A Study of the Generalisability of CNNs for Disease Prediction

Thashil Naidoo
Department of Computer Science
University of Pretoria
Pretoria, South Africa
tash2814@gmail.com

Nelishia Pillay
Department of Computer Science
University of Pretoria
Pretoria, South Africa
nelishia.pillay@up.ac.za

***Abstract*—Convolutional neural networks (CNNs) have proven to be effective for disease prediction from images, namely, CT scans. However, previous research studies the application of CNNs for disease prediction separately for each disease. The aim of this study is to investigative the generalisability of CNNs for disease prediction. The study firstly examines the performance of two neural networks, namely, AlexNet and ResNet34 for disease prediction across three diseases, namely, brain tumour diagnosis, lung cancer detection and skin cancer diagnosis. The study also investigates an incremental neural network approach (INNA) for learning in neural networks when used for disease prediction. The INNA divides the dataset into easy, medium and hard using a difficulty estimation techniqu The performance of these CNNs and the INNA are evaluated independently for each of the three diseases examined, as well as across all three diseases. The study revealed that when evaluated separately for each disease AlexNet performed the best for brain tumour diagnosis and ResNet34 for lung cancer detection and skin cancer diagnosis. When evaluated across all three datasets for the different diseases, ResNet34 outperformed AlexNet and both CNNs with INNA. While the INNA did not outperform AlexNet or ResNet34, it had a lower computational cost and outperformed state of the art approaches for lung cancer detection.**

***Index Terms*—disease prediction, convolutional neural networks, generalisability**

I. Introduction

While convolutional neural networks(CNNs) have proven to be effective for disease prediction, the research done in this area applies and evaluates CNNs independently for disease diagnosis. This study forms part of a larger initiative aimed at improving the generalisability of CNNs for disease prediction. In addition the study also evaluates an incremental approach(INNA) in learning in CNNs for disease prediction to reduce computational cost and possibly improve accuracy. The motivation for the INNA is that if it learns on an easier dataset first this model will form the foundation for the datasets that are more challenging, at a medium level of difficulty and then at the hard level of difficulty. Thus there are two research problems in this study. The first is to investigate the generalisability of CNNs for disease prediction. The second is to investigate an incremental approach of learning in CNNs for disease prediction.

Three disease prediction problems are used for evaluation, namely, brain tumour diagnosis, lung cancer detection and skin cancer diagnosis. Both ResNet34 and AlexNet are used in this study as both these CNNs have performed well in previous studies for disease prediction from CT scans [1] [2].

The study revealed that ResNet34 performed the best across all three problems, outperforming existing state of the art approaches. While the INNA resulted in big improvements with regards to the decrease in computational cost, it did not outperform networks not using incremental learning. Hence, the main contributions of this study are:

- An investigation into the generalisability of CNNs for disease prediction produces results competitive to SOTA.
- An investigation into an incremental approach for learning in CNNs to reduce computational cost.

The following section provides a description of the three disease prediction problems and start of the art approaches for the corresponding datasets. Section III describes the INNA. The experimental setup used to evaluate the generalisability of the networks is presented in section IV. The performance of the networks is discussed in section V. The findings of the study and future research directions are presented in section VI.

II. Related Work

This section describes the three disease prediction problems that the INNA is evaluated on. The datasets used for each of these problems and state of the art approaches for these datasets are presented.

A. Brain Tumour Diagnosis

Brain tumour diagnosis essentially involves classifying images into a category corresponding to the type of tumour and is hence a multiclass image classification problem [3]. This study uses the Figshare brain tumour dataset [4]. The dataset is comprised of 3064 images divided into three classes, namely, 708 glioma images, 1426 meningioma images, and 930 pituitary images. The rest of this section provides an overview of state-of-the-art approaches applied to this dataset.

Abiwinada et al. [5] employed a CNN architecture consisting of 2 convolutional layers with a maxpool layer after each, a flatten layer, a fully connected layer, and finally an output layer with 3 output nodes. The ReLu activation function was used in the convolutional layers which were made up of 32 filters of 3x3 each. The maxpool kernel size was 2x2 and the fully

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00021

connected layer consisted of 64 neurons. The Adam optimizer was used due to its ability for handling noisy problems. The model achieved a training accuracy of 84.19% and a testing accuracy of 98.51%.

