RESEARCH ARTICLE

Multi-modal medical image classification using deep residual network and genetic algorithm

Muhammad Haris Abid, Rehan Ashraf ⓘ *, **Toqeer Mahmood, C. M. Nadeem Faisal** ⓘ

Department of Computer Science, National Textile University, Faisalabad, Pakistan

* rehan_ashraf94@yahoo.com

Abstract

Artificial intelligence (AI) development across the health sector has recently been the most crucial. Early medical information, identification, diagnosis, classification, then analysis, along with viable remedies, are always beneficial developments. Precise and consistent image classification has critical in diagnosing and tactical decisions for healthcare. The core issue with image classification has become the semantic gap. Conventional machine learning algorithms for classification rely mainly on low-level but rather high-level characteristics, employ some handmade features to close the gap, but force intense feature extraction as well as classification approaches. Deep learning is a powerful tool with considerable advances in recent years, with deep convolution neural networks (CNNs) succeeding in image classification. The main goal is to bridge the semantic gap and enhance the classification performance of multi-modal medical images based on the deep learning-based model ResNet50. The data set included 28378 multi-modal medical images to train and validate the model. Overall accuracy, precision, recall, and F1-score evaluation parameters have been calculated. The proposed model classifies medical images more accurately than other state-of-the-art methods. The intended research experiment attained an accuracy level of 98.61%. The suggested study directly benefits the health service.

1. Introduction

In recent years, multiple medical CBIR techniques have been presented. The majority of developed CBIR retrieval mechanisms employ a single imaging modality. The retrieval algorithms can give the chance to choose the image class before similarity comparison which is one technique to retrieve the required medical images across large image libraries. A CBIR system might benefit greatly from good image categorization since it would eliminate the desire to search through irrelevant images, reducing the number of images that the system would have to look [1].

Convolutional neural networks (CNN) established approaches not only improve classification accuracy, although they are also considered good general descriptors of features. CNN extracts features in a hierarchical manner, with lower layers encoding lower characteristics such as edges, forms, texture, and so on, and higher levels encoding semantic level aspects

associated with an image. Because the kernels within these networks have been learned rather than constructed, no preliminary parameterize nor human involvement is required [1].

Non-learning approaches perform better under some situations; nonetheless, the disparity across higher levels semantics with low-level depictions of features in diverse pictures leads in reduced image retrieval efficiency. Existing techniques have also used multiple learning-based strategies to close the gaps in semantics that increase picture retrieval effectiveness. Despite techniques using learning effectively bridge the gap within higher levels semantics alongside low-level visualizations of features across various images, they are dependent upon several kinds of various attributes and lack the ability to perform adequately on all types of images over each feature descriptor. Furthermore, learning-based CBIR approaches are computationally challenging than non-learning CBIR techniques [2].

Convolutional neural networks (CNNs) have already made significant advancements in the computer vision field [3]. There have already been introduced multiple neural network architectures, such as VGGNet, GoogLeNet, ResNet, DenseNet, as well as more recently NASNet [4]. ResNet itself and its variants have drawn the most attention within these deep networks. ResNet has shown exceptional results across both medium and high computer vision applications. This shortcut connection technique, which enables the training about a deeper structure where gradients may directly flow across construction blocks but the gradient vanishing dilemma can be somewhat avoided, is largely responsible for ResNets' exceptional success. Its shortcut connection process, on the other hand, forces every block to concentrate on learning its residual output while somehow ignoring the internal block connectivity and making it therefore that some reusable knowledge generated in earlier blocks is often neglected in subsequent blocks [5].

The current CBIR studies are focused on developing new techniques to better describe visual material that are more relevant to the users. In recent trends for medical image retrieval investigations, the images are described using a set of semantic terms. This set of semantically specified image attributes could be applied to recognize a broad variety of images which increases the user's attention to visual aspects [6, 7]. The advantages of semantic terms for diagnostics decision-making would be that they could enable radiologists to search image databases across instances with similar high-level and enhanced quality [8].

The radiologist's keywords for image observations are the key factor that is appropriate for the implementation of CBIR [9, 10]. To bridge the semantic gap across images and their associated meaning, adding semantics to image description can therefore be a novel technique [11]. This combination with text attribute searching, which depends on the contents of the image, with limited visual features, which are generated directly first from the representation of images, has been discovered to enhance medical semantic search findings [12].

In this study, we stated the issue of semantic gap elimination in image retrieval. To resolve this challenge, semantic language for radiological image contents has been proposed. Feature searches and visual qualities complement each other. Complementary notions and implementation to image databases results in a more will useful result for all users. These visual characteristics of an image transmit a relatively low level of characterization, making it impossible to accurately express using keywords alone.

The major contributions of this proposed work are:

➢ To bridge the semantic gap across user requests with system responses

➢ To implement a genetic algorithm for optimal multi-classification

➢ To apply optimal model training for improving the multi-classification of multimodal medical images

2. Related works

In the last few years, digital image processing and the combination of machine learning has shown good results in various applied domains of computer vision [13–15]. The recent focus of research for image classification-based models is the use of deep learning architectures and frameworks [16–18].

In the work of [19], the primary goal of the modality classification included separate various forms of medical images, such as X-ray, CT, electrocardiography, and PET, as well as generic graphs, from other medical sources for illness diagnosis. Radiologists require an effective categorization system for retrieving associated clinical cases to make accurate illness diagnoses [20]. Similarly [21] used the Faster-RCNN technology in together with SVM-based classifier to provide a unique strategy for the autonomous classification about melanoma lesions. If the challenge is one of classification, it may be preferable to utilize a deep neural network [17].

The development of medical image modality by [22] categorization system was beneficial in narrowing the retrieval query space within a specific modality. For the creation of modality categorization systems, two techniques were frequently utilized i) hand-crafted (traditional) but also ii) Deep neural networks. Medical images, especially as opposed to general images, contain a variety of characteristics such as postural difficulties, texture, and aesthetic elements. The classification of medical image modalities had been mostly based on form, color, and textured features [23]. Similarly [13] discussed positive guiding importance towards multi-modal medical imaging evaluation.

Deep Learning techniques were [24] suggested the best alternative for classifying medical images. Algorithms are used to automatically extract the primary elements from medical images in order to execute the classification procedure. The primary concept is to create feature maps using Conv layers. During the convolution operations, various filter masks having varied orientations are employed to build feature maps. Such maps are then processed using pooling procedures as a feature minimization method. The goal is to employ the most realistic characteristics for image classification and avoid any manually extracted features throughout the process of classification. In work of [16] increased the adaptability about deep inpainting structures to training sets alongside diverse variety, while improving inpainting effectiveness as judged through qualitative as well as quantitative measures for an extensive variety about deep models.

Researchers have presented [25] modality classification algorithms to improve performance on baseline methods published by Image CLEF. It was noted that the effectiveness of previously proposed techniques employing hand-crafted attributes varies but achieves adequate accuracy overall [26]. It was because classification performance is heavily reliant on expert judgment when obtaining acceptable data for modality categorization. This was challenging to determine the amount and kind of retrieved attributes from modality images for efficient categorization [27, 28]. Those techniques were limited by their large computing needs as well as the constraint of conditionality. Mostly as a result, it is essential to design an efficient modalities categorization strategy that enhances performance while requiring less human interaction.

Regarding image retrieval under a multi-class instance, a CBIR technique utilizing a hybrid characteristics descriptor using the genetic algorithm alongside SVM classifier was presented by [2]. The suggested method's performance was evaluated using four benchmark datasets, alongside its comparison to 25 alternative CBIR approaches. Experimental findings show that their technique surpasses prevailing state-of-the-art retrieval algorithms.

Several CNN models with binary along with multi-class categorization of COVID-19 instances were studied by [29]. These models were tested on various CT alongside X-ray datasets using Transfer learning ideas for deep-tuning while fine-tuning settings. Transfer learning

frameworks involving LeNet-5, VGG16, AlexNet, along with Inception nave v1 being deep-tuning frameworks and DenseNet121, DenseNet201, DenseNet169, ResNet50, VGG16, ResNet152, and VGG19 being fine-tuning frameworks have been thoroughly compared. Simulation tests were carried out on a total of 12,032 images from chest CT with X-ray collections (COVID-19 = 2,466, pneumonias = 4,273, and normal = 5,293). Every model was evaluated using a variety of categorization assessment measures. Considering the investigated X-ray with CT images, ResNet152 and DenseNet201 performed better compared to various Transfer learning frameworks. Similarly investigated [14] improved deep learning based LeNetsþþ on softmax, centred integrating i-center loss function, using a variety of standard image recognition stages.

In the work of [30] presented a method for predicting patient survival based on reliability and effectiveness. Furthermore, researchers wanted to show how important it was to use classification then FS algorithms for achieving the greatest outcomes in the quickest period of time, since this is a critical aspect in an individual's survival. Following doing trials and analyzing the findings with regard to of mistake rate and precision, it was revealed the classification algorithms delivers superior results when not combined alongside the FSFA. Therefore, rather than employing FSFAs, an approach based on classification proved more accurate and efficient.

The techniques' scope was centred [31] on illness categorization, early screening, and organ localization, including benign and cancerous detection. Classification, and segmentation, including detection, are common CAD operations. Image classification treats each image only as a separate entity that must be distinguished from other images. Image separation was based upon pixel points, which divide the image over numerous distinct parts with distinct attributes, including image classification with the specific border of the existent objective [32]. Image detection seems to be the retrieval of a particular sub-image from a recognized image, whereas classification involves the retrieval of many items in an image [33].

The fundamental requirement for success mostly in classification explained by the [34] challenge was to identify highly discriminative characteristics about specific classes. This was very simple for categories having excellent internal consistency similarity, however, it may be challenging for domains having low inter-class correlation [35]. For example, mammography classification accuracy was generally poor, while discriminating characteristics for breast cancers are hard to capture in the context of overlapping, diverse fibro glandular structures. Considering the significant inter-class resemblance, the concept that fine-grained visual classification (FGVC), which tries to discover tiny distinctions among visually similar items, may be suitable for learning distinguishing characteristics [36].

As an outcome, techniques developed and assessed by [37] on such datasets could not be easily transferable for medical datasets when only a subset of images demonstrate significant inter-class similarities instead of all of them. Other methods for improving characteristic discrimination power incorporate the use of concentration modules, local and global features, specialized knowledge, and everything else [38]. If just a subset of the such training phase is labelled, the algorithm achieves the feedback connection from the labelled data but is enhanced through learning semantics plus fine-grained characteristics from the unsupervised learning [39]. As a result, the model optimizer was split into two stages: self-supervised pre-training but also supervised fine-tuning. The model was first improved using unidentified images to successfully learn excellent features which are indicative of such image semantics.

