0nÞÿÞŠ h P2 Šö

(ISSN 2784-0735)

Vol. 13

2024

Eximia Journal
Vol. 13, 183-193, January, 2024
ISSN: 2784-0735
www.eximiajournal.com

# Cancer Diseases Prediction Using Multiple Transfer Learning based on CNN Algorithm.

***Maysaa H. Abdulameer*** 1

*Ruwaida M. Yas* 2

*Huda H. Ali3*

*Huda W. Ahmed4*

[1,2,4](*Informatics Institute for Post-Graduation Studies, Iraqi Commission for Computer & Informatics, , Baghdad, Iraq*),
[3](Department of computer engineering techniques, Imam Al-Kadhum College).
Email: burdenmay2018@gmail.com

**Abstract**. For many years, the most common types of cancers that causes death has been lung and brain cancer. In addition, they are very hard to treat once they have been spread in the late stages. Early cancer detection and diagnosis have become crucial for providing immediate and efficient care to those who are afflicted with the disease. To predict cancer, this process needs cancer image classification which is a difficult task in the recent improvement of computer-based diagnostics.

The main objective of this research is to classify Computed-Tomography (CT) images of the lungs and brain as cancerous or not. Therefore, this research shows the possibility of using a convolutional-neural-network (CNN) for lung and brain cancer detection through transfer learning (TL) without the difficulty of specifying features for image-classification.

The proposed TCNN is trained and tested from scratch using two public and one private dataset using the TL technique. The suggested transfer learning-based model attained an accuracy of 92.81%, 97.21%, and 99.40% for Scratch TCNN Model, Pre-Trained TCNN Model, and Transfer Pre-Pre-Trained Model in that order. Hence, the proposed method gives high accuracy and decreases the time of diagnosis and treatment.

**Keywords**. Cancer_ Detection, Transfer_ Learning, C.N.N Algorithm, Lung And Brain Tumors Datasets

Eximia Journal
Vol. 13, 183-193, January, 2024
ISSN: 2784-0735
www.eximiajournal.com

## 1. Introduction

Cancer is a intricate and deadly illness that still presents many obstacles to diagnosis and treatment. It is brought on by irregular, uncontrolled cell growth. [1]. With 1.8 million deaths annually, cancer is the leading cause of death, according to the World Health Organization. By permitting promote intervention and individualized treatment plans, early detection and accurate can-cer-prediction can significantly improve patient out-comes [2].

In order to complete the mission-critical medical diagno-sis, a traditional machine-learning model is trained on a dataset, which frequently experiences a bottleneck re-garding excellent model performance (e.g., sensitivity, specificity, and accuracy). Large-scale datasets might also be unavailable for some applications, making it difficult to train a deep-learning-based model that is accurate. Recent years have seen a remarkable potential for machine-learning-techniques, especially deep-learning-algorithms, in a number of medical do-mains, including cancer-prediction [3] [4].

Transfer-learning is a new research trend that involves moving knowledge from the source domain. These fac-tors are driving this trend. The literature has amply demonstrated the superiority and value of trans-fer-learning in a variety of research applications [5]. A machine-learning technique known as transfer-learning uses learning model developed for one learning task to serve as the foundation for a new learning model devel-oped for a different learning task. it is a field of study in machine-learning that focuses on applying knowledge gained from solving one problem to another that is relat-ed but unrelated [6].

Various methods for the categorization of lung and brain tumor cancer diseases have been designed. The authors [7] developed new method by combining the traditional supervised-learning with integrated-learning by using transfer-learning. Two datasets are used The Can-cer_Genome_Atlas (TCGA), and the Clini-cal_Proteomic_Tumor Analysis Consortium (CPTAC) to decrease pathologists' manual annotation requirements while maintaining accuracy. First, the adenocarcinoma or squamous cell carcinoma status through whole-slide images (WSI) is established. Next, the subtypes of the cancer are further identified by developing weak classi-fiers, and finally, the final classifier is created using in-tegrated learning. The authors achieved a good accuracy value of 96% for lung cancer dataset. In [8] proposed new method that utilizes a deep neural network (D.N.N) as a feature extraction strategy in a computer_aided diagnosis (CAD) system to aid in the detection of lung ailments with high resolution. The proposed method is divided into three stages: first, performing data augmen-tation, followed by classification using the pre-trained C.N.N model, and finally, localization. This method achieved a good accuracy of 97.2% for lung cancer only. The proposed work [9] examines the feasibility of utilizing a convolutional neural network (C.N.N) to identify lung cancer. The suggested network recognizes lung cancer tissues automatically via transfer learning, eliminating the need to define criteria for image categorization. The research includes lung C.T pictures from 12 male and 4 female participants ranging in age from 35 to 77 years. The images are in Bitmap-image-format and have a resolution of 512 512. Two instances, one healthy and one cancerous, are studied. Two cancer cases are included: a poorly differentiated carcinoma and a moderately differentiat-ed carcinoma. 800 lung CT images are included as a training data set and 100 images as a testing data set. This work reach accuracy of 93%. Deep-transfer-learning which is used in the proposed classification system. In [10], which extracts features from brain M.R.I images using a pre-trained GoogLe-Net. To categorize the retrieved features, proven classifier models are incorporated. The experiment uses a patient_level five_fold cross_validation technique on a figshare M.R.I dataset. The suggested work gave a classification accuracy of 98%.