In the study conducted in [6] a variation of GoogleLeNet was used with transfer learning. The model consisted of 2 convolutional layers, 2 pooling layers, 1 fully connected layer and 9 inception layers each with 6 convolutional layers and 1 pooling layer. Filters of sizes 1x1, 2x2, and 3x3 were used. The Adam optimizer was used. 3 different classifiers were used, softmax, SVM, and KNN. The KNN and SVM classifiers achieved the highest accuracy of 98.0% and 97.8% respectively. The SVM classifier also obtained an average of 99.7% in correctly classifying the tumour into one of the three classes.

Rehman et al. [1] proposed a transfer learning model using 3 architectures of convolutional neural networks: AlexNet, GooLeNet, and VGGNet). The VGG16 model achieved the best accuracy of 98.69%. This model consisted of 16 convolutional layers, 3 fully connected layers with ReLu applied in each one, 5 max pooling layers with a kernel size of 2x2, and a softmax layer for output.

A CNN model consisting of 4 convolutional layers with ReLu applied after each one, 4 max pooling layers with kernel size of 2x2, a dropout layer before each max pooling layer, 2 fully connected layers, and a softmax layer, for output was proposed by Bad˘za and Barjaktarovi´c [7]. This model achieved an accuracy of 96.56%, however with only 4.3 million weights, it outperforms models such as VGG16 which has 138 million weights.

Chaki and Wo´zniak [8] proposed the Brain Tumour Segmentation and Classification Network (BTSCNet) to classify brain tumours into the 3 classes described above. The model includes 4 folds: segmentation of brain tumour region, ROI selection using morphological operation, feature extraction using multi-region gray level co-occurrence matrix and lastly a sliding window for classification. The model was able to achieve accuracies of 98.1%, 96.6%, and 95.3% in classifying glioma, meningioma, and pituitary tumours respectively.

EfficientNets and their use in multi-class brain tumour classification was studied by Zulfiqar, Bajwa, and Mehmood [9]. In the study, they found that fine tuning a pre-trained EfficientNet showed the best performance and achieved an accuracy of 98.86%. The proposed model is lightweight and computationally inexpensive.

Due to the small size of the dataset, Gupta et al. [10] proposed a method of using Cycle Generative Adversarial Networks to increase the dataset size. Their full method consisted of a modified InceptionResNetV2 pre-trained model for tumour detection and Random Forest Tree for classifying the tumour into one of the 3 classes. The model achieved an accuracy of 99% for tumour detection and 98% for tumour classification.

Sadad et al. [11] proposed a method of improving brain tumour classification accuracy by applying a contrast-stretching algorithm to obtain high resolution images and data augmentation such as rotating and flipping the images. The researchers made use of a UNet architecture with a backbone of ResNet50 for brain tumour detection and a NASNet model for classification. The model achieved an accuracy of 99.6% in brain tumour classification.

B. Lung Cancer Detection

Lung cancer detection involves determining whether a patient has cancer or not from an image, namely, a CT scan of the patients lungs [12]. Thus, this is a binary classification problem. The dataset used in this study for lung cancer detection is the 2017 Kaggle Data Science Bowl [13]. This section presents an overview of the state-of-the-art approaches for this dataset.

Chon et al. [12] used o a U-Net architecture proposed for image segmentation. The network took 2D images as input and output an image of 1's and 0's indicating whether the pixel contained a nodule or not. The study then looked at a "vanilla" 3D CNN model as well as a model built on the GoogLeNet architecture for classification. Both models made use of the Adam optimizer. The vanilla 3D CNN model achieved an accuracy of 70.5% whilst the GoogLeNet architecture achieved and accuracy of 75.1%. These low accuracies could be attributed to the subset of data that was used to train to models.

Alakwaa et al. [14] proposed the use of a 3D CNN for the detection of nodules in the CT scans. The U-Net architecture was used as a pre-processing step for the CNN. The U-Net model was pre-trained on the LUNA16 dataset. The goal of the U-Net model was to determine exactly where the nodules were if present. Classification was done using the 3D CNN architecture consisting of 2 convolutional layers with a ReLu activation function, a max pool layer after each convolutional layer, and finally 2 dense layers with one being a binary output layer. The Adam optimizer was used. The model achieved an accuracy of 86.6%.

