Biomedical image classification using deep convolutional neural networks – overview

M Bharath Simha Reddy¹, Pooja Rana²

¹School of Computer Science and Engineering, Lovely Professional University Phagwara, Punjab, India. ²School of Computer Science and Engineering, Lovely Professional University Phagwara, Punjab, India.

E-mail: ¹bharathsimha.maram@gmail.com, ²pooja.20992@lpu.co.in

Abstract. Deep Learning is an advanced area of machine learning which gained much interest in the past decades. It has been widely used in a variety of applications and has proved to be an effective machine learning method for many complicated issues. Especially when it comes to the medical field, the classification of biomedical images is a complex task to identify and classify the images manually by the doctors. So, Deep Learning is a key to enhance the classification of biomedical images using various architectures. The biomedical picture classification aims to identify and classify biomedical characteristics efficiently, which have significant advantages to numerous study and development fields. In this paper, the framework focused on the different architectures that were used to classify the medical images along with their performances.

1. Introduction

Over the past decades, significant development has been done using Neural networks on the classification of medical images. Deep Learning (DL) is one of the branches of machine learning dependent on studying various layers of perception, through the development of a hierarchy of characteristics that distinguish lower levels from higher levels and that can lead to identifying several upper-level characteristics. A significant issue in computer vision is the classification of pictures, which can be represented as the categorization of pictures within one of the multiple predetermined groups. These issues were said to be handled by several deep learning methods, such as using multi-layer nonlinear data processing, classification, feature selection, and transforming, and structure identification. Among all other architectures, CNN is the primary architecture to identify, classify, and to monitor most of the medical images.

One of the difficult tasks is the collection of medical images because the gathering and classification of health information both involving security issues hence the need for time-intensive details from the experts. One is to collect further data, like those of mass sourcing or exploring in existing diagnostic information in all specific resolving ways. Another way is to explore how else to enhance the efficiency of a limited dataset that is very necessary while the study experience will apply to a massive- dataset study. In the earlier decades, one of the machine learning techniques which is a traditional SVM (Support vector Methods) were used for biomedical image classification[24]. However, SVM also has disadvantages that include time-consuming for identifying and extracting the features of an image, and performance is lower than many other methods.

This paper is organized as mentioned: Section 2 includes a literature review which is a previously completed related work. Section 3 includes an overview of various Deep learning architectures.
Section 4 includes a description of medical images followed by CNN architecture. Section 5 includes a comparative study of different DL architectures. Section 6 is about future study and applications in the field of Biomedical picture classification. Finally, Section 7 concludes the paper.

2. Literature Review

Afshar, P, et al. 2018 [1] proposed a new Capsule network model which mainly works as neurons that implement on biomedical images. In the architecture, along with the CapsNet model, the Authors combined Convolutional Neural Networks to segment the images based on parameters: segmented tumors and brain images. Dataset deals with 233 subjects and their MRI images of 3,064 that contains three types of brain tumours. After a comparison of different architectures, the highest accuracy of 86.56% has been predicted on the 1-Convolutional layer that has 64 feature maps.

Frid-Adar, M, et al. 2018 [2] discussed the Hybrid model that has GAN (Generative Adversarial Networks), CNN, Synthetic data augmentation which implemented on Liver lesion images. Dataset consists of CT images of liver lesions(182) that categorized into haemangiomas(65), cysts(53), metastases(64). The authors implemented the dataset using different methods such as CNN-AUG, BOVW-MI, CNN-AUG-GAN and compared them. By implementing the CNN-AUG-GAN model, the Authors got the highest sensitivity performance is 85.7% and the highest performance for specificity is 92.4%.

Kermany, DS, et al. 2018 [3] gave an introduction about Transfer learning and implemented on the datasets of pediatric radiograph chest images. But the primary goal is to build the architecture to predict the retinal diagnosis based on the Optical Coherence Topography images using Transfer learning. By consulting six random experts, they examined the 1000 OCT images and the diagnosis for the vision loss. Those 1000 images of 633 patients were given as an input to the developed model and got a sensitivity of 96.6% and weight error has 12.7%. The authors compared the four different stages of retinal infections using the confusion matrix and got an approximation of 99%.

