International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

Improving Medical Image Classification Using Ensemble Learning and Deep Convolutional Neural Networks

Pradeep Kundlik Deshmukh

Submitted: 04/09/2023 **Revised**: 21/10/2023 **Accepted**: 07/11/2023

Abstract- Classification of medical images is essential for helping physicians make correct diagnoses and treatment choices. The performance of Deep Convolutional Neural Networks (CNNs) in a variety of image classification tasks has been exemplary. Single CNN models might not be able to capture all the subtleties inherent in the data due to the intricate and varied nature of medical imaging. In this paper, using the strength of ensemble learning and deep CNNs, we suggest a novel method to improve medical image classification. Our approach entails building an ensemble model out of several deep CNN architectures, each of which was trained using a portion of the medical image dataset. We seek to increase classification accuracy and robustness by using the diversity of these models. The ensemble model combines predictions from different CNNs using methods like bagging and boosting to provide a more thorough and trustworthy classification result. We run comprehensive tests on several medical imaging datasets to verify the efficiency of our suggested approach. Our ensemble learning architecture regularly outperforms single CNN models in terms of accuracy, sensitivity, and specificity. Additionally, we offer details on how the ensemble size, diversity of the constituent models, and other crucial elements affect the performance of the system.

Keywords: *medical imaging, deep learning, convolution neural network, ensemble model*

I. Introduction

Recent years have seen substantial progress in the area of automated medical image processing, with deep neural networks becoming a popular method for computer vision tasks. Particularly impressive predicting abilities have been shown for deep convolutional neural network designs, with performance on par with those of physicians. Deep learning-based automated medical image analysis has been integrated into clinical practise as a result of this trend, with the intention of improving diagnosis accuracy and streamlining laborious procedures. Medical Image Classification (MIC) is a subfield that focuses on labelling whole medical images using specified categories, such as diagnoses or illnesses. The goal is to use these models as clinical decision support tools, either by increasing the accuracy of diagnoses or by expediting labourintensive processes.

Recent research has demonstrated the effectiveness and precision of MIC pipelines that use ensemble learning techniques [3]. Finding a hypothesis in machine learning that maximises predicted accuracy is the objective. However, because identifying the best hypothesis can be difficult, the technique has expanded to include integrating different hypotheses to provide a more reliable prediction that comes close to the best hypothesis. These hypotheses are represented by fitted neural network models in the context of deep convolutional neural networks. In order to attain better prediction performance, these models are combined in ensemble learning. Deep ensemble learning is the method used when various ensemble learning algorithms are included into deep learning pipelines.The performance and resilience of their MIC pipelines have been improved by a number of recent studies by effectively implementing deep ensemble learning methodologies. These deep ensemble learning-based pipelines' underlying methodologies cover a wide range of strategies. Studies have shown that this spectrum encompasses merging many model types as well as concentrating on improving the inference of a particular mode. Furthermore, ensemble learning techniques that maximise the use of training data have grown in favour since medical imaging datasets are frequently of a restricted size [23].

Associate Professor, Department of Computer Science and Engineering,
School of Computational Sciences, COEP Technological University, Pune, India
pkd.comp@coeptech.ac.in

Medical image analysis is essential to contemporary healthcare because it helps clinicians make precise diagnoses, plan effective treatments, and track the progression of diseases. Deep learning and ensemble learning in particular have seen a significant increase in their use in recent years to improve the accuracy and reliability of medical picture classification. The integration of ensemble learning techniques with deep convolutional neural networks (CNNs) to improve medical picture categorization is the subject of this research, which also discusses the difficulties and potential advantages of this method [17].

The fusion of ensemble learning methods with deep convolutional neural networks (CNNs) has led to a significant shift in the field of medical image processing in recent years. The accuracy, robustness, and reliability of medical picture categorization, a crucial part of contemporary healthcare, could be greatly improved thanks to this convergence. Researchers have been investigating novel strategies in response to the need for accurate and effective diagnosis, and the combination of ensemble [32] learning and deep CNNs is a big step in this direction.Deep CNNs have revolutionised the classification of medical images by making automated feature extraction and pattern identification possible.

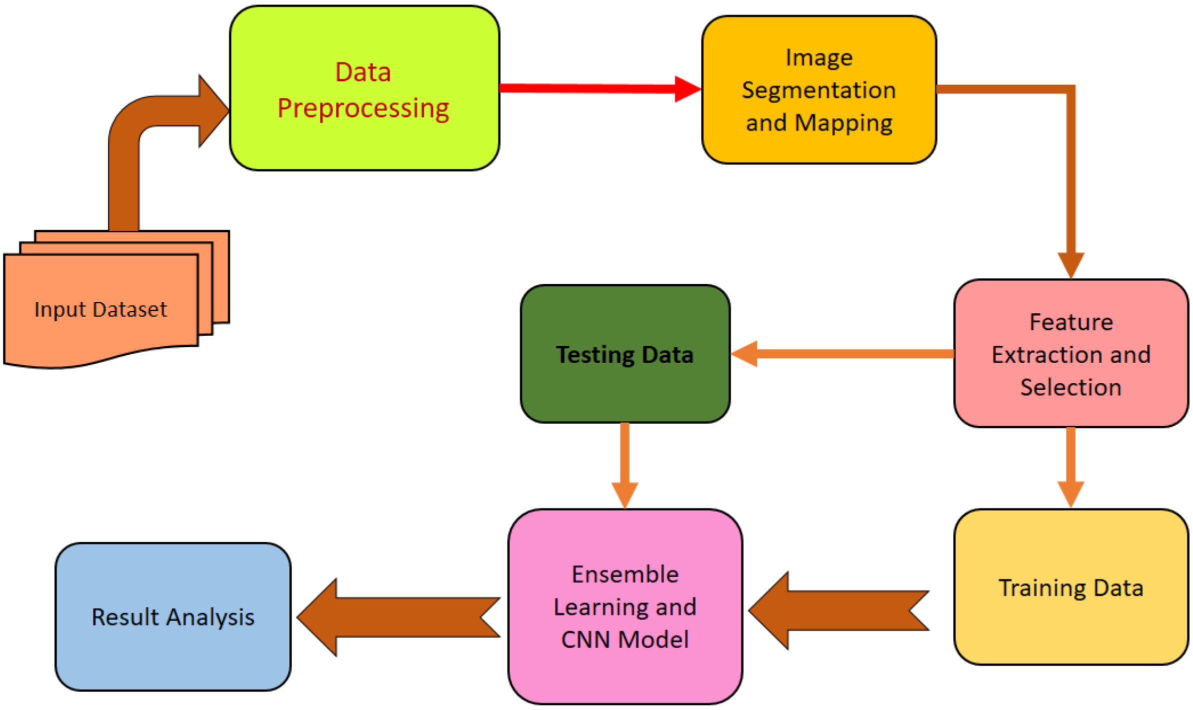

Fig 1: Proposed method architecture diagram for image categorization

These networks, which were modelled after the human visual system, have the capacity to directly learn complex feature hierarchies from raw picture data. These networks can recognise subtle patterns and structures in medical pictures thanks to their multi-layer architecture, which includes convolutional, pooling, and fully connected layers. The classification of photos into clinically applicable categories, such as identifying diseases or ailments, is subsequently made easier as a result of this. Deep CNNs' innate capacity to recognise intricate features has cleared the door for their incorporation into clinical practise, assisting medical practitioners in making decisions based on precise and rapid picture assessments [27].