The authors in [11] have suggested a new method for detecting brain tumors that combines a convolutional neural network with a transfer learning strategy and the dimensionality reduction method. To demonstrate the efficiency of the proposed model, a comparative study of multiple models of transfer learning with and without dimension reduction approaches is given. The authors achieve, with their proposed model, a 97.14% accuracy rate. In [12] the authors focus on the multi-class classi-fication of brain tumors in this research because binary classification has received a lot of attention. They studied the performance of multiple deep-learning (D.L) architectures, including Visual_Geometry_Group 16 (V.G.G16), InceptionV3, V.G.G19, ResNet50, Incep-tionResNetV2, and Xception, to detect malignancies quicker, more unbiasedly, and reliably. They present a multi-class classification model based on transfer learn-ing (T.L) named I.V.X16, which is built on all three of the best-performing TL models. Additionally, they utilize Explainable AI for performance evaluation and validity of each DL model, as well as to incorporate recently announced Vison-Transformer (ViT) models and compare their results to the TL and ensemble models. Additionally, they utilize Explainable AI for the perfor-mance evaluation and validation of each DL model, as well as to incorporate recently announced Vi-son-Transformer (ViT) models and compare their results to the TL and ensemble models. they utilize a dataset that has 3264 images in total. they reach peak accuracy of 95.11%, 93.88%, 94.19%, 93.88%, 93.58%, 94.5%, and 96.94% for V.G.G16, InceptionV3, V.G.G19, ResNet50, Inception_ResNetV2, Xception, and I.V.X16, respectively, through extensive tests. In [13] the researchers examined previously trained deep transfer learning techniques such as ResNet50, ResNet101, V.G.G16, and VGG19 for diagnosing breast cancer by using a dataset of 2453 histopathology pictures. The images in the data set were divided into two separate groups: those with and those without invasive-ductal-carcinoma (IDC). The researchers found that ResNet50 outperformed other models after analyzing the transfer learning model, achieving accuracy rates of 90.2%, Area under the curve (A.U.C) rates of 90.0%, recall rates of 94.7%, and a marginal loss of 3.5. In [14] the authors presented a new model of deep transfer learning for faster detection of brain tumors using MR imaging. For identifying brain tumors, the new model is built using five prominent deep learning architectures which are Xception, Dense.Net201, Dense.Net121, Res.Net152V2, and Inception.ResNetV2. To increase classification accuracy, the last layer of these designs has been updated using the proposed deep dense block and softmax-layer as the output-layer. It obtains 99.67% classification accuracy on the 3-class dataset and 95.87% accuracy on the 4-class dataset.

In all previous works, different machine learning meth-ods and models have been proposed and implemented in this direction for predicting brain tumor or lung cancer. Every existing model performs well in various situations and different datasets. Despite the fact that no model can accurately predict lung cancer and brain tumors together, thus, in this work, a new deep-learning model for effective brain and lung cancer prediction with high accuracy is proposed.

The primary objective of this study is to investigate the efficacy of multiple transfer learning approaches in lung and brain cancer disease prediction using a CNN-based framework. By harnessing the power of transfer learning, we aim to overcome the limitations of limited labeled data in medical imaging, which often hinders the de-velopment of accurate and robust cancer prediction models. By leveraging the pre-trained models' learned representations, we can effectively transfer knowledge from relevant domains to enhance the performance of the prediction model. Main contributions to the work is as follows:

- Created a model that uses the TL process to categorize the extent of lung and brain cancer in patients, making it the first model of its kind to be used with these datasets.
- Using different epochs to learn the suggested CNN from scratch on the Google Colab platform, where it was used the multi-TL methods in this case.

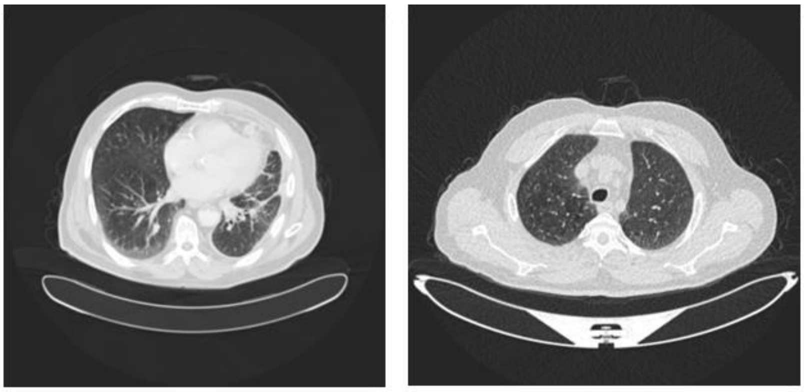
The remainder of the research is structured as follows: Section 2 discusses the background and related work. Section 3 describes the proposed system's approach, including a full description of each phase. Section 4 pre-sents the results with analysis, and it also includes the comparison research. Finally, section 5 comprises the concluding remarks.