Zhang. Yijia, et al. 2019 [4] explains the general CNN model used in the field of Biomedical relation classification to recognize the protein entity and its impact on the patients. The authors also explained about the Hybrid model considering RRN(Recurrent Neural Networks) and SVM(Support Vector Machines) to make the model more accurate. One of the approaches mentioned was reducing the dependencies on training labelled data and the relationship of drug effect concerning other drugs on the patient.

Munir. Khushboo, et al. 2019 [5] discussed the cancer diagnosis for various cancer types such as lung, brain, breast, skin, prostate cancers using deep learning architectures. Some methods used by the authors were ABCD (Asymmetry, Border, Colour, Diameter), Menzies method, seven-point detection, pattern analysis, and predicted based on the specificity, sensitivity, Jaccard index, and accuracy measures. The authors mentioned different CNN architectures and their applications along with datasets that to be implemented.

S. Hoo-Chang, et al. 2016 [6] implemented 5 different CNN architecture based neural networks methods to identify the interstitial lung disease using the dataset of 2D images of CT scan slices. The dataset contains 905 images of 120 patients with 6 types of lung tissue conditions. The authors trained the model using GoogleNet, CifarNet, AlexNet, and ImageNet architectures. The authors also implemented different architectures to detect the Thoracoabdominal Lymph node and made empirical evaluation.
Mohsen. Heba, et al. 2018 [7] discussed brain tumors and its importance to classify the tumor in MRI images. Dataset consists of human brain MRI images of 66 people of which 44 were abnormal and 22 were normal conditions. Fuzzy C-means are used for image segmentation and feature extraction, authors used discrete wavelet transformation. Image segmentation is done by using DNN and got a classification rate of 96.97%. The project was built on the MATLAB R2015a.

Tharani. S, et al. 2016 [8] proposed a CNN-based approach for the classification of diabetes and heart datasets. The approach includes a feature extraction layer and a repetitive feature mapping layer in CNN architecture that contains modified backpropagation. Authors compared CNN with other NN models such as ID3, fast rotational forest, and J48. For the heart disease dataset, the Authors got 91% ROC using the CNN model and for the diabetic dataset, the ROC is 82% when CNN is used.

Y. Song, W. Cai, et al. 2013 [9] proposed a new approach for classifying lung tissue diseases. The dataset contains 113 sets of High-resolution CT images. The proposed feature descriptors were RGLBP (Rotation-Invariant Gabor-LBP) and MCHOG (Multi-coordinate HOG). After feature descriptors, for image patch, the Authors used PASA (Patch-adaptive sparse approximation) algorithm. Features like gradient, texture, and intensity were extracted for each image set.

Mao. Keming, Tang. Renjie, et al. 2018 [10] classified the lung nodules using the Unsupervised Deep Autoencoder algorithm. Using the Superpixel method, the global image is divided into subpixels then, for each subpixel, deep Autoencoder is implemented to extract the features from the local image. The dataset used by the Authors contains 379 lung CT images divided into 8 groups randomly. The authors used different classifier algorithms such as SoftMax, KNN, SVM, and Decision tree and compared the model.

Ueda. Masaru, Ito. Koich, et al. 2019 [11] gave an introduction on 3D-CNN and proposed an age estimation model using 3D-CNN based on the human MRI brain images and the patterns in those images. The dataset contains 1010 human MRI images of both healthy and ill people who got a brain tumour, psychiatric diseases. The accuracy for the proposed 3D-CNN is 96% on brain images whereas, when 2D-CNN is implemented the accuracy is 94% only.

Jia, X., & Meng, M. Q. H. et al. 2016 [14] implemented a deep CNN on endoscopy pictures taken by (WCE) wireless-capsule camera for the purpose of detecting the inner bleedings. For classification, SVM is used on a dataset that contains 10,000 WCE images of which 7150 were normal images and 2850 images were of bleeding. CNN architecture was trained on 8200 images and the rest 1200 images were taken under the testing stage. For the proposed CNN, the Authors got the F1 score of 0.9955.