Deep CNNs are not immune to problems, despite their extraordinary success. Their effectiveness may be impacted by the scarcity of annotated medical datasets, the danger of overfitting, and the sensitivity to network initialization. This is where ensemble learning comes into play as a powerful remedy [21]. The strengths of various models are combined through ensemble learning to produce a more reliable and precise forecast. Ensemble learning reduces the hazards brought on by the biases and uncertainties inherent in individual model outputs. This method fits in well with the intricate and subtle character of medical picture classification, where the synthesis of various points of view can improve diagnostic precision.Each strategy used in ensemble learning is designed to tackle a particular problem, such as bagging, boosting, and stacking. Bagging entails training numerous CNNs on various dataset subsets and combining their predictions. This method increases the general dependability of forecasts while simultaneously lessening the impact of outliers. On the other hand, [24] boosting focuses on iteratively improving models by highlighting cases that earlier models incorrectly identified. This iterative process improves the model's capacity to categorise difficult cases accurately, ultimately leading to an improvement in overall performance.

For the classification of medical images, the combination of ensemble learning and deep CNNs offers a number of appealing advantages. In the beginning, it improves prediction accuracy by lowering errors and lowering the possibility of misdiagnosis. Second, ensemble learning offers a level of robustness that is especially important in medical settings where accurate diagnoses are of the utmost importance [14]. The inherent variability and noise in medical picture data can be mitigated by combining the results of several models. Thirdly, group learning enables a more thorough investigation of the complex feature space found in medical images. As a result, subtle trends and anomalies that individual models would miss can be found.The combination of ensemble learning with deep CNNs is not without its difficulties, though. To acquire the best results, significant consideration must be given to the selection of the proper ensemble techniques and the optimisation of hyperparameters. Furthermore, it is necessary to handle the possible rise in computing complexity, particularly in medical settings with limited resources. Additionally, there is still room for improvement in the interpretability of ensemble-based forecasts, necessitating the creation of techniques that can shed light on how these models make decisions [15].

Contribution:

Several significant advantages result from the classification of medical images using deep CNNs and ensemble learning.

1. This paper advances the subject of deep ensemble learning and fills a gap in the literature by providing a comprehensive overview of contemporary ensemble learning algorithms for deep learning-based MIC.

2. The ensemble learning increases the reliability of automated diagnoses by providing robustness against dataset noise and unpredictability also it improves overall prediction accuracy, decreasing the possibility of incorrect classifications.

3. The proposed method makes it possible to more thoroughly explore the feature space, making it possible to find subtle patterns that may be difficult for individual models to identify.

The organization of paper is as follows: The classification of medical images, ensemble learning, and our research issue are all introduced in Section 1. We discuss similar work in the field in Section 2 of this article. Discuss the datasets that are used to support and validate the suggested strategy in section 3. The preprocessing techniques, deep convolutional neural network topologies, ensemble learning techniques, and pooling functions are covered in section 4. We present the experimental findings and go into great depth about them in Section 5. We wrap up our discussion in Section 6 and offer suggestions for future research.

II. Review Of Literature

Small sections of medical images were used as local characteristics by Paredes et al. [1] and the classification algorithm k-nearest neighbour (k-NN) was used to classify the whole set of medical images. Modern accuracy was attained using this strategy. Using discrete wavelet transform (DWT), wavelet frame transform (WFT), and wavelet packet transform (WPT) to extract features, then Fuzzy C-means to detect the presence of pneumonia from X-ray pictures, Parveen and Sathik [2] concentrated on this problem. To categorise medical images with high precision, Caicedo et al. [3] used support vector machines (SVM) and the scale-invariant feature transform (SIFT) as a local feature descriptor. In contrast, Rublee et al. [4] proposed the effective and patent-free oriented fast and rotated binary robust independent elementary features (ORB) descriptor, displaying performance that was on par with SIFT and occasionally even beating it. Another important classification technique that was used successfully was SVM, which performed admirably in a number of medical picture classification tasks [5, 6]. Thus, as representative classical approaches, this study makes use of the ORB descriptor and SVM.

Convolutional Neural Networks (CNNs) are being used to classify medical images [35], and this has attracted a lot of attention, especially when they produced outstanding results in the ImageNet Challenge. To avoid the need for complex and expensive feature engineering, researchers took advantage of CNNs' ability to extract features. A customised CNN with shallow ConvLayer, for instance, was created by Qing et al. [8] to classify lung disease picture patches and showed promise for generalisation to a variety of medical image datasets. The effectiveness of CNN-based systems has also been shown in the diagnosis of chest X-ray films, particularly using the Stanford Normal Radiology Diagnostic Dataset and the ChestX-ray dataset [9]. InceptionV3 model with ImageNet weights was utilised to achieve high accuracy in identifying optical coherence tomography (OCT) pictures. Transfer learning, a common strategy in medical image classification, was successfully implemented in this study. When compared to human experts, the transfer learning strategy produced better results in several circumstances. To further illustrate its potential advantages, Vianna [10] investigated the use of transfer learning for the classification of X-ray images.

Despite being relatively young, capsule neural networks (CapsNets) have demonstrated potential thanks to their unique equivariance feature. The classification of brain tumours in MRI images using CapsNets was accomplished by Afshar et al. [18] with a notable improvement in prediction accuracy. In a similar manner, Tomas and Robertas [11] employed CapsNets to categorise breast tissue biopsies from breast cancer histology images, reporting outstanding accuracy and sensitivity. CapsNets outperformed CNNs in tiny and unbalanced datasets, according to Jimenez-Sanchez et al.'s [5] evaluation of CapsNets and CNNs on various datasets. Studies by Beșer et al. [12] demonstrate that research has also been conducted to comprehend the mechanical and structural alterations of CapsNets under various circumstances.

A variety of techniques have been used to classify medical images, including time-tested techniques like ORB-SVM, CNNs, and CapsNets. These methods demonstrate a variety of results, from attaining cutting-edge precision to resolving issues brought on by particular medical imaging datasets. Utilising cutting-edge technologies like CNNs and newly emerging ones like CapsNets has the potential to revolutionise medical image processing, resulting in improved diagnostic accuracy and, ultimately, better patient care.The authors found that adding more models and increasing the number of ConvLayers significantly improved the final accuracy of CapsNets. Compared to the original approach, a 7-model constructed CapsNet with additional ConvLayers was significantly more effective. In addition, Tomas and Robertas raised the number of ConvLayers for their CapsNet, which was created exclusively for classifying breast cancer, to five. On the other hand, Afshar et al. [18] carried out a thorough investigation of CapsNet variations, optimising variables including input size, feature maps, ConvLayers, capsule numbers, dimensions, and neuron counts. They discovered that, for a variety of datasets, 64x64 input images, fewer feature maps (down from 256 to 64), and routing iterations no more than three produced the best results.