## 2. The Proposed Method

The methodology of the proposed research depends three types of dataset trained using proposed neural network with two stages and as explain in the following sections

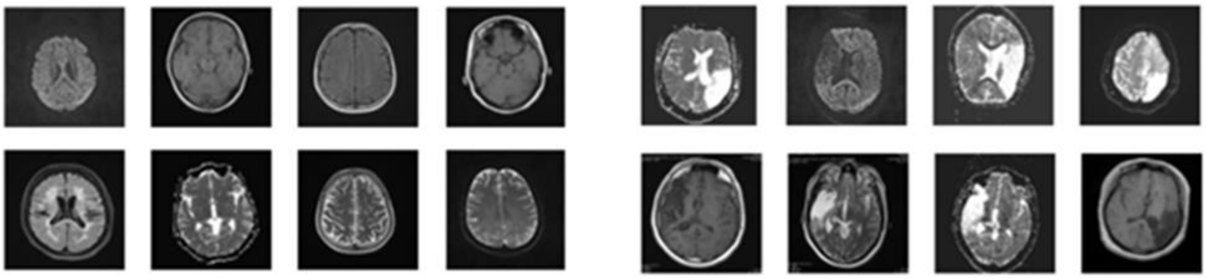
*2.1 Datasets Description*

The lung_cancer_dataset from Iraq-Oncology Teaching Hospital – National_Center for Cancer_ Diseases (IQ-OTH/NCCD) was gathered in 2019. It comprises images from CT scans of patients with lung cancer at various stages as well as healthy participants. The dataset comprises 1190 high-quality JPG pictures (512X512) depicting CT scan slices from 110 patients, which are classified as normal, benign, or malignant. Gender, age, educational_achievement, location-of-residence, and living_situation differ across the 110 cases. Figure (1) presents two classes of lung cancer.



Figuer 1. CT scans of lung cancer

Other public dataset is Brain Tumors image dataset, the best technique to detect brain tumors is Magnetic_Resonance_Imaging (M.R.I). The scan can created a huge amount of image data. This dataset containing 7022 M.R.I images 533 patients with two kinds of brain tumor: brain Tumor image (3622), and no tumor image (3400 slices). A further private cancer dataset is obtained from the Medical City Department / Oncology Teaching Hospital in Baghdad, Iraq. The dataset consists of 30-40 patients from January 2022 to the end of June 2022. The MRI machine generates 80-100 pictures for a full human head scan, selecting between top and side views. Only around 15-20 of these photos are appropriate for processing. Thus, 500 photos were gathered and divided into two folders: 250 in the tumor folder and 250 in the normal folder. Figure (2) shows samples of Real MRIs tumor and non-tumor images.



(a) (b)

Fig. 2 – CT scans of brain tumor, (a) Non-tumor, (b) tumor

*2.2 Method description*

The proposed methodology, which consists of two stages, depends on using different disease datasets to train the proposed CNN using transfer learning, the suggested datasets are Brain Tumor, and Lung Cancer datasets.The block diagram of the proposed model is shown in figure (3). It contains two stages

Eximia Journal
Vol. 13, 183-193, January, 2024
ISSN: 2784-0735
www.eximiajournal.com

