Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 07 Issue 02
2024

Research Article
https://doi.org/10.56979/702/2024

Optimizing Skin Disease Classification: Evaluating the Effectiveness of Hybrid CNN with Batch Normalization and L2 Regularization for Enhanced Accuracy

Abdul Rasheed[1], Muhammad Azam[1], Hafiz Muhammad Afzaal[1*], Muhammad Ashraf[2], and Abid Ali Hashmi[3]

[1]Department of Computer Science & Information Technology, The Superior University Lahore, 54000, Pakistan.
[2]IT Department, Gulab Devi Teaching Hospital, Lahore, 54000, Pakistan.
[3]Project Director, Gulab Devi Educational Complex, Lahore, 54000, Pakistan.
[*]Corresponding Author: Hafiz Muhammad Afzaal. Email: afzaalnawaz17@gmail.com

Received: January 24, 2024 **Accepted:** July 22, 2024 **Published:** September 01, 2024

Abstract: Skin disease classification remains a significant challenge in medical image analysis, demanding robust models for accurate diagnosis. In this research paper, we meticulously explore and compare the performance of various neural network architectures on a curated skin disease dataset. Our study includes traditional models such as Multi-Layer Perceptron (MLPs) with different hidden layers, a powerful deep architecture like ResNet and proposed hybrid Convolutional Neural Networks (CNNs) with Batch Normalization and L2 Regularization. MLP classifiers, equipped with one and two hidden layers, demonstrate promising F1 scores, sensitivity, and specificity, showcasing their effectiveness in skin disease classification tasks. The Sequential Neural Network, a versatile architecture, exhibits commendable accuracy, precision, recall, and F1 score, highlighting its suitability for medical image analysis. Our standout performer is the proposed hybrid CNN model with Batch Normalization and L2 Regularization, achieving an impressive accuracy of 97%. This hybrid architecture outclasses other classifiers, emphasizing the impact of advanced techniques in enhancing model performance. By using various assessment criteria and diagrams, the suggested paper offers crucial information about the advantages and disadvantages of each model, helping researchers and practitioners choose the optimal architecture for skin disease classification. These findings augment the ongoing study in medical image analysis using deep learning techniques and reemphasizes the effectiveness of model selection in diagnosing diseases accurately and comprehensively.

Keywords: Skin Disease Classification; CNNs; MLPs; ResNet; Hybrid Models; Medical Image Analysis.

1. Introduction

Skin diseases are the diseases that are relatively benign to some of the most severe and dangerous diseases and are one of the most common health problems worldwide affecting millions of people. Skin diseases have to be diagnosed as closely to their real nature and as early as possible since it is the key to successful treatment and outcome. There are characteristic types of defect diagnostics known in traditional approaches that are informative but need a lot of skills and time.

First WHO Global Meeting on Skin Related Neglected Tropical Diseases (skin NTDs) March, 2023 in Geneva Switzerland established that skin conditions are approximately computed to affect 1. 1 Hudghton, B. A. D. & MacKenzie, C. D. 8 billion people at any one time. Skin infections which are bacterial, viral, fungal and parasitic are the most frequent disease in the tropical and resource poor countries.In most communities, skin NTDs form about 10% of skin diseases. All the above-highlighted reasons have heightened the importance of endemic countries to embrace broader, community-based strategies to address comprehensively skin NTDs and all other skin ailments as a component of the UHC and SDG – no-one left behind [1].

CCs and, in particular, CNNs have recently been incorporated in different approaches to automate the diagnosis of skin diseases over the last few years. The nature of CNNs architecture to learn feature hierarchically from the medical images will enable the identification of complex patterns in skin lesions to assist in mapping and early detection and classification of a disease.

This paper aims at comparing the prediction performance of different deep learning models: Multi-layer Perceptron (MLP), ResNet, and a CNN with Batch Normalization and L2 Regularization. The importance of each model is determined from considerable values including accuracy, sensitivity, specificity, precision and F1 score models. The goal of this work is to find the best model to classify the skin diseases so that to be a basis for the diagnosis tools reliable and efficient. Defining dermopathological classes goes beyond skin disease classification and can be seen as an exemplary application of deep learning in the field of medicine that opens the doors to the enhancement of medical image analysis and promotion of diagnostic accuracy in numerous healthcare areas.

The main aim of this research can be listed as follows:

Objectives of the Research Paper:

1. Performance Assessment: Evaluate and compare the diagnostic performance of different classifiers, emphasizing the impact of hybrid models, particularly the "CNN model with Batch Normalization and L2 Regularization."
2. Graphical Representation Analysis: Utilize graphical representations, including ROC curves, Precision-Recall curves, and F1 score vs Thresh hold, to visually assess and compare the classifiers' diagnostic capabilities, aiding in the identification of superior models.

2. Literature Review

The study by Paithane and Kakarwal introduces the LMNS-Net, a lightweight multiscale deep learning approach for automatic pancreas image segmentation in CT scan images. The proposed model effectively addresses the challenges of sharp and smooth pancreatic segmentation. LMNS-Net employs 12 layers with 4 convolution layers, incorporating a multiscale block to aggregate required features efficiently. The model achieves a dice similarity index score of up to 88.68%, demonstrating high accuracy and reduced computation time [2].

Mahamed Najeeb and Abdul Majjed Dahl focus on brain tumor segmentation using deep learning networks, employing transfer learning, U-Net, and ResNet architectures. The research emphasizes the importance of early detection for effective treatment planning and survival. The proposed framework outperforms state-of-the-art methods, achieving accuracy rates of 80.0% and 90.2% for U-Net and ResNet, respectively [3].

Yan, Yang, and Zhao propose the Attention-based Deep Residual U-Net (ADRU-Net) for brain tumor segmentation, addressing the challenges posed by complex tumor shapes. ADRU-Net incorporates ResNet50 as a backbone feature extractor for U-Net and introduces Coord Attention modules for fine segmentation of tumor edges. The model demonstrates improved convergence speed and effective segmentation, especially for small objects [4]

Yuvaraj et al. present an ensemble deep learning model for medical image segmentation, evaluating ultrasound breast images with CNN, MaskR CNN, U-Net, and ResNet. The proposed model achieves high accuracy, with 98.6% on training, 94.5% on validation, and a F1 score of 94.32%. Ensemble techniques prove effective in reducing error rates and enhancing segmentation performance [5].

Niyas et al. conduct a survey on 3D deep learning methods for medical image segmentation. The review emphasizes the increasing popularity of 3D deep learning methods in medical image analysis, leveraging advancements in imaging systems and hardware support. The paper identifies research gaps and discusses future directions in 3D medical image segmentation [6].

