligence, *BMC Med* 2021; 19: 76, Available from: <https://doi.org/10.1186/s12916-021-01942-5>
- [11] Davri A., Birbas E., Kanavos T., et al. Deep Learning on Histopathological Images for Colorectal Cancer Diagnosis: A Systematic Review. *Diagnostics, Basel* 2022; 12(4): 837, Available from: <https://doi.org/10.3390/diagnostics12040837>
- [12] Sari C. T., Gunduz-Demir C. Unsupervised feature extraction via deep learning for histopathological classification of colon tissue images, *IEEE Trans Med Imaging* 2018; 38(5): 1139-1149, Available from: <https://doi.org/10.1109/TMI.2018.2879369>
- [13] Sirinukunwattana K., Pluim J. P. W., Chen H., et al. Gland segmentation in colon histology images: The glas challenge contest, *Medical Image Analysis* 2017; 35: 489-502, Available from: <https://doi.org/10.1016/j.media.2016.08.008>
- [14] Awan R., Sirinukunwattana K., Epstein D. Glandular morphometrics for objective grading of colorectal adenocarcinoma histology images, *Scientific Reports* 2017; 7(1): 16852, Available from: <https://doi.org/10.1038/s41598-017-16516-w>
- [15] Sirinukunwattana K., Ahmed S. E., Tsang Y. W., et al. Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images, *IEEE Trans Med Imaging* 2016; 35(5): 1196-1206, Available from: <https://doi.org/10.1109/TMI.2016.2525803>

- [16] Chaddad A., Tanougast C. Texture Analysis of Abnormal Cell Images for Predicting the Continuum of Colorectal Cancer, *Anal Cell Pathol (Amst)* 2017; 2017: 8428102, Available from: <https://doi.org/10.1155/2017/8428102>
- [17] Kather J. N., Krisam J., Charoentong P., et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study, *PLoS Med* 2019; 16(1): e1002730, Available from: <https://doi.org/10.1371/journal.pmed.1002730>
- [18] Yamashita R., Nishio M., Do R. K. G., et al. Convolutional neural networks: an overview and application in radiology, *Insights Imaging* 2018; 9: 611-629, Available from: <https://doi.org/10.1007/s13244-018-0639-9>
- [19] Kather J. N., Halama N., Marx A. 100,000 histological images of human colorectal cancer and healthy tissue, *Zenodo* 2018; Available from: <https://doi.org/10.5281/zenodo.1214456>
- [20] Paszke A., Gross S., Massa F., et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library, *Conf. on Neural Inf. Process. Syst. (NeurIPS 2019)* 2019; 721: 8024-8035, Available from: <https://proceedings.neurips.cc/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf>
- [21] Marco T. R., Singh S., Guestrin C., et al. "Why Should I Trust You?": Explaining the Predictions of Any Classifier, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 2016; 16: 715-722, Available from: <https://doi.org/10.1145/2939672.2939778>
- [22] Szegedy C., Liu W., Jia Y., et al. Going deeper with convolutions, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2015; 1-9, Available from: <https://doi.org/10.1109/CVPR.2015.7298594>
- [23] Szegedy C., Vanhoucke V., Ioffe S., et al. Rethinking the Inception Architecture for Computer Vision, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2016; 2818-2826, Available from: <https://doi.org/10.1109/CVPR.2016.308>
- [24] He K., Zhang X., Ren S. Deep Residual Learning for Image Recognition, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2016; 770-778, Available from: <https://doi.org/10.1109/CVPR.2016.90>
- [25] Howard A., Zhu M., Chen B., et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, *ArXiv* 2017; abs/1704.04861, Available from: <https://doi.org/10.48550/arXiv.1704.04861>
- [26] Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2017; 1800-1807, Available from: <https://doi.org/10.1109/CVPR.2017.195>
- [27] Huang G., Liu Z., Van-Der-Maaten L., et al. Densely Connected Convolutional Networks, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2017; 2261-2269, Available from: <https://doi.org/10.1109/CVPR.2017.243>
- [28] Xie S., Girshick R., Dollár P., et al. Aggregated Residual Transformations for Deep Neural Networks, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2017; 5987-5995, Available from: <https://doi.org/10.1109/CVPR.2017.634>
- [29] Araújo T., Aresta G., Castro E., et al. Classification of breast cancer histology images using Convolutional Neural Networks, *PLoS One* 2017; 12: 6, Available from: <https://doi.org/10.1371/journal.pone.0177544>