Although there were several approaches to constructing feature temples, a generally accepted rule seems to be that robust [40], moderate semantics must be combined alongside high-dimensional maps. Furthermore, when there were a high number of medical images that have structural, textural, but also semantic similarities with the targeted dataset, pre-training

producers and/or classification techniques may help with computational efficiency and enhanced efficiency [41]. Similarly an effective transfer learning approach using the AlexNet framework provided [42] to properly classify and identify melanoma.

Following Table 1 is discussing and clarifying current studies, their limitation and helps us to bridge our research gap.

3. Methodology of proposed work

To assist our research, we used openly accessible medical datasets. This dataset contained five types of medical images (i.e. endoscopy, CT, chest, hand x-ray, and lungs CT). A maximum of 28378 good-quality jpg image formats was utilized within datasets. Images are then resized into 512 X 512 pixels. The model's pre-processing procedure was used for the pre-processing purpose. Only as a consequence of our model's testing using medical images, did researchers focus on establishing their database. Through this heterogeneous dataset, we picked images at irregular intervals from each class. During our research, we have used a dataset of 28378 images across 5 distinct classes. Crucial issues during this data included significant intra-class variance and great inter-class similarities caused by using multiple classes with various imaging technologies. We used 80% of the images during training and 20% throughout the testing. Because of the obtained dataset's complex dimensions and structure, each image from each class was modified to 512×512 again and translated into a consistent jpg file. We used supervised learning to apply a class label.

A possible perspective of multi-medical image classification and assessment is displayed in Fig 1. Images were initially gathered and sorted into classes. Image processing procedures include image shearing, transformations, image flipping, and scaling. These images were again input into the suggested method for model training at the next stage. That recently trained model has been used. Finally, multi-modal medical image identification but also classification had been achieved.

3.1 Images category

In this research, we used several medical images of multi-modal image classes which are shown in Fig 2. Generally, there are several steps of Machine Learning techniques toward medical image identification and classification employing Convolutional Neural Networks. These steps include dataset collection, dataset pre-processing, image segmentation, extraction of features, and classification. Each image was pre-processed and classified using the Kaggle platform. The significant percentage of datasets enhances the effectiveness of learning models and reduces over-fitting. Acquiring a dataset that can be used as input to such a training phase is a time-consuming but difficult task. As just a result, image enhancement expands the overall training data set offered for deep learning algorithms. Image flipping, resizing, rotation, color transformations, color enhancement, and noise reduction, are all deep learning-based intensification methodologies [52]. Automated extraction of features offers a high identification speed and precision. Feature extraction during segmentation converts the images towards a vector containing fixed features. These system-adopted characteristics include color, texture, but also shape. While extracting texture characteristics from some kind of color image, using a grey-scale cross matrix is preferable.

3.2 Genetic algorithm

We applied a Genetic algorithm for optimization. Genetic algorithms, which depend on bio-inspired operators including mutation, crossover, but also selection, are often employed to develop strong solutions for optimization and searching issues. The reason to use a genetic

Table 1. Systematic literature review.

Publication Name & Ref	Contribution	Dataset and Techniques	Limitations and Future Work
Multimodal medical image segmentation using multi-scale context-aware network [43]	86.8% state-of the-art performance on three benchmark datasets of different modalities captures the rich context information with dense skip connection and assigns distinct weights to different channels	Multi-scale context-aware network (CA-Net) for multimodal medical image segmentation	Improve the feature expression ability of the network
Multimodal Medical Image Fusion Based on Pixel Significance Using Anisotropic Diffusion and Cross Bilateral Filter [44]	72% outperforms other algorithms in terms of objective evaluation metrics Edge preserving processing of the original images where it combines linear low pass filter with nonlinear techniques	Multimodal medical image fusion technique based on anisotropic diffusion and cross bilateral filter	Multimodal image fusion by proposing various algorithms
Efficient color image retrieval method using deep stacked sparse auto encoder [45]	Content-based image retrieval system for natural color images using a deep stacked sparse auto encoder	Corel-1K, Corel-10K DSSA model latent features	Efficacy of the latent representation learned
A deep neural network model for content- based medical image retrieval with multi-view classification [46]	Accuracy 92.35% Body part orientation view classification labels, intending to reduce the variance that occurs in different types of scans	A deep neural network-based approach for view classification and content-based image retrieval is proposed and its application for efficient medical image retrieval is demonstrated	Techniques to handle the class imbalance in the dataset with optimization techniques
Cross-Modality Sub-Image Retrieval using Contrastive Multimodal Image Representations [47]	Content-based image retrieval system (CBIR) for reverse (sub-)image search to retrieve microscopy images in one modality given a corresponding image captured by a different modality	Combine deep learning to generate representations that embed both modalities in a common space, with classic, fast, and robust feature extractors (SIFT, SURF) to create a bag-of-words model for efficient and reliable retrieval	Observe the importance of equivariance and invariance properties of the learned representations and feature extractors in the CBIR pipeline.
Content-based medical image retrieval using topic and location model [48]	Outperforms existing medical image retrieval systems in terms of Precision and Mean Average Precision. The proposed method achieved better Mean Average Precision (86.74%) compared	Automated medical image retrieval system using Topic and Location Mode	Retrieval of medical images based on the location and size of the anomaly
Content-based Image Retrieval and the Semantic Gap in the Deep Learning Era [49]	Semantic image retrieval Perform inferior to much less sophisticated and more generic methods	Aggregated convolutional features and opposed to traditional local features	Semantic image retrieval methods are often hardly comparable regarding the task definition and the evaluation data
Convolutional neural network-based dictionary learning to create hash codes for content-based image retrieval [50]	Visual complexity disappears	ResNet-50 architecture is trained with modified COREL dataset images	High retrieval time
MDCBIR-MF: Multimedia data for content-based image retrieval by using multiple features [9]	Qualitative methodology again for CBIR. Color values inside the HSV color space were utilized to extract color features, while texture characteristics were extracted using DWT and Gabor wavelets.	To improve the feature representation, the color, and edge omnidirectional descriptors were generated and incorporated, with widths of 1 250. As bigger the input vectors dimension, the much more precise the retrieval findings, but once longer it lasts to search and compare. The suggested system was evaluated on several datasets Corel and obtained good median accuracy.	High computational time owing to large feature vector dimensions.
An encrypted image retrieval method based on harris corner optimization and LSH in cloud computing [51]	Suggested a CBIR approach for retrieving encrypted images from the cloud that is predicated on an upgraded and the SURF detector with the descriptor.	These authors utilized the Local Sensitive Hash algorithm to generate accessible directories for extracted features to minimize retrieval time and improve	The performance is not great with large image datasets.

https://doi.org/10.1371/journal.pone.0287786.t001

algorithm is that some greyscale medical images such as chest X-rays and CT need to be enhanced for better identification. Better identification will lead us toward optimized classification. By changing pixel values, the developed optimization algorithm will be reproduced dataset images. The implementation steps of the genetic algorithm included 1) reading of images 2) preparation of fitness function 3) implementation of mutation 4) implementation of statistics and results.
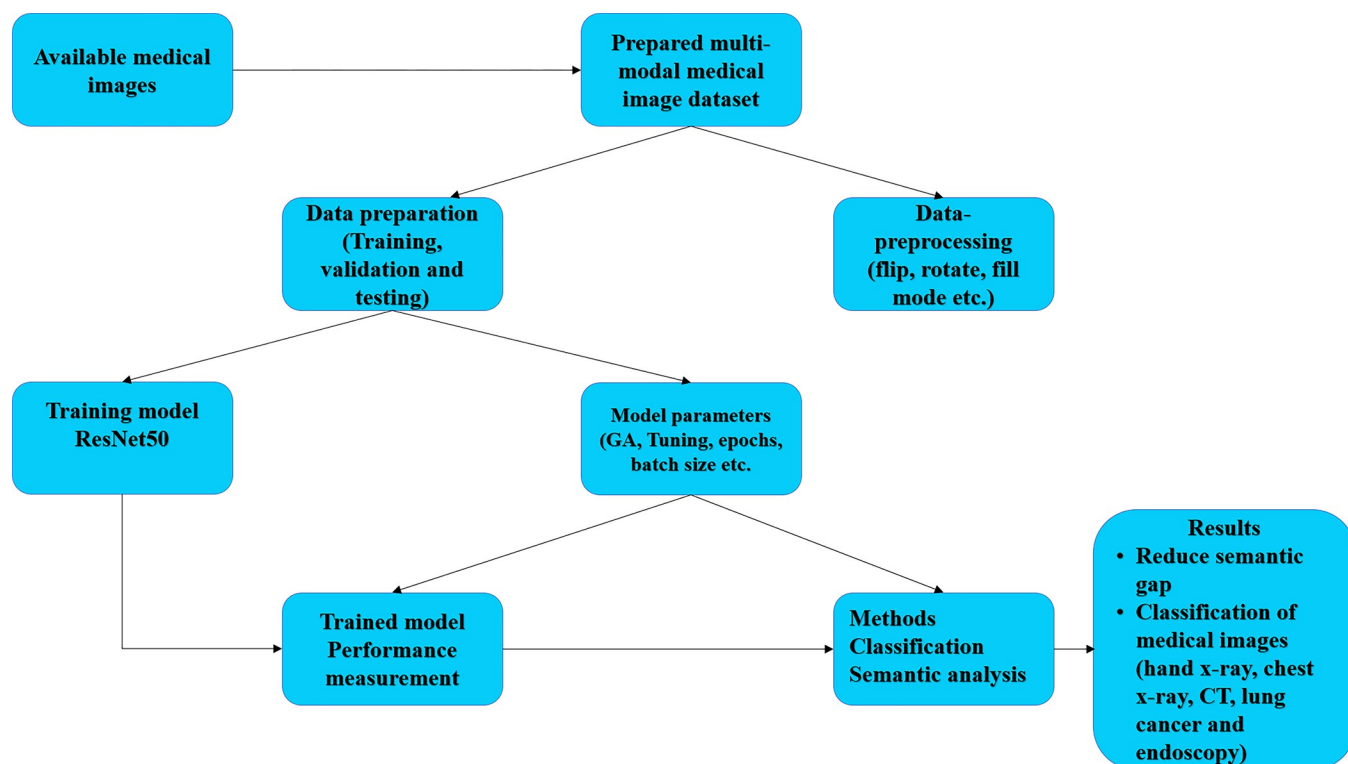
Fig 1. Proposed research flow diagram.