Song et al. explore deep learning-based segmentation in cardiac radiography, emphasizing the critical role of accurate segmentation in computer-aided diagnosis. The paper reviews existing techniques, introduces Attention-based Deep Residual U-Net, and discusses the benefits of deep learning frameworks in enhancing segmentation accuracy [7].

When it comes to the classification of the various planes of maternal fetal ultrasound, Bala Krishna and Kokil employ deep learning models. From the base architectures of AlexNet and VGG-19 some of the deep features are extracted and combined as a multi-layer perceptron. The suggested model reveals a higher percent of knowledge classification compared to the other models available [8].

Consequently, Sathish et al. were focused on applying Deep MLP for the investigation of natural fibre reinforced plastic composites. For the estimation of different parameters of PMC/Plastic Composite, the study has suggested the revised MLP with about 50 hidden layers. On the proposed Deep MLP, higher accuracy, precision and high recall compared to single layer models has been realised [9].

In a more recent work, Amine Ben Slama et al. formulation of an improved segmentation method of tumor in MR images with the use of Res-Net architechture and effective VGG gliomas grading. The work concerns with the issue of distinguishing and measuring brain tumours in medical MRI without radiation. The method described herein uses deep convolutional networks for the segmentation of the dataset and yields better performance as compared to the BRATS 2020 for Brain Tumor Image Segmentation. According to the results of cross-validation, it is possible to state that the accuracy of the proposed method can be considered higher than in previously described works, which is important to emphasize the necessity of fully automated segmentation for the correct measurement of the tumor size [10].

Deepak Jayaprakash Doggalli and Sunil Kumar B S assess the efficacy of the U-Net model in segmenting liver tumors from abdominal CT images. The study proposes a U-Net-based method for liver and tumor segmentation, achieving high accuracy on the LiTS dataset. The U-Net architecture proves to be simple and effective, demonstrating a dice global score of 94% for liver segmentation and 73% for tumor segmentation [11].

Saeed Iqbal and Adnan N. Qureshi introduces BreastUNet, heteromorphous Deep Convolutional Neural Network (CNN) with the feature grafted for evaluating mitotic nuclei in breast histopathology images. The new model of theirs is built on an entirely distinctive structure – four deep CNNs designed to address the structural, textural, and morphological features of mitotic nuclei. The model named as BreastUNet delivers improved F1 score of 0. Supersensitivity, superspecificity of 0. 95. There was also a mean rating of 95, and the area under the precision curve at . To this, figure 95 illustrates how it can be valuable when it is incorporated into the pathologist's practice as a supplementary tool [12].

Automated model for early diagnosis of Alzheimer's disease artefact using functional MRI images is proposed by Aarthi Chelladurai et al. The developed approach includes image pre-processing which is image normalization, segmentation of the brain region using Segmentation by Aggregating Superpixels (SAS), feature extraction utilizing Gabor wavelet and Gray Level Co-Occurrence Matrix (GLCM) technique, reduction of feature space using Honey Badger Optimization Algorithm (HBOA), and classification using Multi-Layer Perceptron (MLP). The model achieves high classification accuracy (99.44%) and demonstrates superior performance compared to conventional segmentation and classification models [13].

As for the issues related to independent identification and classification of brain malignancies by MRI, we can mention the following works by Zahid Rasheed et al. These proposed methodologies were applied to the benchmarked data where it was observed that the proposed approach has better accuracy than the popular pre-trained models such as VGG16, ResNet50, VGG19, InceptionV3, MobileNetV2 with the overall classification accuracy of nearly 97 84%, precision of 97. 85%, recall of 97. The accuracy rate was of 85% and F1-score 97. 90%. The received evaluation proves the high accuracy and the possibility of the extension of the proposed approach to the subtype classification of brain tumor [14].

The study aims to assess the performance of classification algorithms, particularly Naïve Bayes (NB) and K-Nearest Neighbor (K-NN), utilizing the Scopus dataset. The implementation of Bayesian boost and bagging techniques results in notable improvements in classification accuracy, precision, and recall, with K-NN showing a performance increase of over 7% compared to NB, emphasizing the significance of data preprocessing and addressing class imbalance issues in enhancing text classification algorithms [15].

The research evaluates the accuracy of various classification algorithms, including SVM, NB, KNN, and others, proposing a model stacking methodology based on their correlation to enhance accuracy. Neural Networks stood out with over a 25% improvement, achieving the highest accuracy of 95.66%, showcasing the effectiveness of the proposed model stacking approach over individual algorithm performances [16].

The paper proposes a model to enhance the classification accuracy of sentiment data, validated on Amazon, Yelp, and IMDB reviews. By combining outputs from diverse base-level classifiers at the metalevel, the model demonstrates significant accuracy improvement compared to individual classifiers, offering a reliable approach for sentiment analysis on various datasets [17].

The 3D color histogram and Gabor filter describe image properties effectively. Proposed method [16] combines color coherence vector and wavelets for improved retrieval. Feature selection, using discrimination and Genetic Algorithm, refines features. Researcher implemented a retina-inspired key point descriptor to enhanced performance [17].

Structure Elements' Descriptor (SED) for image texture description [18] to effectively extract characteristics related to appearance, Histogram of Image Structure Elements (SEH) is calculated using SED in the HSV color space with 72 bins. SEH combines analytical and architectural methods for pattern depiction, enhancing to capture of spatial correlation. CBIR systems offer a means to extract visuals with comparable semantics from large databases. The academic community is in search of more effective CBIR techniques for time-sensitive applications.

This paper [19] introduced a NN for CBIR, which incorporates an effective feature extraction technique. This technique utilizes intricate pattern assessment via wavelet decompositions and Gabor filters characteristics applied to visual representation. Comparative analysis showcases the superior efficiency of the presented approach. Proposed [20] introduced image database creation, visual feature extraction for retrieval. Validation via color histogram and Euclidean distance on database. System includes design, extraction, distance techniques. Achieves 85% accuracy over 20 iterations. The researchers introduced a proficient description of color-based image features for CBIR.

The descriptor [21] achieves inherent rotation invariance by quantifying color occurrences in local pixel neighborhoods. There are 64 shades within the RGB color space for streamlined representation. Color occurrences are extracted and depicted in dual form to generate binary arrangements based upon local color occurrences, ensuring the effective capture of local color information. A novel approach to image extraction combines form, textural and coloration elements within CBIR. Simulation results demonstrate good precision, particularly for clear and distinct targets in queries, showcasing the method's efficiency. Notably, the texture feature, referred to as CLCM, outperforms GLCM in retrieval. However, there's room for substantial improvement, especially for complex query scenes where low-level visual features may be insufficient. Utilizing high-level visual features like semantic features could further enhance retrieval performance [22].

