**Leveraging Residual and Separable Convolutions for Brain Tumor MRI Analysis: A Hybrid Representation with Explainable AI**

**ABSTRACT**

This study proposes a model for volume detection and classification of brain tumors in MRI images based on the ResNet101 and Xception models. This model implements and achieves the integrated use of residuals convolutions and separable convolutions with self-attention and parallel model integration. The model produced classification accuracies of 95.5%, 98.1%, and 99.6% respectively. The dataset used contains 3000 MRI images from which two classes were obtained one containing 1500 images that exhibited brain tumors and a second one containing 1500 images without been divided into 2400 training images and 600 testing images to maintain balanced data per class. To improve model interpretability, the study integrated explainable AI (XAI) methodologies which included Local Interpretable Model-agnostic Explanations (LIME) and Ablation Class Activation Mapping (Ablation-CAM) that aid in understanding model interpretability. These methods add transparency, which is essential for medical applications of AI. Combining both approaches allows for better performance than individual models. This study provides an insight of how advanced deep learning models can be applied for the detection of brain tumors with an important consideration on explainable deep learning approaches that are instrumental for clinical use. Further work should center on applying the model to bigger pools of images and other imaging techniques to assess its robustness and clinical usefulness.

**Keywords:** AI in healthcare, brain tumor detection, feature fusion techniques, medical imaging analysis, model transparency in healthcare AI, explainable AI, self-attention mechanism, hybrid deep learning frameworks, depth wise separable convolutions, parallel model integration.

1. **INTRODUCTION**

Despite advancements, the timely and accurate detection and grading of brain tumors remain a clinical challenge with high implications for patients’ outcomes and survival. Neurooncological tumors are a heterogeneous group of tumors which require different prognostic and treatment strategies and therefore need to be identified accurately for proper treatment. MRI examinations cum radiologists’ experience have been the backbone of diagnostic and classification methods as they are known today. But these methods are also very time consuming, error-prone, and limited by tumor type and stage[1].

The brain is surrounded by a skull which is spherical in shape although it is undeveloped. It is quite possible to understand how profound changes in this very delicate space can come about due to just a slight increase in volume in the said spaces. The pathological mass which is in any of these cavities poses dangers to the anatomical components of the intracranial cavity and at most times, this may lead to loss of life. Today, these tumors are clear in the US population for both young and adults aged 30-60 years. It is worthy of note that they are among the most prevalent malignant cell types including brain and other neuroepithelial malignancies, contributing significantly to the top ten cancer death statistics. The prognosis for these patients is extremely poor and is determined by the type of the tumor, the form or the location of the tumor among many others. Brain tumor is a condition affecting 700,000 individuals in the United States now, with over 80% cases benign and about 20% cases malignant. As per the projection of the American Cancer Society in 2021, more than 78089 cases of brain tumor were reported with 55150 being benign and around 24900 being malignant out of which approximately 13840 were males and about 10690 were females. Studies have documented brain tumor as the most severe cancer type and comes second after lung malignancy [2].

Today’s medicine makes use of medical images when diagnosing different conditions as well as for education and research purposes. The market of medical images is ever on the increase, thus, for instance, the radiology unit of the University Hospital in Geneva created between 12000 – 15000 images per day in 2002. Hence, such developments call for the use of an efficient and suitable computer-assisted diagnosis system with applications to diagnosis images and compiling medical reports. Other direct disadvantages also come in addition to these methods of looking at people such as age estimation which is subjective to human influences and are quite laborious. Machine Learning in these days is also growing for diagnosis and treatment purposes. Many studies that focus on developing models specifically for brain tumor detection and classification have been reported to achieve great levels of accuracy and low levels of errors. More specifically, the convolutional neural networks technique makes use of already established models such as GoogLeNet, AlexNet, ResNet-34 for medical image classification. Applications of this nature incorporate deep multilayered neural networks to enhance the precision of the processes performed during the medical diagnosis [2].

The paper begins with literature review of existing work which helped in exploring the methodologies and stating the research problem of this paper. Furthermore, the paper covers the recommended approaches to achieve desired results whose performances have been evaluated in results and discussion. Finally, this paper outlines the conclusions drawn from the research and the future scope of the project.