Cui, Z., Yang, J., & Qiao, Y. 2016 [15] proposed a patch-based CNN made up of 7 layers. The proposed CNN was implemented on the MRI images of the human brain. Dataset considered is available publicly that contains 103 MRI images contains 4 categories of brain diseases. For training, 100000 patches were considered from the MRI images. A dice ratio of 95.19% is achieved which is used to measure the accuracy of segmentation whereas CNN architecture accuracy was 90.83%.

Ciresan, D. C., Giusti, A. et al. 2013 [16] proposed a DNN architecture mainly on the feed-forward network along with 2D convolutional layers in the architecture. The proposed method was to implement for detecting breast cancer, especially on the mitosis condition. Dataset MITOS was used that contains 50 images of which divided into 3 subsets. The authors got the performance of 0.782 of the F1 score.

M. Russo, M. Stella et al. 2019 [25] proposed a CNN method on which they compared VGG16 and ResNet50 architectures were compared on the histopathology images of lungs to detect lung cancer. Dataset considered were ACDC@LUNGHP (Automatic Cancer Detection and classification in Whole
slide Lung histopathology). From the dataset, the Authors considered only an ROI part of images. VGG16 gave accuracy for patch classification is 97.9%.

K. Laukamp, F. Thiele et al. 2018 [26] implemented a multiparametric deep learning model for segmenting and detecting the brain tumour, especially in the case of meningiomas type of cancer. The authors considered the MRI dataset images of 136 patients of which DLM detected 55 patients out of 56 patients who have got the tumour. They conducted manual segmentation and Automatic deep learning-based segmentation. They got 98% accuracy on automatic deep learning-based segmentation.

A. Demir, F. Yilmaz et al. 2019 [27] implemented Inception-V3 and ResNet101 on skin cancer datasets to detect cancer at early stages. Dataset consists of 2437 images of size 224X 224 X 3 which contains malignant(1107), benign(1330) and the source for the dataset is from ISIC-Archive. The authors got an accuracy of 84.09% for ResNet101 and 87.42% for Inception-V3.

3. Deep Learning - Overview

In Deep learning, the term “deep” indicates the number of layers that to be implemented to manipulate and transform the data. Deep learning is one of the branches of machine learning dependent on studying various layers of perception, through the development of a hierarchy of characteristics that distinguish lower levels from higher levels and that can lead to identifying several upper-level characteristics. The most advanced methods of deep learning[20] were dependent on artificial neural networks, especially CNN which contains multiple layers for processing the image. Each layer seeks to process the information into a significantly more generalized and structural representation within deep learning. For example, In face recognition, the 1st layer may encode the pixels and extracts the edges, the 2nd layer may design and interpret the edge layouts, the 3rd layer may extract the ears and nose, and the 4th layer may identify that now the picture has a face. Unsupervised learning activities can be handled by deep learning architectures. This represents a major advantage since unspecified data are more common than those identified and tagged. Some of the deep learning architectures were Convolutional Neural Networks, Recurrent Neural Networks, Autoencoders, Deep Reinforcement Learning, Generative Adversarial Networks, Encoder-Decoder Architectures, Feed Forward Neural Networks.

**FFNs** - A FFN (feedforward neural network) is a deep learning architecture[21] that does not have loop relationships between nodes. It was developed as the first and easiest form of deep learning techniques. Throughout this framework, data flows from the input layer to the target layer through the hidden layer in just one way. The infrastructure does not have loops or cycles. It is a supervised learning-based architecture.

**RNNs** - Recurrent Neural Networks[23] are the neural networks that consist of loops and consequently have the state memory. They can handle the patterns in continuous content effectively and efficiently. Several RNN frameworks, which include LSTMs and GRUs, have been designed to help recognize characteristics in sequences. To process the variable sequential inputs, the RNNs can utilize their memory which gives more accurate results. It is a supervised learning-based architecture.
CNN - Convolutional neural networks (CNN) are a class of deep learning. CNN[23] is mostly used for analyzing visual images. The primary goal for using deep learning is to enable the machines to grasp the real scenario as people think and perform the tasks by using the gathered knowledge. It does tasks such as segmentation, image classification, and object detection that can be used in medical images. It is a supervised learning-based architecture.