This paper introduces an effective scheme for enhancing image quality in video surveillance, enhancing face detection and introducing a content-based image retrieval method. The integrated enhancement method combines contrast enhancement and color balancing, showing significant improvements in face detection. The retrieval approach prioritizes results using fuzzy heuristics and employs well-known algorithms for color, texture, and shape analysis. Evaluation metrics demonstrate superior performance compared to existing methods, with future work focusing on real-world data integration and processing speed considerations [23].

Image signatures are derived through the aggregation of interest points across different representation levels, initially gathering shape features by grouping connected pixels based on binary brightness thresholds. The HOG is utilized to characterize attributes for identified points of interest within optimal stability regions. Descriptors are then combined with rotation-invariant texture attributes obtained through uniform patterns following the application of our suggested reordering algorithm. Experimentation on Caltech-101, Caltech-256 and Corel-100, and datasets illustrates the superiority of our technique over existing methods across various image categories. Results indicate that the integration of local and global features enhances the method's capability in foreground and background object retrieval, with feature description employing a sliding window approach, thereby strengthening its robustness in object recognition [24].

Arabic handwritten script recognition, challenging due to diverse styles and large datasets, benefits from effective techniques. BDLF-MLP, using Block Density and Location Feature with MLP, achieves 97.26% accuracy by extracting pixel density and location from letter images, surpassing other algorithms [25].

Convolutional neural networks (CNNs) exploit grid-structures and spatial dependencies in images to detect adjacencies, colors, and patterns. This research integrates GoogLeNet, VGG-19, and ResNet50 with Eigenvalues and convolutional Laplacian scaled object features, using mapped colored channels for high image retrieval rates across diverse benchmarks. The approach enhances deep learning fusion and descriptor creation efficiency, showing remarkable performance on datasets like ALOT, Corel-1000, CIFAR-10, CIFAR-100, Oxford Buildings, and others, compared to state-of-the-art methods, demonstrating significant accuracies across various image types [26].

This study enhances image retrieval by detecting interest points and utilizing their features for content-based analysis, including shape, texture, and color attributes. It improves feature detection with techniques such as non-maximum suppression, edge and corner detection, and symmetric sampling. Evaluation on benchmarks like Corel-1000, ImageNet, and Caltech-101 shows superior performance, comparing favorably with RGBLBP, LBP, SURF, SIFT, DoG, HoG, and MSER in accuracy and retrieval rates [27].

3. Dataset and Methodology

3.1. Dataset

The dataset employed for this research is sourced from the openly accessible on Kaggle platform [18]. It comprises images categorized into 10 distinct classes. Undefined For this research, only two classes were selected thus having a total of two thousand images while each class had one thousand images. The images were divided so that 70% of them were used for the training of the algorithms while the remaining 30% were used to validate the results.

A visual representation of a select few images from these two classes is provided below.

Figure 1. Image class-1

Figure 2. Image class-2

3.2. Classifiers Used

3.2.1. Multi-Layer Perceptron (MLP) with One Hidden Layer

MLP having one-single hidden layer is one of the Feed Forward Neural Network Model. In this context the network holds an input layer, a hidden layer and an output layer only. Each layer has nodes (neurons) linked with nodes within the subsequent and prior layers. Read layer includes the capability of the hidden layer to identify non-linear transformations of the input data. It avails on backpropagation to tune the weights and the biases in a manner that during the training phase, the sum total difference between the expected and the forecasted outputs is as small as possible.

3.2.2. Multi-Layer Perceptron (MLP) with Two Hidden Layers:

As it is done in the previous section with one hidden layer, MLP with two hidden layers is constructed by adding another hidden layer between the input and output layers. This deeper architecture enables the network to capture more complex hierarchical features in the data. The training process involves forward and backward passes, adjusting weights and biases to optimize the model's performance on the given task, such as image classification.