3.3 Transfer learning

The optimization with the training of the model seems to be a difficult yet time-consuming process. The training requires a strong graphics processing unit (GPU) along with thousands of training samples. Transfer learning, which is used in deep learning, meanwhile, eliminates all of these concerns. This transfer-learned per-trained Deep Learning Approach (CNN) is optimized for one activity and transfers information to different patterns [53]. This multi-modal images dataset model has 512 X 512 in size. We required modification in the residual network (ResNet). Its final layer even before softmax across all ResNet50 configurations is indeed a 7 X 7 average-pooling structure. Whenever a pooling size is reduced, a relatively small image may fit through into the network.

3.4 Convolutional neural network

The structure of any Convolutional Neural Network (CNN) is made up of convolutional layers, pooling layers, and fully-connected layers, including dense layers as shown in Fig 3. The descriptions of the layers are presented below.
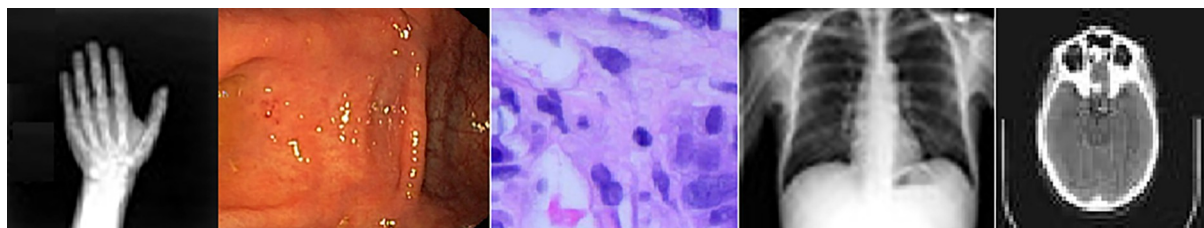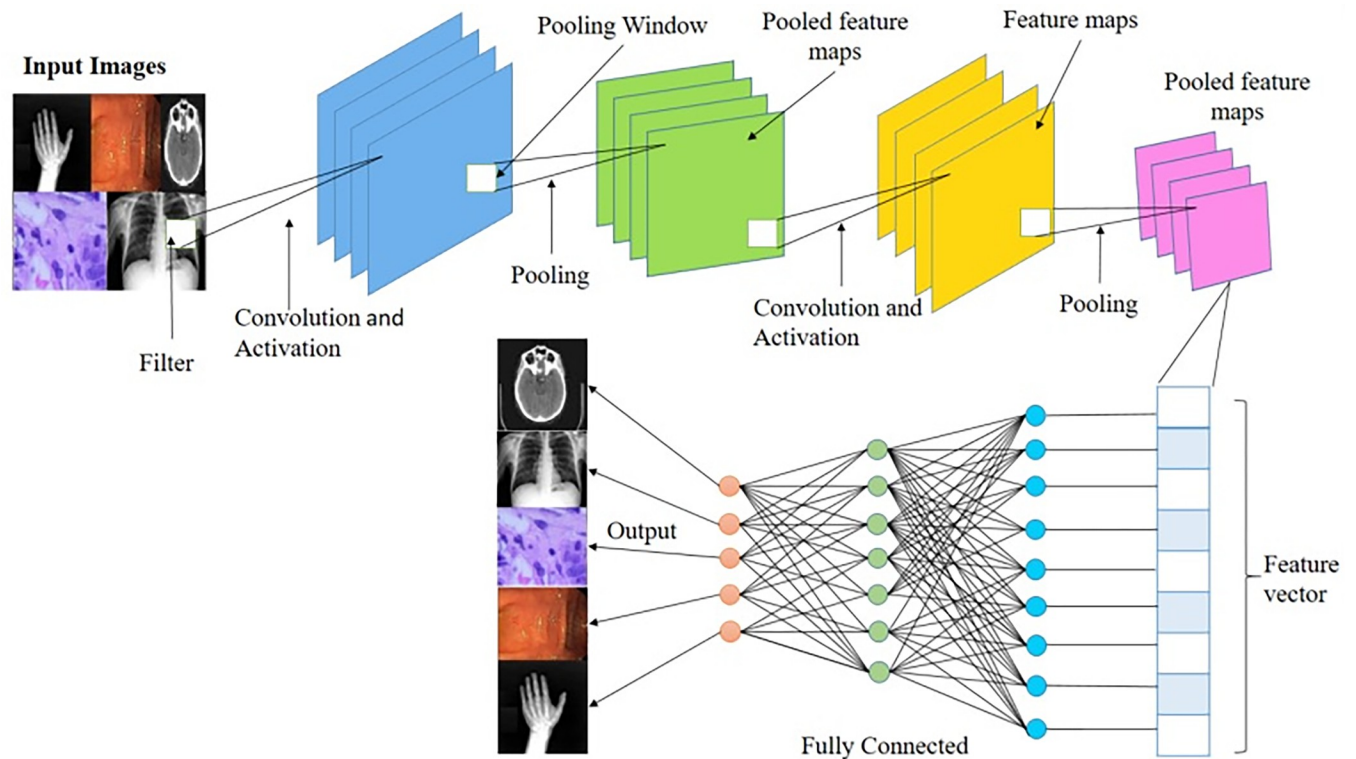

Fig 2. Sample medical images.

Fig 3. Proposed CNN architecture.

3.5 Convolutional layer

The primary function of convolutional layers included extracting distinctive features using images. The need for convolutional layers regularly aids throughout the extraction of input information [54]. The following Formula 1 is used to estimate the features extraction (FE_i) across various layers through CNN.

$$FE_i = \boldsymbol{\omega}(\boldsymbol{FE_{i-1}}\boldsymbol{Wg_i} + \boldsymbol{OFS_i}) \tag{1}$$

Where, FE_i—Feature map, Wg_i–Weight, OFS_i is offset and ω–Rectified Linear Unit (RELU).

3.6 Pooling layers

These pooling layers have become an important part of a Convolutional Neural Network (CNN). They reduce the dimensionality of convolved elements while also reducing the computer resources required for computer vision. Pooling may be divided into two categories maximum pooling plus average pooling. Usually, the highest values of images are returned by max pooling, but the mean values of such image sections are returned by average pooling.

3.7 Drop-out layers

Such dropout layers enhance the performance of a training phase. It offers regularization and inhibits over-fitting by lowering the correlation among neurons. Most activation functions employ the dropout procedure, however, it is enhanced by factor [55].

3.8 Flatten layers

It reduces its spatial dimensions about the mapping pooled characteristics while keeping its channel dimensions intact. This flattened layer includes dimensions before being converted into such a vector. This vectored input to completely linked layers is sometimes referred to here as a dense layer but rather fully connected layers.

3.9 Fully-connected layers

Along with their unique function, retrieved image categorization features require fully linked layers. This softmax function forecasts image properties collected from previous stages. Softmax is an activation function mostly in output layers that performs classification. During knowledge involvement, the neural network layer implements another multiplayer perceptron structure as either a classifier. Variability is induced in the entire vectors through the rectified linear unit (RELU) activated in the system. The depth of the ConvNet architectural design is its most important component. By establishing extra design parameters and continually increasing network depth by adding more convolutional layers, which is possible by employing extremely tiny convolution filters throughout all levels. Mostly as the outcome, have created significantly more precise ConvNet structures which not only achieve state-of-the-art precision on resolved input data classification as well as localization activities, while also being applicable towards other image processing datasets, within which they perform excellently even when used throughout relatively simple flow-lines.

Throughout the training, our ConvNets were provided this fixed-size 512×512 image. In only one pre-processing we have subtracted each pixel from the average value calculated mostly from the training dataset. To transport the image throughout a stacking of convolution operation, we use filters with an extremely tiny receptive field. In several of the setups, we also applied convolution filtering, which represents a linear modification of the inputs. This convolution stride was set to one pixel, and indeed the spatial padding of its convolutional layer inputs is set between one pixel for three convolution operations that maintain spatial resolution during convolutional. Spatial pooled is performed by 5 max-pooling levels that follow the portion of such convolutional layers.

3.10 ResNet50

ResNet50 pre-trained architectures using Convolutional Neural Networks are applied to increase performance but also classify images. This model adequately transfers information across pre-trained ResNet50 networks toward image quality identification and analysis. This CNN model has maintained fresh images learned to produce a model with identification and classification [56]. Using big kernel-sized filtering and convolutional layers besides a kernel filter size, our ResNet50 model improved. This size of the supplied image is set at 512×512. Images were pre-processed and then sent through another convolutional layer. This filter size was estimated with the linear treatment of the network interface (1 x 1). The stride value is taken at one, and the maximum pooling size was two by two. This filter size was specified for the sequential transformation of the channels. The fully—connected layer will use the same structure in the following phases, having 2048 channels within every layer. These Softmax activation structures are the outermost layer, succeeded by such RELU activation mechanisms in Table 2.

4. Experiment, results, and discussion

The model has been fine-tuned to maximize accuracy with minimizing expected loss. On Kaggle, an extensive experimental analysis took place. Python programming packages have been

Table 2. Proposed Resnet50's architecture.

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 512, 512, 3)	0	Conv1
conv1_pad (ZeroPadding2D)	(None, 518, 518, 3)	0	input_1[0][0]
conv1_conv (Conv2D)	(None, 256, 256, 64)	9472	conv1_pad[0][0]
conv1_bn (BatchNormalization)	(None, 256, 256, 64)	256	conv1_conv[0][0]
conv1_relu (Activation)	(None, 256, 256, 64)	0	conv1_bn[0][0]
pool1_pad (ZeroPadding2D)	(None, 258, 258, 64)	0	conv1_relu[0][0]
pool1_pool (MaxPooling2D)	(None, 128, 128, 64)	0	pool1_pad[0][0]
conv2_block1_1_conv (Conv2D)	(None, 128, 128, 64)	4160	pool1_pool[0][0]
conv2_block1_1_bn (BatchNormali	(None, 128, 128, 64)	256	conv2_block1_1_conv[0][0]
conv2_block1_1_relu (Activation	(None, 128, 128, 64)	0	conv2_block1_1_bn[0][0]
conv2_block1_2_conv (Conv2D)	(None, 128, 128, 64)	36928	conv2_block1_1_relu[0][0]
. . . .			