Encoder: Decoder - In the FFNNs, RNNs and CNNs are essential networks that estimate with the use of a dense encoder, recurrent or convolutional encoder. Based on the original dataset this attempts to generate a meaningful representation, these encoders can either be merged or modified. "Encoder-decoder" architectural design is a more advanced analysis based on the decoding phase to produce a high-dimensional result rather than doing an estimation. It is a supervised learning-based architecture.

GANs - GAN is a training network architecture[22] designed for producing new practical representative samples. The training cycle involves two networks in its simplest form. One process, termed a generator, produces new information occurrences to deceive the other network, the discriminator, which categorizes images as false or true. In recent years, a wide variety of GAN variants and modifications have been suggested, such as the ability to produce pictures from a specific category and the ability to transfer pictures from one field with the other. It is an Unsupervised learning-based architecture.

Autoencoders - The learning process of the elements for the autoencoder systems[20] is unsupervised because the labelled data are not being used. Unsupervised autoencoder's basic architecture is to step out with an input layer, often a hidden layer, and a production layer. For pre-training or dimensional reducing, autoencoder may be used while the architecture assumes the shape of the bottleneck. For convenience, assume an invisible layered autoencoder; then, by building hidden
layers, an autoencoder may gain many layers of data representation and works as a feature extraction algorithm.

![Autoencoder Diagram](image)

**Figure 6. Autoencoder**

Deep RL - Reinforcement learning is an arrangement in which an entity is trained on how to respond in the community such that reward is maximized. If the training is performed through a neural network, we relate this as Deep RL. RL architectures have 3 types: model-based, policy-based, and value-based. Deep RL empowers us to implement deep learning whenever sequences of assumptions are expected in virtual or actual-world conditions.

![Reinforcement Learning Diagram](image)

**Figure 7. Reinforcement Learning**

4. CNN - Overview

A CNN is made up of inputs and outputs layer along with several hidden layers. CNN's hidden layers are usually a series of convolutional levels that interconnect with transmission or other level components. The most generally used activation function is RELU and is accompanied by multiple convolutions including pooling layers, interoperability and fully connected layers that can be said as hidden layers, like that of the function activation and final convolution wraps their layer parameters. Firstly, when we consider convolutional layers[19], the input is fragmented into levels and its output is passed towards the next stage. Each convolutional neuron performs just information on its field of transmission. While neural networks could be fully connected having feed-forward capabilities and to distinguish the data but applying this framework on the medical pictures is not feasible. Because of the very huge inputs combined with pixels in which any point is a significant component, a wide range of connections will be required even under a small architectural design. The above problem is given by the convolution procedure which reduces the chances of empty constraints, enabling a deeper network with far fewer metrics. The pooling of non-linear down-sampling is another essential layer of CNNs.

Max pooling - The implementation of maximum pooling is perhaps the most prominent of many non-linear enhancements. It divides the pixel region of an image into sub-regions by slicing the pixel values in a matrix that further gives the maximum output. The idea of max-pooling is to reduce the representation size for extracting features and mainly to avoid the overfitting of the customized model.

Activation function - The most used activation function is the RELU as a layer that removes the negative values from the value matrix by considering those negative values as zero. It decreases the linear functionalities of the activation function[17] without disturbing the subfields of convolutional layers. Because of its faster speed for training the neural networks model to get the standard accuracy, the RELU function is preferred. The most generalized equation is as follows:

\[ F(a) = \max(0, a) \] (1)

Where \( a \) is the matrix containing pixel values of an image. The sigmoid function used is as follows:

\[ \sigma(a) = \left(1 + e^{-a}\right)^{-1} \] (2)
**Dropout layer** - As most constraints are being used by a fully interconnected layer, the overfitting of a model can be possible. One way to eliminate overfitting is to dropout[18]. In the learning process, independent entities were either removed from the network with such a probability of \((1-P)\) or maintained with a probability \(P\). For the respective initial weights, the dropped entities were reconfigured into the network architecture. Throughout the learning phases, the most considered probability is 0.5 for dropping the constraints of input nodes. The dropout helps to reduce the overfitting if all components of the training data were excluded. This approach also enhances training efficiency enormously.

**Stochastic Pooling** - The major limitation of Dropout was that the nodes are not fully connected to those of the convolutional layers. For stochastic pooling, a stochastic technique is often used to replace traditional probabilistic pooling functions, where the activation among each pooling field is identified randomly based on a generalized linear allocation, as determined by activity in the pooling zone.