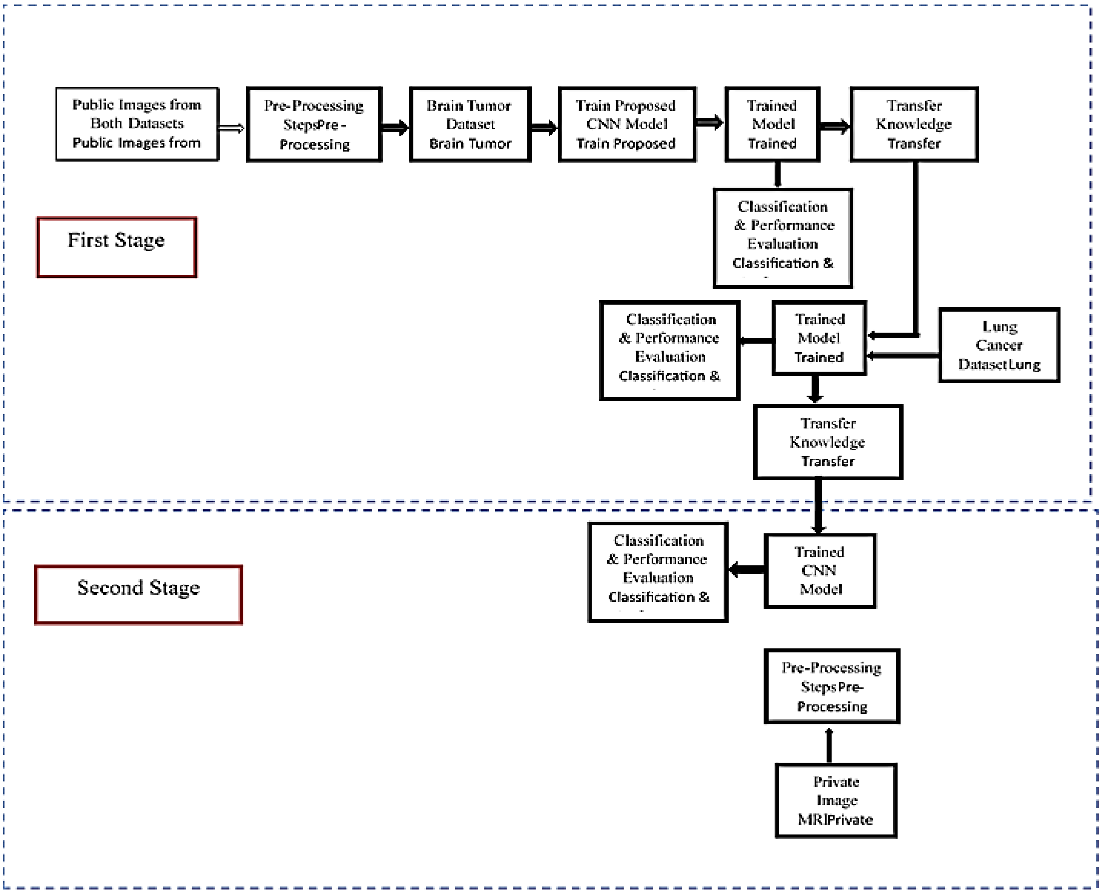



Figure 3. The proposed model

*2.2.1 First Stage*

Its consists of two main steps, First Step, Pre-processing, this stage works to improve the CT scan images of the lung and the brain datasets to improve the efficiency and accuracy of lung and brain cancer diagnosis. Figure 2 shows the preprocessing steps:

1-Image resizing:
Image resizing is crucial because it attempts to standardize the image's dimensions. The M.R.I images of very different sizes, required a change in image size in order to properly implement the process. This number was approved based on an experiment on different sizes and their impact on accuracy, as shown in figure (4) for both public datasets. The original image size was 250 * 250.

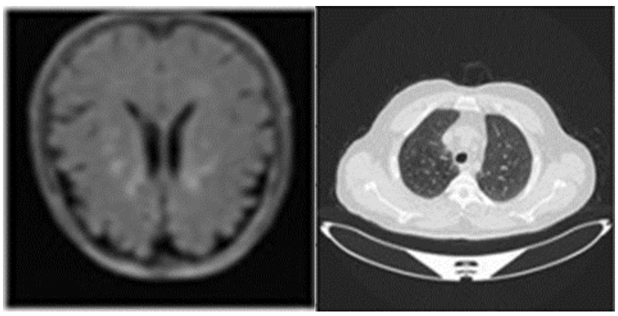


Figure (4) Resize images datasets

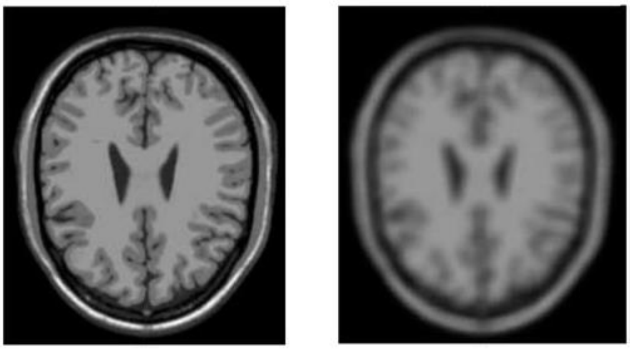
Eximia Journal
Vol. 13, 183-193, January, 2024
ISSN: 2784-0735
www.eximiajournal.com

2- Image enhancement (filtering):

Gaussian blurring is the technique to smooth, reduce the noise, and enhance the characteristics of the image [15]. Gaussian kernels consist of values that are higher in the middle and drop off towards the outer edges of the square array. Equation (1) defines the 2D Gaussian function [16] [17].

