

HYBRID INCEPTIONV3-CNN-RNN FRAMEWORK FOR MULTI-DISEASE CLASSIFICATION USING CHEST X-RAY AND BRAIN MRI IMAGES

KANIKA PAHWA¹, RANJIT SINGH², SARUCHI³

¹Research Scholar, Chandigarh University, Department of Computer Science and Engineering, India

²PROFESSOR, Chandigarh University, Department of Computer Science and Engineering, India

³ASSISTANT PROFESSOR, Chandigarh University, Department of Computer Science and Engineering, India

E-mail: ¹kanikapahwa86@gmail.com, ²ranjit.bhatthal@gmail.com, ³ganpati.saruchi@gmail.com

ABSTRACT

The increasing application of automated medical image-based analysis has underscored the need to come up with effective multi-disease diagnostic systems capable of processing heterogeneous sources of data. The authors present a Hybrid Inception V3-CNN-RNN deep learning network for the classification of a chest X-ray and a brain MRI image in this study. The model integrates spatial feature extraction with gated recurrent feature aggregation to enhance diagnostic accuracy across imaging modalities. The provided methodology consists of the complicated preprocessing pipeline, including the image normalization, the artificial noise simulation and its removal, and the region-of-interest (ROI) segmentation with the assistance of the K-means clustering algorithm. Deep feature extraction is also done with a pre-trained InceptionV3 network to extract rich hierarchical representations, and then Chi-square-based feature ranking is evaluated to identify a compact subset of discriminative InceptionV3 features, with the number of retained features determined through validation-based ablation analysis. The selected features are reshaped into a fixed representation and processed by the hybrid CNN-RNN architecture. The convolutional layers extract local patterns from the grouped feature representation, whereas the recurrent layer performs gated nonlinear aggregation across the feature groups. Experimental assessment of seven diagnostic classes indicates that the presented model has better classification performance than the case of standard CNN and fine-tuned Inception V3 models, with an accuracy of 95%, a recall of 94.37%, a F1-score of 94.68%, and a precision of 95.15%. Combining CNN and RNN blocks enables local feature refinement and gated integration of information across the selected feature groups, which supports better generalization and resistance to noise. The findings support the idea that the hybrid deep learning (DL) model has validated to be effective in addressing challenges of single-modality diagnostic tools and real-life multi-disease conditions and offers a steady diagnostic accuracy across the domains of both chest and brain imaging. The article reports the importance of integrating spatial deep feature extraction with gated recurrent feature aggregation for medical image classification. The future research will be on the concept of multimodal data addition, enhanced explainability, and clinical metadata addition to facilitate useful implementation in healthcare decision-support systems.

Keywords: *Multi-disease Classification, Medical Image Analysis, InceptionV3, Hybrid CNN-RNN, Deep learning*

1 INTRODUCTION

The continuous progress of AI and ML technologies has significantly changed the sphere of medical diagnostics, in particular, the automated processing of medical images. The conventional diagnostic processes may require radiologists who might be subjective in their interpretation of such

work because of workload, fatigue, or subliminal variances in the presentation of the disease. As a consequence, the Computer-aided diagnostic (CAD) models have become useful tools of clinical decision-making support and patient care efficiency [1]. Specifically, Magnetic Resonance Imaging (MRI) and chest X-rays (CXR) are considered as some of the main imaging modalities that help

identify neurological and respiratory disease, which makes the image-based classification systems as automated and accurate as possible.

Deep Learning (DL) methods, especially CNN, have indicated high potential in modelling hierarchical and spatial features of medical images. They have been established to be efficient in the classification of brain tumor where segmentation and localization have an important effect on the analysis and planning of treatment [2]. Studies conducted with a full convolutional architecture have made it possible to detect tumors with better accuracy and perform more region-specific analysis of MRI images [3]. In the same way, the incorporation of optimised deep neural networks has led to an improvement of precision in classifications and uncertainty in diagnosis when analysing brain tumor MRI images [4]. The above developments point out that deep feature extraction methods will continue to play a central role in the creation of effective multimodal medical diagnosis systems.

Nevertheless, the current literature is based on single-domain disease classification, i.e., the brain tumor diagnosis by MRI or the lung disease diagnosis based on the CXR images, without considering the problem of creating a more general framework that can process multiple disease types with heterogeneous imaging domains. The classification of multiple diseases cannot be done correctly without effective spatial feature extraction, but also the capability of capturing sequential or contextual trends in the transformed feature representations. Moreover, the growing popularity of ANNs in healthcare decision making highlights the necessity to create interpretable and flexible models that can help to increase the clinical reliability [5].

Therefore, Hybrid InceptionV3-CNN-RNN multi-classification of CXR and brain MRI images is recommended in this paper. This framework employs well-organised preprocessing, region-of-interest segmentation, deep feature extraction, and gated recurrent feature aggregation to improve diagnostic performance and generalisation. The model addresses the constraints of single-architecture classifiers by integrating spatial feature extraction, convolutional refinement, and gated recurrent aggregation for image-based medical diagnosis.

The sections that follow have been arranged in the following manner to help in the flow of the research. Section 2 offers an overall literature review of medical image analysis and hybrid DL

technologies that apply to the classification of various diseases. Section 3 includes an in-depth description of the proposed research work, preprocessing, feature extraction, feature selection, and the Hybrid InceptionV3-CNN-RNN model design. Section 4 offers the experimental outcomes and assesses the efficiency of the model based on many diagnostic indicators, comparative analysis, and visual analysis, highlighting the advantages, implications, and limitations of the proposed paradigm in Section 5, which is the Discussion. The study's result is presented in Section 6, and it provides recommendations for future research and further improvements.

2 LITERATURE REVIEW

2.1 Category 1: Brain MRI-Based Studies

The article by Sanjay Kumar et al. (2018) [2] discussed the importance of strong, precise, and effective segmentation in the early detection of head lumps. It suggests a convolution-based method of skin tone capture and describes the complicated learning procedure of continuous pixels and semantic images. The study addresses the usage of a bounce structural chart in the collection of semantic requirements, including deep, impolite deposits to low and well-coated layers. This format was used to assess the prediction performance of a brain tumor segmentation 3D full CNN developed by the scientists. The 3D convolutions were handled using UNET architecture to support the large memory requirements, and the images were cropped. The research highlights the possibility of better early head lump detection and analysis with post-processing, which smooths out model-produced segmentations.

According to Gokila Brindha et al. (2021) [3], abnormal cell growths in the brain, which can develop into cancer, are identified as brain tumors. MRI is the most common method of detection as it provides information about aberrant tissue growth. The research articles make use of DL and ML techniques to identify brain tumors. The proposed research will increase the accuracy of the treatment and decision-making procedure of radiologists through the application of Artificial Neural Networks (ANN) and CNN to quickly identify cancerous brain tissues based on MRI images. The self-defined algorithms contribute to the diagnosis of brain cancers and are constantly enhanced with the help of performance analysis.