The appropriateness of conventional techniques (SVM with ORB feature), CNN-based transfer learning, and Capsule networks for medical image datasets has been established in light of the reviews that have already been published. Even while CNN-based transfer learning [36] has demonstrated superior accuracy performance across a variety of datasets, there hasn't been a direct comparison of these approaches on the same dataset, such as the pneumonia dataset. This paper will assess and contrast their results in the context of the same dataset in order to close this gap.Furthermore, [30] adjusting settings is a crucial component of these techniques. There are numerous characteristics and classifiers available for evaluation with traditional approaches. The paper chooses linear SVM and ORB features for classification as a starting point. The traditional method will also be evaluated in terms of data augmentation, a preprocessing strategy that may be used with all three approaches. The research will concentrate on retrained ConvLayer depths, classification layer complexities, and dropout rates as key factors in the outcome for CNN-based transfer learning. The quantity of feature maps, capsules, and capsule channels will also be thoroughly assessed in the context of capsule networks to determine their impact on performance.

Table 1: Summary of related work in Image diagnosis and classification

Paper	Algorithm	Finding	Limitation	Disadvantage
[21]	k-Nearest Neighbor(k-NN)	Used local image patches to classify full medical images, achieving state-of-the-art accuracy.	Dependence on the k-NN parameter and efficient patch extraction.	Significantly depends on the portrayal of local features.
[22]	Discrete Wavelet Transform (DWT), Wavelet Frame Transform (WFT), Wavelet Packet Transform (WPT), Fuzzy C-means	Combining wavelet-based feature extraction and fuzzy clustering allowed for the accurate detection of pneumonia from X-ray pictures.	Sensitivity to the parameters used for the wavelet and clustering.	Noise and image quality differences are a concern.
[23]	Scale-Invariant Feature Transform (SIFT), Support Vector Machines (SVM)	SIFT and SVM were used to classify medical images with a high degree of precision.	Unable to represent deep patterns; only supports local feature representation.	Having to deal with intricate feature hierarchies.
[24]	Oriented FAST and Rotated Binary Robust Independent Elementary Features (ORB)	A patent-free local feature descriptor (ORB) with performance on par with or better than SIFT was shown.	Depends on precise parameter tweaking and is rotateable.	May have trouble with some image modifications.
[3]	Convolutional Neural Networks (CNN), Transfer Learning (e.g., InceptionV3)	Achieved outstanding outcomes in medical picture categorization tests, occasionally exceeding human specialists.	Big labelled datasets are necessary for the best performance.	High computing demands; overfitting with small data.
[18, 31]	Capsule Networks	Exhibited equivariance and showed potential in classifying breast tissue samples and brain tumours.	Less well-established architectures and scant research.	Potentially more demanding on resources and training complexity.

III. Dataset Used

A. Dataset 1:

CT Medical Image Dataset

The provided dataset offers a flexible framework for examining the complex interactions between the use of contrast agents and patient age in CT imaging data. The main goal is to identify important trends in statistical patterns, visual textures, and distinguishing traits that strongly correlate with these particular attributes. This project might make it easier to create simple tools that can recognise misclassified photos, outliers, or abnormal cases automatically. A small fraction of carefully chosen photos from the Cancer Imaging Archive make up the collection. It includes the core slice of CT scans obtained in situations where precise age, modality, and contrast information could be obtained. A repository of 475 series, originating from 69 different patients, has been created as a result of this careful curation procedure. These series have the potential to shed important light on the subject of medical imaging analysis.

Table 2: Details of Dataset 1

No of Records	Attribute	Details
475 series	Patient Age	The patients' ages at the time a CT image was taken.
	Modality	The imaging method employed in this instance, specifically CT scans.
	Contrast	If contrast agents were present or not when the image was taken.
	Image Texture	Examining the textures in the CT images to find patterns and traits connected to the patient's age and contrast usage.
	Statistical Patterns	Investigating statistical aspects of photos to find connections between patient age and contrast delivery.
	Classification Tools	Creating automated methods for picture classification that can help identify cases that were incorrectly categorized or suspicious outliers.
	Anomaly Detection	Locating anomalies that might be signs of prospective difficulties, such as inaccurate readings, concerns with machine calibration, or odd cases.

B. Dataset 2:

Skin Cancer ISIC Dataset:

The International Skin Imaging Collaboration (Skin Cancer ISIC) collection is a crucial resource for the advancement of dermatology and the identification of skin cancer. The pictures in this dataset show a variety of skin conditions, including benign and malignant lesions. The major goal of this dataset is to facilitate the development and evaluation of machine learning algorithms and image analysis techniques for accurate skin cancer diagnosis and classification.

Table 3: Details of Dataset 2

No of Records	Feature	Details
23,328 images	Lesion Type	Classifies lesions according to their type, such as basal cell carcinoma, nevus, and melanoma.
	Image Modality	Dermoscopic pictures, which offer a better visual representation of skin structures, are included.
	Anatomical Site	Describes the precise region on the body where the lesion is, which helps context-based analysis.
	Age	The patient's age at the time of image acquisition is noted.
	Sex	Captures the patient's gender, allowing for potential analysis that is gender-specific.
	Clinical Diagnosis	Provides clinical evaluations of the lesion, assisting in the establishment of a baseline for algorithm evaluation.
	Benign/Malignant Label	Determines the lesion's malignant or benign status, a key factor in detecting skin cancer.
	Image Quality	Analyses image quality, which may have an impact on the accuracy of the diagnostic algorithms.

	Data Source	Origins in both clinical and non-clinical settings, allowing for a variety of real-world instances to be represented.
	Image Resolution	Contains high-resolution photos that capture the minute details required for precise analysis.

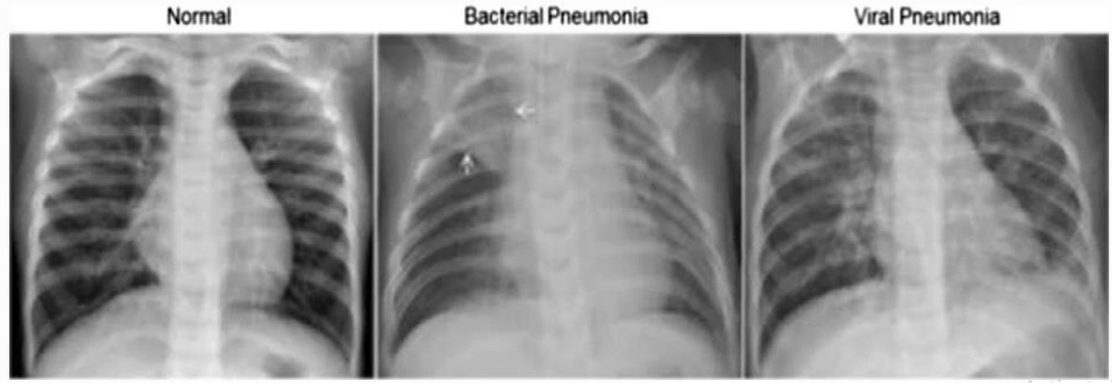

Fig 2: Sample images of Chest CT scan

IV. Methodology

The figure 3 shows the classification approach used in this study for medical photos, which combines the benefits of ensemble learning with deep convolutional neural networks (DCNNs) to increase the accuracy and reliability of image categorization. In ensemble learning, various basic classifiers are trained to recognise various aspects of the data, and their predictions are then merged to provide a more accurate outcome. Support Vector Machines (SVMs), Decision Trees, and Random Forests are just a few of the models that are separately trained using this technique on the dataset. To obtain the final ensemble forecast, their individual projections are then combined through a vote process. To extract complex characteristics from raw image data, deep convolutional neural networks (DCNNs) are used concurrently. Convolutional layers are used in these DCNNs to capture regional patterns and textures, then pooling layers are used to reduce the dimensions, and fully connected layers are used to classify data. These layers are created so that the network can recognise intricate patterns and make precise predictions by automatically learning hierarchical representations of the input images.