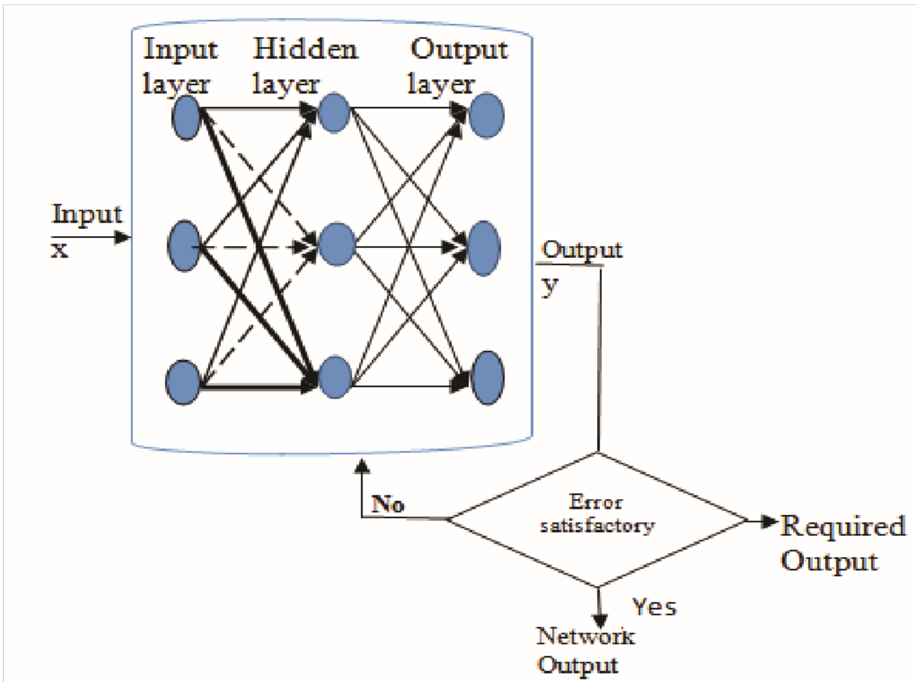

Figure 3. One Hidden Layer MLP Mode

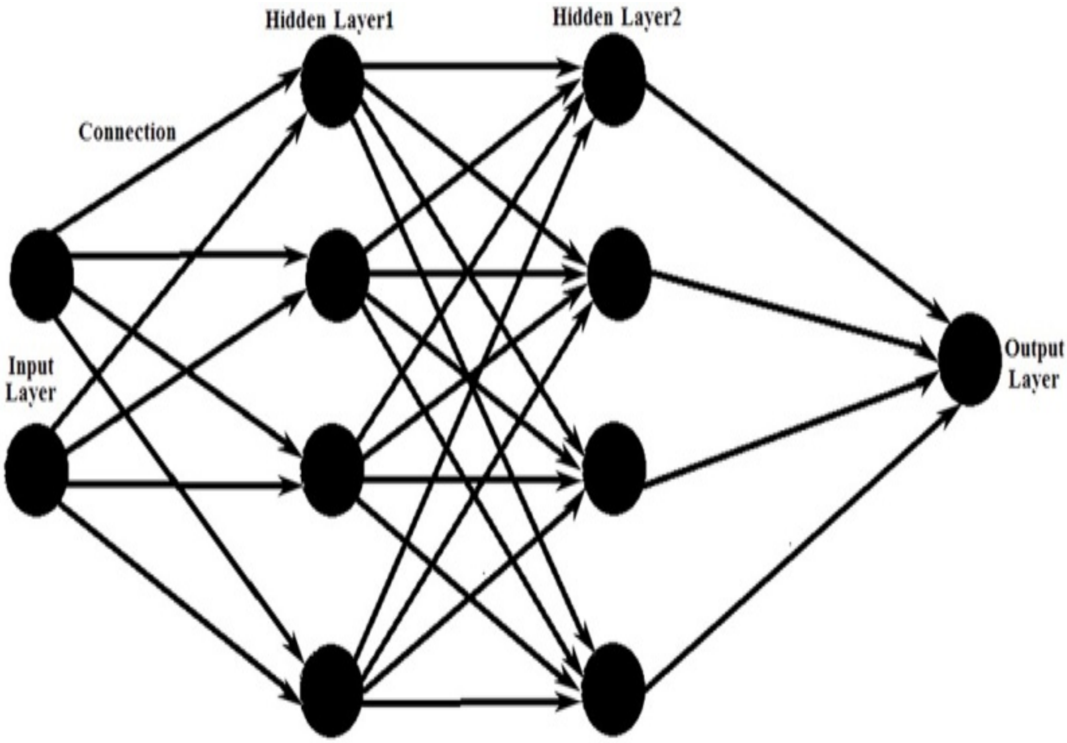

Figure 4. Two hidden layer MLP Model [20]

3.2.3. Sequential Neural Network

A Sequential Neural Network, in an ideal meaning, can be defined as the neural network architecture where the data flow is always unidirectional, that is, from the input layer through the hidden layers and to the output layer. This term encompasses both MLPs and other variants. In image classification, feedforward neural networks process input images through layers of interconnected neurons, where each connection has a weight associated with it. The network is trained to adjust these weights to minimize the difference between predicted and actual class labels.

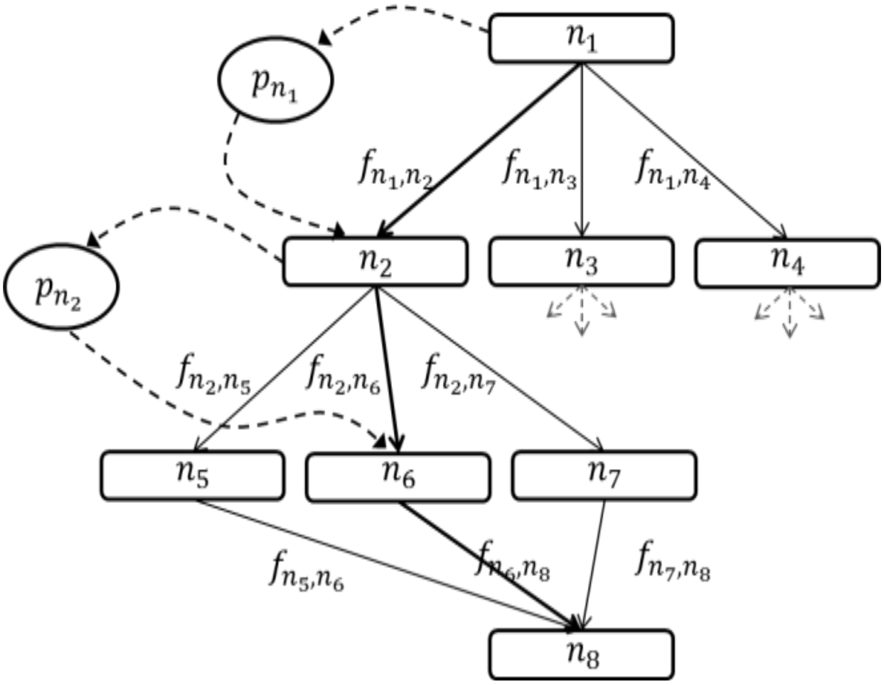

Figure 5. Sequential Neural Network Model [21]

3.2.4. Residual Network (RES-NET)

Residual Networks of a specific architecture of deep neural network which are employed to solve problems of training of very deep networks. The primary change in ResNets is the addition of residual blocks where the network can learn the residual functions by having shortcut connections. This facilitates the training of deeper networks by mitigating issues like vanishing gradients. In image classification, ResNets have shown significant success in learning intricate features and patterns, making them well-suited for tasks with large and complex datasets.

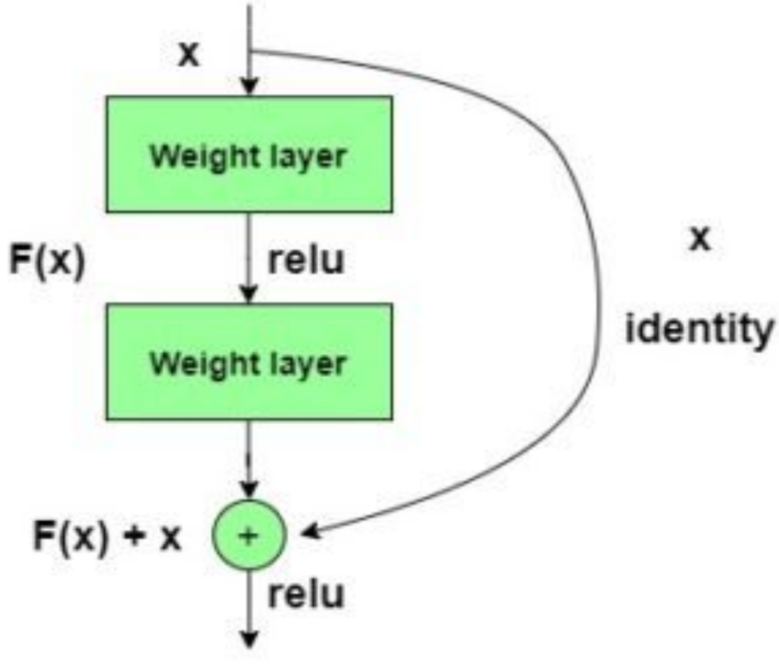

Figure 6. Res-Net Model [22]

3.3. Proposed Methodology

Incorporation of Both Models While CNN model means a convolutional neural network, the presented structure aims at combining Batch Normalization and L2 Regularization.