Jaya Lakshmi Machiraju et al. (2022) [4] suggested a DL-based model of brain tumor classification based on MRI images. Inception-V3

CNN architecture was used in the study to extract the multi-level features to detect malignant and benign brain tumors at an early stage. The Adam optimizer and a softmax classifier were used to optimize the hyperparameters to predict multiple classes. The model has a classification accuracy of 99.34 on training data and 89 on validation data, indicating that the model is very reliable when it comes to MRI-based tumor recognition.

Yakub Bhanothu et al. (2020) [6] described Medical care for brain tumors as necessary depending on the kind and location of these dangerous disorders. MRI pictures are necessary for the diagnosis of malignant cancers, but they take time and require specialized knowledge. In response, a Faster R-CNN DL method was developed to use a Region Proposal Network (RPN) to identify and indicate the tumor's appearance. This method focuses on three primary brain tumors. This approach focuses on identifying three tumor categories: pituitary, meningioma, and glioma. Utilizing the VGG-16 model for classification and feature extraction, the technique recorded average precision scores of 75.18% for glioma, 89.45% for meningioma, and 68.18% for pituitary tumors. The model yielded a mean average precision of 77.60% across the three classes.

Zhiguan Huang et al. (2020) [7] discussed the brain tumor classification utilizing MRI images are demonstrated using a CNN built on complicated networks. To achieve a more precise diagnosis of brain tumors, a network generator maps arbitrarily generated graph algorithms into a computable neural network, which creates the network structure. In comparison to the well-established CNN families ResNet, DenseNet, and MobileNet models, the updated CNN built on a complicated network model reduces test loss and outperforms them with a brain tumor classification accuracy of 95.49%. In addition to improving neural network construction methodology, our model produces promising results for brain tumor image classification.

Himanshu Das et al. (2019) [9] addressed challenges in medical disease prediction using ML methods for conditions such as cardiovascular, diabetic, and hepatic disorders. These challenges depended on many factors, including the type of data, completeness, and inaccurate data, etc. If these factors existed in datasets, then the prediction model was not providing accurate performance. They proposed LNF with an FE model for medical data prediction for disease diagnosis. The proposed linguistic fuzzification (LF) system develops associative values and solves various issues. Due to

these issues, it takes more time for model training, and was not easy to measure. The authors gave better solutions for these issues by using the proposed model. The selection of desired features is achieved by using the composition of FS and NF (neuro-fuzzy) models to enhance the network capacity. The proposed model achieved better improvement in performance and implementation with 8 datasets. The authors used Fredman and HB (Holm-Bonferroni) as a statistical method for testing. The proposed approach achieved better results over real-world challenges.

Emrah Irmak (2021) [12] suggested a model using CNN to find and classify brain tumors in an automated manner. Three different CNN architectures are proposed to address the various classification tasks. In brain tumor detection, the first model performs better than the second. It attained an accuracy of 92.66% in classifying cancers into five types (glioma, meningioma, pituitary, metastatic) and further distinguishing them across three grades (II, III, and IV). For the early categorization of brain cancers, this technique was crucial. The study was the first to multi-classify MRI images of brain tumors and shows good classification results on large clinical datasets by using a grid search optimization approach to create hyperparameters for CNN models.

The framework suggested by Saeedi et al. (2023) [15] refined the brain tumor diagnostic system developed using convolutional DL in combination with chosen ML algorithms. The research involved using MRI data in determining tumor site and in classifying the benign and malignant tumors. The hybrid method outperformed the traditional classifiers (SVM and RF) in terms of sensitivity (SN) and specificity (SP) by combining CNN-based deep feature extraction with these classifiers. These findings showed that handcrafted combined with deep features were more effective than individual DL, which showed that hybrid AI methods have potential applications in clinical tumor detection.

Chatterjee et al. (2022) [16] had a profound presentation of the deep spatio-spatial model of "brain tumor classification" based on the MR images. They also use 3D CNN that do not compromise the spatial coherence or volumetric data to represent features better. The multi-centre MRI data was trained using the model, which yielded a better generalization and accuracy than the traditional CNN architectures. This study has emphasised the significance of spatial context in MRI-based classification, and it has reached a high accuracy level of more than 97 percent, which was

successful in automated tumor recognition and diagnostics. The article by Luo et al. (2019) [17] presented a multimodal DL model of brain tumor segmentation and survival prediction on MRI scans. The technique is a composite of various imaging modalities (T1, T1c, T2, and FLAIR) to enhance the quality of tumor delineation and prediction of patient outcomes. The study results utilised a cascaded CNN with a multi-path architecture, which demonstrated better scores on segmentation activities and credible survival prediction measures. The combination of multimodal data in the framework enabled a very high level of clinical decision support and effective investigation of the trends of tumor progression.

Mahajan et al. [20] have created an ML-based system that combines MRI image segmentation of brain tumors. The framework improved tumor localization accuracy with the help of optimised preprocessing and extraction of the features. The findings of the experiments described that the segmentation and classification accuracy were higher than standard CNN models. The analysis revealed that hybrid ML solutions have the potential to diagnose brain tumors at the initial phases.

The authors of [21] suggested an ensemble DL model consisting of a GAN to increase the brain tumors diagnosis consuming MRI images. The augmented training technique increased data complexity and strength, which minimised the effect of overfitting. The method was able to get high multi-class accuracy in different types of tumors. This paper highlighted the efficacy of generative data augmentation in medical image analysis.

Musthafa et al. [22] introduced VISION, which is an integrated structure that combines both segmentation and classification in brain tumor detection. The system employed a view-specific MRI examination to enhance the diagnostic accuracy and generalization. The hybrid model was found to perform well on benchmark datasets on cross-view. Findings confirmed the use of the framework in real-time clinical interpretation of MRI.

2.2 Category 2: Chest X-Ray (CXR)-Based Studies

Riwjan Khan et al. (2022) [1] proposed Machine Learning (ML) methods in the medical domain for smart healthcare services. The health issues were diagnosed with digital technology using DL and ML

methods. In the medical domain, several hospitals depend on smart machines to diagnose patients' diseases, track hospital and appointments from anywhere, etc. The proposed system was designed as healthcare equipment called Medicolite to resolve health issues through diet and facilitate online doctor appointment booking. The offered ML model collects massive data and fetches the required information regarding patients from the collected data. The comparison analysis is based on the two classifiers to train and test the result of CNN (the 30s), better than SVM.

Nida Shahid et al. (2019) [5] implemented a smart healthcare framework using Artificial Neural Networks (ANN) for multi-disease detection, including conditions such as diabetes, cardiovascular disorders, and respiratory diseases. The model employed ML and DL techniques to automate clinical workflows such as disease diagnosis, medicine preparation, and appointment scheduling. By enhancing ANN efficiency, the system provided low-cost and scalable healthcare solutions. The study highlighted the capability of ANN-based systems to improve early detection and decision-making across multiple disease domains.