Total params: 26,733,446

Trainable params: 3,145,734

Non-trainable params: 23,587,712

Number of FLOPs: 47175424/s

uploaded since installed for scientific purposes. All experiments in our study were conducted under a computer including the following specifications: A CoreTM i7 CPU, 12 GB RAM, and a graphics card. This type of graphics card offers parallel computation throughout these training and testing periods. Upon that Windows 10 platform, Python (Keras plus Tensor Flow) was utilized to implement this whole training but also validation CNN methods. The data set has been structured as a directory containing two sub-directories, classes as well as tests. This classes directory is applied to training while the tests folder has been applied to testing. This class's directory comprises five sub-directories containing various medical images (i.e. Endoscopy, CT, Chest, Hand X-ray, and Lung CT). Images categories were not allocated to the folders' names. The purpose to achieve this is to effectively train set to bridge the semantic gap. The Directory structure can be explained by following Eqs 2 & 3.

$$f(I) = I_{MD} \tag{2}$$

$$I_{MD} = I_{MD} + \sigma \tag{3}$$

Where MD is a medical image collection and image denotes an image including a name but a path. Earlier than the training technique began, every single image within the dataset has been scaled into $512 \times 512 \times 3$ during the pre-processing step. Eq 4 represents the scaling formula.

$$A_{I+1} = A_I + S_X$$

$$B_{I+1} = B_I + S_Y \tag{4}$$

The model has been loaded with adjusted weights after being fine-tuned based on dataset parameters. Every feature vector took into account the ultimate pooling layer's conclusion. This pooling function involves applying a two-dimensional filtration to each channel from the feature map but then summarizing the features which lie within the filter's covering zone. These are the dimensions that the output obtained because a pooling layer was used instead of

a feature map well with dimensions provided in Eq 5.

$$\frac{(fmhi - fil + 1)}{{}_str*(fmwi - fil + 1)/_stride*fmch} \tag{5}$$

Where *fmhi* is feature map height, *fmwi* is feature map width and *fmch* is the number of feature map channels. Similarly, *fi* is the size of the filter and *stride* is the length of the stride.

Given this decreasing gradient barrier, sigmoid and hyperbolic tangent activation has been utilized in multi-layer networks. Its rectified linear activation overcomes that vanishing gradients problem, allowing models to train faster while performing better. Utilizing rectified linear activation is the typical activation for developing multi-layer perceptron and convolutional neural networks. ReLU has been used here for activation functions in neural networks. ReLU is represented in Eq 6.

$$ReLU(Img) = max(0, Img) \tag{6}$$

Whereas if the source becomes negative, then the result of ReLU equals 0; if the source becomes positive, then the result is *Img*.

Adam is one stochastic gradient optimizer. This common solution 'adam' works well on moderately large datasets in respect of both management time plus validation scores. To pick activation or solver, a selected group has been made, i.e. returns a collection at random out of such an array. This random approach takes into account access to a variety of critical functions, including the capacity to generate random options.

In the next step genetic algorithm has been implemented for image reconstruction. The reason to use a genetic algorithm is that some greyscale medical images such as chest X-rays and CT need to be enhanced for better identification. Better identification leads us toward optimized classification. By changing pixel values, the developed optimization algorithm reproduced dataset images. The pixel levels varied within 0–255, 0–1 scale based upon that chromosomal description. This pixel-computed value influences other factors such as the range through which probabilities are chosen during mutation or the set of values utilized in the current population.

The code constructs one fitness function which will be used to calculate the overall fitness value with each solution within a population. Each function needs to be a maximizing function that receives two parameters, one indicating a solution while the second expressing its index. This gives back a value that represents the optimal solution. This fitness value can be calculated by adding the absolute differences in gene levels between the initial and replicated chromosomes. Since this genetic algorithm could work using 1D chromosomes, this function has been run before the actual fitness function should represent the image as such a vector. The fitness functions are represented in Eq 7.

$$\textbf{\textit{Fitness Function}} = {}^1/_{|}\textbf{\textit{x}} + \textbf{\textit{y}} + \textbf{\textit{z}} - \textbf{\textit{t}}| \tag{7}$$

Consider the following three factors: x, y, as well as z. The goal is to discover the optimum collection of parameters for x, y, but also z such that whose total value equals t. We must keep the total of x+y+z from departing from t, namely |x + y + z—t| must be zero. Only as result, the fitness value may be thought of as the inversion of |x + y + z—t|.

It is critical to employ random mutation but also set its mutation by replacement parameter to True. These bases for selecting towards the range low, range high, random mutation mini val, but also random mutation maxi val factors should be obtained based mostly on the range available pixel values. Whereas if image pixels are between 0 and 255, leave the range low and random mutation mini val at 0, but increase the range high with random mutation maxi val to

255. Mutation can be explained by Eq 8.

$$\mu = {}^{m}/_{N} \tag{8}$$

Where N denotes the mean quantity of cells each cultured.

Following the completion of the run procedure, actual fitness values among all generations may be observed in Fig 4.

The findings can even be improved by modifying the arguments given to such class's function Object. Fig 5 below is showing a sample of source images which shows how it transformed after a few iterations.

Following that, Fine-tuned ResNet50 subsequently trained upon that basis all the preceding phases. The checkpoint has been set for the said model so that the best fitness results could be saved and the most recent best accuracy could be used. Finally, classification was performed by supplying the query image and then converting it with an array. An argmax function was used, that returns this index of the largest number within the given row and column, also with rows or columns selected based on both the argmax method's axis property. This predict function describes the type of function provided that assists in generating output predictions using the specified sample of parameters onto a model.

As a result of matching prediction with input arrays about image classification, the semantic gap significantly decreased. Overall training loss vs accuracy including both degrees for cross-validation for each epoch showed in Fig 6. After a certain epoch, the total loss has been 0.3304 across all configurations, but the prediction accuracy has hit 98.61%, suggesting that our ResNet50 CNN has also been properly trained to utilize training data. Moreover, after completing a set of CNN model training testing, we noticed that fine-tuning our model produces more accuracy versus standard training from the start.

5. Performance measure

To assess classification performance, metrics F1 score and precision matrix have been utilized. Evaluation metrics have been used to assess the classifier's efficiency.
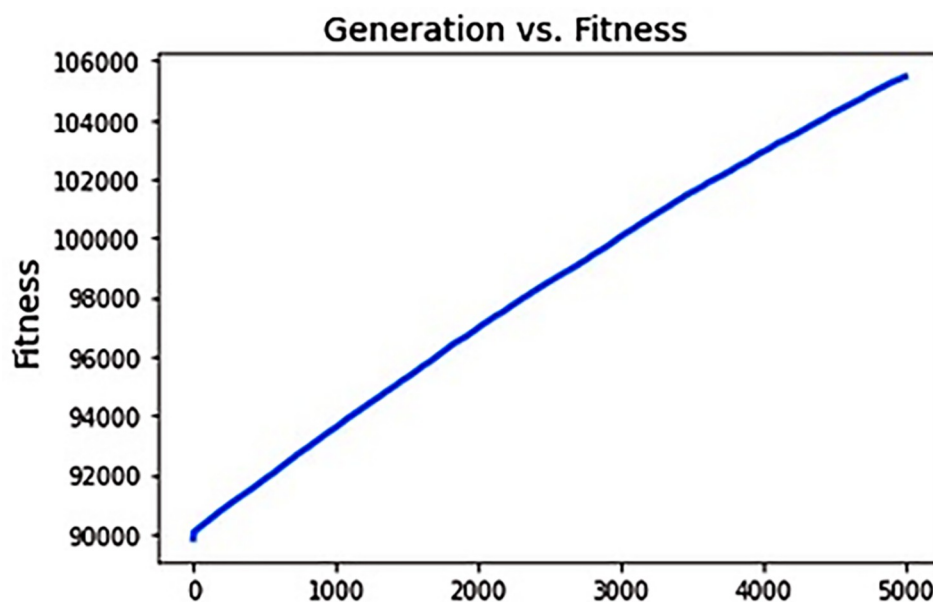

Fig 4. Fitness function.

https://doi.org/10.1371/journal.pone.0287786.g004

Fitness value of the best solution = 105454.69348013483
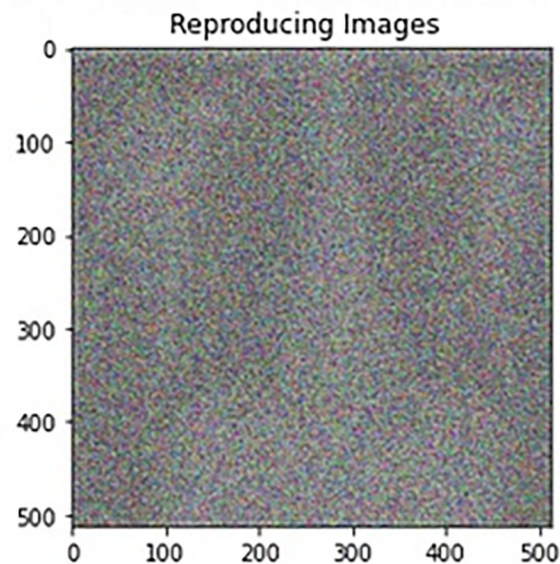Best fitness value reached after 4997 generations.

Fig 5. Image reconstruction.

5.1 Accuracy metrics

The performance of the model including all classes is precisely measured. Overall accuracy is measured by adding the overall number of relevant guesses to the overall number of forecasts. Precision, recall, but also F1-Score have been calculated as performance parameters. The
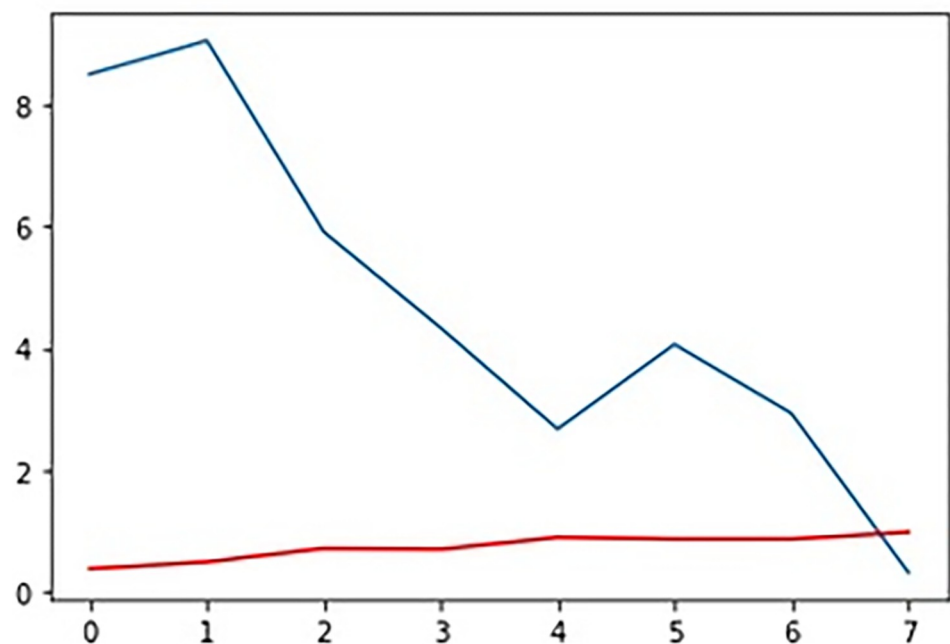

Fig 6. Accuracy vs loss function (Red = Accuracy, Blue = Loss function).