1. **RELATED WORK**

Deep learning systems put some ease on the task involving the diagnostics classification by letting the deep neural networks learn the features and patterns of the datasets on their own. The growing use of deep learning as an approach for classification problems has an endorsement of powerful computing hardware architectures, particularly GPUs and TPUs. Also, some other free-source coding frameworks like Keras with TensorFlow backend and PyTorch are contributing to this process. In addition, a variety of such neural network deep architectures as VGG, AlexNet, Inception, ResNet, Xception and other models are available for incorporation in deep computer vision applications in other [3].

MRI images aided in the construction of an advance classification which supports in recognizing brain tumor. Tumor magnitudes such as VGG16, ResNet50, DenseNet and VGG19 enabled the use of transfer learning so as to efficiently identify most typical cases of brain tumors. The Deep Transfer Learning models were trained and validated on 3,000 MRI scans available in the Figshare data repository. It achieved 99.02% accuracy which is very favourable when comparing it with results obtained from ResNet50[4].

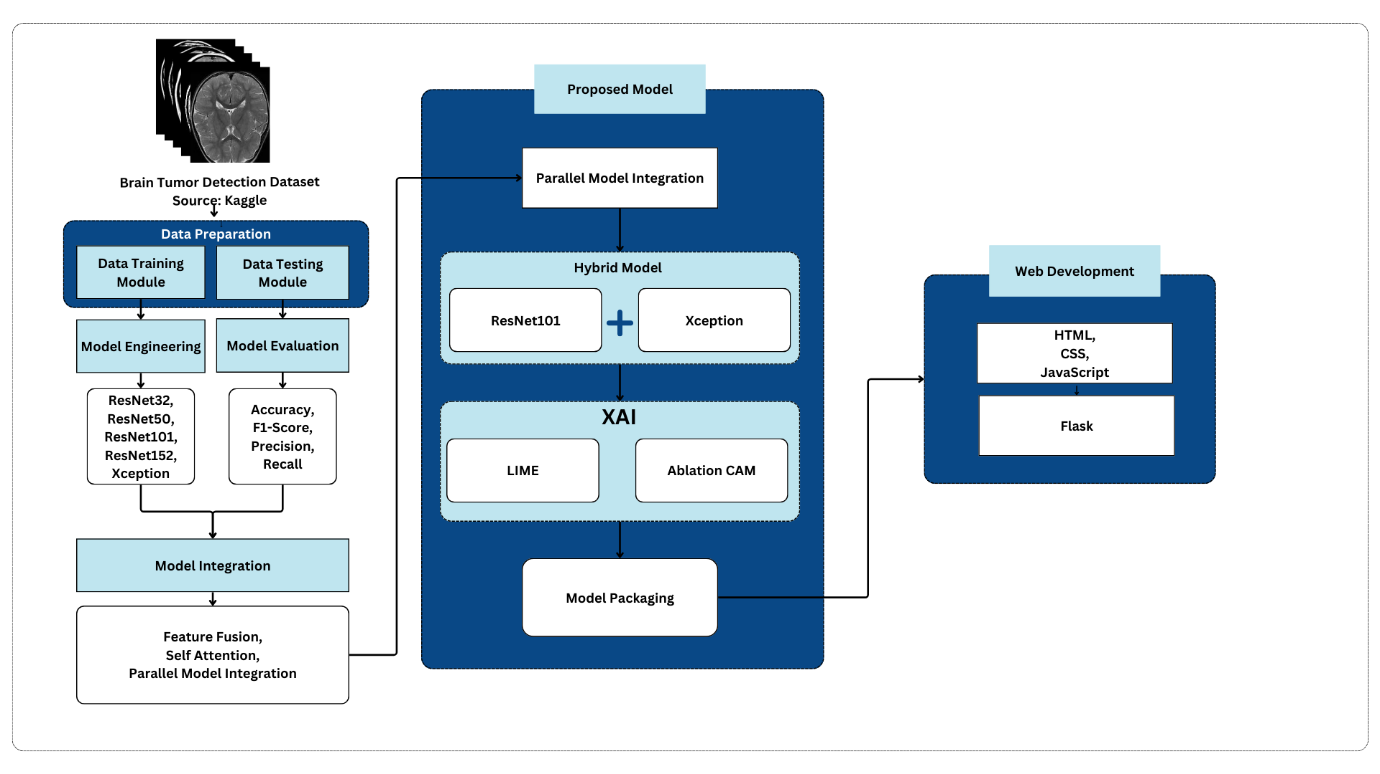
The ongoing increased adoption of deep learning (DL) approaches, and most specifically convolutional neural networks (CNNs) has made many contributions in medical image interpretation. In contrast to the traditional machine learning (ML) techniques, CNNs employ a top-down feature learning strategy which analyses images in their natural form with no input on feature design. This ability has tremendously enhanced the accuracy and efficiency of brain tumor detection systems. Most of the studies have concluded that deep learning algorithms outperform classical machine learning algorithms particularly in image analysis involving complex spatial features and tumor shape deformation. However, despite these achievements, the application of DL into real world clinical applications is hampered by the fact that large annotated data sets, high performance computational resources and model explainability among others are required[1].