Saranya et al. (2021) [8] discussed and reduced-dimensional data as a data pre-processing step towards AI efficiency, mitigating unwanted data, increasing the efficiency of accuracy, and improving performance metrics. However, the data measurement process provided a very fast process for representing various feature collections of features and extraction techniques as related to effectiveness and productivity. However, the ML technique was a development field, yet based on data mining (DM) and prediction techniques for data pre-processing. Most of the selection of feature methods faced some issues regarding efficiency and accuracy due to earlier enhancements. Several types of feature selection (FS) techniques were accessible in research based on filtering, wrapping, and hybridization.

Arsalan Khan et al. (2023) [10] described cardiovascular disease as a dangerous syndrome growing significantly worldwide. Early recognition and forecast of cardiovascular disease and other heart illnesses protect several lives. This system involved experimental statistical analysis. The probability of analytical ML processes to grow the treatment center awareness was important to all investors in the health zone. The authors utilized an ML method to deploy a major clarification for precise calculation and decision-making for cardiovascular disease patients. Random sampling was employed to obtain patient data related to heart

disease from the “Khyber Teaching Hospital” and the “Lady Reading Hospital”, Pakistan. The authors executed tentative and investigational production analyses for all methods, such as SVM, RF, LR, etc. The projected ML procedure was presented in many environments to identify the superlative ML method in the class of models. The implemented method attained a performance-based accuracy of 92.11%.

Wang et al. (2023) [11] discussed the approach of predicting COPD in the case of unbalanced data. The author used a questionnaire to collect data for disease prediction. SMOTE has been used by the author to deal with unbalanced data, and ML-based traditional ML classifiers were used for disease prediction. LR, along with balanced data, performed better than other classifiers.

Sungyeup Kim et al. (2022) [13] developed a DL approach proposed in this research that uses the transfer learning methodology to identify lung disorders from chest X-ray radiography (CXR) images. This improves the diagnostic efficiency of CADs by increasing their efficiency and accuracy. To obtain informative features for classifying disease categories, in the end-to-end approach, raw CXR images are directly input into the EfficientNet v2-M DL model. Three lung disease classes in the US were used to test the approach. Utilizing datasets obtained from the NIH and Cheonan Soonchunhyang University Hospital, Validation was performed on samples belonging to the normal, pneumonia, pneumothorax, and tuberculosis classes. The validation performances were 82.15%, 81.40%, and 94.48%, respectively.

Shamrat et al. (2022) [14] suggested a fine-tuned CNN model from frontal chest X-ray imaging to propose a multiclass classification of lung illness. Using a combined dataset from several sources, this classification is done on 10 lung disease classes. Eight augmentation strategies were utilized to preprocess and analyze the information, and 80,000 X-ray pictures were classified. Out of the eight pre-trained CNN models, VGG16 had the best accuracy, coming in at 92.95%. To improve classification accuracy, LungNet22 was built. The study employed an ablation study to find hyperparameters.

Al-Qaness et al. [23] explored DL algorithms to diagnose lung diseases using chest X-rays. The review examined CNN, hybrid, and transformer models in the prevention of pneumonia, tuberculosis, and COVID-19. It reviewed such major issues as the imbalance of the dataset and explainability. The article provided important

insights into the future development of the CXR-based diagnostic systems.

Sharma et al. [24] proposed that computer vision be used in identifying medical images with regard to diseases. The article outlined CNN and transformer-based automated X-ray and MRI diagnosis. It emphasised early disease detection, workflow productivity, and model transparency. The article highlighted the AI, imaging, and healthcare analytics interdisciplinary integration.

The effectiveness of medical image-based disease classification remains limited by a series of problems, even though progress has been demonstrated in many studies. Most of the existing research focuses on single-disease or single-modality data, and this restricts their application in real-life clinical practice, where multimodal analysis is commonly required. Moreover, unfitting and reduced generalizability could be caused by the common inability of DL models to handle high-dimensional feature redundancy, imaging quality variability, and limited dataset sizes. Although these hybrid models attempt to enhance the accuracy of classification, they typically concentrate on either sequential context or spatial features learning, but not both within a single framework. To enhance the diagnostic reliability of different imaging modalities, a robust hybrid DL architecture that is capable of integrating temporal pattern learning and robust spatial feature extraction with effective preprocessing and feature selection is yet to be developed. This gap is the reason behind the proposed Hybrid InceptionV3-CNN-RNN multi-disease classification framework in this study.

3 MATERIAL AND METHOD

A multi-stage hybrid DL system for the automated classification of medical pictures, including categories for brain MRI scans and chest X-rays, is proposed in this paper. Systematic data preprocessing, noise simulation and suppression, region-of-interest segmentation, statistical feature selection, sequence-driven feature reshaping, deep feature extraction using a pertained Inception network, and a novel hybrid CNN-RNN model for final classification are all included in the methodology. The authors' previously published workflow illustrates the entire experimental processing pipeline.

3.1 Dataset Description

Two conventional medical imaging datasets were used in this work. The first dataset consists of the images of chest X-rays that are obtained on the

Kaggle clinical imaging platform and are categorised into various diagnostic groups: bacterial pneumonia, viral pneumonia, and healthy cases [18]. The second dataset contains of MRI scans of the brain, which are identified as glioma and meningioma, and the third dataset consists of the kinds of tumors of the pituitary gland [19]. The two datasets have different anatomical noise, imaging modality features, and intensity distributions. Cross-domain interoperability was ensured by ensuring that all preprocessing and normalization of features were consistent. The combined information was used to generate seven diagnostic classes, which were used as the basis of multi-class categorization.

3.2 Preprocessing

Preprocessing was necessary to minimize inconsistencies in acquisition and increase the fidelity of input representation.

3.2.1 Resizing and Normalization

To meet the input dimensional requirements of the pre-trained InceptionV3 model, all photos were downsized to $224 \times 224 \times 3$. Pixel intensities were adjusted to the range $[0, 1]$ $[0, 1]$ $[0, 1]$ in order to stabilize gradient updates during training:

$$I_{norm} = \frac{I_{resized}}{255} \quad (1)$$

3.2.2 Artificial Noise Contamination and Denoising

Salt-and-pepper noise was introduced artificially to mimic the disruptions of images in the real world and ensure that it is resilient. A 3x3 median philtre was then applied to re-establish continuity whilst preserving boundaries of anatomy

$$I_{filtered}(x,y) = \text{median}\{I_{noisy}(i,j)\} \quad (2)$$

3.2.3 Data Augmentation

To prevent overfitting due to small medical sample sizes, geometric expansion changes such as random horizontal flips, $\pm 90^\circ$ rotations, and zoom operations were employed. These changes enhanced the resilience of the learnt representations and artificially enhanced the variability of datasets.

To improve generalization and reduce overfitting, the dataset was augmented by the use of augmentation techniques that included flipping, rotating, and zooming of the input samples, as illustrated in Figure 1.