Serj et al. [15] proposed a deep CNN architecture consisting of four convolutional layers which follow two max pooling layers, a full-body convolutional layer, and one fully connected layer. The ReLu activation function was used after each convolutional layer. The cross-entropy loss function was used in the training model. The model achieved a sensitivity of 87%, a specificity of 99%, and an F1 score of 95%.

Zhang et al. [16] proposed a 3D deep CNN model for classifiction of pulmonary nodules as malignant or benign. Segmentation of the image data was done beforehand to remove unimportant features such as bones and surrounding air. The contrast of the images was increased to highlight the lung tissue. Training was done in two phases. The first phase trains a nodule detection network whilst the second phase fine tunes the network. The model achieved a sensitivity of 84.4% and a specificity of 83%.

Vijh et al. [17] proposed a hybrid bio-inspired algorithm. The algorithm was built using the whale optimization algorithm and the adaptive particle swarm optimization. A CNN was then used for classification. The CNN consists of three densely connected layers with the ReLu activation function

after each layer. The images are preprocessed and segmented. The model achieved an accuracy of 97.18%, a sensitivity of 97%, and a specificity of 98.66%. Although the performance was quite good, the study only made use of 120 images from the dataset.

C. Skin Cancer Diagnosis

Skin cancer diagnosis is a multi-class classification problem involving classifying an image as one of the types of skin cancer [18]. The ISIC 2018 challenge HAM10000 dataset [18] is used in this study. State-of-the-art approaches that have been applied to this dataset are presented in this section.

Chaturvedi et al. [8] proposed an efficient skin cancer classification model. This model made use of a pre-trained MobileNet that was then fine tuned on the HAM10000 dataset. The model achieved an accuracy of 83.1%, a precision of 89%, a recall of 83%, and an f1-score of 83%. These metrics matched the performance of expert dermatologists.

Nugroho et al. [19] proposed a CNN architecture for classifying skin cancer images. The proposed model consisted of 4 convolutional layers with a ReLU activation layer after each one. Max-pooling with a kernel size of 2 was used after every 2 convolutional layers. The Adam optimizer was used with a learning rate of 0.001. The model achieved a testing accuracy of 78%.

Garg et al. [2] proposed their own CNN model for skin cancer classification. The model achieved a precision of 0.88, a recall of 0.74, and an F1-score of 0.77. The authors compared their model with ResNet and VGG16 models that made use of transfer learning. The ResNet model achieved the best accuracy of 90.5%

Gajera et al. [20] proposed a deep CNN model to classify the images in the HAM10000 dataset. Deep features were extracted from 8 different CNN models which were then fed into a group of classifiers for final classification. The model achieved an accuracy of 81% on the HAM10000 dataset.

Pai and Giridharan [21] proposed the use of a VGGNet for skin cancer classification. The proposed model was trained using the Adam optimizer and an initial learning rate of 0.001 which was reduced by a factor of 0.5 every 5 epochs. The model achieved an accuracy of 78%. The accuracy achieved could be attributed to the class imbalance in the data and could thus be increased by adding more data.

From the survey of the literature in this section it is evident that each study has focussed on just one disease and the studies have been done in isolation of each other. There appears to be no research on comparing the performance of CNNs across problems or assessing the generalisability of CNNs for disease prediction.

III. Incremental Neural Network Approach (INNA)

The INNA is based on the hypothesis that the neural network will be more accurate if it learns incrementally on subsets of the datasets rather than the entire dataset. The dataset is divided into easy, medium and hard and the neural network learns incrementally over these subsets of data. The overall approach is depicted in Algorithm 1:

Algorithm 1 INNA Algorihtm

```
Divide the data into n subsets based on difficulty
for i = 1 to n do
    if i=1 then
        Initialize weights to random values
    else
        Initialize weights to the best weights
        from iteration n − 1
    end if
end for
Evaluate the CNN from the last iteration on the testing set
```

The INNA firstly divides the data into subsets of data corresponding to different levels of difficulty. This is done using a probe network [22]. The probe network was chosen for difficulty estimation due to its low computational cost. The network consists of 3 convolutional layers and uses a ReLu activation function. The accuracy of the probe network is used as the measure of difficulty of the data. The probe network is used to calculate the difficulty for each image in the dataset. The images are then ranked based on the difficulty value and divided into n subsets according to difficulty. In this study n is 3 so each of the subsets corresponding to easy, medium and hard are allocated a third of the data instances in order of rank. The instances are split into thirds within each class to alleviate the potential problem of class imbalance. This ensures that the ratio of each class is maintained in each level of difficulty. The test set is created by randomly selecting an equal number of images of each difficulty level (i.e. easy, medium and hard). This is done to ensure that the test set accurately represents each of the difficulty levels.