Table 3. Performance analysis of proposed and other models.

Model	Accuracy	F1.Score	Precision	Recall
AlexNet [58]	69.71	69.63	70.31	68.97
VGG16 [59]	77.36	78.22	79.06	72.27
VGG19 [59]	77.46	78.25	79.08	77.41
GoogleNet [60]	78.66	79.51	80.23	78.81
ResNet50 [61]	80.42	81.41	82.01	80.83
InceptionV3 [62]	80.43	81.63	82.13	81.14
Enhanced residual network [63]	80.62	81.50	82.24	80.79
Proposed	98.61	99.30	100	98.6

precision is stated as follows.

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{9}$$

Where *TP* stands for True Positive, *TN* stands for True Negative, *FN* stands for False Negative, whereas *FP* is for False Positive. Our classifier performance is measured using evaluation measures.

Another essential statistic for assessing the algorithms is the F1 score. This is the fundamental accuracy and recall which is given as follows:

$$F1.score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{10}$$

The influence about transfer learning has been examined through fine-tuning top 3 CNN models based on Table 3 outcomes using our improved deep residual modelling. The earlier ImageNet dataset had previously pre-trained those models [57]. This chosen collection fine-tuned the final few convolutions along with all of the FC layers over transfer learning, whereas the dataset provided by ImageNet optimized filter weights within the early convolutional layer training. In terms of precision, F1 score, accuracy, and Recall, our improved deep residual strategy outperformed the other models by using transfer learning. Table 3 and Fig 7 below show a comparison of various previous research and our model.

A t-test approach was used to investigate the relevance of our suggested model compared to the second-highest method, Enhanced residual network [63]. The Table 4 displays the t-test results for our suggested model with the second-highest method. The t-test study was predicated upon a test of null hypothesis, of which assumed that we have no significant difference regarding performance comparing our suggested model with the second-highest method. The results presented in Table 4 demonstrate that the significance levels of accuracy as well as F1. score during this test remained 0.0269 (below 0.05) as well as 0.02189 (below 0.05), respectively, while running a t-test. These findings indicating the null hypothesis regarding accuracy was not accepted at a 95% confidence level, indicating that there was indeed a significant disparity for accuracy comparing our model with the second-highest model. Furthermore, the null hypothesis regarding F1. score was not accepted given 95% confidence, demonstrating the significant improvement of our proposed model over the second-highest model.

Data pre-processing approaches including random rotation flips, and scale transformation, along with associated pre-processing operations, are utilized to expand the training set to ensure the variety of sample images and prevent over-fitting. These procedures are detailed further down.
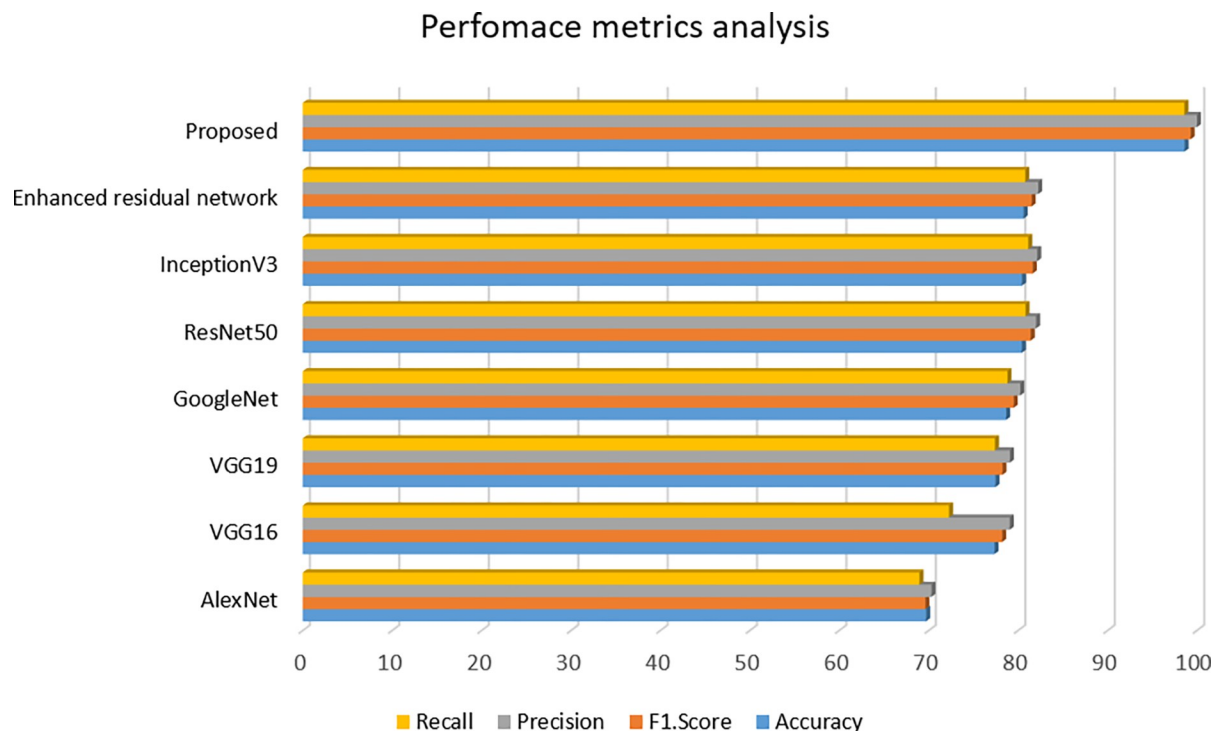
Perfomace metrics analysis

Fig 7. Performance metrics.

https://doi.org/10.1371/journal.pone.0287786.g007

1. Image resizing: For said model fitting, whole images were resized to 512 by 512-pixel resolution.

2. Image per-processing: used to reduce the various sequences of image data to bring these into ratios despite keeping the original image's knowledge construction and striving to minimize image distortion.

3. Dataset separation and training. That section contains a collection of randomly sampled images for suggested tests and computed results.

4. Testing and validation. The images are used to assess the model being tested, and additional images from certain modelling are utilized to validate the model's efficacy.

6. Comparison with chest X-ray dataset

We tested our proposed model with existed model and dataset [64] in terms of creating a more direct comparison across our technique and cutting-edge approaches in real medical applications. Through this experiment, we executed multiclass classification on just a dataset containing Chest X-rays to replicate the treatment of our technique for identifying juvenile pneumonia. There were 5232 images in the dataset including normal and pneumonia. The

Table 4. T-test for proposed and 2nd-best model (Enhanced residual network [63]).

Models	P value for Accuracy	P value for F1.Score
Enhanced residual network	$0.0269 < 0.05$	$0.02189 < 0.05$
Proposed		

https://doi.org/10.1371/journal.pone.0287786.t004

Table 5. Chest X-ray dataset performance comparison with various existing models.

Model	Accuracy	F1 Score
Proposed	90.0	93.67
AlexNet [65]	74.20 ±1.48	82.57 ±1.4
Net in Net [66]	77.40 ± 0.91	86.57±0.95
VGG [59]	78.53±0.96	86.68±1.01
GoogLeNet [67]	79.65± 0.91	86.34±1.47
ResNet [68]	81.25± 0.98	86.40±1.05
DenseNet [69]	84.53±0.83	89.02±0.94
NIN-based [64]	84.46±0.69	89.86±0.70
GoogLeNet-based [64]	86.70±0.82	91.26±0.87

https://doi.org/10.1371/journal.pone.0287786.t005

Table below displays the results of the multiclass exercise performed on the dataset. Table 5 and Fig 8 showed that our technique outperforms existing approaches in general.

7. Comparison with the Invasive Ductal Carcinoma (IDC) dataset

IDC [70] represents the most prevalent form of breast cancer, accounting for over 80% of all cases. Unfortunately, because of the absence of distinguishing characteristics, it is challenging to identify IDC as just a distinct histological type such as lobular with tubular carcinoma. Throughout this study, we measure the effectiveness of our methods against the performances of CNNs mostly on the BHI dataset [71]. This collection contains 274 histopathology presentation images showing IDC tissue areas from 274 individuals, which were scanned using thought the entire scanner. Each experiment is run on subgroups of the complete BHI dataset using ready-to-use image patches. The very first collection includes 269 patient presentations with 272494 patched, comprising 76303 successful but also 196191 negatives. According to Table 6 and Fig 9, our strategy outperforms the other techniques in terms of efficiency and F1 measure. From several viewpoints, it has been demonstrated that CNNs achieve higher statistical
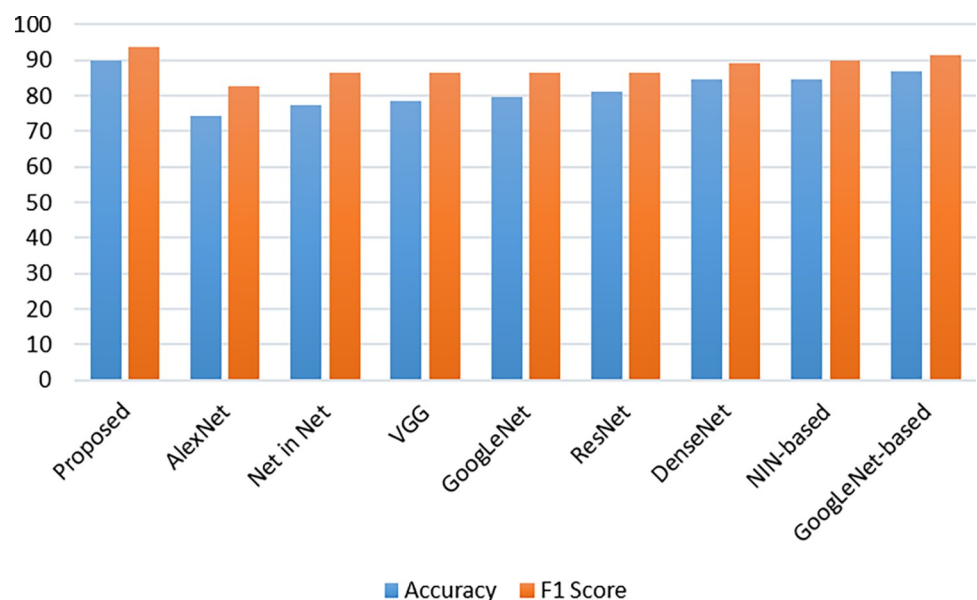

Fig 8. Performance comparison of multi-classification with chest X-ray dataset.

https://doi.org/10.1371/journal.pone.0287786.g008

Table 6. IDC dataset performance comparison with various existing models.