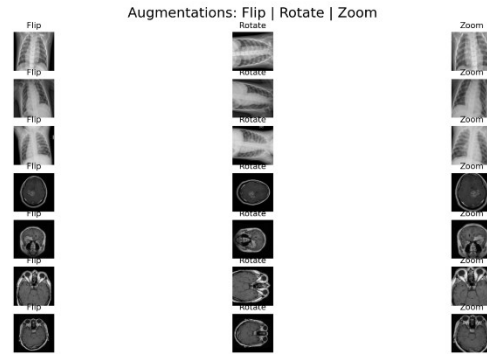


Figure 1: (a): flips augmentation, (b): rotations, (c): zoom augmentation

3.3 Segmentation Using K Means Clustering

In order to remove the unwanted background areas and highlight the pathological tissue, K-means clustering ($k = 2$) was conducted on a pixel-by-pixel basis. The pixels were clustered sequentially in terms of Euclidean similarity in feature space:

$$\text{cluster}(x, y) = \text{argmin}_k \|I(x, y) - u_k\| \quad (3)$$

The resulting foreground cluster was kept as the ROI, which enhanced the discriminative ability of the next layer of feature extraction.

Figure 2 gives a graphic representation of the preprocessing pipeline, which illustrates the sequence of operations performed on the raw medical image to extract the ROI on which features are to be extracted and classified.



Figure 2: Preprocessing pipeline including (a) Actual image, (b) Resized image, (c) Noise-induced image, (d) Filtered image using a median filter, and (e) ROI extracted via K-means segmentation.

3.4 Deep Feature Extraction Using Pre-trained InceptionV3

The extraction of features was performed with the help of the pre-trained InceptionV3 network that was selected due to its demonstrated ability to extract hierarchical and multi-scale spatial features. Once the first classification head was removed, the 2048-dimensional deep feature embedding of each image was obtained with the global average pooling output

layer. This approach provides generic visual priors suitable for medical imaging problems using less training data through the use of large ImageNet pre-training.

3.5 Statistical Feature Selection Using Chi-Square Test

Redundancy may compromise the model generalisation as the feature space that has been extracted is high-dimensional. This led to the use of a statistical dependency test based on the chi-square:

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad (4)$$

Where O denotes the observed values, and E denotes the expected values of the frequencies of activation of class-dependent features.

To optimise the classification efficiency, the 2000 most significant features were selected by ranking the features.

3.6 Data Splitting and Feature Reshaping

80 percent of the extracted and optimised features were used in the training process, and 20 percent of them were tested. The feature vector of 2000 dimensions was converted to a 40 x 50 sequence that could be learned sequentially and could be modelled with a hybrid CNN-RNN

3.7 Hybrid CNN–RNN Classification Model

Table 1 presents the 1D-CNN block in greater detail, which is used in detecting localised spatial patterns in the reshaped sequence of features. The input of the CNN block is a 40x50 sequential feature map, each segment of which is an organized part of the deep features collected. A 1D Conv layer (64 filters, kernel size 3) is used to determine short-range spatial feature connections. The most important patterns are retained, and dimensionality and computation load are reduced with the help of the MaxPooling layer. In order to reduce overfitting, we use a Dropout layer with a dropout rate of 0.3, which randomly deactivates neurons during training. This block enhances the discriminative power of the input characteristics before they are sent to the RNN component.

Table 1: Parameters used in the Generalize CNN model

Layer	Description
Input Layer	40×50 fixed feature-group representation
1D Conv (64 filters, k=3)	Local feature-group refinement
MaxPooling1D	Dimensional reduction
Dropout (0.3)	Overfitting prevention

Table 2 describes the RNN/LSTM processing block used for gated recurrent aggregation of the pooled feature representation. The 128-unit LSTM integrates information across the fixed feature groups through its gating operations. Because the input is derived from static images and the feature-group order is computationally imposed, the LSTM is used as a gated feature aggregator rather than as a temporal model. In this role, it provides nonlinear integration of information distributed across the selected feature groups. A Dropout layer (0.3) is introduced again to reduce overfitting before final classification.

Table 2: Generalize RNN parameters, layer-wise description

Layer	Description
LSTM (128 units)	Sequential dependency modeling
Dropout (0.3)	Regularization

The LSTM output was passed into a dense layer followed by a softmax classification layer.

$$y = \text{Softmax}(Wh + b) \quad (5)$$

This setup takes advantage of spatial as well as sequential discriminative clues. CNN approach, uses a number of convolution and pooling layers to learn edges, textures, and patterns of areas to extract spatial features of medical images. Arrangement is then done by fully linked layers. The RNN (LSTM) model, in its turn, is expected to handle sequential or time-dependent data, and the LSTM units are expected to capture contextual relationships, as well as long-term dependencies between sequences of features. The HybridIncep-CNN-RNN architecture Figure 3 utilises the strengths of both models by first obtaining high-level spatial features of images with a pre-trained CNN (InceptionV3) and then combining them with the RNN. The most discriminative ones are then selected, restructured to look like a sequence, and fed through 1D-CNN and LSTM layers to acquire the local spatial variation and temporal dependencies between features. The outcome of such a hybrid integration is a higher accuracy and strength of the multi-disease medical picture classification, and it enables the system to better represent the deeper contextual representations as well as structural disease patterns.

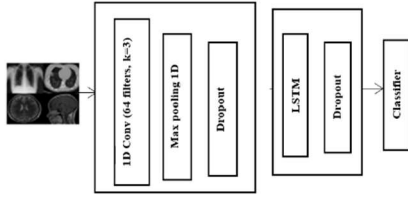


Figure 3: Architecture of Proposed Hybrid Incep-CNN-RNN Model

3.8 Model Training Procedure

Table 3 provides the optimization configuration and training hyperparameters of the integrated DL architecture. The AdamW optimizer was to be interacted with a learning rate (LR) of 0.001 to give stable and adaptive gradient updates. The 64 and 100 training epochs were chosen to balance between the effectiveness of the computations and the model performance. The Sparse Definite Cross-Entropy loss was applied to allow the prediction of various classes. To ensure the best performance on inference, model checkpointing was also used to automatically store the best-performing model during training.

Table 3: Model Training Procedure

Parameter	Values
Optimizer	AdamW
LR	0.001
Batch Size	64
Epochs	100
Loss Function	Sparse Categorical Cross-Entropy

Model checks ensured retention of the best-performing epochs.

3.9 Performance Evaluation

The research system's diagnostic presentation was quantified using:

- *Accuracy*: Evaluates the ratio of appropriately classified instances to the total number of predictions made (e.g., as exposed in Equation .

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

- *Precision (Pre)*: Determines the reliability of positive predictions by counting the number of positively predicted samples that are really positive.