As a starting point, the authors proposed a deep wavelet autoencoder-based DNN for extracting high-level features from brain tumor images. The images were segmented so that different regions could be processed through the deep wavelet autoencoder. The processed brain tumor images were further analyzed with a deep CNN. In this case, a few other classifiers were compared with the deep autoencoder classifier, which was shown to be the most reliable as well as efficient with an accuracy of 96%[5].

A DNN network with an auto-encoding module was further used to classify brain tumor in the next study. The images were pre segmented and features were extracted from the images while running through the DNN layers. Intensity based features and textures were obtained and captured by using GLCM and DWT respectively. Finally, two modules of autoencoder were included in the architecture of DNN and a softmax layer was also included to classify[5].

In this research paper we will be recommending an approach which will solve the problem of lack of explainability and interpretability of the models to ameliorate the trust on AI by using explainable AI on the model’s prediction which will provide assisstance in retraining and enhance the results which the above related work have not addressed. Also, to abstain from any kind of bias in model’s predictions, a balanced dataset is used.

1. **METHODOLOGY**



**Figure 1: Project Roadmap**

* 1. **DATASET DESCRIPTION**

The dataset used is Magnetic Resonance Imaging (MRI) images which has two classes yes and no which contain images of tumored and non-tumored brain MRI images.Figure 1 depicts the distribution of the dataset which is important to train model without allowing any bias to occur in the model’s training. Followed by this are the sample images from both the classes which are in Figure 2 and Figure 3.

A close-up of a brain scan

Description automatically generatedA close-up of a brain scan

Description automatically generated

**Figure 2: Sample images of tumored MRI images**

A close-up of a brain scan

Description automatically generatedA close-up of a brain scan

Description automatically generatedA close-up of a brain scan

Description automatically generated

**Figure 3: Sample images of non-tumored MRI images**

The dataset is made up of 3000 MRI images that have been divided into two classes: 1500 tumorous images and 1500 non-tumorous ones. It is also further divided into training and testing sets, and each set is divided into two subsets likely referred to as “Yes” and “No” for classes representative of the respective sets. 2400 images are included in training set and the classes supporting these images are 1200 for both classes. 600 images also support two classes in the testing set with 300 for each set. The design of the dataset is critical for ensuring maximum model efficiency as well as improving its strength.

* 1. **MODEL ARCHITECTURE**

Deep learning refers to a method that seeks a hierarchy of features from a given input through a neural network with several layers. It is a popular method since, in contrast to traditional machine learning techniques where features of an image are hard coded, this one does it on its own. A variety of different deep-learning algorithms have been used in order to achieve algorithms based on various objectives[6]. It is important to explain that such architectures as ResNet101 or Xception, for example, are deep and good at capturing complex features so that they are needed for complicated work such as processing of medical images.

**ResNet101**

A diagram of a diagram

Description automatically generated

**Figure 4: Architecture of ResNet101**

ResNet101 was created in a way which solves the problem of rapid decay of gradients in deep networks. In deep networks, this is a common problem. This is made possible by its highly innovative structure of residual learning which incorporates identity mappings through skip connections and thus, allows gradients to flow freely during the backpropagation cycle. The model consists of 101 such convolutional layers and thus, its depth is such that it can capture the hierarchical features in a very complex way. This makes the model useful in capturing nuances and outliers such as brain tumor segmentation. In addition, ResNet101 employs a design strategy in which computational resource efficiency and representational efficacy are integrated since bottleneck structures are deployed to reduce complexity without impacting performance. Its pre-trained versions come from image data sets such as the ImageNet which empowers it high functional features which are important in many tasks. All these trait’s of depth, residual learning, and transferability makes ResNet101 fit for medical imaging-related researches where precision and scale of implementation is required[7].

**MATHEMATICAL CONCEPTS BEHIND ResNet101**

ResNet101 is a state-of-the-art variant of convolutional neural network(CNN) which implements the idea of residual mapping instead of direct mapping. The formulae for residual mapping is as follows:

: Output of the residual block

: Input to the block

: Residual function parameterized by weights

Batch normalization normalizes activations to improve convergence.