3.3.1. Batch Normalization

It is practice that is used in order to standardize the input field of each layer in the neural network. It works in mini-batches: it normalizes the activations to control the internal covariate shift. This helps in training deeper networks more efficiently by stabilizing and accelerating the training process. It can lead to faster convergence and allows the use of higher learning rates.

Below, in (1) we provide an exact definition of the effect of the batch norm output on its input. The input for BatchNorm for instance at a local neighborhood can be as we have defined above convolution weights can be defined as the previous activations by a Bias. We can therefore express in (3) the value of the convolution input in terms of the BatchNorm output which we can express in equation (4) in terms of new weights W' and bias b' revealed in equation (5) and equation (6) respectively.

$$y_i = \frac{\gamma}{\sqrt{\sigma^2+\epsilon}} x_i - \frac{\gamma}{\sqrt{\sigma^2+\epsilon}} \mu + \beta \tag{1}$$

$$x_i = W.z_{ci} + b \tag{2}$$

$$y_i = \frac{\gamma}{\sqrt{\sigma^2+\epsilon}} (W.z_{ci} + b) - \frac{\gamma}{\sqrt{\sigma^2+\epsilon}} \mu + \beta \tag{3}$$

$$y_i = W'.z_{ci} + b' \tag{4}$$

$$W' = W \frac{\gamma}{\sqrt{\sigma^2+\epsilon}} \tag{5}$$

$$b' = \frac{\gamma}{\sqrt{\sigma^2+\epsilon}} (b - \mu) + \beta \tag{6}$$

Following figure graphically represents the process of batch normalization.

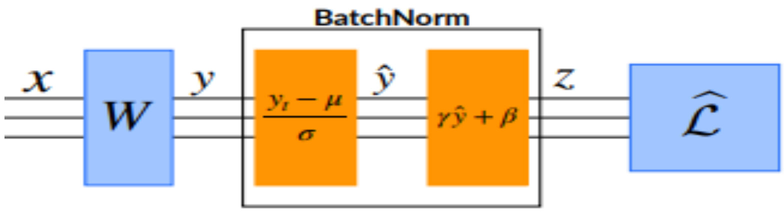

Figure 7. Batch Normalization [23]

3.3.2. L2 Regularization

L2 Regularization, also known as weight decay, is a regularization technique that penalizes large weights in the model. It is added to the loss function, and during training, it discourages the network from assigning too much importance to any one feature. L2 regularization helps prevent overfitting by promoting smoother weight distributions and can improve the generalization performance of the model. Mathematically cost is defined by following expression.

$$\underbrace{\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij} b_j\right)^2}_{Loss\ term} + \underbrace{\lambda \cdot \left(\sum_{j=1}^{p} b_j\right)^2}_{Regularizer\ term}$$

Figure 8. Mathematically Cost is Defined

Following is pictorial representation of L2 regularization.

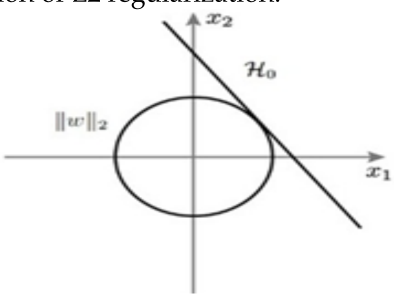

Figure 9. L2 Regularization [24]

3.3.3. Hybrid Model

Implementation of Batch Normalization and L2 Regularization in CNN results in the formulation of a complex model with the merits of both procedures. It enables to solve the problem of internal covariate shift and speeds up the training process; L2 Regularization provides an opportunity to carry out a regularization, thus preventing overfitting.

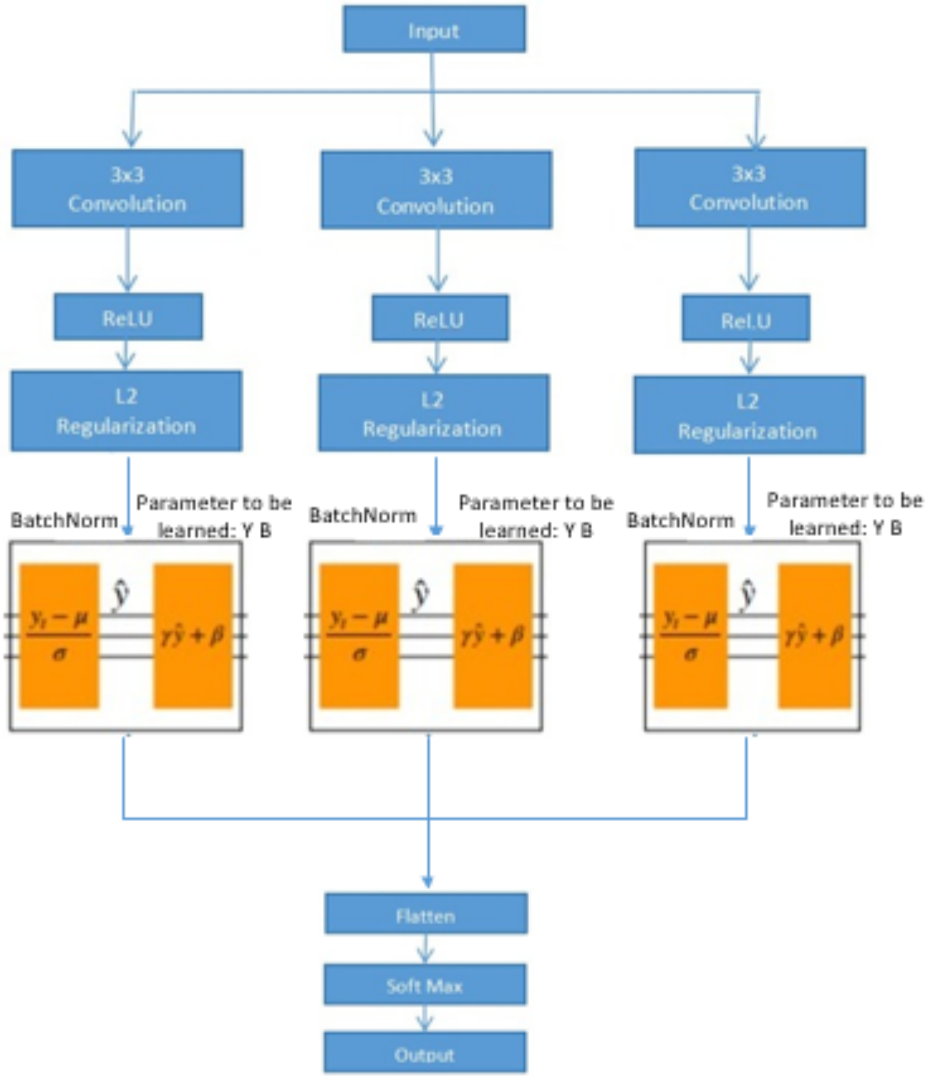

Figure 10. Proposed Hybrid CNN Model

3.3.4. Performance Metrics

We used F1 score, Accuracy, Precision, and Recall to elaborate our findings. Following are mathematical formulas for above Performance Metrics.

$$\text{Accuracy} = \frac{TP+TN}{P+N}$$

$$\text{F1-Score.39} = \frac{2TP}{2TP+FP+FN}$$

$$\text{Recall or True Positive Rate or sensitivity or probability of detection} = \frac{Tp}{P}$$

$$\text{Specificity or True negative rate} = \frac{TN}{N}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

Here TP represents True Positive samples, TN is for True Negative samples, T is for all positive samples and N represents all negative samples.