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

- *Recall (SN)*: It computes the capacity of the research model to identify the actual positive instances in an appropriate manner, which is a measure of the completeness of the detection

$$Recall = \frac{TP}{TP+FN} \quad (8),$$

- *F1-Score*: calculates the accuracy and recall's harmonic mean to provide a balanced figure that may be used to explain both false positives (FPs) and false negatives (FN), for example, as shown in Equation

$$F1 = 2 \frac{Pre * recall}{Pre + recall} \quad (9)$$

- *Dice Coefficient*: Measures how closely the predicted regions or class distributions match the ground-truth data (e.g., as shown in Equation.

$$Dice = \frac{2TP}{2TP+FP+FN} \quad (10)$$

- *Jaccard Similarity Index*: Measures the intersection-over-union between predicted and actual classes (e.g., as shown in Equation.

$$Jaccard = \frac{TP}{TP+FN+FP} \quad (11)$$

- *Confusion Matrix Visualization*: Provides a class-wise comparison of predicted vs. actual labels to analyze misclassification patterns
- *ROC-AUC Analysis*: Evaluates the discriminative ability of the classifier by computing the area under the ROC curve.

3.10 Computational Time

The suggested hybrid CNN-RNN framework's computational performance was assessed in terms of per-image inference latency, average epoch processing time, and overall training length. The workstation used for all of the trials included an NVIDIA GPU, which allowed for faster tensor operations.

Let is the time taken to make one epoch of forward-backward propagation cycles over the training data. We shall specify the amount of training time and the number of training epochs that were completed. The computation time is represented as:

$$T_{train} = T_{epoch} * N_{epochs} \quad (12)$$

In the same way, the mean time of inference per test sample, also known as prediction latency, can be defined as:

$$T_{inference} = \frac{T_{total_test}}{N_{test}} \quad (13)$$

Where T_{total_test} denotes the time required to generate predictions for the complete test set and N_{test} is the number of test samples.

Table 4 summarises the computing environment used to develop and validate the proposed model. The studies were conducted in a workplace with an Intel Core i7 CPU, an NVIDIA RTX 3060 or T4 processor that has more matrix operations, and 16 GB of RAM to support efficient memory utilisation during training. The model was implemented with TensorFlow/Keras with CUDA GPU acceleration. This hardware setup is not only a demonstration of the fact that the model can be applicable in the real-world research or clinical computing setting, but also guarantees the repeatability of the experimental performance.

Table 4: Experimental Hardware Configuration

Component	Specification
Processor	Intel Core i7 / Ryzen 7 class CPU
GPU	NVIDIA RTX 3060 / T4 CUDA-enabled GPU
Memory	16 GB RAM
Framework	TensorFlow / Keras + CUDA acceleration

Table 5 presents the computation time of every important step in the processing and training workflow. The speed of pre-trained forward-pass computation was proved by the fact that it required approximately 1.8 seconds to extract features with InceptionV3 on 100 images. Since Chi-Square only does a single operation on the retrieved feature set, the feature selection process required between 14 and 18 seconds. The hybrid model required 42-55 seconds per epoch to train, which is approximately 70-90 minutes in 100 epochs. The per-image inference time ranged between 0.015 and 0.031 seconds, which means that it can be used in clinical and real-time decision support.

Table 5: Measured Computation Times

Stage	Computation Time	Description
Feature Extraction (InceptionV3)	~ 1.8 seconds per 100 images	Due to the pre-trained model forward pass only
Feature Selection (Chi-Square)	~ 14–18 seconds total	Statistical scoring and ranking
Hybrid CNN–RNN Model Training	~ 42–55 seconds per epoch	Depends on batch size and GPU load
Total Training Time (100 Epochs)	~ 70–90 minutes	Final optimized model stored at

		the best epoch checkpoint
Inference per Image	0.015–0.031 seconds	Suitable for real-time clinical use

4 RESULTS

This section presents the experimental results in a progressive manner. After visualising the extracted and selected features, the proposed Hybrid InceptionV3-CNN-RNN model is evaluated separately on the Brain MRI and Chest X-ray modalities, followed by representative multi-image predictions and modality-specific confusion matrices. The analysis then reports the combined seven-class, multiple-modality performance, ROC-AUC behaviour, and comparison with baseline models.

4.1 Feature Extraction Visualization

The pre-trained feature extraction model using the InceptionV3 model was chosen for the deep feature extraction step because of its demonstrated capacity to learn hierarchical and multi-scale spatial representations from medical pictures. All the images were fed through the InceptionV3 convolutional layers to produce a 2048-dimensional feature embedding after preprocessing, denoising, and ROI segmentation. These feature maps, illustrated in Figure 4, have activation areas that are associated with meaningful anatomical structures, e.g., inflammatory opacities in chest X-ray images or tumor margins in MRIs. The model manages to focus on therapeutically meaningful structures and not background noise, as the visual outputs testify. This makes sure that the retrieved features are very discriminative, and they can be further classified using the proposed hybrid CNN-RNN architecture.

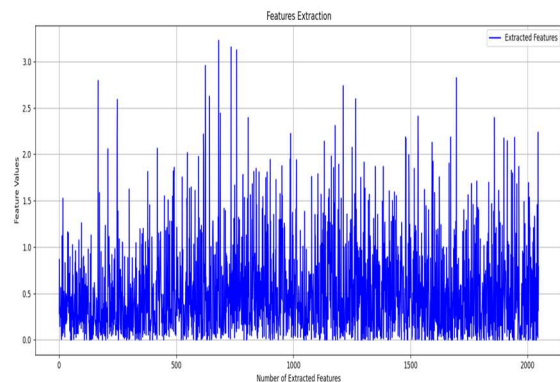


Figure 4 Deep feature extraction output obtained using the pre-trained InceptionV3 network.

These high-level descriptors are extracted and used as the basis of further learning and classification.

4.2 Feature Selection Result

Not every extracted feature is useful for classification, despite the fact that InceptionV3 has many spatial feature representations. Consequently, the Chi-Square statistical feature selection method was used to determine the significance of each feature w.r.t the class labels. Based on their reliance score, the 2000 most discriminative features of the original 2048 characteristics were kept, as Figure 4.2 indicates. This dimensionality reduction step is used during model training to prevent overfitting, reduce redundancy, and reduce computational costs. To facilitate spatial and temporal learning with the hybrid CNNRNN model, the selected features formed a high-quality optimal representation that was transformed into a 4050 sequential format.

Figure 5 greatly enhances stability in classification and minimises model complexity.

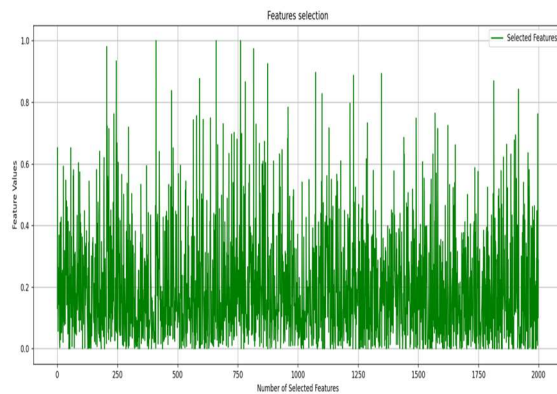


Figure 5: Feature selection workflow and reduced feature vector representation.