The key restructuring done in ResNet101 is the addition of skip connections, which add the input *x* directly to the output of the residual block. The formulae for skip connection are as follows:

ResNet101 uses ReLU activation function which is represented as given below:

The convolutional layers are used to extract features from the input data. The convolutional operation is:

**Xception**

**A diagram of multiple cubes

Description automatically generated**

**Figure 5: Architecture of Xception or Extreme Inception**

Considered an extreme perfection in design and computation, the Xception model (of the extreme inception architecture) is best known for the feature of depthwise separable convolutions for feature extraction where it makes the most impact. Using this advanced convolutional mechanism, the inputs information in the spatial and channel wise pieces are compressed into a variety of operations which classify and considerably reduce the amount of compute requirements while preserving the depth of representation. Xception on the other hand, replaces standard convolutional layers with depthwise and pointwise convolutional layers and this way it provides the opportunity for the model to learn even more complex patterns without any flattery[8]. Its distinct advantage is that from large dimensional datasets it can obtain detailed information thus making it optimal for brain tumor segmentation and other domains that require detail. Further, its application simply entails transfer learning as it is already trained on a large data set similar to that of the ImageNet data set. The reasons which make Xception the model of choice for complex image analysis problems especially those involved in pattern recognition include a compromise between computational efficiency and model interpretability as well as better performance in differential features[3].

**MATHEMATICAL CONCEPTS BEHIND Xception**

Xception is an improvised version of Inception-V3 model with a slight enhancement in how the network processes a data.

Xception utilizes the concept of depthwise seperable convolutions instead of standard convolutions which applies the following formulae:

Instead of fully connected layers using generally, Xception uses Global Average Pooling to reduce each feature map:

Xception utilizes residual mapping to collect features in a better way which is represented mathematically by this formulae:

**MODEL INTEGRATION**

While inspired by ResNet101 and Xception architectural concepts, the synergy stems from harnessing the strengths of both models and addressing the deficiencies in the two models. The reasons for the strategies of combining these two architectures, ResNet101 and Xception, are rather synergetic in nature with respect to the fact that the models complement each other in the given process. Practically, the architecture of ResNet101 in association with its architecture that contains numerous layers and residual connections is outstanding for capturing hierarchical and wide features while Xception is able to efficiently capture channel dependent and spatial features with inbuilt depthwise separable convolutions. The strength of these approaches is that they allow for wider and more comprehensive imaging, accounting for both global and key local features which enable the tumor to be more evidently illustrated. Moreover, the combination employs the mechanism lowering the probability of overfitting – the efficiency of Xception’s parameters compensates for the given depth of the ResNet101 thus allowing better performance with new data. Such a strategy improves diagnostic accuracy and robustness of the model as it takes advantage of the ensemble-like trait of the combination thus enhancing the performance of the model under noise and distortions of medical images.

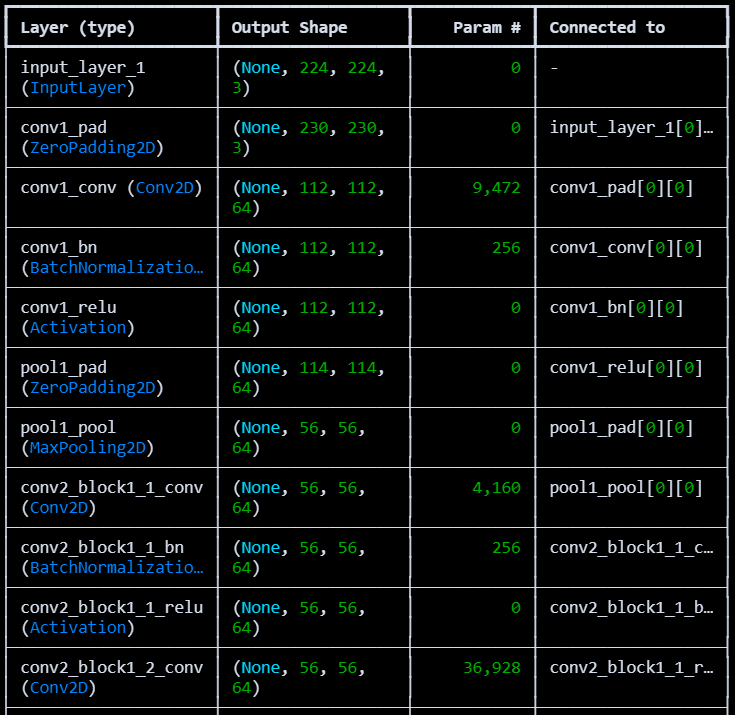
In addition, brain tumor identification requires the ability to detect both large-scale features, such as the locating of the tumor, and small scale features such as rough textures which the model built is able to do more efficiently than the use of individual models. The proof of this effectiveness is also performance metrics improvement, say accuracy and precision, which are performance indicators this integration approach clearly demonstrated its advantage in the ability to extract and leverage multiple distinct features that are needed to address the task of the correct tumor classification. This all encompassing approach makes it possible for the integrated model, which is particularly useful for brain tumor detection, as significantly greater than that of ResNet101 or Xception alone.