Once the dataset is divided into subsets based on level of difficulty a CNN is trained on each subset iteratively. The initial values of weights of the CNN trained on the first subset are randomly generated. However, on subsequent iterations the best weights of the CNN from the previous iteration are used as the initial weights of the CNN for the current iteration. The CNN of the final iteration is then evaluated on the test set.

IV. Experimental Setup

A. Data Preprocessing

The images are resized according to the requirements of the model used. AlexNet requires images to be of size 227x227 whilst the ResNet34 requires images to be of size 224x224. The mean and standard deviation of the dataset is calculated and used to normalize each pixel value in the dataset to reduce computational cost. The dataset is divided into training, validation and test sets using the following ratio 70:10:20.

B. Experiments

The study involves two experiments:

- Experiment 1: Compares the performance of ResNet34, AlexNet without and with INNA for each of the problems

separately. The performances metrics in section IV-C are used to compare performance.

- Experiment 2: Compares the performance of ResNet34, AlexNet without and with INNA across all three problems. Formula 1 ranking, used to assess the performance of cross-domain hyper-heuristics [23], is used to assess performance. Each algorithm is assigned a rank based on its performance for each problem and the ranks are summed.

  Due to the stochastic nature of the approaches 30 runs, each using a different random number seed, is performed for ResNet34, AlexNet without and with INNA and performance is reported over these runs. Hypothesis tests using the Z statistic is used to test the statistical significance of the results.

C. Performance Metrics

The following metrics are used to assess the CNNs and INNA independently:

- Accuracy : (TP + TN)/(TP + FP + FN + TN)
- Precision: TP/(TP + FP)
- Recall: TP/(TP + FN)
- F1 Score: 2*(Recall * Precision)/(Recall + Precision)
- Specificity: TN/(TN + FP)

where:

TP: True Positive- Model correctly predicted disease; TN: True Negative - Model correctly predicted healthy; FP: False Positive - Model incorrectly predicted disease; FN: False Negative- Model incorrectly predicted healthy.

D. Problem Details and Parameters

The dataset details and parameters values for each of the problems is listed in Table I. The parameter values were determined by trial and error, testing different values for the number of epochs and learning rate.

TABLE I
PROBLEM PARAMETERS

	Brain Tumour Diagnosis	**Lung Cancer Detection**	**Skin Cancer Diagnosis**
Subset 1 size	715	1167	2336
Subset 2 size	715	1167	2336
Subset 3 size	716	1168	2339
Test set size	918	1498	3004
Number of classes	3	2	7
Number of epochs	30	30	30
Batch size	32	16	16
Learning rate	0.00005	0.00005	0.00005
Loss function	Cross entropy	Cross entropy	Cross entropy
Optimizer	Adam	Adam	Adam

E. Technical Specifications

Python was used to implement the CNNs and INNA. The computer used to run the models consisted of the following specifications:

- Ryzen 5 3600 6-Core CPU
- Nvidia RTX 3080 10GB GPU
- 16GB DDR4 RAM

V. Results and Discussion

This section compares the performance of ResNet34, AlexNet without and with INNA for brain tumour diagnosis, lung cancer detection and skin cancer diagnosis. Section V-A discusses the results for Experiment 1 and section V-B for Experiment 2. The performance comparison with state of the art approaches for the datasets is presented in section V-C.

A. Experiment 1 Results

This section discusses the performance of ResNet34, AlexNet and INNA evaluated independently for each problem. Table II presents the comparison of ResNet34 and AlexNet with and without the INNA for brain tumour diagnosis. AlexNet outperforms ResNet34 and both CNNs with INNA. These results are statistically significant at a 99% level of confidence.