4.3 Single Modality Results

To examine whether the unified framework remains reliable within each imaging domain, the test predictions were partitioned into Brain MRI and Chest X-ray subsets. Accuracy, precision, recall, F1-score, and error rate were calculated independently for each modality. This analysis distinguishes modality-specific behaviour from the aggregate results obtained from the combined seven-class test set.

4.3.1 Brain MRI Results

The Brain MRI subset achieved the strongest single-modality performance. As reported in Table 4.1, the model obtained 96.88% accuracy, 97.06% precision, 96.88% recall, and a 96.89% F1-score, while limiting the error rate to 3.12%. The close agreement among the four principal metrics

indicates balanced recognition of the MRI classes and only a small number of false predictions.

Table 6: Single-Modality Performance for Brain MRI

Modality	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Error Rate (%)
Brain MRI	96.88	97.06	96.88	96.89	3.12

Table 6 shows that all four primary classification measures exceed 96%, whereas the error rate remains close to 3%. This narrow spread indicates that the model does not improve one metric at the expense of another.

It confirms the high and nearly uniform accuracy, precision, recall, and F1-score achieved for Brain MRI. The substantially shorter error-rate bar visually reinforces the low frequency of incorrect MRI predictions.

4.3.2 Chest X-Ray Results

The Chest X-ray subset produced 92.00% accuracy, 92.59% precision, 91.04% recall, and a 91.73% F1-score, with an 8.00% error rate. Although these values are lower than those for Brain MRI, the four principal measures remain above 91%, showing stable classification despite the greater visual overlap among the chest disease patterns.

Table 7: Single-Modality Performance for Chest X-Ray

Modality	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Error Rate (%)
Chest X-ray	92.00	92.59	91.04	91.73	8.00

Table 7 indicates that precision is the highest chest-specific measure, while recall is comparatively lower. This pattern suggests that the model makes relatively few false-positive chest predictions but misses a small number of positive cases.

4.3.3 Modality-Specific Confusion Matrices

The confusion matrices were also separated by imaging modality to identify domain-specific error patterns. In the Chest X-ray matrix, the diagonal counts are 26, 44, and 22, with eight off-diagonal predictions concentrated among the three chest classes. In the Brain MRI matrix, the diagonal counts are 37, 39, 39, and 40, with only five off-diagonal predictions, demonstrating stronger class separation for the MRI subset.

Figure 6 corroborates the numerical results in Tables 4.1 and 4.2: Brain MRI exhibits fewer errors,

whereas most Chest X-ray errors occur among visually related chest disease categories. The strong diagonal structure in both matrices confirms that the

proposed framework remains effective when each modality is evaluated independently.

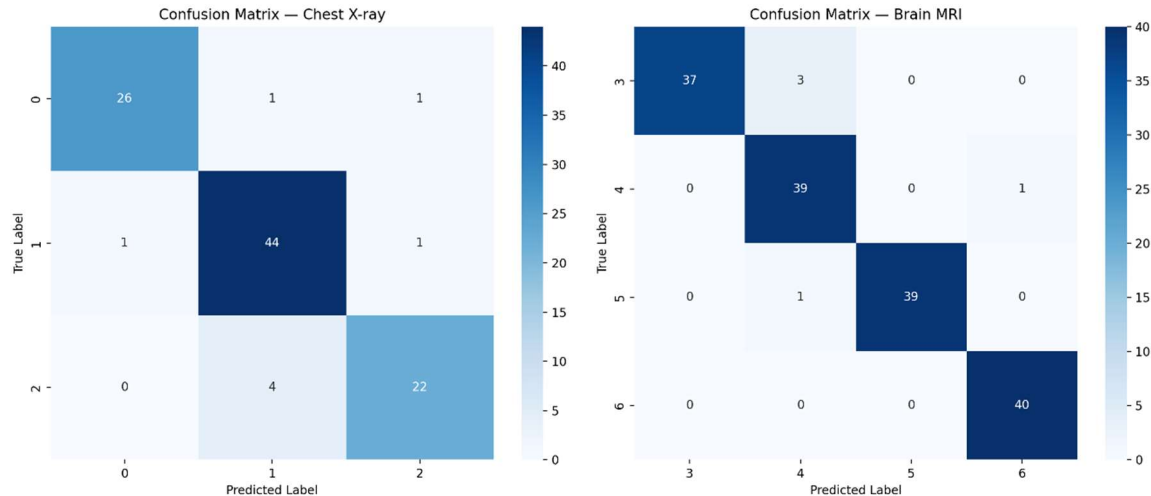


Figure 6: Single Modality confusion matrices for Chest X-ray and Brain MRI classification.

4.4 Evaluation Metrics of the Hybrid Inception CNN–RNN Model

Typical classification parameters, including recall, precision, accuracy, and f1-score, along with overlap-based measures including the Dice Coefficient and Jaccard Index, were engaged to calculate the hybrid model's performance. The research model demonstrated good recognition capabilities across the seven diagnostic classes, as shown in Table 8, with an overall accuracy of 95%. Reliable positive case detection as reflected by the high precision and recall values of 95% and 94%, respectively. A stable link between recall and precision is confirmed by the F1-score of 95%, indicating minimal FPs and FNs, and the significant degree of overlap between the forecast labels and the ground truth is further validated by the Dice Coefficient and Jaccard Index scores. These findings show that in both the MRI and X-ray imaging domains, the hybrid spatial–temporal learning

strategy performs better than single-stream CNN-based models. The high-scoring metrics in Figure 7 demonstrate the strong classification reliability of the model.

Table 8: Evaluation Metrics of the Hybrid Incep-CNN–RNN Model

Parameters	Value (%)
Accuracy	95%
Precision	95%
Recall	94%
F1-Score	95%
Dice Coefficient	95%
Jaccard Index	90%
Error Rate	5%

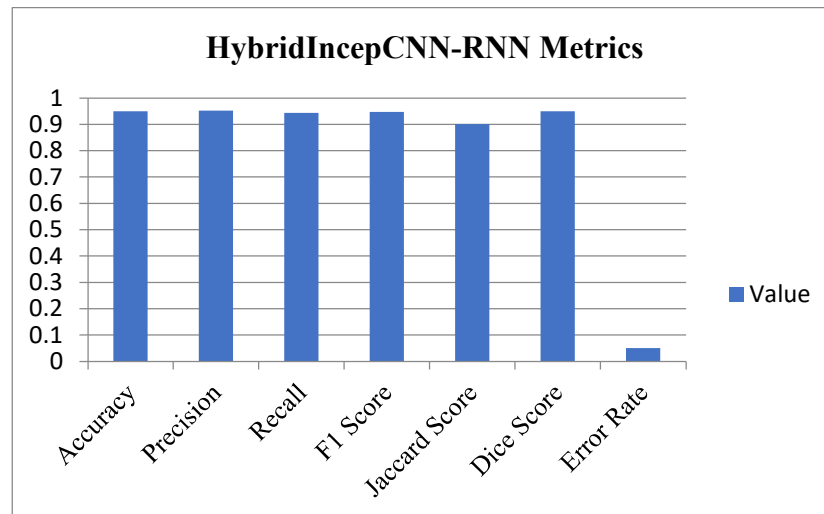


Figure 7: Comparative metric performance of the Hybrid Inception CNN-RNN model.