---

**Flowchart 1. CNN Architecture**

---

**5. Comparative Study**

From the comparison table 1, Datasets includes the images related to 4 brain tumor, 1 heart disease, 2 lungs related diseases, 1 skin, 1 liver lesion, and 1 retinal dataset on which certain various deep learning techniques include CNN, SVM, GAN, Capsule Net, ID3 and transfer learning was implemented to get the appropriate accuracy without getting into overfitting the model. Type of images considered were CT, MRI, clinical Photography, and OCT images. For deep CNN, transfer learning and fuzzy C – means techniques got the highest accuracy when compared with all other techniques concerning their datasets. Especially, by using the 3D – CNN, detecting the brain tumour or disease has got the highest accuracy.
Table 1: Comparison of various methods along with their medical dataset images

| REFERENCE | DATASET          | IMAGE TYPE | METHOD USED            | ACCURACY  |
|-----------|------------------|------------|------------------------|-----------|
| [1]       | Brain Tumour     | MRI        | Capsule Network model  | 86.56%    |
| [2]       | Liver lesion     | CT         | GAN + CNN              | 92.4%     |
| [3]       | Retina           | OCT        | Transfer learning      | 96.6%     |
| [7]       | Brain Tumour     | MRI        | Fuzzy C - means        | 96.97%    |
| [8]       | Heart disease    | CT         | ID3, CNN               | 91%       |
| [8]       | Diabetic         | CT         | ID3, CNN               | 91%       |
| [10]      | Lung disease     | CT         | AutoEncoder            | 93.9%     |
| [11]      | Brain Images     | MRI        | 3D - CNN               | 96%       |
| [11]      | Brain Images     | MRI        | 2D - CNN               | 94%       |
| [12]      | Lung Nodules     | CT         | Deep Fusion feature    | 96.02%    |
| [13]      | Skin lesion      | Clinical Photography | SVM + RSurf | 82.6%    |
| [25]      | Lung disease     | Histopathology | VGG16, ResNet50      | 97.9%     |
| [26]      | Brain tumor      | MRI        | DLM(Multiparametric)   | 98%       |
| [27]      | Skin cancer      | Clinical Photography | ResNet101, Inception-V3 | 84.09% , 87.42% |

6. Future Scope
Implementation of a Convolutional neural network with transfer learning in the field of biomedical images may lead to an increase in the performance of the model and could achieve better accuracy. Especially, One of the CNN architecture which is developed for estimating the tumours in medical images is the U-Net. It was developed for the segmentation process in the biomedical images along with computer vision techniques. For each biomedical image, pre-processing of the data is very important, as the images may contain noisy data which is needed to be removed for better analysis of images. For better accuracy, customized CNN with transfer learning can be implemented on medical images to get better performance in image segmentation and classification.

7. Conclusion
This paper explains several CNN frameworks. The key objective of this paper has been on the introduction of deep learning algorithms that includes CNN, RNN, GAN, Autoencoders for brain, lung, retina, and skin cancer along with performance metrics were provided. Various techniques were described in this review to determine various kinds of cancer. DCNNs had seen considerable improvement in the aspects of picture labelling, the state of the art has been focused on different complicated classification standards. Customized CNN with transfer learning may improve the accuracy of medical images by including image processing methods before implementing the model.

Acknowledgments
I would like to convey my most sincere gratitude to my guide Pooja Rana of Lovely Professional University, for her valuable guidance and advice. Her willingness to motivate me contributed tremendously to achieve the goal successfully.
[26] Laukamp, K. R., Thiele, F., Shakirin, G., Zopfs, D., Faymonville, A., Timmer, M., … Borggreve, J. (2018). Fully automated detection and segmentation of meningiomas using deep learning on routine multiparametric MRI.

[27] A. Demir, F. Yılmaz, and O. Kose, "Early detection of skin cancer using deep learning architectures: resnet-101 and inception-v3," 2019 Medical Technologies Congress (TIPTEKNO), İzmir, Turkey, 2019, pp. 1-4, doi: 10.1109/TIPTEKNO47231.2019.8972045.