TABLE II
PERFORMANCE COMPARISON FOR BRAIN TUMOUR DIAGNOSIS

	ResNet34	AlexNet	ResNet34+ INNA	AlexNet+ INNA
Accuracy	94%	**99%**	79%	91%
Precision	94%	**99%**	80%	90%
Recall	93%	**99%**	83%	91%
F1	93%	**99%**	82%	91%
Specificity	97%	**99%**	91%	96%

Table III lists the performance of the CNNs and INNA for lung cancer detection. ResNet34, AlexNet and AlexNet with INNA perform comparatively and there is no statistical significance in their performance. However, all three networks outperform AlexNet with INNA at the 99% confidence level.

TABLE III
PERFORMANCE COMPARISON FOR LUNG CANCER DETECTION

	ResNet34	AlexNet	ResNet34+ INNA	AlexNet+ INNA
Accuracy	**100%**	99.99%	99.33%	97.73%
Precision	**99%**	**99%**	**99%**	98%
Recall	**99%**	**99%**	**99%**	98%
F1	**99%**	**99%**	**99%**	98%
Specificity	**99%**	**99%**	**99%**	98%

As can be seen from Table IV ResNet34 outperforms the other networks for skin cancer diagnosis at the 99% confidence level. It is interesting to note that INNA did not perform well for this problem irrespective of the CNN used with it. Future work will investigate this further.

In terms of computational cost, the CNNs with INNA had a lower computational cost for all three problems as can be seen from the average runtimes (in seconds) in Table V.

TABLE IV
PERFORMANCE COMPARISON FOR SKIN CANCER DIAGNOSIS

	ResNet34	AlexNet	ResNet34+ INNA	AlexNet+ INNA
Accuracy	**96%**	95%	69%	72%
Precision	**91%**	**91%**	43%	48%
Recall	**91%**	90%	43%	49%
F1	**90%**	**90%**	43%	48%
Specificity	**99%**	97%	89%	88%

TABLE V
RUNTIME COMPARISON

	ResNet34	AlexNet	ResNet34+ INNA	AlexNet+ INNA
Brain tumour diagnosis	450	390	299	257
Lung cancer detection	1170	615	242	460
Skin cancer diagnosis	1710	1410	1178	913

B. Experiment 2 Results

This section compares the performance of ResNet34, AlexNet, ResNet34+INNA and AlexNet+INNA across the three problems. Based on the performance in terms of accuracy a rank is assigned to each network based on its performance on each of the three problems and the ranks are summed. Table VI lists the sum of ranks for each of the networks. ResNet34 performs the best over all three problems followed by AlexNet and the INNA networks.

TABLE VI
PERFORMANCE COMPARISON ACROSS PROBLEMS

Network	Rank Sum
ResNet34	4
AlexNet	5
ResNet34+INNA	10
AlexNet+INNA	11

C. Comparison with State of the Art Approaches (SOTA)

For completeness the performance of the CNNs with and without INNA are compared to state of the approaches. A description of these approaches are provided in section II. Table VII compares the performance with SOTA for brain tumour diagnosis. The AlexNet employed in this study outperforms all of the SOTA approaches except the approach employed by Sadad et al. [11].

From Table VIII it can be seen that the ResNet34 and AlexNet with and without the INNA outperform the SOTA approaches for lung cancer detection.

For skin cancer detection ResNet34 and AlexNet outperform the SOTA. The INNA did not perform well for this dataset and future research will investigate this further.

TABLE VII
BRAIN TUMOUR DIAGNOSIS COMPARISON WITH SOTA

Network	Accuracy
CNN [5]	98.51%
GoogleNet [6]	98.00%
VGG16 [1]	98.69%
CNN [7]	96.56%
EfficientNet [9]	98.86%
InceptionResNetV2+Random Forest [10]	98.00%
Unet+ResNet50+NASNet [11]	99.60%
ResNet34	93.93%
AlexNet	99.36%
ResNet34+INNA	79.19%
AlexNet+INNA	91.07%

TABLE VIII
LUNG DIAGNOSIS COMPARISON FOR SOTA

Network	Accuracy
GoogleNet [12]	75.10%
3D CNN [14]	86.60%
CNN [17]	97.18%
ResNet34	100%
AlexNet	99.99%
ResNet34+INNA	99.33%
AlexNet+INNA	97.73%