4.5 Confusion Matrix Evaluation

For every disease category, Figure 8 offers a class-wise evaluation of prediction accuracy. Strong diagonal intensities show that most samples were accurately categorized using both MRI and X-ray modalities. Since there were overlapping radiological patterns, some minor misclassification was observed between visually similar types of illnesses, e.g., bacterial and viral pneumonia. These misclassification regions are quite small, but that is why the discriminative potential of the hybrid model is high. Consequently, the confusion matrix validates the clinical diagnostic application of the system and supports the fact that it can be used to distinguish between the complicated disease categories reliably.

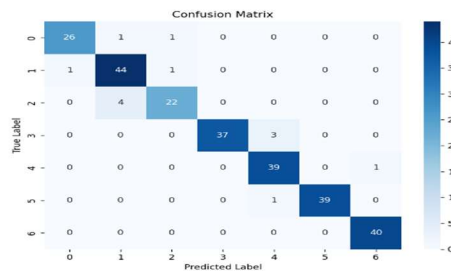


Figure 8: Performance Evaluation Using the Confusion Matrix for the Proposed Inception-Driven CNN-RNN Architecture

Large diagonal cell intensities indicate that most of the samples were properly classified with minimum misclassifications, particularly in the visually related classes of diseases.

4.6 ROC-AUC Curve Performance

To determine the discriminative power of the classifier, Receiver Operating Characteristic (ROC)

curves were plotted against each of the seven diagnostic classifications. Most of the classes have ROC curves close to the upper-left boundary, indicating high sensitivity and specificity, as indicated by Figure 9. In most types of diseases, the Area under the Curve (AUC) scores are far above 0.90, which means that there is a good correlation between positive and negative class prediction. This great AUC result indicates that despite the low visual characteristics of illness or the overlapping nature of these characteristics, the proposed model can still offer a high level of classification. Consequently, the ROC-AUC analysis provides strong evidence of the diagnostic validity of the model and possible clinical decision support.

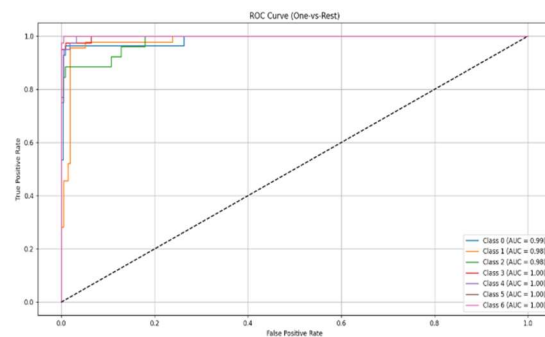


Figure 9: ROC Curves for all disease classes with corresponding AUC values.

The strong discriminative ability of the classifier is confirmed by curves that are located near the upper-left segment of the plot.

4.7 Comparative Analysis with Existing Models

Table 9 compare the overall proposed framework with the standard CNN and ResNet50-ViT-B/16 baselines. The modality-wise comparison shows that Brain MRI yields the highest individual scores,

while the combined model retains a strong 95% accuracy across all seven classes.

This demonstrates that the unified framework does not rely on a single imaging domain and retains

balanced precision and recall when both modalities are evaluated together.

Table 9: The Comparison Analysis With Proposed and Existing Models

Model	Accuracy (%)	Precision (%)	Recall (Sensitivity) (%)	F1-Score(%)	Dice Coefficient (%)	Jaccard Index (%)	Error Rate(%)
Standard CNN [1]	85.7 ~ 86	86.61~87	85.07	86	86	77	14.29
ResNet50-ViTb16 Model [25]	91	91	91	91	90	84	9.0
Proposed Hybrid Inception CNN-RNN (Ours)	95	95.15	94.37	94.68	95	90.10	5

This hybrid approach produced a better accuracy (95%), precision, and recall, and a lower error rate as compared to the baseline models. These findings indicate that the hybrid feature learning approach strategy is extremely effective in enhancing diagnostic reliability and generalisation ability, particularly where different medical imaging modalities are considered.

The Hybrid model outperforms the other methods due to:

- Multi-scale deep feature learning from Inception
- Gated feature aggregation via LSTM
- Robust preprocessing and feature selection

5. DISCUSSION

The Hybrid InceptionV3-CNN-RNN model proposed in this research showed high performance when applied for multi-disease recognition from CXR and brain MRI scan images, obtaining 95% accuracy rate, 95.15% precision, an F1-score of 94.68% and 94.37% recall. These results indicate that there is high diagnostic reliability and generalization capability with the various imaging modalities.

The operational efficiency of the introduced system can be seen in relation to the current DL - based classification frameworks, as an indication of significant improvements in characteristics extraction, noise-resilient preprocessing, and gated recurrent feature aggregation.

The past studies of brain MRI-based tumor classification have indicated that deep CNN can be useful in distinguishing the types of tumors. The study by Irmak (2021) obtained high multi-class tumor recognition with a fully optimised CNN framework, which validates the applicability of CNN-based models to MRI feature learning [12]. On the same note, Saeedi et al. (2023) showed that convolutional DL coupled with the selected machine learning classifiers enhanced the accuracy of the diagnosis in detecting brain tumors [15]. The operational ability of the proposed model is in line with the outcomes observed in earlier studies and extends them by adding LSTM-based gated aggregation, which provides nonlinear integration of the extracted feature groups before classification.

Kim et al. (2022) [13] emphasized that the use of DL architectures is very successful when it comes to discriminating various types of lung diseases, yet in certain cases, the performance level is associated with the variability of the data and the quality of imaging [13]. Shamrat et al. (2023) enhanced the detection of multi-class lung diseases employing a customized MobileNetV2 architecture, emphasizing the importance of lightweight feature extraction networks in clinical applications [14]. The suggested hybrid model has the advantage of having good visual hierarchical characteristics of InceptionV3 and does not suffer from the performance decline characteristic of lightweight architectures, thus leading to improved classification resilience without compromising inference speed.

To interpret brain MRI, the current investigation by Chatterjee et al. (2022) [16] and Luo et al. (2019)

[17] showed that the use of multi-scale spatio-spatial patterns and multimodal MRI channels enhances the results of classification and segmentation [16] [17]. Although these studies are devoted to modality-specific enhancements, the proposed model is unique in its ability to combine two non-homogeneous imaging areas (X-ray and MRI) into the pipeline of the unified classification, which is made possible by the regular preprocessing, feature normalization, and hybrid features learning [25]. This ability to cross domains is one of the main advantages, as most of the past studies concentrate on either measuring the chest or the brain.