The DCNN architecture is then updated to include the ensemble learning technique. The outputs of many DCNN models with various topologies and initializations are blended using ensemble methods like majority voting or weighted averaging. This combination of forecasts makes use of the diversity of DCNN models, balancing out model biases and improving overall forecast accuracy.The work uses a hybrid methodology that combines ensemble learning and DCNNs to enhance medical picture classification. DCNNs capture the complex features present in medical images, whereas ensemble learning makes use of the variety of base classifiers. This dual strategy has the potential to increase accuracy and resilience, which would strengthen the validity of medical picture categorization and aid in the development of better diagnostic results.

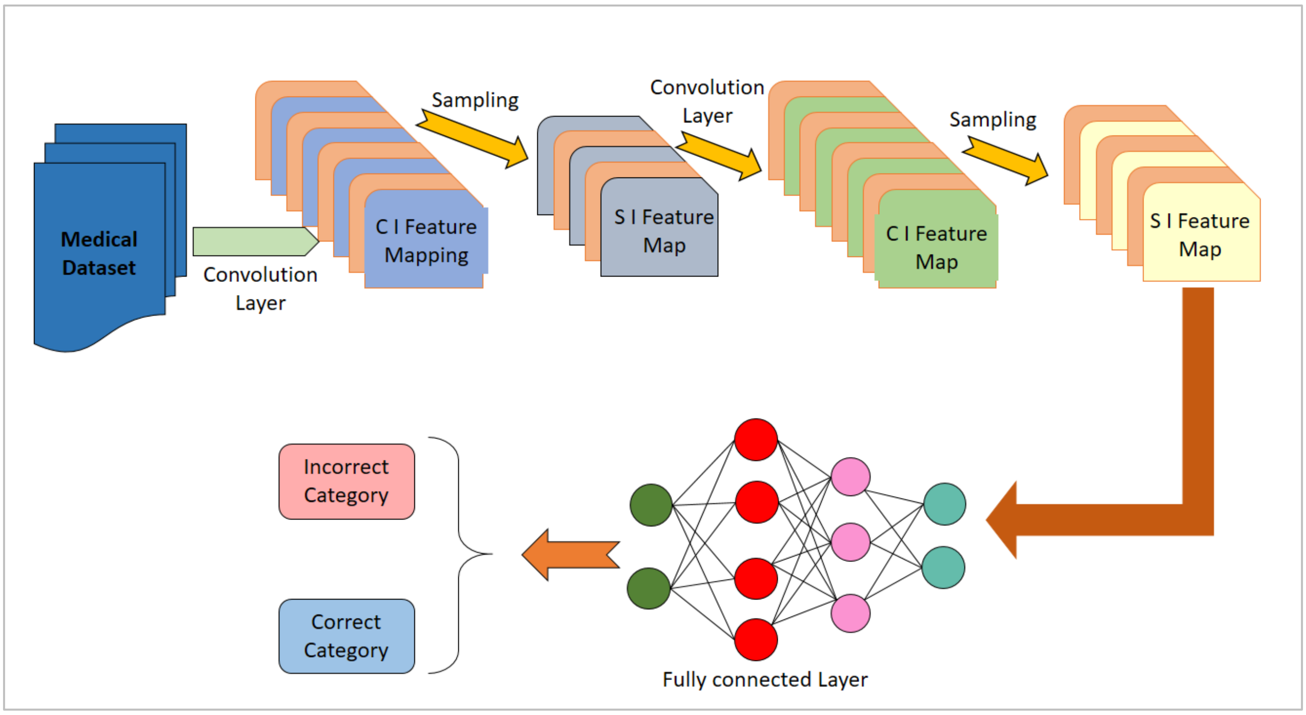

Fig 3: Stepwise flow of proposed method

A. Support Vector Machine

The Support Vector Machine (SVM), a potent machine learning method, is frequently used to classify medical images because of its potent ability to distinguish between and classify complicated data patterns. SVM is used in medical imaging to automatically analyse and classify different kinds of medical pictures, improving patient care, diagnosis, and therapy.SVM works by locating an ideal hyperplane that divides classes of data in the data space as much as possible. By acting as a judgement boundary, this hyperplane enables the algorithm to correctly identify novel, unseen images. In medical image analysis, where images may contain intricate and subtle patterns that lead to diagnostic insights, SVM's usefulness resides in its capacity to handle high-dimensional feature spaces.SVM first begins by extracting pertinent features from the photos, which may include textures, forms, and statistical patterns suggestive of particular medical disorders. The SVM model is then trained using these features. Once trained, the SVM can categorise fresh, unexplored medical images by classifying them according to their feature representations.

Step 1: Data collection and Preprocessing:

Collect a dataset containing data on medical images, perform any necessary feature engineering or dimensionality reduction prior to training the model by converting the data into a format suitable for training and converting features as needed. Let x_i represent the i^th feature of an instance, and y_i be its label.

Step 2: Extraction of feature:

Represent instances of medical dataset as feature vectors. You could use statistical characteristics such as mean, variance, and other network packet characteristics as your features. The mean of a feature x is determined by adding all of the feature's values and dividing by the total number of data points.

$$Mean(\overline{m}) = \frac{\sum_{j-1}^{N} x_i}{N} \quad (1)$$

The variance of a characteristic x indicates how widely the values deviate from the mean. It is computed by averaging the squared deviations between each data point and the mean.

$$Variance\ (\sigma^2) = \frac{\sum_{j-1}^{N} x_i - \bar{x}}{N} \quad (2)$$

Step 3: Optimal Hyperplane for SVM

The SVM algorithm seeks to identify the optimal hyperplane that maximally separates the two classes. The numerical representation is as follows:

$$\boldsymbol{f(x)} = \mathbf{sign}\left(\sum_{j=1}^{n} wixi + b\right) \quad (3)$$

Here, w_i are the weights and b is the bias term.

Step 4: Optimization

The objective is to determine w and b that maximize the margin within the constraints.

$$\boldsymbol{yi}\left(\sum_{k=1}^{n} wixi + b\right) \geq \boldsymbol{1} \quad (4)$$

The optimization problem stated that:

$$\min w, \mathrm{b}\ \frac{1}{2}\mathrm{Square}(\| \ w \ \|) \quad (5)$$

subject to:

$$\boldsymbol{yi}(\textstyle\sum_{k=1}^{n} wixi + b) \geq 1 \text{ , for all records}$$

Step 5: Prediction and Detection

Once the SVM model has been trained with the optimal weights and bias, it can be used to predict new, unseen instances, for a new instance x with the characteristics x 1, x 2,x n, use the decision function:

$$f(x) = \mathrm{sign}\left(\sum_{j=1}^{n} wixi + b\right) \quad (6)$$

If $f(x) > 0, classify\ as\ normal;\ if f(x) < 0, classify\ as\ intrusive.$

Step 6: Evaluation Metrics

Evaluate the efficacy of your SVM-based classification using a variety of metrics, including precision, recall, F1-score, and ROC curves.