4. Results and Discussion

Various curves and performance metrics were employed to comprehensively analyze and present the results of our study. It provides a multiple-perspective comparison of the various curves and measures of performance, and then, presents the evaluation on the classifiers under consideration. Among these the

Receiver Operating Characteristics curves –ROC, Precision Recall curves, the F1 versus Threshold graphs and quantitative measures such as accuracy, precision, sensitivity and specificity. Thus all of these metrics and curves can in turn offer different perspectives of how the classifiers are doing, and their strengths and weaknesses. Such approach provides opportunity to get rest and get the comprehensive picture of characteristics of presented models, and based on that decide on their applicability for the certain cases.

4.1. Receiver Operating Characteristic Curves

In the binary classification models the measures are compared by constructing the ROC curves which stands for Receiver Operating Characteristic. The ROC curves are also used in the measurement of Discrimination of a classification model. They help in graphing the TPR and FPR at various decision thresholds with a view of measuring the trade space between two of them. Following figure shows ROC curves of all classifiers used in this study.

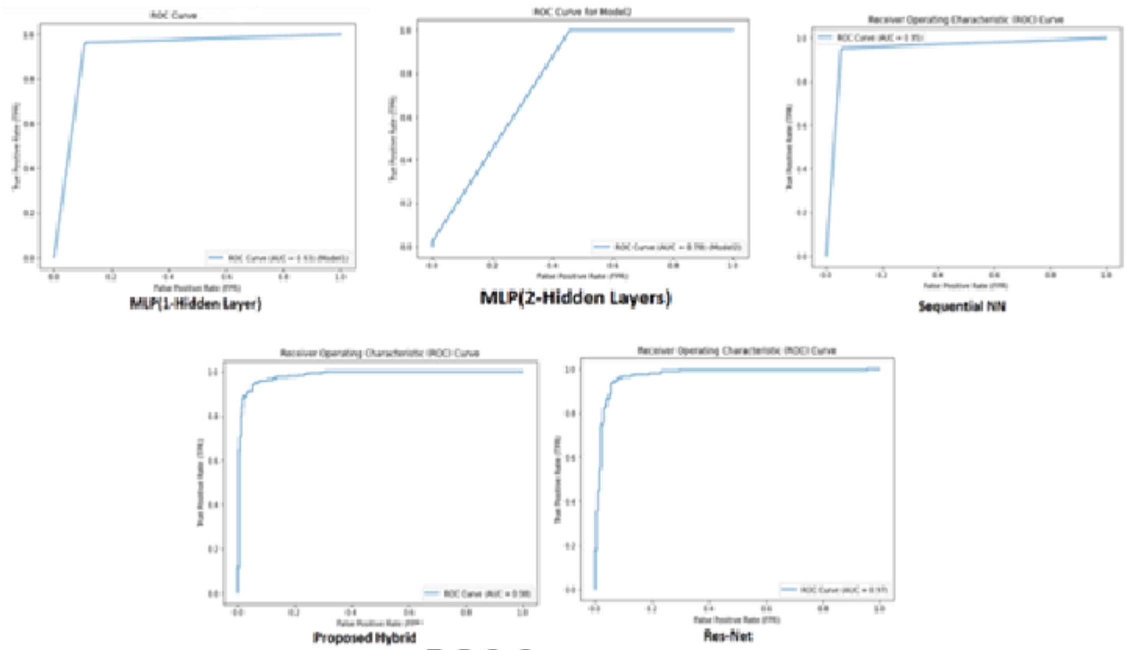

Figure 11. ROC Curves

The depicted figure unmistakably illustrates the superiority of the CNN model with Batch Normalization and L2 Regularization over all other classifiers. Its positioning in the plot, notably outclassing its counterparts, is evident, emphasizing the model's exceptional performance. The results visually highlight the effectiveness of incorporating Batch Normalization and L2 Regularization techniques in enhancing the classification capabilities of the CNN model, making it a standout choice in the comparison among diverse classifiers.

4.2. Precision vs Recall Curves

The precision-recall curve gives a finer distinction between a model's performances, particularly when classes are uneven, or if the important rates in false positives and false negatives differ. Precision-recall curve means plotting of precision with other factors of recall in different classification threshold.

Following figure 12 compares the performance of different classifiers.

In the evaluation of various classifiers, the precision-recall curves offer a nuanced perspective on their performance in distinguishing between positive and negative instances. Notably, the CNN model with Batch Normalization and L2 Regularization exhibited superior performance across different thresholds. Its precision-recall curve consistently outshone those of other classifiers, showcasing a remarkable balance between precision and recall.

4.3. F1 Score vs Threshold Curves

The F1 vs Threshold curves provide valuable insights into the trade-off between precision and recall at different classification thresholds. These curves allow us to define the values of the F-measure of performance, being a weighted mean of precision and recall, at which it is maximum for each classifier. This is used to compute out the appropriate operating point of the models thereby enabling one to determine the acceptable false positive rate and the acceptable false negative rate.