Overall, the findings indicate that combining deep spatial representation extraction (InceptionV3) with gated recurrent feature aggregation (LSTM-RNN) enables the system to extract localised anatomical descriptors and integrate information distributed across the selected feature groups. This combination helps the high performance that is realised as compared to the traditional CNN-only methods. The findings prove the importance of hybrid architectures in the medical image classification issues, where the visual pattern variations of the disease categories and imaging modalities are significant.

6. CONCLUSION

This study shows a hybrid DL model, in which the InceptionV3 feature extractor was used with CNN-based feature refinement and RNN-based gated feature aggregation to perform multi-class classification in CXR and brain MRI images. The preprocessing pipeline, comprised of noise reduction, ROI segmentation, and data augmentation, was useful in making features consistent when a heterogeneous dataset was used. The Chi-square feature selection further minimized repetition so that the model is capable of generalizing better. The outcomes of the experiments revealed that the Hybrid InceptionV3-CNN-RNN proposed model delivered reliable outcomes, reflecting its high accuracy of 95%, precision of 95.15%, recall of 94.37%, F1-score of 94.68%, Dice Coefficient of 95%, Jaccard Index of 90.10%, and error rate of 5%, as well as high stability of the model with all evaluation metrics, showing an improvement over the traditionally used CNN-based models.

The findings show that the integration of spatial deep feature extraction and gated recurrent feature aggregation is important for learning subtle disease-relevant representations. The model also has low

inference latency, but this implies it is suitable for real-time application healthcare support settings.

Future research can look at multimodal combinations of clinical metadata, enhancements of model explainability, and adapting to larger medical datasets to enhance clinical adoption and performance.

REFERENCES:

- [1] Khan R, Srivastava AK, Gupta M, Kumari P, Kumar S. Medicolite–Machine learning-based patient care model. *Computational Intelligence and Neuroscience*. 2022; 2022:1-10.
- [2] Kumar S, Negi A, Singh JN. Semantic segmentation using deep learning for brain tumor MRI via fully convolutional neural networks. In *Information and Communication Technology for Intelligent Systems 2018* (pp. 11-19). Springer; 2019.
- [3] Brindha PG, Kavinraj M, Manivasakam P, Prasanth P. Brain tumor detection from MRI images using deep learning techniques. *IOP Conference Series: Materials Science and Engineering*. 2021; 1055(1):012115.
- [4] Lakshmi MJ, Nagaraja Rao S. Brain tumor magnetic resonance image classification: A deep learning approach. *Soft Computing*. 2022; 26(13):6245-6253.
- [5] Shahid N, Rappon T, Berta W. Applications of artificial neural networks in healthcare organizational decision-making: A scoping review. *PLoS ONE*. 2019; 14(2):e0212356.
- [6] Bhanothu Y, Kamalakannan A, Rajamanickam G. Detection and classification of brain tumor in MRI images using a deep convolutional network. In *Proceedings of 6th International Conference on Advanced Computing and Communication Systems 2020* (pp. 248-252). IEEE, 2020.
- [7] Huang Z, Du X, Chen L, Li Y, Liu M, Chou Y, Jin L. Convolutional neural network based on complex networks for brain tumor image classification with a modified activation function. *IEEE Access*. 2020; 8:89281-89290.
- [8] Saranya G, Pravin A. Feature selection techniques for disease diagnosis system: A survey. In *Artificial Intelligence Techniques for Advanced Computing Applications 2021* (pp. 249-258). Springer; 2021.
- [9] Das H, Naik B, Behera HS. Medical disease analysis using neuro-fuzzy with a feature extraction model for classification. *Informatics in Medicine Unlocked*. 2020; 18:100288.

- [10] Khan A, Qureshi M, Daniyal M, Tawiah K. A novel study on machine learning algorithm-based cardiovascular disease prediction. *Health & Social Care in the Community*. 2023; 2023:1-10.
- [11] Wang X, et al. Machine learning-enabled risk prediction of chronic obstructive pulmonary disease with unbalanced data. *Computer Methods and Programs in Biomedicine*. 2023; 107340:1-10.
- [12] Irmak E. Multi-classification of brain tumor MRI images using a deep convolutional neural network with a fully optimized framework. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*. 2021; 45(3):1015-1036.
- [13] Kim S, Rim B, Choi S, Lee A, Min S, Hong M. Deep learning in multi-class lung diseases classification on chest X-ray images. *Diagnostics*. 2022; 12(4):915.
- [14] Shamrat FJM, Azam S, Karim A, Ahmed K, Bui FM, De Boer F. High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images. *Computer Biology and Medicine*. 2023; 155:106646.
- [15] Saeedi S, Rezayi S, Keshavarz H, Kalhori SRN. MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. *BMC Medical Informatics and Decision Making*. 2023; 23(16):1-13.
- [16] Chatterjee S, Nizamani FA, Nürnberger A, Speck O. Classification of brain tumors in MR images using deep spatio-spatial models. *Scientific Reports*. 2022; 12(1505):1-11.
- [17] Luo L, Sun L, Zhang S, Chen H. Brain tumor segmentation and survival prediction using multimodal MRI scans with deep learning. *Frontiers in Neuroscience*. 2019; 13(810):1-12.
- [18] Diamantaras K. Chest X-Rays: Bacterial, Viral Pneumonia & Normal. *Kaggle*. Accessed 11 February 2025.
- [19] Nickparvar M. Brain Tumor MRI Dataset. *Kaggle*. 2021.
- [20] Mahajan S, Sahoo AK, Sarangi PK, Rani L, Singh D. MRI image segmentation: Brain tumor detection and classification using machine learning. In *International Conference on Data Analytics & Management 2023* (pp. 125-139). Springer Nature; 2023.
- [21] Sahoo S, Mishra S, Panda B, Bhoi AK, Barsocchi P. An augmented modulated deep learning based intelligent predictive model for brain tumor detection using a GAN ensemble. *Sensors*. 2023; 23(15):6930.
- [22] Musthafa N, Masud MM, Memon Q, Farhad M. VISION: View-specific integrated segmentation-classification framework for accurate brain tumor detection in MRI scans. *PLOS ONE*. 2025; 20(10):e0332395.
- [23] Al-Qaness MA, Zhu J, Al-Alimi D, Dahou A, Alsamhi SH, Abd Elaziz M, Ewees AA. Chest X-ray images for lung disease detection using deep learning techniques: A comprehensive survey. *Archives of Computational Methods in Engineering*. 2024; 31(6).
- [24] Sharma R, Kumar N, Sharma V. Computer vision for disease detection—An overview of how computer vision techniques can be used to detect diseases in medical images. In *AI in Disease Detection: Advancements and Applications 2025* (pp. 77-98).
- [25] Hadhoud Y, Mekhaznia T, Bennour A, Amroune M, Kurdi NA, Aborujilah AH, Al-Sarem M. From binary to multi-class classification: A two-step hybrid CNN-ViT model for chest disease classification based on X-ray images. *Diagnostics*. 2024; 14:2754.