Model	Accuracy	F1 Score
Proposed	90.0	94.49
AlexNet [65]	84.55 ±0.71	78.78 ±0.79
Net in Net [66]	85.16 ± 0.54	79.08±0.58
VGG [59]	83.77±0.69	78.42±0.65
GoogLeNet [67]	85.23± 0.49	78.74±0.52
ResNet [68]	87.54± 0.51	85.37±0.50
DenseNet [69]	87.62±0.45	85.22±0.56
NIN-based [64]	87.03±0.30	85.60±0.30
GoogLeNet-based [64]	88.88±0.25	86.33±0.22

https://doi.org/10.1371/journal.pone.0287786.t006

parameters, indicating that their performances are un-stabilized because they are less sensitive to modeling initialization versus our techniques. Based on these findings, we may conclude that our model provided a viable technique for classification tasks using medical image datasets.

8. Comparison with COVID19-CT dataset

This COVID-19-CT dataset [72] includes various medical images gathered by [73]. It contained 349 COVID-19 realistic CT scan images with 397 normal but rather negative CT scan images from different disorders. The images in this collection varied in size between 143×76 through 1637×1225. Our assessment findings from the proposed method and many of the most sophisticated classification methods upon that COVID-19-CT dataset are shown in Table 7. Our experimental results confirmed whether deeper or broader networks typically exhibited better classification performance, which was aided by the complicated network topology. The suggested model's efficiency and F1 value achieved the greatest performance throughout this dataset, according to Fig 10 below.
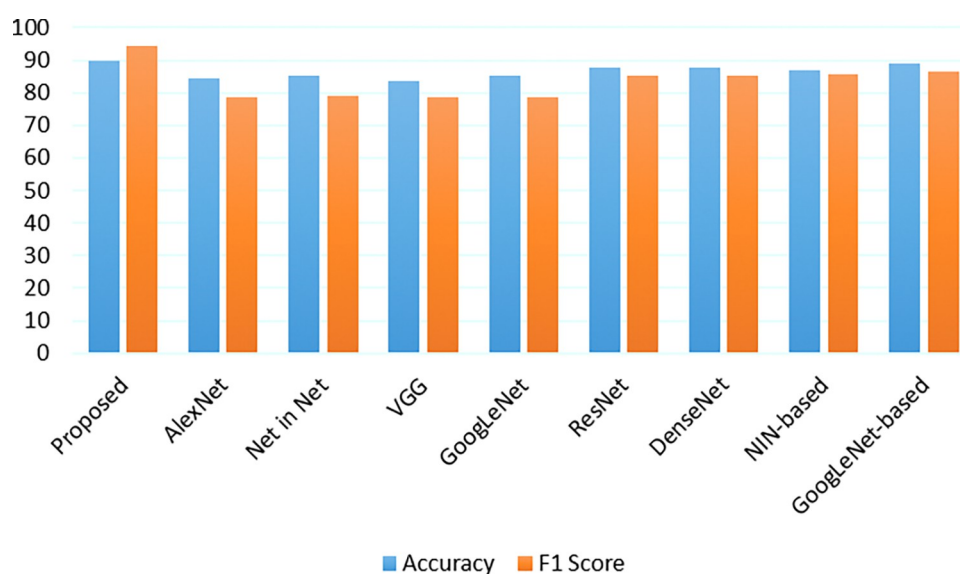

Fig 9. Performance comparison of multi-classification with IDC dataset.

https://doi.org/10.1371/journal.pone.0287786.g009

Table 7. COVID19-CT dataset performance comparison with various existing models.

Model	Accuracy	F1 Score
Proposed	88.89	95.35
VGG-16 [59]	66	58
ResNet-18 [74]	67	66
ResNet-50 [74]	72	73
DenseNet-121 [69]	76	77
DenseNet-169 [69]	80	79
EfficientNet-b0 [75]	72	71
EfficientNet-b1 [75]	70	62
ShuffleNet1.0 X (G = 4) [76]	72	72
ShuffleNet1.0 X (G = 8) [76]	71	72
CRNet [73]	72	76
ShuffleNetV2 (1.0X) [76]	74	74
ShuffleNetV2 (1.5X) [76]	73	76
SENet-50 [77]	76	77
CBAM-50 [78]	78	80
ResNeXt-50 [79]	72	75
Res2Net-50 [80]	73	74
ECA Net-50 [81]	75	74
SKNet-50 [82]	77	76
ResGANet-50(G = 2) [83]	80	81
ResGANet-101(G = 2) [83]	78	81

https://doi.org/10.1371/journal.pone.0287786.t007

9. Comparison with ISIC2018

The ISIC2018 [84] skin lesion diagnostic dataset has been used. This dataset has a total number of 10,015 images divided into seven subcategories. Melanocytic nevus (6705), melanoma (1113), dermatofibroma (115), benign keratosis (1099), actinic keratosis (327), vascular lesion, and basal cell carcinoma (514) become the names of these conditions (142). To make comparisons between ResNet and its derivatives [76, 77, 82, 85]. In each task, we explored 50- as well as 101-layer models. Furthermore, we compared several lightweight techniques. The difficulty level with Shuffle Net [86] is 1.0. These findings demonstrated in Table 8 and Fig 11 that the
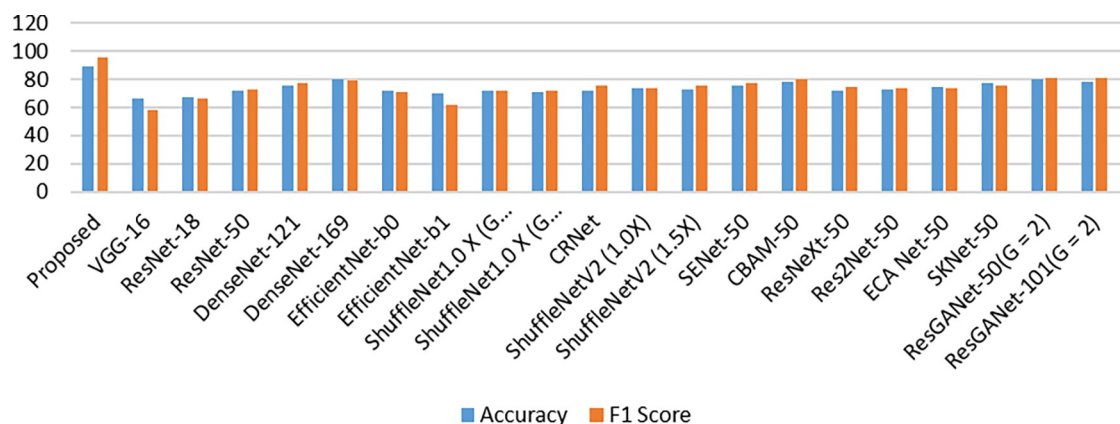

Fig 10. Performance comparison of multi-classification with COVID19-CT dataset.

https://doi.org/10.1371/journal.pone.0287786.g010

Table 8. ISIC2018 dataset performance comparison with various existing models.

Model	Precision
Proposed	**98.82**
ResGANet-50(G = 2) [83]	81.66
ResGANet-101(G = 2) [83]	81.13
SENet-50 [77]	77.97
CBAM-50 [85]	78.47
SKNet-101 [82]	76.61
ShuffleNetV2 [76]	77.7

https://doi.org/10.1371/journal.pone.0287786.t008

suggested model outperforms the most sophisticated ResNet variants network upon this ISIC2018 dataset for medical image categorization.

10. Conclusion and future work

In this research, the multi-modal medical image collection was generated using available public images. The convolutional neural network-based ResNet50 framework is subjected to data enhancement, database pre-processing, training, then testing approaches. The suggested model was developed and tested to enhance the performance that is assessed and compared.

When compared to most accessible datasets and approaches, the evaluation measurement parameters are relatively high and enhanced. As a result, our recommended research study significantly improved by 98.61%. Regularly enhancing the quality of multi-modal medical image evaluation and classification has become an essential part, but this model attained the maximum performance, assisting in the success of certain health sectors. The primary goal of the study is to enhance the health service. The future goal is to acquire and prepare actual datasets to be utilized in deep learning models including adversarial attacks. It is expected that various CNN models will be applied in the future with deeper image evaluation. Our work fosters and stimulates the health industry, which leads to an increase in medical education.
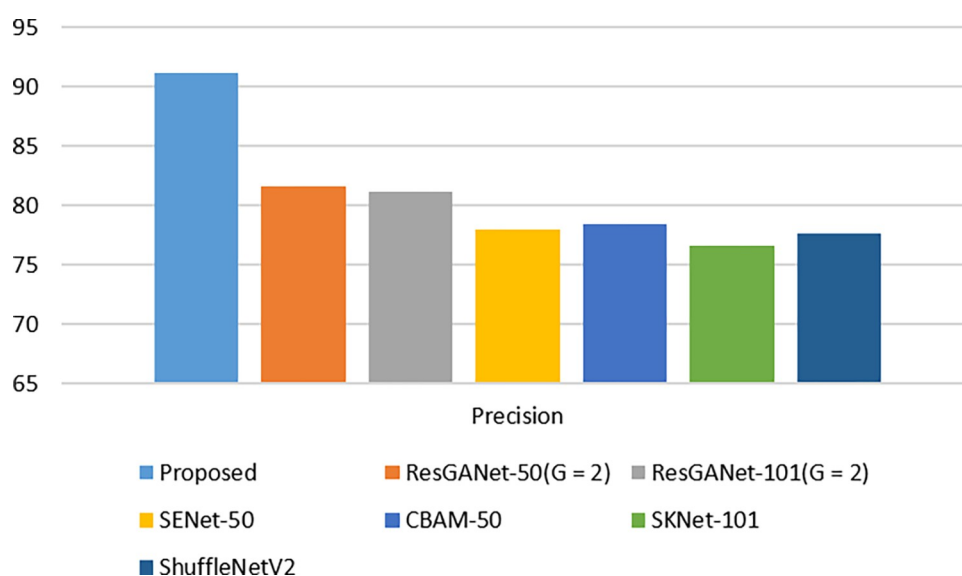

Fig 11. Performance comparison of multi-classification with ISIC2018 dataset.

https://doi.org/10.1371/journal.pone.0287786.g011

11. Limitations

The current research work is related to classification of multi modal medical images. Our dataset contained five types of medical images (i.e. endoscopy, CT, chest, hand x-ray, and lungs CT). Our model is only optimally trained for said five types of images and its accuracy can be affected if different image class included in dataset. Furthermore, our model is only optimally trained, and it is not a robust model that counters adversarial image attacks.