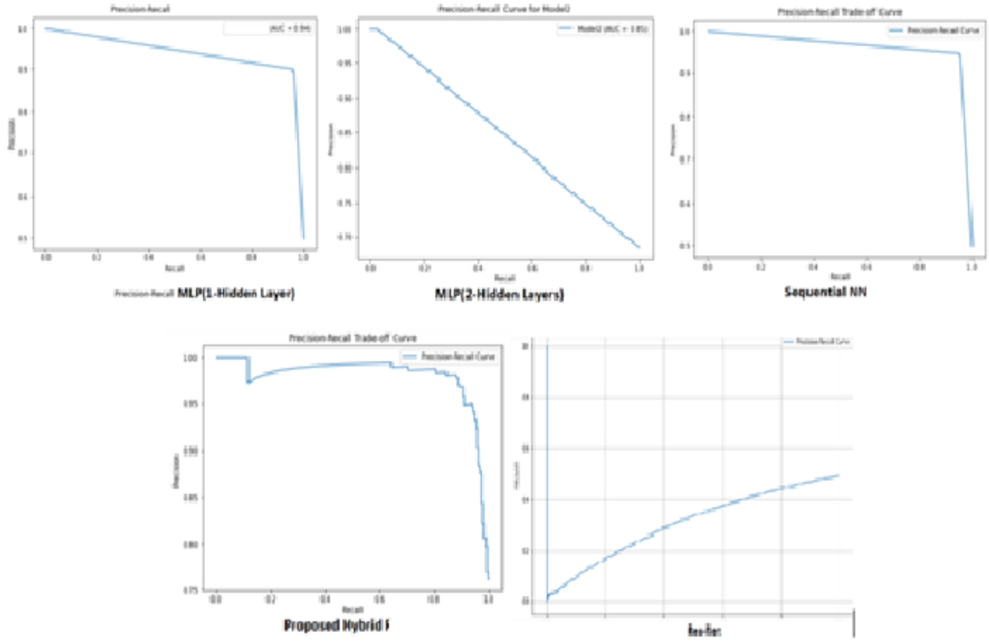

Figure 12. Precision Vs Recall

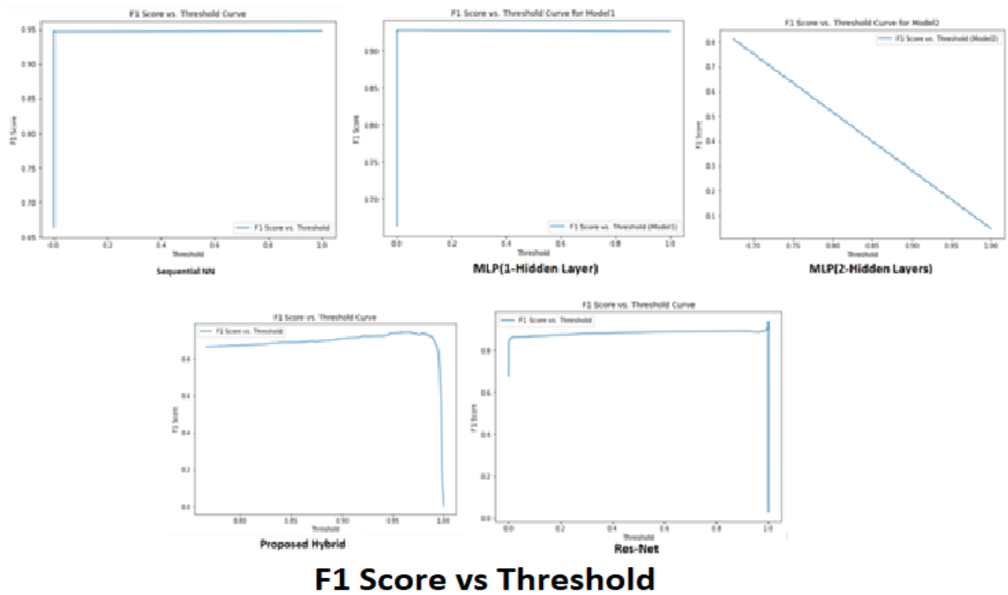

Figure 13. F1 Score Vs Threshold

5. Performance Measures Comparison of Different Classifiers

The subsequent table provides a comprehensive overview of the performance metrics employed for comparing the outcomes of various classifiers in this investigation. A detailed examination and discussion of the results are presented following the table, offering a thorough analysis of the classifiers' performance in the study.

Table 1. Result Comparison

Classifier	Precision	Sensitivity (Recall)	Specificity	F1-Score	Accuracy
MLP(One Hidden Layer)	0.89	0.95	0.89	0.92	0.93
MLP(Two Hidden Layer)	0.54	1.00	0.54	0.81	0.77

Res-Net	0.75	0.98	0.80	0.85	0.86
Sequential Neural Network	0.94	0.95	0.94	0.94	0.94
Proposed Hybrid CNN	0.94	0.99	0.94	0.97	0.97

The Sequential Neural Network and the CNN Model with Batch Normalization and L2 Regularization show the highest accuracy, indicating strong overall performance. ResNet and the MLP Classifier with Two Hidden Layers have lower accuracy.

Out of all the models presented, the CNN Model with Batch Normalization and L2 Regularization has the best F1 score, which means it preciously balances between its precision and recall. When it comes to F1 score the Sequential NN Classifier and MLP with One Hidden Layer does not lag much behind. Nevertheless, the F1 scores of ResNet and the MLP Classifier with two hidden layers is relatively low.

Subsequently, an impressive precision and recall have been observed in the Sequential Neural Network, MLP Classifier with One Hidden Layer, and the CNN Model with Batch Normalization and L2 Regularization. ResNet and MLP Classifier with Two Hidden Layers have lower values of precision and recall.

5.1. High Discriminative Power

The hybrid model showcases high discriminative power, as evident from its impressive F1 Score. This indicates its ability to effectively balance precision and recall, crucial for accurate classification in skin disease scenarios where both false positives and false negatives are significant concerns.

5.2. Robustness against Overfitting

L2 Regularization and Batch Normalization are applied for the CNN model, so, the training process will not over fit the model. This is evidenced by the low specific error occurring in the model, an indication of its high level of generalization on unseen data, a factor that plays a very important role in the reliability of skin disease classifiers.

5.3. Comprehensive Understanding of Features

The interpretation of the hybrid model indicate that CNN is able of perceiving fine spatial patterns in skin images. This feature is rather important for elaborating on nuances of patients' skin lesions or symptoms of each disease as the fine details are essential for the model.

5.4. Optimized Precision-Recall Tradeoff

In the instance of the hybrid model, the measures of Precision and Recall show that this model is favorable in both Precision and Recall; it is perfect for getting the Nilima without yielding to excessive false positive or false negative. This is especially true in a medical diagnosis where mis-diagnosis will lead to an incorrect treatment and thus result in more harm to the patient.

6. Confidence in Predictions

The high precision, recall, and F1 Score collectively instill confidence in the reliability of the hybrid model's predictions. This is crucial for medical applications, where accurate identification of skin diseases is paramount for appropriate treatment decisions.

7. Conclusion

In conclusion, the hybrid model, as represented by the CNN with Batch Normalization and L2 Regularization, emerges as a robust and versatile solution for skin disease classification. Its ability to harness the complementary strengths of different architectures, mitigate overfitting, and strike an optimal balance between precision and recall positions it as a leading candidate for accurate and reliable skin disease diagnosis.