TABLE IX
SKIN CANCER DETECTION COMPARISON WITH SOTA

Network	Accuracy
MobileNet [8]	83.0%
CNN [19]	78.0%
CNN [20]	81.0%
CNN [2]	90.5%
VGGNet [21]	78.0%
ResNet34	95.8%
AlexNet	94.5%
ResNet34+INNA	72.3%
AlexNet+INNA	69.5%

VI. CONCLUSION

The main aim of the research presented in this paper was to examine the ability of CNNs to find acceptable solutions to more than one disease prediction problem. Three such problems, namely, brain tumour diagnosis, lung cancer detection and skin cancer detection, were used fin this study. The study firstly performed a comparison of networks applied individually to each problem. This revealed that different networks worked well for the different problems, with ResNet34 producing the best results for two of the problems. The study then examined how well the networks are able to perform over the three problems, using Formula 1 ranking to assess performance. ResNet34 was able to generalize better than the other networks.

The study also investigated an incremental approach for learning in CNNs. This approach when used with ResNet34 and AlexNet performed comparatively to the networks without incremental learning with a lower computational cost for brain tumour diagnosis and lung cancer detection. However, the

INNA did not work well for skin cancer diagnosis and future work will investigate the reasons for this. It is hypothesised that the performance is possibly related to the number of classes. The INNA performed well for binary classification and not as well for three classes. It performed the worse for skin cancer diagnosis which had seven classes.

For brain tumour diagnosis ResNet34 outperformed SOTA. For lung cancer both ResNet34 and AlexNet, with and without the INNA, outperformed SOTA. For skin cancer detection both ResNet34 and AlexNet outperformed SOTA. It can be seen from this study that some CNNs, like ResNet34, can generalise better than others. While the INNA did not outperform any of the networks without incremental learning, it had a much lower computational cost and produced results that outperformed the SOTA for lung cancer detection.

Future extensions of this research will include investigating other neural networks as well as ensemble learning for generalisability in CNNs for disease prediction. Furthermore a performance metric for evaluating networks for generalisability, similar to that derived for hyper-heuristics [24], will be investigated. Future work will also investigate improving the INNA by firstly investigating other methods for difficulty estimation. The use of ensembles, containing classifiers for the different levels of difficulty, will also be investigated.

REFERENCES

[1] A. Rehman, S. Naz, M. I. Razzak, F. Akram, and M. Imran, "A deep learning-based framework for automatic brain tumors classification using transfer learning," *Circuits Syst. Signal Process.*, vol. 39, no. 2, pp. 757–775, feb 2020. [Online]. Available: https://doi.org/10.1007/s00034-019-01246-3

[2] R. Garg, S. Maheshwari, and A. Shukla, *Decision Support System for Detection and Classification of Skin Cancer Using CNN*, 01 2021, pp. 578–586.

[3] A. Shafi, M. B. Rahman, T. Anwar, R. S. Halder, and H. E. Kays, "Classification of brain tumors and auto-immune disease using ensemble learning," *Informatics in Medicine Unlocked*, vol. 24, p. 100608, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352914821000988

[4] J. Cheng, "Brain tumor dataset," 2017. [Online]. Available: https://www.kaggle.com/datasets/mvhssriraj/brain-tumor-dataset

[5] N. Abiwinanda, M. Hanif, S. T. Hesaputra, A. Handayani, and T. R. Mengko, "Brain tumor classification using convolutional neural network," in *World Congress on Medical Physics and Biomedical Engineering 2018*, L. Lhotska, L. Sukupova, I. Lacković, and G. S. Ibbott, Eds. Singapore: Springer Nature Singapore, 2019, pp. 183–189.

[6] S. Deepak and P. Ameer, "Brain tumor classification using deep cnn features via transfer learning," *Computers in Biology and Medicine*, vol. 111, p. 103345, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482519302148

[7] M. M. Badža and M. Č. Barjaktarović, "Classification of brain tumors from mri images using a convolutional neural network," *Applied Sciences*, vol. 10, no. 6, 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/6/1999

[8] J. Chaki and M. Woźniak, "A deep learning based four-fold approach to classify brain mri: Btscnet," *Biomed. Signal Process. Control.*, vol. 85, p. 104902, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:257908680

[9] F. Zulfiqar, U. Ijaz Bajwa, and Y. Mehmood, "Multi-class classification of brain tumor types from mr images using efficientnets," *Biomedical Signal Processing and Control*, vol. 84, p. 104777, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809423002100

[10] R. K. Gupta, S. Bharti, N. Kunhare, Y. Sahu, and N. Pathik, "Brain tumor detection and classification using cycle generative adversarial networks," *Interdisciplinary Sciences, Computational Life Sciences*, vol. 14, no. 2, pp. 485–502, February 2022.