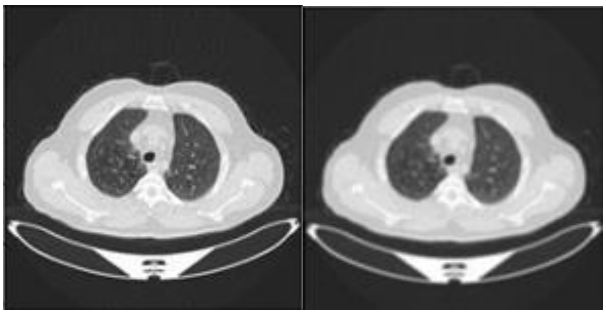
$$\text{Gauss}\ (x, y) = \frac{1}{2n\sigma^2}\, e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad \ldots\ldots. (1)$$

After applying the filter on the datasets images the result is shown as in Figure (5) and figure (6) :



(a) (b)

Figure (5): (a) Original image; (b) Gaussian Blur



(a) (b)

Figure (6): (a) Original image; (b) Gaussian Blur

3-image normalization

A statistical technique called normalization is used to scale data of various sizes. It limits the data to a normalized range and improves their precision while ignoring changes in their magnitude [18]. Normalization aims to convert the dataset's numerical attributes' values to a standard scale while maintaining the ranges' disparities [17]. Therefore, wise to normalize the pixel values using Min-Max normalization, to give each one a value between 0 and 1 using equation (2).

$$X_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad \ldots\ldots\ldots\ldots. (2)$$

Where $X_{max}$ and $X_{min}$ are expressed as pixel values of 255 and 0, $X_{norm}$ is the normalized images [19], figure (7) explain the normalization process on the dataset images:

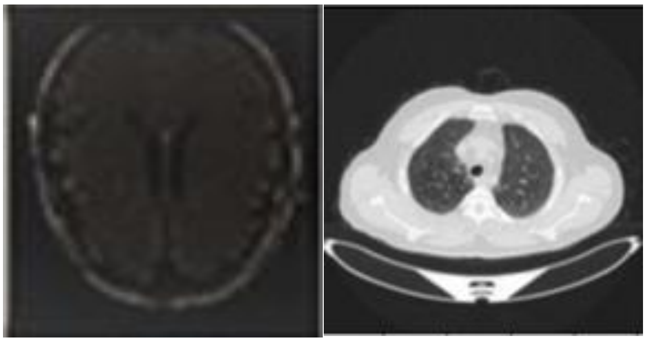


Figure (7) Normalize images

Second Step: Training: the major goal of the model modification is to make an excellent choice for detecting brain cancer and lung cancer. The model's ability to extract information from images is improved by increasing the number of layers and the filter size on each layer. The most major regularization techniques have been applied to deal with the overfitting issue brought on by the depth of the model. Training the proposed CNN algorithm using transfer learning techniques. This step responsible of training the preprocessed images by splitting or divided them into a training (60%), validation (20%) and testing (20%) sets. The architecture of the proposed CNN explained as in figure (8) below:

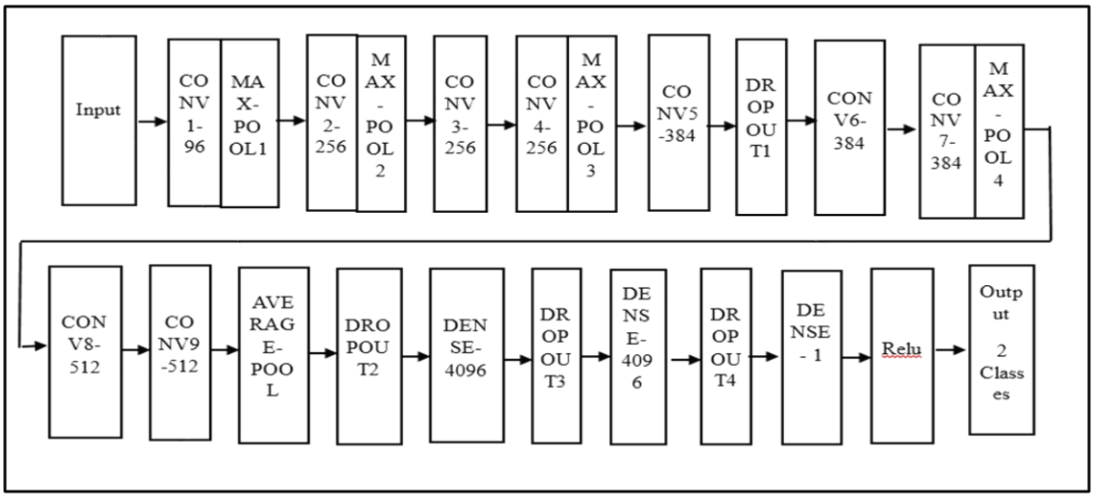



Figure (8) the Proposed Deep CNN Network

The above figure explain the proposed CNN neural network trained from scratch with the first public dataset brain tumor then transfer the knowledge to train the other public dataset lung cancer then the proposed CNN will be ready to train the private dataset with transferring the knowledge from the trained of the two previous datasets which means second stage.