Supporting information

S1 File.
(RAR)

Acknowledgments

I am thankful to everyone who has allowed me to work on this research, as well as different relevant projects. Every member of my Review Committee has given me considerable individual as well as professional advice and has imparted considerable knowledge regarding scientific research as well as all aspects of life.

Author Contributions

Conceptualization: Muhammad Haris Abid, Rehan Ashraf, Toqeer Mahmood.

Formal analysis: Muhammad Haris Abid, Rehan Ashraf.

Investigation: Muhammad Haris Abid.

Methodology: Muhammad Haris Abid.

Supervision: Rehan Ashraf, Toqeer Mahmood.

Validation: Rehan Ashraf, Toqeer Mahmood, C. M. Nadeem Faisal.

Writing – original draft: Muhammad Haris Abid.

Writing – review & editing: Muhammad Haris Abid, Rehan Ashraf, Toqeer Mahmood.

References

1. Bibi R, Mehmood Z, Munshi A, Yousaf RM, Ahmed SS. Deep features optimization based on a transfer learning, genetic algorithm, and extreme learning machine for robust content-based image retrieval. PLoS One. 2022; 17: 1–30. https://doi.org/10.1371/journal.pone.0274764 PMID: 36191011

2. Khan UA, Javed A, Ashraf R. An effective hybrid framework for content based image retrieval (CBIR). Multimed Tools Appl. 2021; 80: 26911–26937. https://doi.org/10.1007/s11042-021-10530-x

3. Muhammad S, Muhammad A, Adnan M, Muhammad Q, Majdi A, Khan MK. Medical image analysis using convolutional neural networks a review. J Med Syst. 2018; 42: 1–13.

4. Xu J, Pan Y, Pan X, Hoi S, Yi Z, Xu Z. RegNet: Self-Regulated Network for Image Classification. IEEE Trans Neural Networks Learn Syst. 2022; 1–6. https://doi.org/10.1109/tnnls.2022.3158966 PMID: 35333722

5. Hassan M, Ali S, Alquhayz H, Safdar K. Developing intelligent medical image modality classification system using deep transfer learning and LDA. Sci Rep. 2020; 10: 1–14. https://doi.org/10.1038/s41598-020-69813-2 PMID: 32732962

6. Ashraf R, Ahmed M, Jabbar S, Khalid S, Ahmad A, Din S, et al. Content Based Image Retrieval by Using Color Descriptor and Discrete Wavelet Transform. J Med Syst. 2018;42. https://doi.org/10.1007/s10916-017-0880-7 PMID: 29372327

7. Nazir A, Ashraf R, Hamdani T, Ali N. Content based image retrieval system by using HSV color histogram, discrete wavelet transform and edge histogram descriptor. 2018 Int Conf Comput Math Eng

Technol Inven Innov Integr Socioecon Dev iCoMET 2018—Proc. 2018;2018-Janua: 1–6. https://doi.org/10.1109/ICOMET.2018.8346343

8. Kumar S, Singh MK, Mishra MK. Improve Content-based Image Retrieval using Deep learning model. J Phys Conf Ser. 2022;2327. https://doi.org/10.1088/1742-6596/2327/1/012028

9. Ashraf R, Ahmed M, Ahmad U, Habib MA, Jabbar S, Naseer K. MDCBIR-MF: multimedia data for content-based image retrieval by using multiple features. Multimed Tools Appl. 2020; 79: 8553–8579. https://doi.org/10.1007/s11042-018-5961-1

10. Zafar B, Ashraf R, Ali N, Iqbal MK, Sajid M, Dar SH, et al. A novel discriminating and relative global spatial image representation with applications in CBIR. Appl Sci. 2018; 8: 1–23. https://doi.org/10.3390/app8112242

11. Dureja A, Pahwa P. Integrating CNN along with FAST descriptor for accurate retrieval of medical images with reduced error probability. Multimed Tools Appl. 2022. https://doi.org/10.1007/s11042-022-13991-w

12. Sharma V. Mammogram Image Retrieval System Using Texture and Semantic Features. J Phys Conf Ser. 2022;2267. https://doi.org/10.1088/1742-6596/2267/1/012071

13. Zhou T, Cheng QR, Lu HL, Li Q, Zhang XX, Qiu S. Deep learning methods for medical image fusion: A review. Comput Biol Med. 2023;160. https://doi.org/10.1016/j.compbiomed.2023.106959 PMID: 37141652

14. Yang Y, Song X. Research on Face Intelligent Perception Technology Integrating Deep Learning under Different Illumination Intensities. 2022; 32–36. https://doi.org/10.47852/bonviewJCCE19919

15. Xu Y, Qiu TT. Human Activity Recognition and Embedded Application Based on Convolutional Neural Network. J Artif Intell Technol. 2020; 1: 51–60. https://doi.org/10.37965/jait.2020.0051

16. Fang B. Deep Generative Inpainting with Comparative Sample Augmentation. 2022; 1: 174–180. https://doi.org/10.47852/bonviewJCCE2202319

17. Gheisari M, Ebrahimzadeh F, Rahimi M, Moazzamigodarzi M, Liu Y, Dutta Pramanik PK, et al. Deep learning: Applications, architectures, models, tools, and frameworks: A comprehensive survey. CAAI Trans Intell Technol. 2023. https://doi.org/10.1049/cit2.12180

18. Rasheed A, Ali N, Zafar B, Shabbir A, Sajid M, Mahmood MT. Handwritten Urdu Characters and Digits Recognition Using Transfer Learning and Augmentation With AlexNet. IEEE Access. 2022; 10: 102629–102645. https://doi.org/10.1109/ACCESS.2022.3208959

19. Abdou MA. Literature review: efficient deep neural networks techniques for medical image analysis. Neural Comput Appl. 2022; 34: 5791–5812. https://doi.org/10.1007/s00521-022-06960-9

20. Hussain DM, Surendran D. Retraction Note: Content based image retrieval using bees algorithm and simulated annealing approach in medical big data applications (Multimedia Tools and Applications, (2020), 79, (3683–3698), 10.1007/s11042-018-6708-8). Multimed Tools Appl. 2022; 13859. https://doi.org/10.1007/s11042-022-13859-z

21. Nawaz M, Masood M, Javed A, Iqbal J, Nazir T, Mehmood A, et al. Melanoma localization and classification through faster region-based convolutional neural network and SVM. Multimed Tools Appl. 2021; 80: 28953–28974. https://doi.org/10.1007/s11042-021-11120-7

22. Li Z, Zhang X, Müller H, Zhang S. Large-scale retrieval for medical image analytics: A comprehensive review. Med Image Anal. 2018; 43: 66–84. https://doi.org/10.1016/j.media.2017.09.007 PMID: 29031831

23. Liu TY, Mahjoubfar A, Prusinski D, Stevens L. Neuromorphic computing for content-based image retrieval. PLoS One. 2022; 17: 1–13. https://doi.org/10.1371/journal.pone.0264364 PMID: 35385477

24. Algarni AD, El-Shafai W, El Banby GM, Abd El-Samie FE, Soliman NF. An efficient CNN-based hybrid classification and segmentation approach for COVID-19 detection. Comput Mater Contin. 2022; 70: 4393–4410. https://doi.org/10.32604/cmc.2022.020265

25. Gupta D, Loane R, Gayen S, Demner-Fushman D. Medical Image Retrieval Via Nearest Neighbor Search on Pre-Trained Image Features. SSRN Electron J. 2022. https://doi.org/10.2139/ssrn.4240384

26. Ran Q, Zhou Y, Hong D, Bi M, Ni L, Li X, et al. Deep transformer and few-shot learning for hyperspectral image classification. CAAI Trans Intell Technol. 2023. https://doi.org/10.1049/cit2.12181

27. Sezavar A, Farsi H. Content-based image retrieval by combining convolutional neural networks and sparse representation Content courtesy of Springer Nature, terms of use apply. Rights reserved. Content courtesy of Springer Nature, terms of use apply. Rights reserved. 2019; 20895–20912.

28. Zhu X, Ding M, Zhang X. Free form deformation and symmetry constraint-based multi-modal brain image registration using generative adversarial nets. CAAI Trans Intell Technol. 2023. https://doi.org/10.1049/cit2.12159

29. El-Shafai W, Algarni AD, Banby GME, El-Samie FEA, Soliman NF. Classification framework for COVID-19 diagnosis based on deep cnn models. Intell Autom Soft Comput. 2022; 31: 1561–1575. https://doi.org/10.32604/IASC.2022.020386

30. Masood F., Masood J., Zahir H., Driss K., Mehmood N. and HF. Novel Approach to Evaluate Classification Algorithms and Feature Selection Filter Algorithms using Medical Data. Jcce. 2022; 2, no: 57–67. https://doi.org/10.47852/bonviewJCCE2202238

31. Li X, Li C, Rahaman MM, Sun H, Li X, Wu J, et al. A comprehensive review of computer-aided whole-slide image analysis: from datasets to feature extraction, segmentation, classification and detection approaches. Artificial Intelligence Review. Springer Netherlands; 2022. https://doi.org/10.1007/s10462-021-10121-0

32. Shaukat F, Raja G, Ashraf R, Khalid S, Ahmad M, Ali A. Artificial neural network based classification of lung nodules in CT images using intensity, shape and texture features. J Ambient Intell Humaniz Comput. 2019; 10: 4135–4149. https://doi.org/10.1007/s12652-019-01173-w

33. Sajjad M, Ullah A, Ahmad J, Abbas N, Rho S, Baik SW. Integrating salient colors with rotational invariant texture features for image representation in retrieval systems. Multimed Tools Appl. 2018; 77: 4769–4789. https://doi.org/10.1007/s11042-017-5010-5

34. Chen X, Wang X, Zhang K, Fung KM, Thai TC, Moore K, et al. Recent advances and clinical applications of deep learning in medical image analysis. Med Image Anal. 2022;79. https://doi.org/10.1016/j.media.2022.102444 PMID: 35472844

35. Zafar B, Ashraf R, Ali N, Ahmed M, Jabbar S, Chatzichristofis SA. Image classification by addition of spatial information based on histograms of orthogonal vectors. PLoS One. 2018; 13: 1–26. https://doi.org/10.1371/journal.pone.0198175 PMID: 29883455

36. Camalan S, Niazi MKK, Moberly AC, Teknos T, Essig G, Elmaraghy C, et al. OtoMatch: Content-based eardrum image retrieval using deep learning. PLoS One. 2020; 15: 1–16. https://doi.org/10.1371/journal.pone.0232776 PMID: 32413096

37. Azizi S, Mustafa B, Ryan F, Beaver Z, Freyberg J, Deaton J, et al. Big Self-Supervised Models Advance Medical Image Classification. Proc IEEE Int Conf Comput Vis. 2021; 3458–3468. https://doi.org/10.1109/ICCV48922.2021.00346

38. Ashraf R, Habib MA, Akram M, Latif MA, Malik MSA, Awais M, et al. Deep Convolution Neural Network for Big Data Medical Image Classification. IEEE Access. 2020; 8: 105659–105670. https://doi.org/10.1109/ACCESS.2020.2998808

39. Raza A, Nawaz T, Dawood H. Square texton histogram features for image retrieval Content courtesy of Springer Nature, terms of use apply. Rights reserved. Content courtesy of Springer Nature, terms of use apply. Rights reserved. 2019; 2719–2746.