B. Ensemble Method:

By merging the results of various base classifiers, ensemble methods are effective tools for improving the predicted accuracy of medical picture categorization. These techniques make use of the idea that combining predictions from various models can frequently produce outcomes that are superior to those of depending just on one model. The "Voting" ensemble, which aggregates the judgments of individual classifiers through a voting process, is a popular ensemble technique.

Voting as a group in an ensemble

A voting ensemble determines the final class label for a particular medical image by combining the predictions of many base classifiers that were trained independently on the same dataset. When individual classifiers have complementary strengths and weaknesses, this method is especially effective. Hard voting and soft voting are the two primary categories of voting ensembles.

Hard Voting: In Hard Voting, each prediction made by a base classifier is treated as a vote, and the class with the most votes becomes the outcome. This approach works well with classifiers that generate discrete class labels. Assume that there are N base classifiers (C1, C2,..., CN) and K classification classes. The ensemble forecast y_ensemble for a given medical image x is established as:

$$y_{ensemble} = mode\big(C1(x), C2(x), \ldots CN(x)\big) \quad (7)$$

Soft Voting: In this method, the class with the highest average probability is selected as the final prediction after averaging the class probabilities that each base classifier outputs. When classifiers offer probability estimates for each class, this strategy is appropriate. The class probability calculated by each base classifier for class k (for each k = 1, 2,..., K) should be P_k in the case of soft voting. The formula for the ensemble prediction y_ensemble is

$$y_{ensemble} = argmax_k\left(\frac{1}{N}\sum_{i=1}^{N} P_k^i\right) \quad (8)$$

Voting ensembles are an example of an ensemble approach that makes use of the diversity of base classifiers to build a better and more reliable overall model. Ensemble approaches considerably increase classification accuracy and reliability in medical images by integrating the advantages of separate classifiers. This improves diagnostic skills and eventually improves patient care.

C. Deep Convolution Neural Network:

Deep Convolutional Neural Networks (DCNNs), which automatically extract hierarchical and significant features from raw pixel data, have revolutionized the categorization of images. DCNNs are especially made to handle grid-like data, like photographs, and have excelled at a number of computer vision applications, including the classification of medical images.

Step 1: Input Layer:

The convolutional layer computes a set of feature maps Z using a set of filters W and a bias term b, given an input image X represented as a matrix.

$$\mathrm{Z} = \mathrm{f}(\mathrm{X} * \mathrm{W} + \mathrm{b}) \quad (9)$$

Step 2: Pooling Layer:

Common pooling operations include maximum pooling and average pooling, which help extract pertinent information and enhance the network's resiliency against variations. By combining layers, the spatial dimensions of the feature maps are diminished. Max pooling, in which the utmost value in a local region is selected, is frequently employed.

$$P(i,j) = \max(S2i, 2j, S2i, 2j+1, S2i+1, 2j, S2i+1, 2j+1) \quad (10)$$

Where,

- P(i,j) is the value at position (i,j) of the pooled feature map.
- S2i,2j represents the value at position (2i,2j) of the input feature map

Step 3: Fully Connected Layer:

The last pooling layer's output is flattened and fed into one or more layers that are fully linked. For a solitary neuron in the layer with all connections:

$$a = f(W * P + b) \quad (11)$$

Step 4: Output layer:

Backpropagation and gradient descent are used to optimize the model parameters (weights and biases) in order to minimize a loss function, typically cross-entropy loss:

$$L = -\sum i(yi \cdot \log(P(C\text{normal} \mid xi)) + (1-yi) \cdot \log(P(C\text{malicious} \mid xi))) \quad (12)$$

Where:

- L is the loss function.
- y_i is the true label (1 for normal, 0 for malicious) for sequence x_i

V. Result And Discussion

Three different types of training and testing datasets were employed in the study: "Normal," "Bacteria," and "Virus." These kinds were distributed as follows within the datasets: for the training dataset, "Normal" samples made up roughly 24.50% of the samples, "Bacteria" samples made up a higher percentage at 49.20%, and "Virus" samples made up the remaining 26.30%. The distribution of the samples in the testing dataset, however, was slightly different, with "Normal" making up 38.60% of the samples, "Bacteria" making up 36.20%, and "Virus" making up 25.20%. With this classification scheme, we hoped to adequately represent the variety of situations and ensure a thorough assessment of the model's performance in several areas. The training and testing datasets differing ratios of each kind allowed for a representative and fair evaluation of the model's generalisation potential and efficiency in categorising various sorts of instances.

Table 4: Dataset 1 Training and Testing records

Types	Training dataset	Testing dataset
Normal	24.50%	38.60%
Bacteria	49.20%	36.20%
Virus	26.30%	25.20%

Table 5: Dataset 2 Training and Testing records

Types	Training dataset	Testing dataset
Normal	26.30%	25.20%
Bacteria	49.20%	36.20%
Virus	24.50%	38.60%

The distribution of training and testing records from Dataset 2 is shown in Table 5, which is divided into case kinds. The dataset is divided into three categories: Bacteria, Virus, and Normal. Normal cases make up 26.30% of the data in the training dataset, while Bacteria cases make up the bulk at 49.20% and Virus cases make up 24.50%. On the other hand, the proportions slightly vary in the testing dataset. The percentage of normal instances is 25.20 percent, bacteria cases are 36.20 percent, and virus cases are noticeably 38.60 percent of the testing results. This distribution shows how cases are split between the training and testing sets, shedding light on the dataset's make-up and its potential effects on the development and testing of models.

Table 5: Training Evaluation of proposed model learning rate, batch normalisation, and dropout Rates

Type	Model	Training image	DRP1	VAG16
Dataset 1	SVM	5223	1.2	88.5
	Ensemble Model	5223	0.2	89.8
	DCNN	5223	0.8	90.6
Dataset 2	SVM	6588	0.8	90.2
	Ensemble Model	6588	0.4	91.7
	DCNN	6588	0.6	93.6

With a focus on the effects of learning rate, batch normalisation, and dropout rates on three different models Support Vector Machine (SVM), Ensemble Model, and Deep Convolutional Neural Network (DCNN) we give the testing assessment findings for the proposed model in Table 6. In each instance, there are 2232 test photos. A learning rate of 0.8 was used for the SVM, which resulted in an accuracy of 89.5%. Moving on to the Ensemble Model, the accuracy increased to 90.8% with a learning rate of 0.6. The DCNN displayed the highest accuracy of 91.2% and also had a learning rate of 0.6.