### *2.2.2 Second Stage*

This stage depend on train the private dataset with using the pervious knowledge through the trained CNN network, the third dataset is small (750 MRI images) for lung and brain cancers. All the dataset

Eximia Journal
Vol. 13, 183-193, January, 2024
ISSN: 2784-0735
www.eximiajournal.com

images will be preprocessed before it used in the training step these processes as mentioned before (image resizing, enhancement, and normalization), then the dataset is split into two sets. The first is used to fit the model and consists of 80% of the entire dataset, including 600 images (Training Set). The remaining 150 images are for testing the model (the Testing Set). Table (1) details the training and testing of MRI image data.

Table (1) Private Dataset splitting Information

| Data subset | Number of images | Yes | No | Image percentage |
|---|---|---|---|---|
| **Training** | 600 | 299 | 301 | 80% |
| **Testing** | 150 | 81 | 69 | 20% |
| **Total** | **750** | **380** | **370** | **100%** |

Real brain and lung MRI data were obtained from the local medical center in DICOM format, and once the image format had been changed, it was stored as JPG digital images.

## 3. RESULTS AND DISCUSSION

The proposed CNN used the brain tumor dataset of 7022 MRI images, training from scratch using 60% for training, 20% for validation, and 20% for testing. All variables are summarized in Table (2).

Table 2 The Proposed Variable for Brain Tumor

| Hyper parameters | Optimizer | Learning rate | Decay | Epoch | Batch size | Random state | Seed | Loss |
|---|---|---|---|---|---|---|---|---|
| **values** | Adam | 0.0001 | 1e-4 | 100 | 32 | 0 | 324 | cross-entropy |

The classification accuracy results for the proposed CNN was 92.46%, which is included in figure (9) Prediction confusion matrix for the proposed model of brain tumor

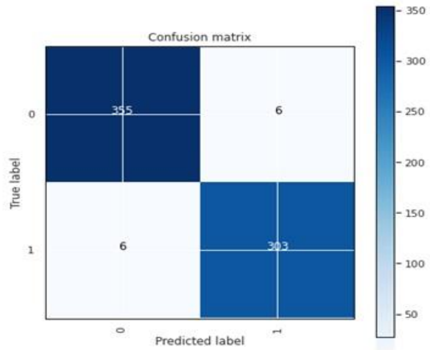



Figure (9) Prediction confusion matrix (Brain Tumor)

The knowledge of the trained model will be transfer to the next level to train the lung cancer dataset using the hyper parameter as in table (3). The classification accuracy results for the proposed CNN was 97.46% which is included in figure (10) Prediction confusion matrix for the proposed model for lung cancer

Eximia Journal
Vol. 13, 183-193, January, 2024
ISSN: 2784-0735
www.eximiajournal.com

Table3 The Proposed Variable for Lung Cancer

| Hyper parameters | Optimizer | Learning rate | Decay | Epoch | Batch size | Random state | Seed | Loss |
|---|---|---|---|---|---|---|---|---|
| **values** | Adam | 0.001 | 1e-4 | 100 | 16 | 0 | 324 | cross-entropy |

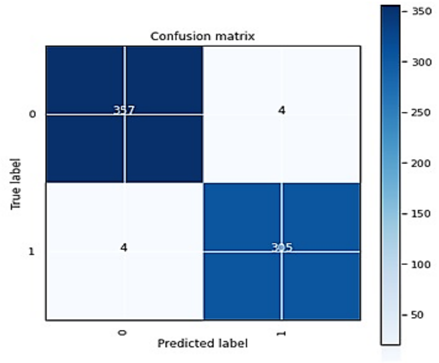



Figure (10) Prediction confusion matrix (Lung Cancer)

Figure (11)-(a) below Shows the accuracy obtained when the developed model is applied to the training and validation sets. The best score was observed when the model was pre-trained, obtaining %97.90 for training accuracy and 97.40% for validation. Moreover, figure (11)-(b) Represents loss values obtained from the pre-trained model with a training loss obtained is 0.0833 and a validation loss of 0.0918. The same, when pplied to the testing set, gives a 0.0892 loss value.