[11] T. Sadad, A. Rehman, A. Munir, T. Saba, U. Tariq, N. Ayesha, and R. Abbasi, "Brain tumor detection and multi-classification using advanced deep learning techniques," *Microscopy Research and Technique*, vol. 84, no. 6, pp. 1296–1308, 2021.

[12] A. Chon, N. Balachander, and P. Lu, "Deep convolutional neural networks for lung cancer detection," Department of Computer Science, Standford University, Tech. Rep., 2017.

[13] A. Buckeye, J. Kriss, J. BoozAllen, J. Sullivan, M. O'Connell, and N. W. Cukierski, "Data science bowl 2017," 2017. [Online]. Available: https://kaggle.com/competitions/data-science-bowl-2017

[14] W. Alakwaa, M. Nassef, and A. Badr, "Lung cancer detection and classification with 3d convolutional neural network (3d-cnn)," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 8, pp. 409–417, 2017.

[15] M. F. Serj, B. Lavi, G. Hoff, and D. P. Valls, "A deep convolutional neural network for lung cancer diagnostic," 2018, arXiv:1804.08170.

[16] C. Zhang, X. Sun, K. Dang, K. Li, X.-W. Guo, J. Chang, Z.-Q. Yu, F.-Y. Huang, Y.-S. Wu, Z. Liang, Z.-Y. Liu, X.-G. Zhang, X.-L. Gao, S.-H. Huang, J. Qin, W.-N. Feng, T. Zhou, Y.-B. Zhang, W.-J. Fang, M.-F. Zhao, X.-N. Yang, Q. Zhou, Y.-L. Wu, and W.-Z. Zhong, "Toward an expert level of lung cancer detection and classification using a deep convolutional neural network," *The oncologist*, vol. 24, no. 9, pp. 1159—1165, September 2019. [Online]. Available: https://europepmc.org/articles/PMC6738288

[17] S. Vijh, P. Gaurav, and H. M. Pandey, "Hybrid bio-inspired algorithm and convolutional neural network for automatic lung tumor detection," *Neural Computing and Applications*, vol. 35, no. 33, pp. 23 711–23 724, 2023. [Online]. Available: https://doi.org/10.1007/s00521-020-05362-z

[18] P. Tschandl, C. Rosendahl, and H. Kittler, "The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, no. 1, p. 180161, 2018. [Online]. Available: https://doi.org/10.1038/sdata.2018.161

[19] A. Nugroho, I. Slamet, and Sugiyanto, "Skins cancer identification system of haml0000 skin cancer dataset using convolutional neural network," vol. 2202, 12 2019, p. 020039.

[20] H. K. Gajera, D. R. Nayak, and M. A. Zaveri, "A comprehensive analysis of dermoscopy images for melanoma detection via deep cnn features," *Biomedical Signal Processing and Control*, vol. 79, p. 104186, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809422006401

[21] K. Pai and A. Giridharan, "Convolutional neural networks for classifying skin lesions," in *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, 2019, pp. 1794–1796.

[22] F. Scheidegger, R. Istrate, G. Mariani, L. Benini, C. Bekas, and C. Malossi, "Efficient image dataset classification difficulty estimation for predicting deep-learning accuracy," *The Visual Computer*, vol. 37, no. 6, pp. 1593–1610, 2021. [Online]. Available: https://doi.org/10.1007/s00371-020-01922-5

[23] E. K. Burke, M. Gendreau, M. Hyde, G. Kendall, B. McCollum, G. Ochoa, A. J. Parkes, and S. Petrovic, "The cross-domain heuristic search challenge – an international research competition," in *Learning and Intelligent Optimization*, C. A. C. Coello, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 631–634.

[24] N. Pillay and R. Qu, "Assessing hyper-heuristic performance," *Journal of the Operational Research Society*, vol. 72, no. 11, pp. 2503–2516, 2021. [Online]. Available: https://doi.org/10.1080/01605682.2020.1796538