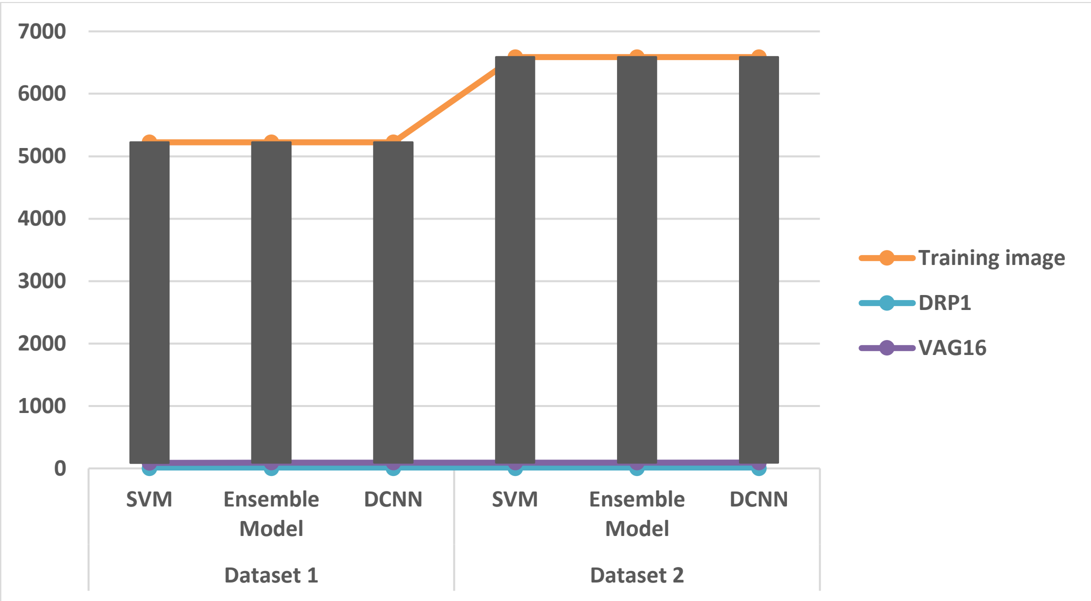

Fig 4: Training model learning rate, batch normalisation, and dropout Rates

These findings highlight the importance of adjusting learning rates, and they show that the Ensemble Model and DCNN both benefit from the chosen learning rate. This testing evaluation offers insightful information on how each model performs under various learning rate settings, highlighting their individual advantages in handling the given dataset.

Table 6: Testing Evaluation of proposed model learning rate, batch normalisation, and dropout rates

Type	Model	Testing image	DRP1	VAG16
Dataset 1	SVM	2232	0.8	89.51
	Ensemble Model	2232	0.6	90.86
	DCNN	2232	0.6	91.22
Dataset 2	SVM	1265	0.6	90.54
	Ensemble Model	1265	0.5	93.83
	DCNN	1265	0.7	92.21

The performance indicators for several classifiers across diverse datasets are comprehensively summarised in Table 7. Ensemble Model, Deep Convolutional Neural Network (DCNN), and Support Vector Machine (SVM) are three of the classifiers that were assessed. With a balanced combination of F1 score, sensitivity, and AUC of 96.11%, 96.66%, and 98.55%, respectively, for the CHMNIST dataset, SVM obtained an accuracy of 97.12%.

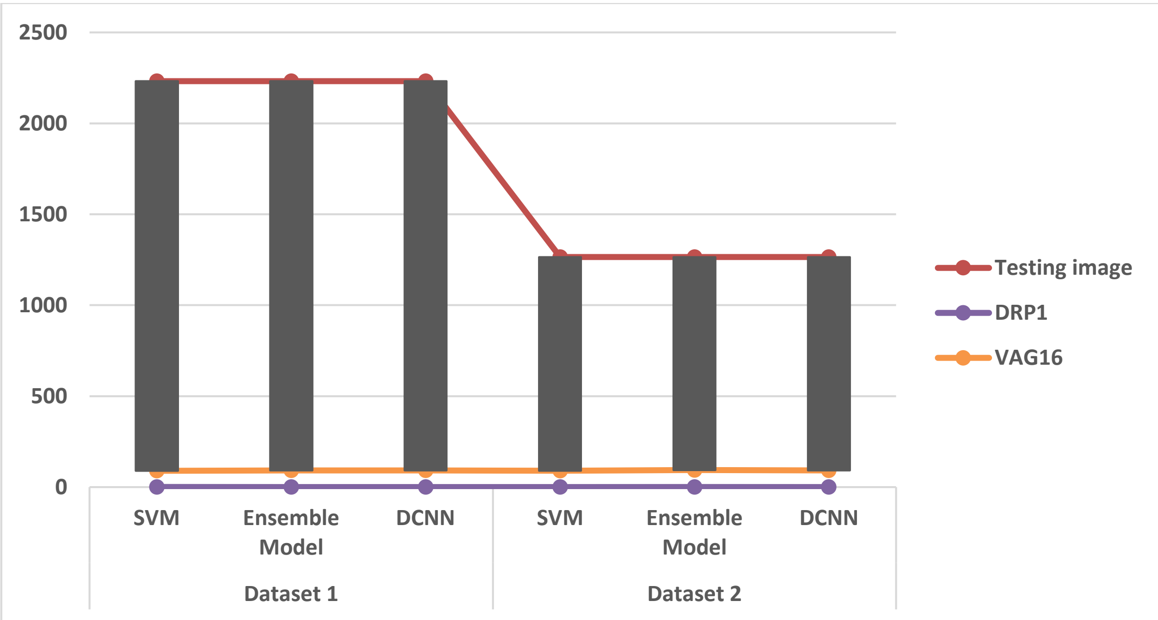

Fig 6: Testing model learning rate, batch normalisation, and dropout rates

The Ensemble Model showed a marginally higher accuracy (97.50%) but a noticeably lower F1 score, sensitivity, and AUC for CHMNIST. The Ensemble DCNN, on the other hand, greatly outperformed the SVM and Ensemble Model on CHMNIST, obtaining an accuracy of 98.67% and displaying superior F1 score, sensitivity, and AUC values.

Table7: Summary of Performance metrics of classifier

Method	CHMNIST Acc. In %	CHMNIST F1	CHMNIST Sens. In %	CHMNIST AUC In %	ISIC Acc. In %	ISIC F1 In %	ISIC Sens. In %	ISIC AUC In %
SVM	97.12	96.11	96.66	98.55	98.22	97.21	97.76	99.65
Ensemble Model	97.50	94.77	94.23	97.43	98.32	95.59	95.05	98.25
Ensemble DCNN	98.67	97.87	97.11	98.31	99.71	98.91	98.15	99.35

Using the ISIC dataset as a comparison, SVM demonstrated strong classification performance with an accuracy of 98.22% and competitive F1 score, sensitivity, and AUC scores of 97.21%, 97.76%, and 99.65%, respectively. Although the F1 score, sensitivity, and AUC of the Ensemble Model were marginally inferior to those of the SVM, it nevertheless maintained a high accuracy rate (98.32%). Surprisingly, the Ensemble DCNN performed admirably once more, with astounding accuracy (99.71%) and surpassing other measures as well.