We have implemented three different techniques to integrate ResNet101 and Xception based on the model’s architecture which are as follows:

1. Feature Fusion Technique

The Feature Fusion approach consists of integrating more than one feature representation model to heighten the performance of the model over the different parameters of the input data. In this case, various models are built to view the image separately; thus, every model can capture different input image or target features. When the singular models get their features independently, they usually get fused together during the confinement stage or blended to create one single unit that is more robust. This fusion makes it possible for the model to incorporate the most functional model components, allowing for a more effective feature set that captures both essential and intricate aspects of the image. By fusing multiple models, the prediction errors location, generalizability and precision of the model is successfully enhanced by the feature fusion in a way that prevents feature extracting errors.

For circumstances like image recognition, this method comes in quite handy for the reason that it allows performing difficult tasks such as distinguishing whether a brain had (or did not have) tissue cancer.



The adopted model utilizes Feature Fusion technique incorporating ResNet101 and Xception for binary brain tumor classification. To determine the model architecture, the authors first performed data generation methods by preprocessing the input images, where images are resized to 224 x 224 and the pixel values are normalized. A common input layer ensures that identical images are provided to both ResNet101 and Xception, both initialized with imagenet weights, which are the two pre-trained models' architectures. These models are loaded without their top layers so that new features can be learned, and they are bottlenecked to protect their pre-trained information. A global average pooling was applied to all models for an increase of their activation as well as reduce the spatial dimensions of the models to compact feature vectors. These vectors are joined together into a single representation, which allows ResNet101 to extract features at different levels while Xception captures spatial and channel-wise features in a more effective manner. The integrated features were crunched down into a Simplified Dense layer with ‘sigmoid’ to yield a score that can predict whether tumor is present or not. The model was compiled with Adam, binary cross entropy, and accuracy as the performance measure. This architecture combines the advantages of ResNet101 and Xception which creates a strong framework to enhance the brain tumor detection.

1. Self Attention Mechanism

The Self – Attention Mechanism technique has the potential to improve a model by emphasizing the most salient parts of an input image through assigning importance to certain features. This technique therefore allows the model to dynamically adjust the weighting of a feature in terms of its relevance to all other features and look for significant areas or patterns within the image to make the required predictions accurately. Self-attention processes the input by first generating query, key, and value representations from the input feature map and use them to compute attention scores. These scores suggest the relative need of the strength of focus on a feature as compared other features. The sounds have predictive implications therefore these predictive implications are used to alter the feature map. Features which are irrelevant or not predictive are suppressed. Such a mechanism assists the model in capturing the long range interaction as well the contextual relationship among the features within the image, important for recognition of both local and global structures such as even small or faint elements in medical images. With self – attention it is expected that the model will perform better for classifying images as it will concentrate more on the essential components of the image’s predicted ones yielding more accurate and robust predictions.

A screenshot of a computer program

Description automatically generated

The picture detection model presented above employs Self-Attention Mechanism technique to use the best features of ResNet101 and Xception for brain tumor identification. Using Tensorflow’s image\_dataset\_from\_directory feature, the pictures are loaded, resized to 224 × 224 and training and validation sets created. Both the ResNet101 and Xception models have been loaded with Imagnet pre-trained weights; however, their top layers have been removed in order to allow feature extraction. To further ensure weight updating within training does not affect the models, these models are also frozen. Global Average Pooling is used to pool the feature maps obtained from both models and they are integrated to create a joined feature vector. To these combined features, self attention is added so the model can focus on important parts of the image. The attention layer is outputted then flattened and passed along with a dense layer for a final time where the sigmoid activation function provides the model eye for probability of the picture containing a brain tumor. This model is trained using the Adam optimizer with binary cross entropy loss on the given dataset for 20 epochs. The architecture uses the unique advantages of both ResNet101 and Xception but it is the self attention mechanism that augments the representation of the features hence enabling the model achieve better performance on classification of brain tumor.