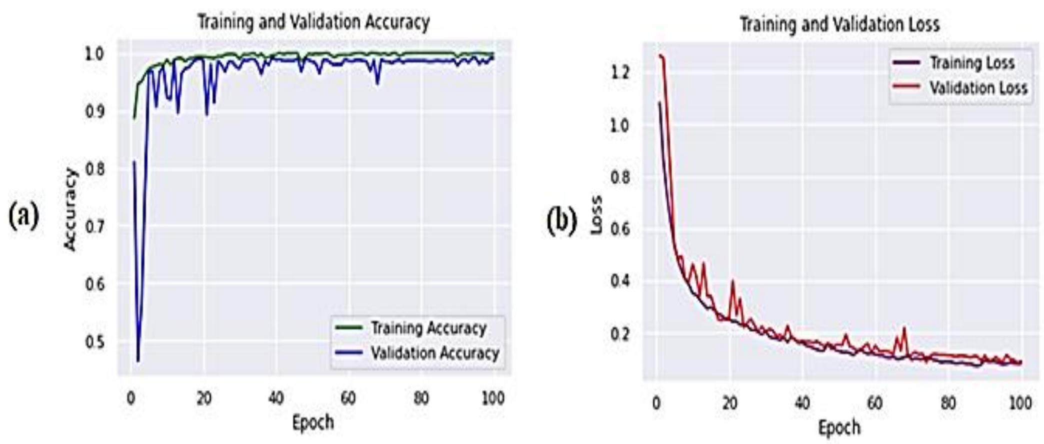



Figure (11): (a) The constructed paradigm that distinguishes between training and validation accuracy in 100 epochs (b) The comparison between the two values of training and validation loss

Using the same architecture of the proposed trained CNN model with its knowledge for training and classifying the two public datasets will be transferred to classify the new private datasets to be more accurate in detecting the tumor for both lung and brain images, the hyper parameter of the final trained model as in table (4):

Table 4. The Hyper Parameters of the Trained Model

| Hyper parameters | Optimizer | Learning rate | Decay | Epoch | Batch size | Random state | Seed | Loss |
|---|---|---|---|---|---|---|---|---|
| **values** | Adam | 0.0003 | 1e-2 | 50 | 16 | 0 | 442 | Binary cross-entropy |

Eximia Journal
Vol. 13, 183-193, January, 2024
ISSN: 2784-0735
www.eximiajournal.com

Figure (12) shows, the model was trained for 50 epochs with 25 iterations per epoch. The best batch size value was 16. The loss value was significantly reduced after using the value of weight decay. It can also be observed, Loss and Accuracy values began off at a reasonable level, then reached a minimum error value, indicating the model's strong ability to learn, which is a slight improvement over the prior model's initial low-value level before gradually improving.

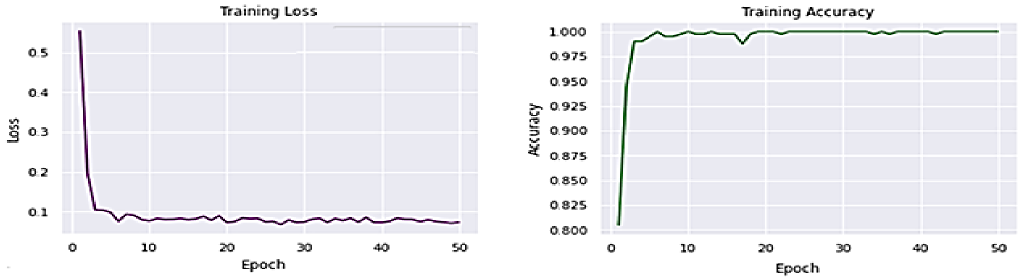



Figure (12): The Designed Paradigm for Training accuracy and Loss value after 50 Epochs

The final transferred CNN model can recognize the pattern and predict new images. So, in figure (13), the final model was tested using the testing set consisting of 100 comparable MRIs. The pre-pre-trained model was successful in producing 99% accurate predictions. This model achieves 98.07%, 100%, and 99.02% for precision, recall, and F1 score.

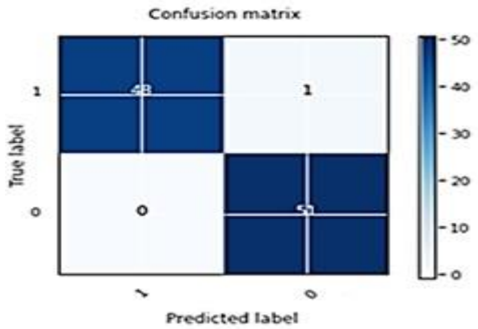



Figure (13): The Testing Set Results for the pre-pre-trained Model

In table (5) which explain the results of all the trained datasets used in the proposed model:

Table 5. The Final Results of the Trained Model

| CNN Model | Dataset Types | Precision | Recall | F1 score | Validation accuracy | Testing accuracy |
|---|---|---|---|---|---|---|
| **Scratch CNN Model** | Brain tumor | 0.9270 | 0.9270 | 0.9270 | 92.81% | 92.81% |
| **Pre-Trained CNN Model** | Lung cancer | 0.9705 | 0.9705 | 0.9705 | 97.36% | 97.21% |
| **Transfer Pre-Pre-Trained Model** | Private images | 0.9967 | 0.9902 | 99.35 | 99.10% | 99.40% |

From the table above which explain the best result calssification with the pre-pre-trained CNN model, which gave more accurate results through using the transfer learning technique.