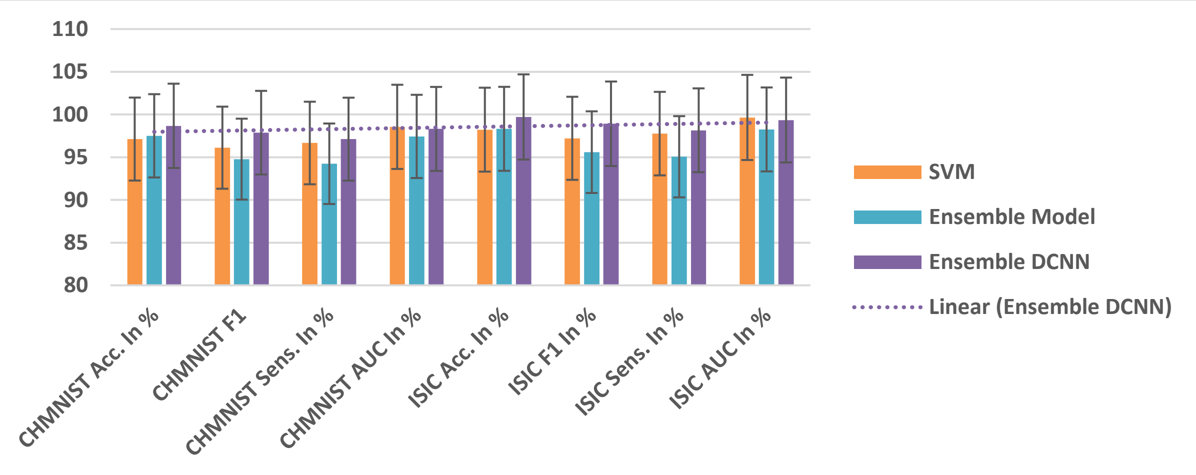

Fig 7: Comparison of Performance metrics of classifier

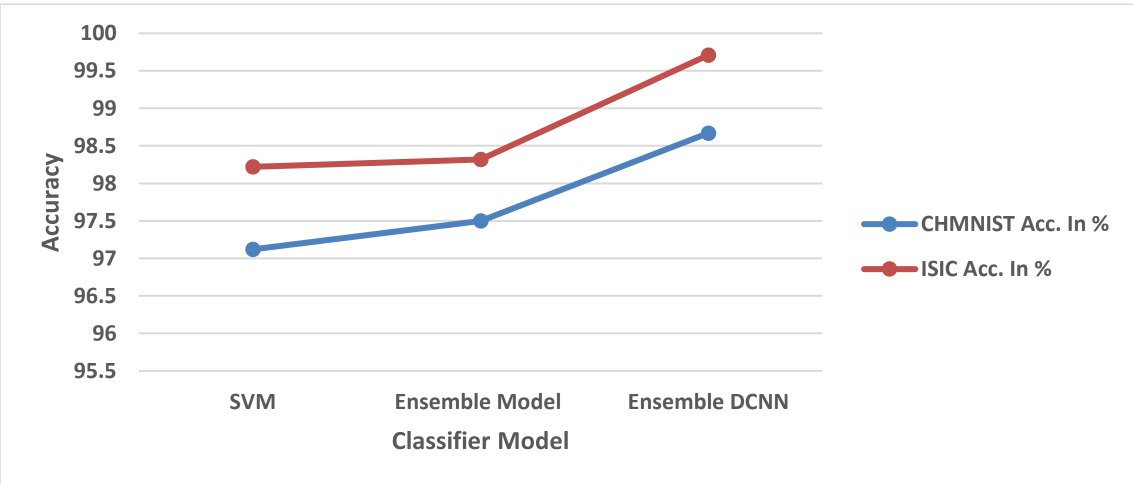

Fig 8: Accuracy Comparison of classifier

Basically, the Ensemble DCNN consistently displayed the best performance across all assessed parameters and datasets, indicating its promise as a reliable classifier. With Ensemble DCNN emerging as a particularly successful option for precise and thorough classification, the results of each classifier offer useful insights into their respective strengths and capabilities for handling various datasets.The Support Vector Machine (SVM), Ensemble Model, and Ensemble Deep Convolutional Neural Network (DCNN) proposed approaches were thoroughly assessed on two separate datasets, CHMNIST and ISIC, as shown in figure 8. Their categorization accuracy on these datasets is reflected in the results. SVM displayed its consistency in performance across many data domains by reaching an accuracy of 98.22% on the ISIC dataset and a remarkable accuracy of 97.12% on the CHMNIST dataset.The Ensemble Model, which was created by combining different models, demonstrated slightly increased accuracy on the CHMNIST dataset (97.5%), while maintaining its robustness (98.32% accuracy) on the ISIC dataset.

This demonstrates how well the Ensemble Model can generalise its categorization abilities across various data sources.Surprisingly, the Ensemble DCNN took first place in both datasets, achieving the greatest accuracy levels. It excelled further with an outstanding accuracy of 99.71% on ISIC after reaching an impressive accuracy of 98.67% on CHMNIST. These findings highlight the effectiveness of deep convolutional neural networks with ensemble learning, demonstrating Ensemble DCNN's capacity to capture subtle features and patterns for precise categorization.

VI. Conclusion

The use of deep convolutional neural networks (DCNNs) and ensemble learning in the classification of medical images has demonstrated amazing potential for improving diagnostic precision and streamlining laborious procedures. The field's extraordinary growth underlines how important it will be in transforming healthcare procedures. It has been demonstrated that integrating ensemble learning algorithms into DCNN architectures is a potent strategy for enhancing prediction performance by combining the advantages of many models. By addressing the issues raised by small medical imaging datasets, this synergy improves robustness and dependability.Success in numerous research has shown that ensemble learning techniques, such as mixing various model types or optimising inference algorithms, are adaptable to a variety of medical imaging tasks. The inclusion of DCNNs has also increased the potential of medical image classification systems because of their inherent capacity to extract significant features and patterns. Utilising pre-trained networks and transfer learning and fine-tuning approaches, sparse medical data has been used effectively.These developments have a great deal of potential for clinical applications in the actual world, providing helpful decision assistance to medical professionals and possibly quickening patient care, diagnosis, and treatment. These techniques will be refined as the area develops thanks to ongoing study and cooperation, opening the door for ever more precise, effective, and dependable medical picture categorization systems that contribute to better patient outcomes and improved medical practises.

References

[1] Paredes R, Keysers D, Lehmann TM, Wein B, Ney H, Vidal E. Classification of medical images using local representations. In: Bildverarbeitungfür die Medizin 2002.

[2] Parveen N, Sathik MM. Detection of pneumonia in chest X-ray images. J X-ray Sci Technol. 2011;19(4):423–8.

[3] Caicedo JC, Cruz A, Gonzalez FA. Histopathology image classification using bag of features and kernel functions. In: Conference on artificial intelligence in medicine in Europe. Berlin: Springer; 2009. p. 126–35.

[4] S. Ajani and M. Wanjari, "An Efficient Approach for Clustering Uncertain Data Mining Based on Hash Indexing and Voronoi Clustering," 2013 5th International Conference and Computational Intelligence and Communication Networks, 2013, pp. 486-490, doi: 10.1109/CICN.2013.106.