40. Panda S, Jangid M, Jain A. A Comprehensive Review on the Significance and Impact of Deep Learning in Medical Image Analysis. Proc Int Conf Technol Adv Innov ICTAI 2021. 2021; 358–366. https://doi.org/10.1109/ICTAI53825.2021.9673189

41. Garg M, Dhiman G. A novel content-based image retrieval approach for classification using GLCM features and texture fused LBP variants. Neural Comput Appl. 2021; 33: 1311–1328. https://doi.org/10.1007/s00521-020-05017-z

42. Ashraf R, Afzal S, Rehman AU, Gul S, Baber J, Bakhtyar M, et al. Region-of-Interest Based Transfer Learning Assisted Framework for Skin Cancer Detection. IEEE Access. 2020; 8: 147858–147871. https://doi.org/10.1109/ACCESS.2020.3014701

43. Wang X, Li Z, Huang Y, Jiao Y. Multimodal medical image segmentation using multi-scale context-aware network. Neurocomputing. 2022; 486: 135–146. https://doi.org/10.1016/j.neucom.2021.11.017

44. Lepcha DC, Dogra A, Goyal B, Chohan JS, Koundal D. Multimodal Medical Image Fusion Based on Pixel Significance Using Anisotropic Diffusion and Cross Bilateral Filter.

45. Kale M, Mukhopadhyay S. Efficient color image retrieval method using deep stacked sparse autoencoder. J Electron Imaging. 2022; 31: 1–23. https://doi.org/10.1117/1.jei.31.2.023003

46. Karthik K, Kamath SS. A deep neural network model for content-based medical image retrieval with multi-view classification. Vis Comput. 2021; 37: 1837–1850. https://doi.org/10.1007/s00371-020-01941-2

47. Breznik E, Wetzer E, Lindblad J, Sladoje N. Cross-Modality Sub-Image Retrieval using Contrastive Multimodal Image Representations. 2022. Available: http://arxiv.org/abs/2201.03597

48. Shamna P, Govindan VK, Abdul Nazeer KA. Content based medical image retrieval using topic and location model. J Biomed Inform. 2019; 91: 103112. https://doi.org/10.1016/j.jbi.2019.103112 PMID: 30738189

49. Barz B, Denzler J. Content-based image retrieval and the semantic gap in the deep learning era. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 2021; 12662 LNCS: 245–260. https://doi.org/10.1007/978-3-030-68790-8_20

50. Öztürk S. Convolutional neural network based dictionary learning to create hash codes for content-based image retrieval. Procedia Comput Sci. 2021; 183: 624–629. https://doi.org/10.1016/j.procs.2021.02.106

51. Qin J, Li H, Xiang X, Tan Y, Pan W, Ma W, et al. An Encrypted Image Retrieval Method Based on Harris Corner Optimization and LSH in Cloud Computing. IEEE Access. 2019; 7: 24626–24633. https://doi.org/10.1109/ACCESS.2019.2894673

52. Arun Pandian J, Geetharamani G, Annette B. Data Augmentation on Plant Leaf Disease Image Dataset Using Image Manipulation and Deep Learning Techniques. Proc 2019 IEEE 9th Int Conf Adv Comput IACC 2019. 2019; 199–204. https://doi.org/10.1109/IACC48062.2019.8971580

53. Nevavuori P, Narra N, Lipping T. Crop yield prediction with deep convolutional neural networks. Comput Electron Agric. 2019; 163: 104859. https://doi.org/10.1016/j.compag.2019.104859

54. Chen J, Chen J, Zhang D, Sun Y, Nanehkaran YA. Using deep transfer learning for image-based plant disease identification. Comput Electron Agric. 2020; 173: 105393. https://doi.org/10.1016/j.compag.2020.105393

55. WU W, YANG T le, LI R, CHEN C, LIU T, ZHOU K, et al. Detection and enumeration of wheat grains based on a deep learning method under various scenarios and scales. J Integr Agric. 2020; 19: 1998–2008. https://doi.org/10.1016/S2095-3119(19)62803-0

56. Alencastre-Miranda M, Johnson RM, Krebs HI. Convolutional Neural Networks and Transfer Learning for Quality Inspection of Different Sugarcane Varieties. IEEE Trans Ind Informatics. 2021; 17: 787–794. https://doi.org/10.1109/TII.2020.2992229

57. Fei-Fei L, Deng J, Li K. ImageNet: Constructing a large-scale image database. J Vis. 2010; 9: 1037–1037. https://doi.org/10.1167/9.8.1037

58. Qayyum A, Anwar SM, Awais M, Majid M. Medical image retrieval using deep convolutional neural network. Neurocomputing. 2017; 266: 8–20. https://doi.org/10.1016/j.neucom.2017.05.025

59. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 3rd Int Conf Learn Represent ICLR 2015—Conf Track Proc. 2015; 1–14.

60. Takiyama H, Ozawa T, Ishihara S, Fujishiro M, Shichijo S, Nomura S, et al. Automatic anatomical classification of esophagogastroduodenoscopy images using deep convolutional neural networks. Sci Rep. 2018; 8: 1–8. https://doi.org/10.1038/s41598-018-25842-6 PMID: 29760397

61. Sangeetha V, Prasad KJR. Syntheses of novel derivatives of 2-acetylfuro[2,3-a]carbazoles, benzo[1,2-b]-1,4-thiazepino[2,3-a]carbazoles and 1-acetyloxycarbazole-2- carbaldehydes. Indian J Chem—Sect B Org Med Chem. 2006; 45: 1951–1954. https://doi.org/10.1002/chin.200650130

62. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2016;2016-Decem: 2818–2826. https://doi.org/10.1109/CVPR.2016.308

63. Owais M, Arsalan M, Choi J, Park KR. Effective diagnosis and treatment through content-based medical image retrieval (CBMIR) by using artificial intelligence. J Clin Med. 2019;8. https://doi.org/10.3390/jcm8040462 PMID: 30959798

64. Yang Y, Hu Y, Zhang X, Wang S. Two-Stage Selective Ensemble of CNN via Deep Tree Training for Medical Image Classification. IEEE Trans Cybern. 2022; 52: 9194–9207. https://doi.org/10.1109/TCYB.2021.3061147 PMID: 33705343

65. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM. 2017; 60: 84–90. https://doi.org/10.1145/3065386

66. Shin HC, Orton MR, Collins DJ, Doran SJ, Leach MO. Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data. IEEE Trans Pattern Anal Mach Intell. 2013; 35: 1930–1943. https://doi.org/10.1109/TPAMI.2012.277 PMID: 23787345

67. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2015;07-12-June: 1–9. https://doi.org/10.1109/CVPR.2015.7298594

68. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 2016; 9908 LNCS: 630–645. https://doi.org/10.1007/978-3-319-46493-0_38

69. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. Proc - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017. 2017;2017-Janua: 2261–2269. https://doi.org/10.1109/CVPR.2017.243

70. Ma J, Jemal A. Breast cancer statistics. Breast Cancer Metastasis Drug Resist Prog Prospect. 2013; 9781461456: 1–18. https://doi.org/10.1007/978-1-4614-5647-6_1

71. https://www.kaggle.com/datasets/paultimothymooney/breast-histopathology-images.

72. Yang X, He X, Zhao J, Zhang Y, Zhang S, Xie P. COVID-CT-Dataset: A CT Scan Dataset about COVID-19. 2020; 1–14. Available: http://arxiv.org/abs/2003.13865

73. He X, Yang X, Zhang S, Zhao J, Zhang Y, Xing E, et al. Sample-efficient deep learning for COVID-19 diagnosis based on CT scans. IEEE Trans Med Imaging. 2020; XX: 10.

74. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2016;2016-Decem: 770–778. https://doi.org/10.1109/CVPR.2016.90

75. Louis M. 20:21. Can J Emerg Med. 2013; 15: 190. https://doi.org/10.2310/8000.2013.131108 PMID: 23663470

76. Ma N, Zhang X, Zheng HT, Sun J. Shufflenet V2: Practical guidelines for efficient cnn architecture design. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 2018; 11218 LNCS: 122–138. https://doi.org/10.1007/978-3-030-01264-9_8

77. Hu J, Shen L, Sun G. Squeeze-and-Excitation Networks. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2018; 7132–7141. https://doi.org/10.1109/CVPR.2018.00745

78. Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional block attention module. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 2018; 11211 LNCS: 3–19. https://doi.org/10.1007/978-3-030-01234-2_1

79. Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. Proc - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017. 2017; 2017-Janua: 5987–5995. https://doi.org/10.1109/CVPR.2017.634

80. Gao SH, Cheng MM, Zhao K, Zhang XY, Yang MH, Torr P. Res2Net: A New Multi-Scale Backbone Architecture. IEEE Trans Pattern Anal Mach Intell. 2021; 43: 652–662. https://doi.org/10.1109/TPAMI.2019.2938758 PMID: 31484108

81. Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2020; 11531–11539. https://doi.org/10.1109/CVPR42600.2020.01155

82. Li X, Wang W, Hu X, Yang J. Selective kernel networks. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2019;2019-June: 510–519. https://doi.org/10.1109/CVPR.2019.00060

83. Cheng J, Tian S, Yu L, Gao C, Kang X, Ma X, et al. ResGANet: Residual group attention network for medical image classification and segmentation. Med Image Anal. 2022; 76: 102313. https://doi.org/10.1016/j.media.2021.102313 PMID: 34911012

84. Codella N, Rotemberg V, Tschandl P, Celebi ME, Dusza S, Gutman D, et al. Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC). 2019; 1–12. Available: http://arxiv.org/abs/1902.03368

85. Zhu L, Yang Y. for Few-Shot Video Classification. Springer International Publishing; 2018. https://doi.org/10.1007/978-3-030-01234-2

86. Zhang X, Zhou X, Lin M, Sun J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2018; 6848–6856. https://doi.org/10.1109/CVPR.2018.00716