## 4. CONCLUSION

Lung and brain tumors are among the major causes of death globally. Early and precise detection of these tumors can enhance therapy results and survival rates significantly. The study's purpose was to accurately and efficiently diagnose lung and brain cancer. For this identification, transfer learning is used on tw [20] [20]o datasets (public and private). The public dataset contains 1190 images for lung and 7022 images for brain which used for training the proposed model. The private dataset which contain 500 images for both lung and brain is used testing.

Our proposed scratch CNN model accuracy was 92.81%, the pre-trained CNN model accuracy was 97.21% , after applying the proposed transfer pre-pre-trained model, the accuracy was significantly enhanced and reached 99.40%

The proposed TL model not only exceeded current methods for lung and brain cancer diagnosis in terms of accuracy, but it also saved time and money. We believe that our proposed technique may be used to effectively diagnose various illnesses.

**REFERENCES**

[1] P. A. S. K. &. S. I. M. Dutta, "Multi-classification of Brain Tumor Images Using Transfer Learning Based Deep Neural Network," *International Conference on Artificial Intelligence for Smart Community,* pp. 927-933, 2020.

[2] K. T. a. G. B. B. a. J. R. H. a. C. H. R. a. A. V. a. A. A. a. N. A. Chui, "Multiround transfer learning and modified generative adversarial network for lung cancer detection," *International Journal of Intelligent Systems,* vol. 2023, pp. 1--14, 2023.

[3] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering,* vol. 22, no. 10, pp. 1345-1359, 2010.

[4] K. K. T. M. W. D. Weiss, "A survey of transfer learning," *Journal of Big Data,* vol. 3, no. 1, p. 9, 2016.

[5] F. a. Q. Z. a. D. K. a. X. D. a. Z. Y. a. Z. H. a. X. H. a. H. Q. Zhuang, "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE,* vol. 109, no. 1, pp. 43-76, 2021.

[6] S. Bozinovski, "Reminder of the first paper on transfer learning in neural networks," *Informatica,* vol. 44, no. 3, pp. 1854-3871, 2020.

[7] B. K. J. J. A. M. N. Y. Z. Y. M. L. Z. W. L. a. I. S. Qiao, "A novel transfer-learning based physician-level general and subtype classifier for non-small cell lung cancer," *Heliyon,* 2022.

[8] S. R. A. S. J. N. Humayun M, "A Transfer Learning Approach with a Convolutional Neural Network for the Classification of Lung Carcinoma," *Healthcare (Basel),* vol. 6, no. 1058, p. 10, 2022 Jun 8.

[9] K. P. T. M. A. D. Lakshmi, "Convolutional neural network in the detection of lung carcinoma using transfer learning approach," *International journal of imaging systems and technology,* vol. 30, no. 2, pp. 445-454, 2020.

[10] S. A. P. Deepak, "Brain tumor classification using deep CNN features via transfer learning," *Computers in biology and medicine,* vol. 111, p. 103345, 2019.

[11] S. V. Priyanka Modiya, "Brain Tumor Detection Using Transfer Learning with Dimensionality," *International Journal of,* vol. 10, no. 2, pp. 201-206, 2022.

[12] S. Hossain, A. Chakrabarty, T. R. Gadekallu, M. Alazab and M. J. Piran, "Vision Transformers, Ensemble Model, and Transfer Learning Leveraging Explainable AI for Brain Tumor Detection and Classification," *IEEE Journal of Biomedical and Health Informatics,* 2023.

[13] M. J. Piran, "Vision transformers, ensemble model, and transfer learning leveraging explainable ai for brain tumor detection and classification," *IEEE Journal of Biomedical and Health Informatics,* 2023.

[14] S. Z. M. T. F. &. Z. Y. Asif, "An enhanced deep learning method for multi-class brain tumor classification using deep transfer learning.," *Multimedia Tools and Applications,* pp. 1-28, 2023.

[15] K. a. Y. M. R. A. a. A. F. H. Kusrini, "The effect of Gaussian filter and data preprocessing on the classification of Punakawan puppet images with the convolutional neural network algorithm," *International Journal of Electrical and Computer Engineering,* vol. 12, no. 4, p. 3752, 2022.

[16] B. a. P. M. a. N. K. K. Desai, "Study on Image Filtering--Techniques, Algorithm and Applications," *arXiv preprint arXiv:2207.06481,* 2022.

[17] I. Petras, "Novel low-pass filter with adjustable parameters of\~{} exponential-type forgetting," *arXiv preprint arXiv:2210.01195,* 2022.

[18] S. a. K. B. a. E. Y. Yaman, "A novel normalization algorithm to facilitate pre-assessment of Covid-19 disease by improving accuracy of CNN and its FPGA implementation," *Evolving Systems,* vol. 14, no. 4, pp. 581--591, 2023.

[19] Y. a. H. W. a. G. W. a. S. H. Tao, "Research on Ndvi Normalization Method Based on gf Images," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* vol. 3, pp. 209--215, 2022.