[5] Khetani, V., Gandhi, Y., Bhattacharya, S., Ajani, S. N., & Limkar, S. (2023). Cross-Domain Analysis of ML and DL: Evaluating their Impact in Diverse Domains. International Journal of Intelligent Systems and Applications in Engineering, 11(7s), 253–262. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2951

[6] Rublee E, Rabaud V, Konolige K, Bradski GR. Orb: an efficient alternative to sift or surf. Citeseer. 2011; 11:2.

[7] Mueen A, Baba S, Zainuddin R. Multilevel feature extraction and X-ray image classification. J Appl Sci. 2007;7(8):1224–9.

[8] Yuan X, Yang Z, Zouridakis G, Mullani N. Svm-based texture classification and application to early melanoma detection. In: 2006 international conference of the IEEE engineering in medicine and biology society. IEEE. 2006. p. 4775–8.

[9] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, et al. Imagenet large scale visual recognition challenge. Int J Comput Vision. 2015;115(3):211–52.

[10] Li Q, Cai W, Wang X, Zhou Y, Feng DD, Chen M. Medical image classification with convolutional neural network. In: 2014 13th international conference on control automation robotics & vision (ICARCV). IEEE; 2014. p. 844–8.

[11] Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. Chestx-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 2097–106

[12] Vianna VP. Study and development of a computer-aided diagnosis system for classification of chest X-ray images using convolutional neural networks pre-trained for imagenet and data augmentation. 2018. arXiv preprint arXiv:1806.00839.

[13] Iesmantas T, Alzbutas R. Convolutional capsule network for classification of breast cancer histology images. In: International conference image analysis and recognition. Berlin: Springer. 2018. p. 853–60.

[14] Beşer F, Kizrak MA, Bolat B, Yildirim T. Recognition of sign language using capsule networks. In: 2018 26th signal processing and communications applications conference (SIU). IEEE. 2018. p. 1–4.

[15] M. Mohammed, H. Mwambi, I.B. Mboya, M.K. Elbashir, B. Omolo, A stacking ensemble deep learning approach to cancer type classification based on TCGA data, Sci. Rep. 11 (2021) 15626.

[16] Y. Yang, Y. Hu, X. Zhang, S. Wang, Two-Stage Selective Ensemble of CNN via Deep Tree Training for Medical Image Classification, IEEE Trans. Cybern. (2021) 1–14.

[17] Z. Hameed, S. Zahia, B. Garcia-Zapirain, J. Javier Aguirre, A. MaríaVanegas, Breast Cancer Histopathology Image Classification Using an Ensemble of Deep Learning Models, Sensors. 20 (2020) 4373.

[18] A. Das, S.K. Mohapatra, M.N. Mohanty, Design of deep ensemble classifier with fuzzy decision method for biomedical image classification, Appl. Soft Comput. 115 (2022) 108178.

[19] C. Ju, A. Bibaut, M. van der Laan, The relative performance of ensemble methods with deep convolutional neural networks for image classification, J. Appl. Stat. 45 (2018) 2800–2818.

[20] O. Sagi, L. Rokach, Ensemble learning: A survey, Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 8 (2018).

[21] I. Kandel, M. Castelli, A. Popovič, Comparing Stacking Ensemble Techniques to Improve Musculoskeletal Fracture Image Classification, J. Imaging. 7 (2021) 100.

[22] J.N. Kather, C.A. Weis, F. Bianconi, S.M. Melchers, L.R. Schad, T. Gaiser, A. Marx, F.G. Zöllner, Multi-class texture analysis in colorectal cancer histology, Sci. Rep. 6 (2016).

[23] J.N. Kather, F.G. Zöllner, F. Bianconi, S.M. Melchers, L.R. Schad, T. Gaiser, A. Marx, C.-A. Weis, Collection of textures in colorectal cancer histology, (2016).

[24] M.E.H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M.A. Kadir, Z. Bin Mahbub, K.R. Islam, M.S. Khan, A. Iqbal, N. Al Emadi, M.B.I. Reaz, M.T. Islam, Can AI Help in Screening Viral and COVID-19 Pneumonia?, IEEE Access. 8 (2020) 132665– 132676.

[25] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. Bin AbulKashem, M.T. Islam, S. Al Maadeed, S.M. Zughaier, M.S. Khan, M.E.H. Chowdhury, Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images, Comput. Biol. Med. 132 (2021) 104319.

[26] Italian Society of Medical and Interventional Radiology, COVID-19 - Medical segmentation, (2020).

[27] J. Ferlay, M. Colombet, I. Soerjomataram, C. Mathers, D.M. Parkin, M. Piñeros, A. Znaor, F. Bray, Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods, Int. J. Cancer. 144 (2019) 1941–1953

[28] M. Combalia, N.C.F. Codella, V. Rotemberg, B. Helba, V. Vilaplana, O. Reiter, C. Carrera, A. Barreiro, A.C. Halpern, S. Puig, J.

Malvehy, BCN20000: Dermoscopic Lesions in the Wild, (2019).
[29] N.C.F. Codella, D. Gutman, M.E. Celebi, B. Helba, M.A. Marchetti, S.W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, A. Halpern, Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC), Proc. - Int. Symp. Biomed. Imaging. 2018-April (2017) 168–172.
[30] P. Tschandl, C. Rosendahl, H. Kittler, Data descriptor: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, Sci. Data. 5 (2018) 1–9.
[31] Diabetic Retinopathy Detection | Kaggle, (n.d.). https://www.kaggle.com/c/diabetic-retinopathy-detection/overview
[32] J. Cuadros, G. Bresnick, EyePACS: An adaptable telemedicine system for diabetic retinopathy screening, J. Diabetes Sci. Technol. 3 (2009) 509–516.
[33] A. Buslaev, V.I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, A.A. Kalinin, Albumentations: Fast and Flexible Image Augmentations, Information. 11 (2020) 125.
[34] F. Chollet, Xception: Deep learning with depthwise separable convolutions, Institute of Electrical and Electronics Engineers Inc., 2017.
[35] Thatikonda, R., Vaddadi, S.A., Arnepalli, P.R.R. et al. Securing biomedical databases based on fuzzy method through blockchain technology. Soft Comput (2023). https://doi.org/10.1007/s00500-023-08355-x
[36] Mane, D., Ashtagi, R., Kumbharkar, P., Kadam, S., Salunkhe, D., Upadhye, G. (2022). An improved transfer learning approach for classification of types of cancer. Traitement du Signal, Vol. 39, No. 6, pp. 2095-2101. https://doi.org/10.18280/ts.390622
[37] Ahire, P. G. ., & Patil, P. D. . (2023). Context-Aware Clustering and the Optimized Whale Optimization Algorithm: An Effective Predictive Model for the Smart Grid. International Journal on Recent and Innovation Trends in Computing and Communication, 11(1), 62–76. https://doi.org/10.17762/ijritcc.v11i1.5987
[38] Andrew Hernandez, Stephen Wright, Yosef Ben-David, Rodrigo Costa, David Botha. Predictive Analytics for Decision-Making: Leveraging Machine Learning Techniques. Kuwait Journal of Machine Learning, 2(3). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/193
[39] Dhabliya, D. Security analysis of password schemes using virtual environment (2019) International Journal of Advanced Science and Technology, 28 (20), pp. 1334-1339.