8. Future Directions

In future research endeavors, one can aim to enhance the proposed model's performance through strategic improvements. One avenue for refinement involves the application of fine-tuning and feature extraction functions. By fine-tuning the model on relevant datasets, we anticipate a more precise alignment of abstract representations within the network layers. This targeted adjustment aims to optimize the model's ability to discern intricate patterns critical for accurate segmentation.

Explore the application of transfer learning techniques in skin disease classification. Pre-trained models on larger datasets, especially in the medical imaging domain, can be fine-tuned to achieve better results with limited labeled data. Focus on developing models that provide explainable and interpretable results. Understanding the decision-making process of the model is crucial, especially in medical applications, where interpretability is essential for gaining trust from healthcare professionals. Investigate advanced data augmentation techniques to enhance the diversity of your dataset. This can help in training more robust models by exposing them to a wider range of variations in skin disease images. Develop models with a focus on real-time diagnosis. This could involve optimizing models for deployment in healthcare settings where quick and accurate diagnoses are crucial.

References

1. https://www.who.int/news/item/31-03-2023-who-first-global-meeting-on-skin-ntds-calls-for-greater-efforts-to-address-their-burden
2. Paithane, P., &Kakarwal, S. (2023). LMNS-Net: Lightweight Multiscale Novel Semantic-Net deep learning approach used for automatic pancreas image segmentation in CT scan images. Expert Systems with Applications, 234, 121064.
3. Mahamed Najeeb, R. S., & Abdul Majjed Dahl, I. O. (2022). Brain Tumor Segmentation Utilizing Generative Adversarial, Resnet AndUnet Deep Learning.
4. Yan, S., Yang, W., & Zhao, J. (2022). Attention-based Deep Residual U Neural Network on Brain Tumor Segmentation Algorithms.
5. Yuvaraj, D., Sampath, A. K., Shanmugapriya, N., Samatha, B., Arun, S., & Thiyagarajan, R. (2022). Medical Image Segmentation of Bio-medical Images with Deep Convolution Neural Networks using Ensemble Approach.
6. Niyas, S., Pawan, S. J., Anand Kumar, M., & Rajan, J. (2022). Medical image segmentation with 3D convolutional neural networks: A survey. Neurocomputing, 493, 397-413.
7. Song, Y., Ren, S., Lu, Y., Fu, X., & Wong, K. K. L. (2022). Deep learning-based automatic segmentation of images in cardiac radiography: A promising challenge. Computer Methods and Programs in Biomedicine, 220, 106821.
8. Bala Krishna, T., & Kokil, P. (2023). Automated classification of common maternal fetal ultrasound planes using multi-layer perceptron with deep feature integration. Biomedical Signal Processing and Control, 86(Part C), 105283.
9. Sathish, T., Sunagar, P., Singh, V., Boopathi, S., Sathyamurthy, R., Al-Enizi, A. M & Sehgal, S. S. (2023). Characteristics estimation of natural fibre reinforced plastic composites using deep multi-layer perceptron
10. Amine Ben Slama, Hanene Sahli, Yessine Amri, Hedi Trabelsi. "Res-Net-VGG19: Improved Tumor Segmentation Using MR Images Based on Res-Net Architecture and Efficient VGG Gliomas Grading." University of Tunis EL Manar, Higher Institute of Medical Technologies of Tunis (ISTMT), Laboratory of Biophysics and Medical Technologies, LRBTM, Tunis LR13ES07, Tunisia. (2022).
11. Deepak Jayaprakash Doggalli, Sunil Kumar B S. "The Efficacy of U-Net in Segmenting Liver Tumors from Abdominal CT Images." RV Institute of Technology and Management, Bangalore, India. (Year not provided).
12. Saeed Iqbal, Adnan N. Qureshi. "A Heteromorphous Deep CNN Framework for Medical Image Segmentation Using Local Binary Pattern." Department of Computer Science, Faculty of Information Technology, University of Central Punjab, Lahore 54782, Pakistan. (2022).
13. Aarthi Chelladurai, Dayanand Lal Narayan, ParameshachariBidareDivakarachari, Umasankar Loganathan. "Automated Diagnosis of Alzheimer's Disease Using fMRI Images: A Hybrid Approach." Sengunthar Engineering College, Tiruchengode, Tamil Nadu, India. (Year not provided).
14. Zahid Rasheed, Yong-Kui Ma, Inam Ullah, et al. "Automated Classification of Brain Malignancies Using MRI and Deep Learning." Harbin Institute of Technology, Harbin, China. (Year not provided).
15. Azam, M., Ahmed, T., Sabah, F., & Hussain, M. I. (2018). Feature extraction based text classification using k-nearest neighbor algorithm. IJCSNS Int. J. Comput. Sci. Netw. Secur, 18(12), 95-101.
16. Azam, M., Ahmed, T., Hashmi, M. U., Ahmad, R., Manan, A., Adrees, M., & Sabah, F. (2019). Improvement in Classification Algorithms through Model Stacking with the Consideration of their Correlation. (IJACSA) International Journal of Advanced Computer Science and Applications, 10(3).
17. Azam, M., Ahmed, T., Ahmad, R., Ur Rehman, A., Sabah, F., & Asif, R. M. (2021). A Two-Step Approach for Improving Sentiment Classification Accuracy. Intelligent Automation & Soft Computing, 30(3).
18. https://www.kaggle.com/datasets/ismailpromus/skin-diseases-image-dataset
19. Ebhota, V. C., Isabona, J., & Srivastava, V. M. (2018). Improved adaptive signal power loss prediction using combined vector statistics based smoothing and neural network approach. Progress In Electromagnetics Research C, 82, 155-169.
20. Borah, T. R., Sarma, K. K., & Talukdar, P. (2016). Biometric identification system using neuro and fuzzy computational approaches. In Handbook of Research on Emerging Perspectives in Intelligent Pattern Recognition, Analysis, and Image Processing (pp. 335-366). IGI Global
21. Denoyer, L., & Gallinari, P. (2014). Deep sequential neural network. arXiv preprint arXiv:1410.0510.
22. https://www.geeksforgeeks.org/introduction-to-residual-networks/
23. https://krishnaik.in/2022/02/14/what-is-l1-and-l2-regularization
24. Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning (pp. 448-456). pmlr.

25. Chen, G., Chen, P., Shi, Y., Hsieh, C. Y., Liao, B., & Zhang, S. (2019). Rethinking the usage of batch normalization and dropout in the training of deep neural networks. arXiv preprint arXiv:1905.05928.
26. Ejaz, F., Ahmad, A., & Hanif, K. (2020). Prevalence of diabetic foot ulcer in lahore, Pakistan: a cross sectional study. Asian Journal of Allied Health Sciences (AJAHS), 34-38.