1. Parallel Model Integration

The Parallel Model Integration technique consists of a set of models that operate on the input data simultaneously and each one of them is responsible for a different characteristic of the image. Pivotal to this direction is the reason as to why this direction appears to be reasonable at first that different models could view similar data differently and hence there’s heterogeneity that could be missed by just one model only. After performing the transformations on the input made by all the prevailing models, their features are usually combined to form a single representation of the features, being concatenation in most cases. This set of features and relevant attributes are then used in the final classification or regression task, which is the task under concern. However, because of the parallel processing, the models are said to have performed better as they have enjoyed the strengths of individual models leading to better comprehension of complexity and hence generalization. Parallel model systems that incorporate the models’ outputs help to enhance the system’s tolerance to data variations and the risk of error, bias, or limitation of a single model working on its own. Such Progression of Integrated Parallel Modeling is beneficial especially in the Classification act as it improves the accuracy of the models by focusing on the same techniques.

A screenshot of a computer

Description automatically generated

The Parallel Model Integration technique in the code above integrates two models RezNet101 and Xception, where input image is passed through both models first then its feature outputs are combined, by merging the features K Bangar on both. In this case, the two models which are both trained separately from both models trained images on ImageNet, were able to concentrate on learning different perspectives of the same data. ResNet101's features were the ones with deep structures which help efficiently capture hierarchical patterns while Xception features applied depthwise separable convolutions for good fast feature learning then merged through the Concatenate layer. Since the combination of the two model's features offers different perspectives, it helps the model provide a better and more robust representation of the input image by leveraging different architectures. The features were then pushed through a final sigmoid layer that gave a probability score that further improves featured overall extraction and enhances the models ability to perform the task of classifying once brain tumors better than either of the models alone.

**RESULTS AND DISCUSSIONS**

**REFERENCES**

[1] G. S. Rajput, K. Kumar Baraskar, S. Telang, M. Ingle, J. Surana, and D. Padma, “Brain Tumour Detection and Multi-Classification Using Advanced Deep Learning Techniques,” 2024.

[2] V. K. Dhakshnamurthy, M. Govindan, K. Sreerangan, M. D. Nagarajan, and A. Thomas, “Brain Tumor Detection and Classification Using Transfer Learning Models,” *Engineering Proceedings*, vol. 62, no. 1, 2024, doi: 10.3390/engproc2024062001.

[3] S. A. Joshi, A. M. Bongale, P. O. Olsson, S. Urolagin, D. Dharrao, and A. Bongale, “Enhanced Pre-Trained Xception Model Transfer Learned for Breast Cancer Detection,” *Computation*, vol. 11, no. 3, Mar. 2023, doi: 10.3390/computation11030059.

[4] H. ZainEldin *et al.*, “Brain Tumor Detection and Classification Using Deep Learning and Sine-Cosine Fitness Grey Wolf Optimization,” *Bioengineering*, vol. 10, no. 1, Jan. 2023, doi: 10.3390/bioengineering10010018.

[5] N. Noreen, S. Palaniappan, A. Qayyum, I. Ahmad, and M. O. Alassafi, “Brain Tumor Classification Based on Fine-Tuned Models and the Ensemble Method,” *Computers, Materials and Continua*, vol. 67, no. 3, pp. 3967–3982, Mar. 2021, doi: 10.32604/cmc.2021.014158.

[6] S. Krishnapriya and Y. Karuna, “Pre-trained deep learning models for brain MRI image classification,” *Front Hum Neurosci*, vol. 17, 2023, doi: 10.3389/fnhum.2023.1150120.

[7] S. Nawaz *et al.*, “Deep Learning ResNet101 Deep Features of Portable Chest X-Ray Accurately Classify COVID-19 Lung Infection,” *Computers, Materials and Continua*, vol. 75, no. 3, pp. 5213–5228, 2023, doi: 10.32604/cmc.2023.037543.

[8] B. Gülmez, “A novel deep neural network model based Xception and genetic algorithm for detection of COVID-19 from X-ray images,” *Ann Oper Res*, vol. 328, no. 1, pp. 617–641, Sep. 2023, doi: 10.1007/s10479-022-05151-y.